Recognition of Man-made Objects from Range Data

Optimisation of Surface Based (Quadric) Representation

Guoyu Wang

CIP-GEGEVENS KONINKLIJKE BIBLIOTHEEK, DEN HAAG

ISBN 90-365-14452

Address: University of Twente P.O.Box 217, 7500 AE Enschede, The Netherlands

RECOGNITION OF MAN-MADE OBJECTS FROM RANGE DATA

OPTIMISATION OF SURFACE BASED (QUADRIC) REPRESENTATION

PROEFSCHRIFT

ter verkrijging van
de graad van doctor aan de Universiteit twente,
op gezag van de rector magnificus,
prof. dr F.A. van Vught,
volgens besluit van het College voor Promoties
in het openbaar te verdedigen
op donderdag 18 mei 2000 te 13.15 uur,

door

Guoyu Wang
geboren op 27 januari 1962
te Qingdao, China

Dit proefschrift is goedgekeurd door

Prof. dr ir P.P.L.Regtien, promotor, en
Associate prof. dr ir Z. Houkes, assistent-promotor

To Li and Ya-Ting

Abstract

The subject of this thesis is object recognition from range data. Being developed for applications of machine vision in engineering tasks, the discussion of this subject is restricted to the class of man-made objects where object surfaces can be modelled with quadric primitives. The study aims at developing optimal representation and recognition of the objects from range data.

Two different range finders are utilised to provide the range data for the work of this thesis: The "frequency-difference" laser scanning sensing system that is being developed by the Ocean University of Qingdao (OUQ), China, and the ranging system developed by University of Twente (UT), The Netherlands. Firstly, the "frequency-difference" laser scanning sensing system is particularly introduced. Based on the newly developed sensing technique and its implementations, this thesis deals with modelling the 3D representation of the measurement system. After configuration of the system parameters, a new fitting algorithm is proposed to acquire the accuracy of parameter estimation in calibration. To avoid the non-linearity of error propagation in 3D representation, a "sensor-driven" technique is proposed for parameter estimation to improve the reliability of the solutions.

Subsequent parts of this thesis are concerned with the investigation of methodological aspects of optimisation of surface representation and recognition. Surface parameters are estimated through fitting of range data to a quadric representation. The bias-corrected renormalisation algorithm is applied for surface fitting because of its property of unbiasedness. However, due to the inexact implementation of the gradient weighted least-squares criterion, the solution is deficient in accuracy. By reformulating the implementation of the estimation criterion, a new method is proposed, which improves the reliability of the bias-corrected solution.

As a focus of this thesis, the uncertainty problem in parameter estimation and feature representation is discussed analytically. In conventional parametric descriptions of objects, only the estimates of parameters (the first order statistics) are used as the explicit representation to model the objects. The uncertainty problem in parameter estimation was not well studied. In this thesis, the second order statistics of errors in parameter estimates are modelled with a covariance matrix. Assuming low level Gaussian noise present in the raw range data, an explicit formulation for the covariance matrix is obtained. Therefore the optimal description of the quadric primitive is represented by the estimate and its covariance. A probabilistic model of normal pdf is used to describe the statistical properties of the estimated parameters. This consequently enables the direct application of optimal classification of surface primitives. Moreover, the probabilistic representation can be applied to establish an optimal model base obtained from measurements.

The feature descriptor of a surface is defined on the invariants of quadric representations. These invariants are extracted through pose transformation between the viewer-specific co-ordinate system and the object-specific co-ordinate system. An advantage of

such a feature definition is that the pose information can be extracted simultaneously. This thesis also deals with pose description of surface-based object representation by compounds of quadrics, especially their relational pose formulation. Combination of feature description and relational pose is expected to provide a generic approach for modelling 3D objects using quadric primitives.

To extraction of the surface primitives, segmentation of range images is investigated in this thesis. The proposed optimal description of a surface primitive is applied for the region-based segmentation approach. Because estimates of a surface primitive are represented with a probabilistic model, the homogeneity criterion is directly established on an optimal framework. Moreover, because the uncertainties of region estimates are explicitly expressed, the order of region merging can be solved in the sense of reliability of region representation.

The proposed optimal description of man-made objects by surface-based quadric primitive and the segmentation approach have been tested experimentally. Both the range data, acquired by the range finder of UT and of OUQ, as well as the synthetic data, are used for surface classification. The results show that the classification, based on an optimal description, is more reliable than the one based on conventional descriptions. The range images acquired by the ranging system of UT and the synthetic data are used to test the segmentation approach. The experimental results show that the proposed segmentation algorithm is practical and reliable.

Acknowledgements

First of all, I wish to thank the University of Twente for providing me the opportunity to pursue a Ph.D. degree that has been a desirable target since I began my academic career.

I start to thank prof. dr ir Paul P.L.Regtien. His encouragement, guidance and academic manner strongly supported me to finish this thesis. I also wish to express my deep gratitude to associate prof. dr ir Zweitze Houkes. Starting from the initial stage of my work, his helps at each aspect and fruitful discussions have accompanied the total work of this thesis. Without his contributions, I could not have reached the final stage by now.

I want to thank dr Maarten J. Korsten. His intellectual remarks improved a great number of elements of this thesis. I am also grateful to Alfred P. de Vries and ing. Egbert Nijland for their consistent helps whenever I needed technical support. Particularly, I wish to express my gratitude to Joan H. Manohar-Mollevanger, who is always very kindly to offer miscellaneous assistance. I would like to thank dr ir Ferdi van der Heijden, Erik van Dop and Jaap Glas, for a lot of academic helps from them. Additionally, thanks to ir Kees de Rooij, Nicolae Damean and Zoran Zivkovic who let me feel a nice working atmosphere.

Particularly, I wish to express my deep gratitude to prof. Guoxing Zheng, Xin Li, Bing Zheng and my other colleagues in the Ocean University of Qingdao, China. It is their remarkable research work, their solid and voluntary supports that initialised the possibilities to start the work of this thesis. Their related contributions are introduced in chapter 2 of this thesis. I also thank prof. Tianjin Feng for his friendly concern.

Finally, I sincerely thank my parents for their voluntary supports. I am deeply indebted to my wife Li. She made a lot of sacrifices while keeping her consistent encouragement, support and love. I am happy to mention my daughter Ting ting. Her sweet concern and lovely patience always seemed like the nice star lights accompanying me when I was working in the late night.

Furthermore, I would like to thank the members of the promotion committee:

Prof. dr H. Wallinga, University of Twente
Prof. dr ir P. P. L. Regtien, University of Twente
Associate prof. dr ir Z. Houkes, University of Twente
Dr M. J. Korsten, University of Twente
Prof. dr ir C. H. Slump, University of Twente
Prof. dr ir J. Meijer, University of Twente
Prof. dr ir P. P. J. van den Bosch, Eindhoven University of Technology

Guoyu Wang, April 2000

Contents

Chapter 1

Introduction

Three-dimensional sensing modalities have been extensively studied to improve the quality of vision systems for machine intelligence. In the past two decades, a variety of range finder systems have been developed for various applications, for example, navigation of autonomous vehicles (on land, in the air, or under water), inspection or measurement of the geometric integrity of manufactured parts, measurement for recycling of industrial products, identification for mechanical assembly/disassembly, or specific designs of medical aids.

A particular example of a newly developed range finder is the so-called "frequency-difference laser scanning 3D sensing system", being under construction by the Opto-electronic Information Laboratory of the Ocean University of Qingdao (OUQ), China. This ranging system is based on a novel "frequency-difference scanning" technique to acquire three-dimensional data of a scene. The working principle and system set-up, as well as some features of the system, are described in chapter 2 of this thesis. This system has been designed mainly for inspection and carrying out engineering tasks, on land or under water, for which reliable recognition and object classification is of paramount importance. The objectives of this thesis originate from this particular requirement: — a study of man-made object recognition from range data.

Another important reason for this research work emanates from the work on a ranging system as part of a multi-sensor recognition system for electronic components on scrap PCB's. This system has been designed and constructed by the Measurement and Instrumentation Laboratory of the University of Twente (UT), The Netherlands ([18]). Although both systems operate in essentially different ways, their output is a set of range data, from which the shape of the observed objects should be extracted unambiguously.

However, since any measurement system is prone to uncertainty, errors exist in the obtained raw range data. Moreover, in most cases, data points are only partially sampled from the objects due to the limitation of sensing resolution, field of view or occlusions. Therefore, when shape is estimated from the raw range data, uncertainties in descriptions are inevitable. Coping with the uncertainty problem in shape extraction from range data is important for the reliability of recognition. This aspect has not been well approached up to now.

A vision system as applied in machine intelligence comprises three main functions: data collection (or data acquisition), data processing and interpretation. Most vision systems intended for industrial tasks utilise radiometric information (colour, intensity), i.e., a two-dimensional (2D) image of the scene, yielding "visual" data about the scene. Then a data

processing module transforms the input images to an intermediate representation, recovering some or all of the features of the scene. Finally, this information is subjected to high level analysis, resulting in the identification or recognition of the objects.

Essentially in 2D image analysis, the processing part should perform the extraction of some physical and (3D) geometrical features from the raw 2D images, such as depth, orientation or reflectance. From the resulting information, so-called *intrinsic images* ([4]) can be derived. These images are more useful for further processing because they are in fact the first version of the mapping from raw 2D images to actual physical characteristics of the scene. However, the problem of getting the intrinsic characteristics of the scene from its 2D images is inherently ill-posed because of the dimension reduction during the 2D image formation. To avoid such difficulties in low-level vision processing with 2D images, three-dimensional (3D) data (range data or range images) that are directly collected using three-dimensional sensing modalities, could be used instead. Since a range image provides essentially 3D information about the scene, it facilitates the "shape" descriptions of the scene. In other words, the geometry of the intrinsic image is embedded in the range data.

However, the interpretation of range data, too, encounters difficulties, arising from:

- scene complexity for shape descriptions;
- errors in the obtained range data;
- limitation of shape information inferred only from the range data.

The last problem can be solved by extension of the sensing system, for example the combination of 2D images and 3D range data or another multi-sensing strategy (sensor fusion). Such approach could also reduce the other two problems. However, once the range data have become available, three-dimensional object representation and recognition appears to be a non-trivial problem and has occupied many researchers for many years.

The next sections review some aspects of object recognition from range data, the background of 3D vision, acquisition of 3D data, and some existing methods for 3D object representation and recognition. Finally, an overview of the work of this thesis is given.

1.1 3D vision

An important aim of computer vision is the recovery of three-dimensional geometric information from optical images, to understand the physical structure of the scenes observed. Marr ([45]) defines 3D vision as follows: '*from an image (or a series of images) of a scene, derive an accurate three-dimensional geometric description of the scene and quantitatively determine the properties of the objects in the scene*'. Here 3D vision has been formulated as a 3D object reconstruction task. More general in Marr's theory, a computer vision system is just an example of an information processing device that could be understood at three levels:

1. **Computational theory**. The theory should define the functional device — the information it can provide (output) given other provided information (input), and the logic of the strategy performed.

2. **Representation and algorithm**. The kind of information representations and the algorithms to carry out the computation must be designed.

3. **Implementation**. The algorithms are finally realised physically, usually in software and hardware.

According to the first level, 3D vision requires an output of three-dimensional descriptors of the geometry of visible objects in the scene.

In early work on 3D vision, 3D geometric characteristics of objects were usually derived from 2D intensity images, following Marr's idea of a "2.5D sketch". The 2.5D sketch recovers the relative distances from the 2D image, which is often represented as a *depth map* in the form of a dense field, i.e., the depth information is available at all points over the 2D visual field. The depth map can be considered as an intrinsic image, which is one of the important outputs in low-level vision. The main approaches to derive the depth map from 2D images are known as "shape from X" (shape from shading, shape from contour, shape from texture, shape from optical flow, shape from motion, etc.), by estimating the local surface orientation rather than absolute depth. However, most of these approaches are generally ill-posed tasks.

In 3D vision based on 2D intensity images, a more tractable and sophisticated paradigm is stereo vision. Stereo vision is sometimes regarded as a mimicry of human visual perception (strictly, only partially in distance perception), where a pair of (or more) cameras are utilised to measure the distance by point correspondence in the two images obtained from different viewpoints. (Shape from motion is similar to stereo vision, because the measurement is based on changes of viewpoints over time). The main dilemma in stereo vision is the inherent correspondence problem. As an extreme case, for example, the scene contains a non-textured flat object with uniform brightness filling the whole imaging field. In that case finding corresponding points in the two images is not possible. Excluding such rare cases, the ambiguity in stereo correspondence can be reduced by adding some constraints, according to the properties of the image acquisition process, to radiometric properties of the scene or to prevailing properties of the objects in the real world. Stereo vision is still an important topic in the field of 3D vision, mainly because the system can be implemented in a simple way.

More recently, studies of *active perception* ([3] and [41]) and *purposive vision* ([1]) have been published. Active vision uses a more sophisticated strategy in data collection. In an active vision system, data collection is dynamically controlled by higher level outputs, i.e., data acquisition is controlled by the measured scene parameters, to obtain better convergence behaviour. In purposive vision, only a specified part of the information in the scene is needed, which justifies simpler solutions.

Generally, the state of the art of 3D vision is characterised by the extraction of 3D information of the scene at a low-level stage. Direct 3D data collection in the sensing process is a straightforward and effective approach to solve the problems encountered in distance recovery from 2D images.

1.2 Formation of range data

Range sensing systems provide three-dimensional information directly as inputs of classification systems. Characteristic for range sensing techniques is that both the sensors and the illumination are actively controlled. It can be considered as an active imaging process, in distinction to stereo vision where the illumination is not controlled (passive imaging process).

Range data can be achieved by various kinds of sensors: optical (electromagnetic) sensors, acoustic (sonar), and thermal or tactile sensors. The choice of sensing modality depends on the application. Optical sensors (CCD, radar sensors) are the most commonly used. The range data from a vision system is often called a "range image". In this thesis the two terms, "range image" and "range data", are used indiscriminately in the meaning of collecting three-dimensional points in the scene (usually the "range image" refers to the whole collection, while "range data" can imply a subset). Some typical optical sensing techniques are described in chapter 2 of this thesis.

The points in a range image are characterised by their three-dimensional co-ordinates with respect to a specified co-ordinate system. If the distance is measured in a Cartesian co-ordinate system and no ordering of points is given, it is called the general *xyz* form (in this thesis, such kind of range data is also referred to as "scattered points"). If the sampling points are equally spaced in the horizontal and vertical directions on the surface, then the range image can be represented in a matrix with rows and columns of scaled and quantised range values. Often such a range image can be converted to a visual form as a 2D greylevel image, in which the position of a pixel corresponds to its *x*-*y* co-ordinates and the greylevel value corresponds to the *z* co-ordinate (distance). The range images used in this thesis have both forms: those from the UT are in matrix form, whereas those from the OUQ are in fact in the *xyz* form. Sometimes we use the notation r_{ij} to represent a regularly organised range image, where *r* indicates the range measurement associated to a specified co-ordinate system and (i, j) is a pair of integers as the indices. In case of a Cartesian co-ordinate system, r_{ij} can be given as a matrix where (i, j) represent the indices of the rows and columns corresponding to the sampling intervals at *x* and *y* directions. If the points are sampled with respect to a spherical co-ordinate system with equal intervals of angles, (i, j) represents the indices of the two angles and *r* is the distance between the point and the origin.

Range data can further be classified as *dense* or *sparse*, according to the space resolution of point sampling relative to the size of objects in the scene. In fact the difference between these two classes is ambiguous. More strictly, a dense range image should refer to a sampling resolution based on the applicable maximum sensing resolution of the modality itself. However, because of the limited field of view as well as the limited sensing resolution, the collection of range data, irrespective of being dense or sparse, only "partially" conveys the three-dimensional information of the geometry of the scene. Moreover, errors in distance measurement are inevitable. The errors could be caused by mechanical instability of the scanning system, electronic noise, deviation of illuminating conditions, extraordinary physical properties of the scene, or just the limited resolution of the sensing system itself. When these errors occur randomly, they are commonly represented as a stochastic process in the low-level processing of range data.

Next to the acquisition of raw range data, the representation of the 3D geometry from the collected data and the interpretation of the data are major tasks in 3D vision.

1.3 Object representation and recognition with range data

The collection of range data is hard to be utilised to render an interpretation directly. Thus a *representation* scheme must be applied to describe the geometry recovered from the range data. Briefly, representation means a deterministic, structural function describing the shape of the object(s). One essential characteristic of representation is that it renders the features of the object(s) suitable for higher level cognitive visual purposes. Another requirement of the representation is its discriminative property, i.e., the representation enables to distinguish between different shapes. The representation scheme is closely related to the recognition strategy. It dictates the robustness and efficiency of the higher-level recognition task, while the control strategy of recognition guides the choice of the representation approach.

Associated to the representation scheme, a low-level task in 3D vision system is the segmentation of range images. The role of segmentation is to extract geometric primitives relevant for higher level processes from the raw range image. These primitives have a common geometric representation (or property). Usually, such a low-level vision task is *data-driven*. This means data processing without any assumptions regarding the semantic content of the scene. Only generic constraints on the scene surface, geometry in formatting co-ordinate representation and physical properties of the sensing devices are employed. In another words, it is desired to avoid making any application-specific constraints of the scene. Although the purpose of the low-level segmentation seems simple, it has turned out to be the most difficult task in computer vision. Human vision, on the other hand, is remarkably adept at detecting surface discontinuities, smoothness and spatial coherence. Obviously, human vision relies for a great part on "experience", "hidden" knowledge in the so-called *perceptual organisation*. This intelligence aspect of human vision seems very difficult to emulate in computer vision. Perhaps, the "representation" of human vision is essentially different to machine vision (this is not well understood).

According to the idea of [50], a representation scheme is desired to have the following characteristics:

- Completeness (or Unambiguity: no two or more different objects have the same representation).
- Uniqueness (for a certain object, there is just a single description).
- Richness (it is able to model a large class of objects).

Unfortunately, up to now most representation schemes fail to satisfy one or more of the above features. In fact, there exist a variety of representations, each of which is associated to a specific recognition strategy. There is no universal representation scheme in 3D vision that copes with any kind of vision tasks. Representation schemes could be classified as generic or distinctive. Generic representation schemes aim at a description of a wide range of

objects (e.g., for free-form objects), while distinctive representation schemes focus on specific classes of objects under consideration.

The goal of a 3D vision system is to recognise the concerned objects in the scene (including localisation of the objects in a specified co-ordinate system). Different to the *reconstruction* task, which aims to estimate a continuous function of a 3D surface, *recognition* means the identification of an instance of the object under the assumption that the object classes are known in advance, i.e., a model-based process. Generally, a model representation is given in an *object-oriented* co-ordinate system, allowing the model description to be viewpoint independent. Thus the model description is *feature*-based. This nature of a model description implies that the representation scheme for geometric recovery of the objects should be ultimately performed at feature-level. Hence, feature descriptions should also be viewpoint independent. Usually the quantitative components of a feature representation are denoted as *invariants*.

Once we have proper feature descriptions for the objects in the scene and the models, the task of recognition is to match these two descriptions. In the sense of scene interpretation, the recognition can be considered as a *constraint satisfaction* or *consistent labelling* problem, i.e., to find a correspondence between the entities of the scene and those of the models, using generic knowledge and domain constraints. Therefore the matching process is in fact a constraint-directed search process.

Two typical matching strategies are:

- Model-driven search (top-down). The search is controlled by the nature of the representation of the object model. Usually the features are defined at the level of primitives. The matching process starts from the correspondence between local primitive features and the model features (satisfying the local consistence). Then the local consistence of matching propagates to a global consistent match between the scene and the model. A commonly used searching approach is the *tree search* ([26], [27] and [28]). Starting from the first level of the tree, each node represents a hypothesised local match and the propagated paths represent local consistence. A global match is achieved if a path of the tree satisfies the global constraints. In the test process, a transformation between the viewer-centred co-ordinates and the object-centred co-ordinates is utilised to check the geometric constraints. Such approach is called "recognition-via-localisation", meaning that object localisation is computed during the procedure of recognition. There exist a variety of searching approaches such as Hough clustering ([54]), geometric hashing ([20] and [40]) and alternatives of indexing strategy.
- Data-driven search (bottom-up). In this strategy, the features are defined at higher level semantic descriptions, often object-specific, rather than primitive-specific. The constraints used in the search are more generic, so it is more suitable for general scene interpretation. The major characteristic of the data-driven approach is that it processes through a bottom-up sequence: image processing — segmentation — feature extraction — recognition. The scene and the model are represented by graph-theoretic descriptions. Thus the search is typically based on such paradigms as graph matching and subgraph isomorphism. The match is co-ordinate-independent. Because the localisation

phase is separated from the recognition phase, such a searching approach is also called "recognition-followed-by-localisation".

The difference between these two strategies can be illustrated with a case where the scene consists of a cubic block. In the model-driven strategy, the features could be represented by the corner points, which are derived in a low-level stage. The object is also modelled with the corner points, including geometric constraints about the cube. The matching relies on the model representation and takes place early in the whole recognition process.

In the data-driven strategy, the features could be defined in terms of the whole object, with a semantic representation, such as "a cubic block in height, width, length of quantified values". This is usually referred to as a *volumetric* description. Obviously, such higher-level representation enables the matching simpler, but the representation is more complicated than the model-driven case.

In fact, the difference between the low-level and high-level representations is not strictly defined. For example, between the point-based and the volumetric object-based representations, there can be alternatives for feature representation, such as line-based or surface-based. Often the representations or feature extraction are thought to be at intermediate level from the low-level processing to the high-level semantic description. Feature representation has been a critical issue in 3D recognition.

In either the model-driven or the data-driven approach, most model-based recognition systems are domain-oriented and do not provide a general solution. Usually the design for representation and recognition schemes is more or less application-dependent. The model-driven strategy is commonly used for goal-driven recognition such as bin-picking tasks. In general, a complex representation scheme allows the use of a simpler control strategy in matching, and vice versa. Different approaches to 3D recognition system could be looked upon as different ways of making a trade-off between the complexity of representation and the complexity of matching schemes.

One of the most important goals of 3D vision systems is the robustness of the recognition process. In general, the recognition system should be able to cope with occlusion, noise, incomplete data, or changes of viewpoints of the scene. The reliability of quantitative measurements of features will largely effect the robustness of the recognition process. Uncertainties in low-level and intermediate-level quantitative measurements inherently propagate into the high-level matching phase. Usually a certain tolerance has to be incorporated or accepted in the matching process. Although techniques in the sense of *robust estimation*, e.g., the Hough transform, can be applied, the reliability of a control strategy could still suffer from the lack of the explicit representation of the uncertainties in the low-level and intermediate-level measurement. The problem of uncertainties in low-level and intermediate-level processing has not yet been well treated. The main objective of this thesis is to arrive at an optimal recognition system using proper modelling of uncertainties in parameter estimation and feature descriptions from range data, with emphasis on *reliability* of the object representation.

1.4 The scope of the thesis

This thesis aims at the representation and recognition of man-made objects from range data, as obtained by the ranging systems of the OUQ and the MI group. Its core is a methodological study on optimisation of surface-based quadric representations with range data. The purpose of the work is to improve the reliability of 3D vision systems.

The expected applications of the ranging systems concerned (either the laser-scanning system of the OUQ or the structured light ranging system of the UT) allows the restriction to man-made objects, where the surface of the objects are assumed to be smooth and regularly shaped.

Second order polynomials (quadrics) are employed for shape description, which has been a popular representation for smooth curved surface primitives. Although the class of object surfaces is limited to the quadrics family, quadric primitives are powerful to model also more complicated objects, using *boundary representations* (B-reps). In such a representation a complicated object is described by a list of surface primitives.

The following particular aspects are considered in this thesis:

1. Utilising a surface-based quadric representation, pose parameters can be recovered simultaneously with feature extraction. A new concept to improve the estimation results is introduced.
2. Uncertainties in quadric representations are investigated and a covariance model is proposed to describe probabilistic characteristics of the estimates. The shape representation is optimised through an explicit probabilistic model. Based on the modelling of uncertainties, the surface feature defined on invariants of representations is also described with a probabilistic model.
3. Consequently, the feature-based surface recognition, as well as the range image segmentation, is based on a Bayesian framework.
4. A new method of calibration of 3D measurement for the “frequency-difference” laser scanning 3D sensing system is presented. It is shown that, without point correspondence, a surface fitting algorithm is applicable for optimal estimation of the parameters of the 3D measurement system.

1.5 Overview of this thesis

After this introduction, in chapter 2 the “frequency-difference” laser-scanning 3D sensing system is introduced. Its basic principle of operation is explained, and a new method of calibration of the measurement system is presented.

Chapter 3 focuses on the investigation of parameter estimation in surface-based quadric representation. The parameters of a quadric primitive are estimated through fitting a set of range data to a second order polynomial in implicit form. From the various approaches for quadric fitting, the *bias-corrected renormalization* method has been chosen. The property

of unbiased statistics provides the possibility to model the estimates with a general analytical representation, i.e., the *normal* probabilistic density function (pdf). However, the estimation with the renormalization method shows up some deficiencies. The problem is induced by inexact formulation in the eigenvector-based solution of the *gradient-weighted least square* criterion. An approach is proposed to improve the reliability of the renormalization method, based on an exact formulation of estimation with the eigenvector method.

In chapter 4, the quadric representations to describe the class of surface primitives of man-made objects are discussed. Using implicit quadric representation, the surface feature, defined by the invariants of the representation, can be obtained through transformation from viewer-specific co-ordinates to the surface-specific co-ordinates, i.e., the standard form. Invariants extraction and recovering the pose parameters (rotation and translation) are performed simultaneously. In order to optimise the parameter and feature representation, this chapter explores the propagation of uncertainties from the estimate in surface fitting into the feature representation. The uncertainties are represented with an explicit form, the covariance matrix. Thus the primitive representation is described with a *covariance model*, consisting of the first and second order moments (expectation and covariance matrix) of the estimate. Consequently, an analytical normal pdf is employed to establish the Bayesian framework for feature-based primitive classification. Representation of pose parameters is also discussed in this chapter.

An important low-level vision task is segmentation of the range image, required to render the surface-based representation of primitives. In chapter 5, a region-based approach for segmentation of a range image is proposed. Since the parameter estimates have been explicitly represented with the probabilistic model, the clustering algorithm is based on an optimal criterion in a homogeneity test. However, as pointed out by [33] and [66], one of the difficulties in range image segmentation is the lack of standard evaluation measures for the segmentation results. Without further quantitative assessment of the "efficiency" or "validity" of the segmentation results, the proposed algorithm, in the sense of *homogeneity* definition, focuses on the reliability of representation of a homogeneous surface primitive.

In chapter 6, experimental results of optimal classification of surface primitives and range image segmentation with real range images are presented.

Finally, summaries and prospective research are given in chapter 7.

Chapter 2

The "Frequency-Difference" Laser Scanning Sensing System

Range-sensing systems have been extensively used in autonomous systems for computer vision as a powerful modality for 3D interpretation. There exist a multitude of range sensing systems designed for various applications with different sensing techniques. One of them is the "frequency-difference" laser-scanning sensing system, which has been developed at the Ocean University of Qingdao, China. This system was designed and applied for inspection of underwater engineering tasks, where man-made objects were studied/inspected. For example, it has been applied for detection of pipelines for underwater oil transmission, inspection during repairing some mechanical elements of vessels in water, etc. A picture of this system is given in figure 2.1. This chapter introduces the characteristics of the "frequency-difference" laser-scanning sensing system and its principle of range sensing. A calibration method based on the formalisation of a 3D representation in a Cartesian co-ordinate system is proposed.

2.1 Background of active range sensing techniques

In contrast with "passive" sensing, which is the case when the source of radiation is not controlled, "active" sensing occurs when the illumination (in a wider sense) is either time or space controlled. Among the modalities of active sensing (optical, ultrasonic, tactile, etc.), the optical sensing system is used most frequently. Surveys of range finders and their applications can be found in [7], [12] and [22]. [7] gives also examples of samples from various optical sensing systems, together with a classification and a comparison.

Optical ranging techniques can be divided into two types: monocular and binocular systems. The monocular approaches are those used in laser-radar range finders. The measurement is based on the propagation time of light, the so-called "time-of-flight" (TOF). There are three basic modes of operation. A pulse detector measures the time-of-flight of discrete light pulses: this time is proportional to the distance between the laser emitting point and the reflecting target. The frequency-modulated continuous-wave (fm-cw) measures the time-of-flight indirectly by detecting the beat frequency of the emitted laser beam and its reflection. The amplitude-modulated continuous-wave (am-cw) measures the time-of-flight directly by detecting the phase shift between the emitted beam and the reflection. Generally, the pulse detection is suitable for remote applications while the am-cw or fm-cw methods are usually used in a restricted working volume and are more attractive in machine

vision. The advantages of these techniques are the direct availability of range information without subsequent computation and the fact that the measurement accuracy will not degrade when used over a larger range. In order to get a sufficiently large set of data points over the whole working space, a scanning mechanism has to be attached, which results in higher costs of implementation.

The active binocular technique is in fact the "triangulation" method. In cotrast to the passive stereo approach, the triangulation technique uses controlled light, so-called "structured light", by which the correspondence between projection and sensing has been predefined. Thus the distance can be measured with the well-known triangulation algorithm. The controlled light pattern can be a set of points, lines or grids, etc, combined with a specific sensing configuration. Triangulation ranging systems are quite popular for 3D measurements, because of their suitability for 3D surface measurement and scene profiling. At large working ranges the triangulation method is less accurate and the "missing part" problem is unavoidable because the projected and detected light beam are not co-axial. However, at moderate ranges, triangulation systems perform accurate and fast measurement and are easy to implement ([7]). In active triangulation methods, the illuminating pattern and the mode of light detection determine the formulation of 3D measurement.

Being able to configure a structured light pattern, laser-scanning techniques have been paid attention for alternative applications. In a laser-scanning triangulation system, the light pattern can be either a set of points, directly projected by the laser beam onto the target,

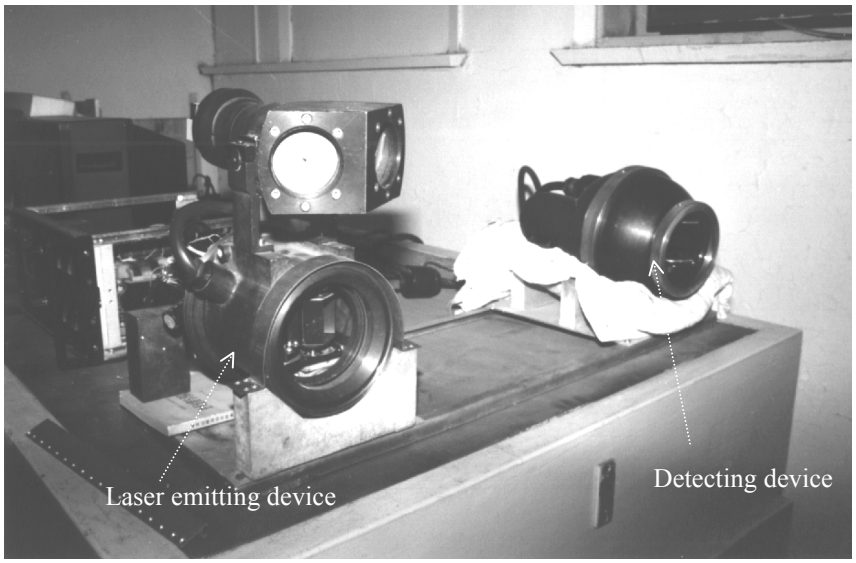


Fig. 2.1. The "frequency-difference" laser scanning sensing system. (The picture was taken in the opto-electronic information laboratory of OUQ.)

with a 2D scan mechanism, or a set of lines, created from a laser beam using a strip (slit-beam) and a cylindrical lens, with a 1D scanning procedure.

A commonly used detector in most triangulation systems is the CCD camera. However, other opto-electronic sensors might be applied as well. An example is the photo-multiplier tube that was used in the laser-scanning system as described in this chapter.

Most ranging systems, no matter what sensing techniques were used, were developed for on-land applications. When applying a ranging system in another environment, i.e., underwater, difficulties will occur in the sensing process so that a well-performed system on land would be useless in such special applications. Because the system specified here is assumed to work in water, the influence on the optical sensing process should be investigated first.

2.2 Optical sensing in an underwater environment

Compared to applications in air, absorption and scattering during light propagation in water induce some particular difficulties. The mechanisms of absorption and scatter are different. Whereas absorption causes only reduction of light energy, scattering obstructs the sensing process in a more complicated way, because it not only reduces the intensity for illumination, but also causes a back-scattering background, both of which severely degrade the signal-to-noise ratio of the sensor signals.

Scattering originates from different physical mechanisms; it depends on the composition and condition of the water, and in particular on the size of particles interacting with photons. Propagation of scattered light is omnidirectional and scattering occurs wherever the light passes through the water. In case of turbid water, the forward-scattered and back-scattered light will strongly influence the illumination and detection process in the entire field of view.

The imaging systems for underwater applications are mainly used for marine mobile robots, i.e. ROV (Remotely Operated Vehicle System) and UUV (Unmanned Undersea Vehicle System). They can also be applied for a variety of underwater engineering tasks, such as monitoring parts of offshore oil-platforms, inspection of underwater pipelines or other constructions. Since underwater imaging systems (2D or 3D) have the same functionality as those for use on-land, they must be able to overcome the influence of the scattering effect somehow. The reason that many underwater sensing systems utilise acoustic modality (sonar) in stead of light is just to avoid the absorption and scattering problems. But the drawbacks of sonar sensing systems, i.e. poor resolution and low sensing speed, limit their applications as imagers.

Several techniques have been developed to solve the scattering problem for optical sensing systems, all of which are based on the idea of reducing scattered light in the sensed signal. The polarisation method separates the target light and scattered light through polarising filters in the projection and detection parts, according to the different polarisation properties of scattered light and the diffuse target light. The main drawback of the polarisation method is the loss of light energy during polarisation. The gating method uses pulsed light and controls a gate that opens only at the moment the target light arrives, so most of the scattered light spread over the

whole projecting path is blocked during the remaining time. A difficulty of this method is that the gate is hard to correctly control by predicting the target distance from the pulse-illumination pattern. The synchronism method uses a synchronous scanning pattern for projection and detection to reduce the amount of effective scattered light. Figure 2.2 illustrates the principle of the synchronous scanning method. The light beam from a continuously emitting source scans the target, while the detector scans synchronously the reflected light within a small field-of-view. In this way, only the scattering in the overlapping region of the light beam and the detected field-of-view (the so-call "scattering volume") will effect the signal-to-noise ratio. The smaller the scattering volume, the less scattered light will be received. Compared with other methods, the synchronous scanning method is more effective and suitable for imaging systems working in a scanning way.

Although these methods were developed for 2D imaging systems, the basic concept can be applied for 3D sensing systems as well. The "frequency-difference" laser scanning sensing system, which is introduced here, was designed based on the synchronism technique, to solve the scattering problem in underwater applications. The mode of illumination is the laser point scanning. Although a laser slit is common in many laser-ranging systems, it is not recommended for an underwater environment because the slit-beam will yield a too large scattering volume, which conflicts with the requirement of a small scattering volume. Moreover, the "frequency-difference" laser scanning sensing system uses a photo-multiplier tube as detector in stead of a CCD camera, working in a time-sequential detection mode to comply with the point-scanning pattern. Since the sensitivity of a multi-photo-tube is much higher than of CCD elements, the detection range, which is usually small in water because of low visibility by absorption and scattering, can be increased.

The ranging system described below has been designed particularly for underwater applications. A novelty in this design is the "frequency-difference" scanning technique ([29] and [69]). Its advantages include automatic and fast sensing, adjustable resolution and reduction of the scattering effect all in one. Moreover, by incorporating a 2D greylevel imager, real-time 3D display on a 2D monitor can be realised just by combining the outputs of the two sensors. The real-time data-saving modality at optical sensing speed is still in development.

2.3 Principles of the "frequency-difference" laser-scanning sensing technique

In this section, the implementation of the laser scanning triangulation measurement is discussed and the principle of "frequency-difference" scanning is described.

2.3.1 Principle of laser-scanning triangulation measurement

Figure 2.3 is a simplified model to illustrate the principle of the laser-scanning sensing system. F denotes the laser emitter and S denotes the detector. S_0 is the distance between these components, and is called the baseline. The laser point-beam performs scans in lon-

gitudinal direction (denoted as *line* scanning), and latitudinal direction (called *frame* scanning). A slit-shaped detecting plane scans the scene in latitudinal direction. Assume at a certain time a point P is sensed at a line-scanning angle γ, a frame scanning angle α and a slit-scanning angle β. The position of P can now be calculated using the concept of triangulation.

Laser-scanning triangulation can be implemented in different ways, for example, a scanning slit-beam (strip of light) with an array detector such as a line scan CCD camera.

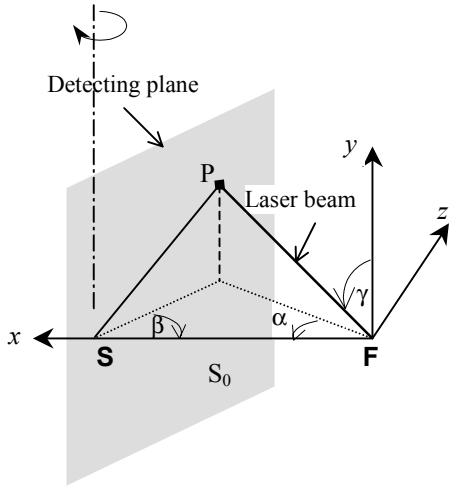


Fig.2.3 working principle of laser scanning triangulation

γ *line*-scanning angle
α *frame*-scanning angle
β *slit*-scanning angle
F laser emitter
S detector
S_0 baseline
P point of the object being sensed

2.3.2 Implementation of the scanning measurement system

The 3D sensing system consists of an optical sensing part and an electronic processing part. The optical part contains two rotating mirror-faced polyhedrons, deflecting a laser beam as to realise the two-dimensional scanning (line- and frame- scanning). A sketch of the implementation is shown in figure 2.4(a). The detection device is mechanically independent to the laser-emitting device. A slit in front of the multi-photo-tube sensor is located at the focus of a lens so that a scanning "detecting plane" is generated in space. Scanning of the de-

tecting plane is also carried out by a mirror-faced rotating polyhedron, as shown in figure 2.4(b).

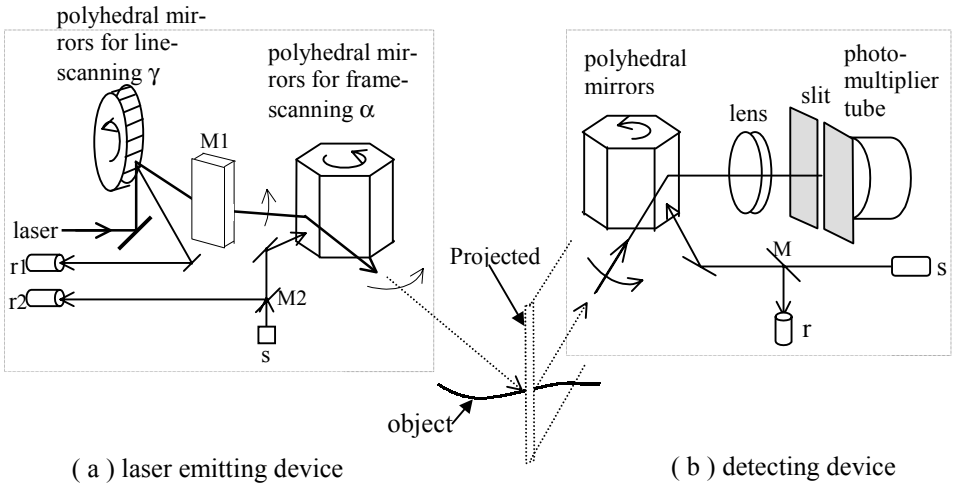


Fig.2.4 Implementations of optical scanning and sensing scheme.

(a) illustrates *line*- and *frame*-scanning implementation. **M1** is a fixed mirror transforming *line*-scanning to *frame*-scanning. The synchronising signal for *line*-scanning is generated by laser reflection at a certain pose of the mirror, detected by detector **r1**, while the synchronising signal for *frame*-scanning is generated also by reflected light at a defined position of the *frame*-scanning mirror, but an active illuminating source **s** is used. **M2** is a semitransparent mirror by which reflected light is detected by **r2**.

(b) illustrates the *slit*-scanning implementation. A slit positioned at the focus of a lens is imaged onto space through the lens and the *slit*-scanning mirror. Then an imaged (virtual) "detecting plane " is scanned using the mirror rotation. Only diffusion of the point that lies on the detecting plane can be sensed by the sensor behind the slit. The synchronising signal is generated by illuminating source **s** and a detector **r**, coupled by the semitransparent mirror **M**.

2.3.3 The principle of "frequency-difference" scanning

In the scanning laser point-beam system, a slit in front of the detector is projected onto the scene to be observed. A virtual "detecting plane" is thus created, whose angle β can be positioned while keeping the field-of-view in a fixed longitudinal direction. For any angle β, all visible illuminated object points on this detecting plane will be sensed. The frequency of the *line* scanning with respect to γ is much higher than that of the *frame* scanning with re-

spect to α and β. So during one *line* scan, the track of the laser beam projected on the detecting plane is almost a straight line perpendicular to the *x-z* plane. Only points on this track can be sensed by the detector. We call this track the "measurement track". Thus one *frame* scan contains a number of *line*-scans, resulting in a set of sequential measurement tracks moving along the depth direction (*z*-axis direction). All sensed points are the intersecting points of these tracks and the object surface (see figure 2.5).

If the scanning frequency of the detecting plane is exactly the same as (or a multiple of) the scanning frame, these measurement tracks will be repeated on the same location for all frame-scans. However, if there is a small difference between the frequency of the scanning detecting plane and the scanning frame, the location of the measurement tracks change in space with respect to the different frame-scanning periods. Consequently, during the scanning process, the measurement tracks move over the scanning field so that the range sensing is processed automatically. This is the concept of "frequency-difference" scanning sensing technique. Figure 2.6 illustrates the spatial changes of the measure tracks in different frame-scanning periods.

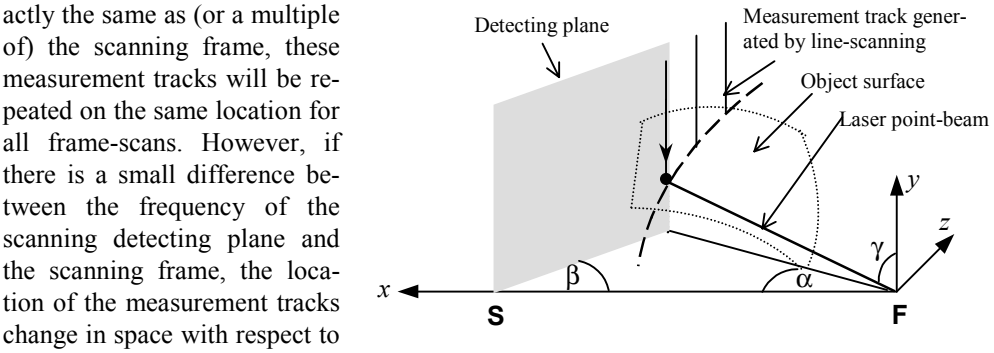


Fig.2.5 Measurement tracks generated in one *frame*-scanning period

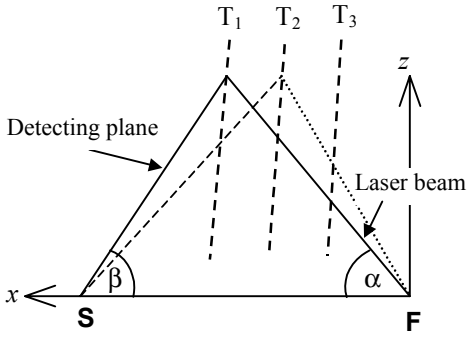
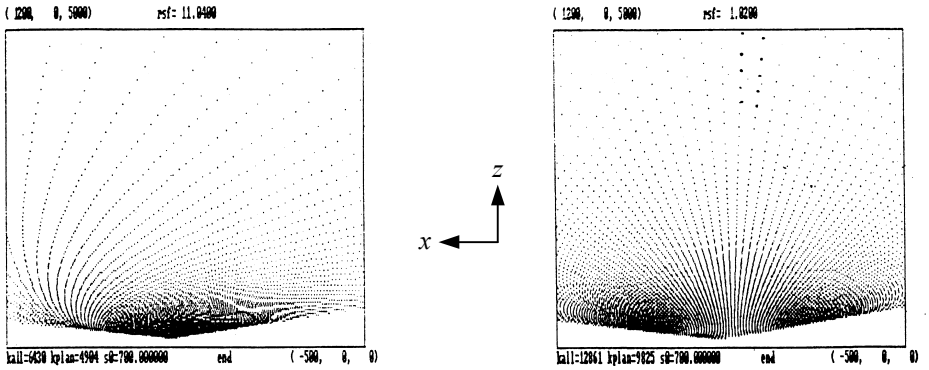


Fig.2.6 Top view of spatial changes of measurement tracks in different *frame*-scanning periods.
(T_1,T_2, T_3 , …) indicate the measurement tracks during different *frame*-scanning periods, viewed at *x-z* plane.

To quantify the "frequency-difference", a so-called *frequency-difference* coefficient is defined ([29]) as $R_{sf} = \omega_s / \omega_f$; where ω_s is the *slit*-scanning frequency and ω_f the *frame*-scanning frequency. Figure 2.7 shows two instances of the pattern of the measurement tracks with two different *frequency-difference* coefficients (viewed



(a) R_{sf}=11.04, S_0=700mm (b) R_{sf}=1.02, S_0=700mm

Fig 2.7 Two instances of simulated pattern of measurement tracks viewed at *x*-*z* plane, corresponding to different *frequency-difference* coefficients R_{sf}.

at the *x*-*z* plane), as generated by a computer simulation ([29]). The parameters are taken from the real system.

Some of the characteristics of this sensing technique are summarised below.

- By laser point-beam scanning, points are sensed through a time-sequential process.
- The measurement tracks move periodically, i.e., after a few periods of the frame scanning, the measurement tracks will cover the whole scanning space and this process is repeated. The time needed to complete such a change is defined as the "measurement period". This period is determined by the frequencies of both the scanning detecting plane and the scanning frame. The measurement period is in the order of a few seconds.
- Along the *z*-direction, the spatial sensing resolution depends on the ratio of the *line*-scanning frequency and the *frame*-scanning frequency while in the *x*-direction the spatial sensing resolution depends on the ratio of the *slit*-scanning frequency and the *frame*-scanning frequency (R_{sf}). Along the *y*-direction, the resolution is determined by the size of the laser spot and the sampling speed.

The (optical) sensing speed is defined as the maximum number of points to be sensed in one frame-scanning period. It depends on the point spatial resolution and the frame-scanning frequency. The sensing speed ([29]) can be up to 0.46μs/point.

Once the scanning frequencies of α, β and γ be designed, the *frequency-difference* coefficient R_{sf} can be used to control the sensing resolution.

2.4 System set-up

As a reference, some standard parameters of this system are listed below.

Laser: wavelength — 532 nm; power — 80 mW
Scanning frequency: line —10800Hz; frame — 25Hz; slit — 300Hz ± (an adjustable range)
Scanning range: vertical — ±20°; horizontal — 120° (both in air medium)
Baseline: about 100cm

To get some more insight in the system set up, a schematic diagram is presented in figure 2.8.

As mentioned in the beginning of this chapter, this system was designed for underwater applications. Advantages of this system when used underwater are:

(1) The instantaneous field-of-view of the detection is slit-sized while the illuminated region is spot-sized, so the "scattering volume" is small and the scattered light can be effectively reduced, improving the signal-to-noise ratio.
(2) The laser-emitting device and detecting device are mechanically independent and the baseline is adjustable. For targets at different distances, changing the baseline enables us not only to improve the resolution, but also to reduce the scattering volume.
(3) In stead of a CCD camera, a photo-multiplier tube with high sensitivity was used as light sensor. The purpose is to improve the working distance in case of poor visibility in water.

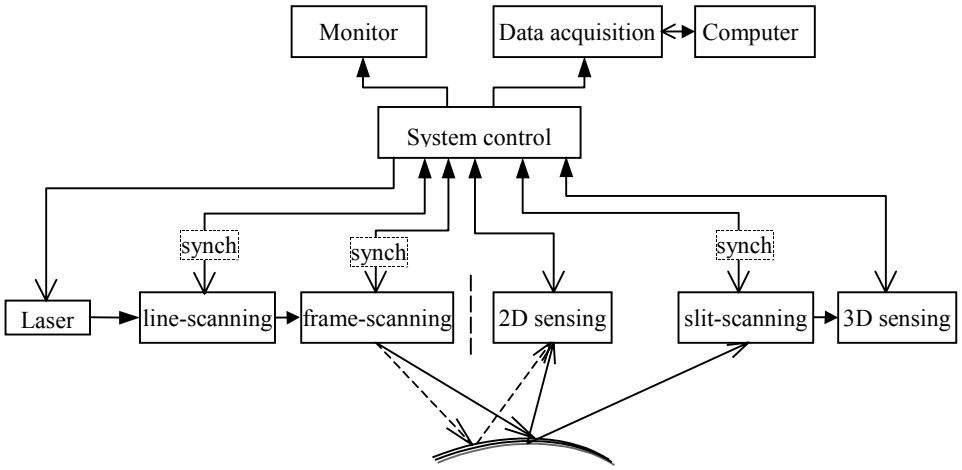
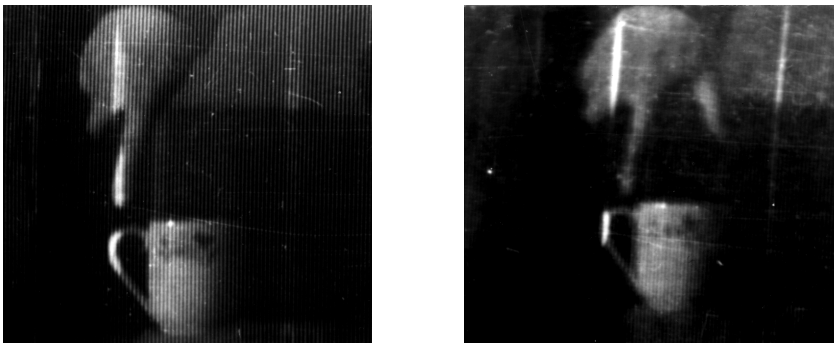


Fig. 2.8 Set-up of the "frequency-difference" laser scanning sensing system.

2.5 3-D visualisation technique

The "frequency-difference" laser scanning sensing system poses a significant advantage of the 3D-visualisation facility. Combined with a 2-D imaging system, the 3-D sensing system is capable of displaying real-time depth information on a video monitor. This possibility has been implied in the system set-up of figure 2.8. A 2-D sensing system mounted to the 3-D sensing system generates a grey-level image during the laser-scanning process, and provides a sequential video signal to a monitor. The output of the 3-D sensing system is added as an overlay to the 2-D video signal and is displayed in a highlight mode, i.e., the 3-D signal is shown in binary format. The video signal is synchronised at the start of each *line*- and *frame*- scanning period. According to the working principle, in one scanning frame, different scanning lines correspond to different depths in the formation of measurement tracks (see figure 2.5). Therefore, for 2-D video signals, the monitor displays a grey-level image corresponding to the *x*-*y* plane. When 3-D signals are displayed on the monitor, two neighboured lines represent the depth difference along the *z*- axis. In other words, the screen can be interpreted as *y*-*z* co-ordinates of the 3-D video signal. Therefore, relative depth information can be "visualised" on a simple screen by combining the 2-D grey-level video signals and the 3-D binary video signals.

As an example, figure 2.9 shows some pictures taken from the monitor screen to illustrate the 3D visualisation. The grey-level images were derived from the 2D-video signal, while the white lines indicate the output of the 3D sensing system. In (a), the white line (the 3D scanning line) covers the ear of the cup, a man's hand and finger, meaning that the ear of the cup and the hand and the finger are at the same depth. In (b), only the ear of the cup and the hand are covered by the white line, meaning that the finger has moved to a different depth relative to the hand and cup.



(a) (b)

Fig. 2.9. Two examples of observation of 3D depth information from the monitor screen.

This novel 3D-display technique can be applied for real-time inspection tasks, for example, to monitor the relative distance between a gripper and the manipulated object in a man-interfered mobile manipulation. More details about this characteristic can be found in [70].

2.6 Implementation of range data acquisition

The acquisition of range data points using a CCD camera is rather straightforward. In this "frequency-difference" laser scanning sensing system, the 3-D signal is sensed in a different, time-sequential way. Since no ready-made facilities are available, a special data acquisition system has been designed.

According to the principle, the position of a sensed point is determined by the three scanning angles α, β and γ. These scanning angles are expressed by $\alpha=\alpha_0+\omega_f t_f$, $\beta=\beta_0+\omega_s t_s$ and $\gamma=\gamma_0+\omega_l t_l$, where ω_f, ω_s, ω_l are the frequencies for *frame* scanning, detecting plane and *line* scanning, respectively. Therefore, acquisition of a data point equals recording the three time values of t_f, t_s, t_l at the moment a point is sensed.

Obviously, given the inputs t_f, t_s, t_l, an explicit representation of 3D co-ordinates of the sensed points can only be given after the definition of a reference co-ordinate system and the determination of the system parameters involved. The model of this 3D measurement system, including the definition of the 3D co-ordinate system, the estimation of system parameters and the computation of 3D co-ordinates, are described in section 2.8. A brief model of the 3D representation is given in figure 2.10.

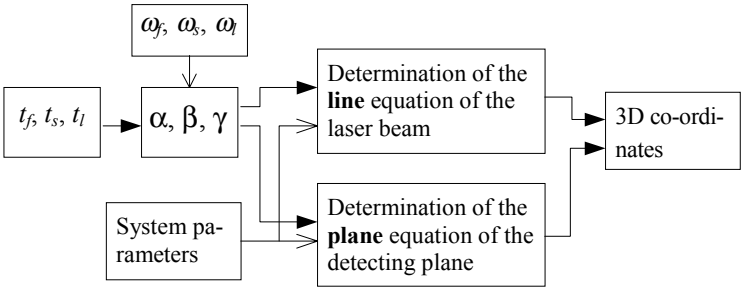


Fig.2.10. The principle of computation of 3D co-ordinates.

Theoretically, the "frequency-difference" sensing technique results in a high-sensing speed. For example, for the parameters presented in 2.3, and assuming that the number of sensed points in a line-scanning period is 200, the sensing speed ([29]) can be up to 0.46 μs/point.

Such a high sensing speed is applicable for real-time data acquisition only when the electronic processing (A/D transfer, communication, etc.) can work at the same speed. Be-

cause of limitations in the present hardware facilities, this has not yet realised in the current system. However, the 3-D data acquisition was implemented as a man-interfered facility rather than a real-time acquisition facility. Currently, it is applied only for experimental studies, while research on data acquisition is going on. The principle of data-acquisition is explained below, without giving details of the hardware implementation, just for basic understanding of the 3D measurement system.

Acquisition of the time values t_f, t_s, t_l is performed by a "cursor-positioned" mechanism, controlled by the 2-D image. A computer-generated cursor is added to the video signal and displayed on the monitor with the 2-D image. Setting the cursor to a certain position, once the point of observation corresponding to this position is sensed, the three times are saved by dedicated computer-interfaced hardware. In the realisation, the values of t_f and t_l are set by the cursor position; only t_s is recorded instantaneously. The synchronising signals, which are generated at the beginning of a scanning period, are used to trigger the time counters. By moving the cursor over the whole region of interest, a set of data point becomes available.

Clearly, the measurement precision is effected by the stability of the scanning frequency. To compensate for errors caused by this instability, the three scanning periods are also recorded simultaneously. In the next section, causes of measurement errors are clarified.

2.7 Causes of errors

The major errors of the ranging system are summarised in the following:

- Errors in recording of the scanning time

In contrast to CCD imaging, the output of this system is the recorded scanning time (t_f, t_s, t_l) at which a point is sensed. The clock frequency of the counter is high enough (~10MHz) so that the counting errors can be neglected. However, due to non-ideal shapes of the synchronising signal that triggers the counter (bandwidth limitation, noise, interference etc.), uncertainty in the determination of the start and end moments of the scanning process occurs. Of course, improvement in circuit implementation can reduce these errors.

- Extension of the laser point-beam

Since the size of the laser spot is not infinitely small, point ambiguity in the intersecting area of the laser beam and the surface yields uncertainty in the measured detecting angle β.

This problem becomes significant when the laser beam hits the surface at a small angle, causing a large spot on the surface. Therefore, object parts with surfaces along the projection direction may yield larger errors than other parts.

- Extension of the detector slit

The slit scanning detection was realised by imaging a slit through a rotating mirror onto the scene, as illustrated in figure 2.4(b). The slit is located at the focus point of a lens so that

the virtual image of the "slit plane" performs the scanning process. Only the light reflected from the object surface and passing through this slit will be sensed. Due to possible out-of-focus location, the effective width of the slit, imaged onto the scene, will be expanded. This also causes uncertainty in the scanning angle β. Obviously, the narrower the effective slit width, the less is the error. But if the slit width is too small, the light power to be received decreases so the signal-to-noise ratio degrades. An optimal choice is to let the strip-covered area be the same as the projecting laser spot. But in reality both the illuminated area and the detected area change during the measurement, so a practical slit width can only be determined empirically, according to the application.

- Instability of scanning frequencies

A precise determination of the three scanning frequencies is important for the computation of the scanning angles α, β and γ, which determine the co-ordinates of the object. To eliminate the effects of unstable scanning frequencies, the scanning periods for α, β and γ are also recorded simultaneously with t_f, t_s and t_l. Then for each record of t_f, t_s and t_l, the corresponding frequencies were used to compute the three angles.

However, the motor speed is irregular so the recorded frequency value is only an average during a period (e.g., the variation of the slit-scanning frequency is within 10^{-3}). Therefore, there are still errors in the calculated scanning angles due to frequency instability.

A solution of this problem relies on the improvements of the mechanical performances that can be obtained, but, this will lead to increased costs.

- Other errors

Besides the errors discussed above, a direct consequence of the system implementation, other errors may originate from the interactive illumination or the sensing process, for instance:
- inter-reflections due to irregularities of the surface,
- specular reflection that causes the extent of the illuminated area,
- inter-reflection below transparent surfaces,
- non-diffuse reflections causing too small signals.

Since such errors are commonly encountered in many ranging systems, a detailed discussion is omitted here.

Particularly, in case of underwater applications, scattered light may cause erroneous sensing when the scattering effect is strong. However, the "scattering volume" is reduced greatly with the mentioned system, and it can also be controlled by extending the baseline and reducing the slit width. So it is expected to effectively reduce occurrences of such kind of errors.

For example, in the experiments using real range data for object recognition (see figure 6.3 in chapter 6), the variance of noise in the obtained range data was within 0.6~1cm^2, where the objects were measured at a range within 0.5~1m.

2.8 Model of the 3D co-ordinate system and calibration

In a broad sense, calibration of a measuring system is to formalise data representation under a defined co-ordinate system. The focus of a calibration is to determine the system parameters which are involved in the computation of the output data. A variety of calibration methods for triangulation-based ranging systems has been presented in [12], [14], [15], [16] and [49]. All of them were related to CCD imaging modalities. Those for laser-scanning range finders were quite sparse among the available publications, and moreover, they were also limited to slit-beam pattern and CCD detection ([14], [15] and [49]). In fact, no available calibration routines can be directly applied to the "frequency-difference" laser-scanning ranging system because of the following differences:

- the light pattern in this system is based on points instead of lines.
- the sensing process is time-dependent rather than geometry-dependent in case of a CCD camera.

However, the basic concept of co-ordinate formalisation and parameter estimation is shared by all kind of calibration actions. In general, the measurement co-ordinate system can be defined arbitrarily. It can be fixed to world co-ordinates in case of a static view, or to robot-centred co-ordinates in case of autonomous vehicles and robots. In all cases the measuring output of the range sensor itself must be configured and calibrated by coupling it to another related co-ordinate system.

Calibration of the "frequency-difference" laser scanning system includes modelling the measurement system and finding the system parameters. In this application, the measurement co-ordinates are defined with respect to the system, that is, they are independent of either the world co-ordinate system or a kinematic co-ordinate system to which it is related when it is mounted on a mobile vehicle. Thus the parameters involved in the calculation of data points are regarded as intrinsic or internal parameters.

In line with other self-calibration methods ([19], [22], [44] and [62]), an optimal approach is proposed in this chapter to estimate the parameters through planar fitting, by which the estimation can be evaluated in the sense of *accuracy*. Moreover, the calibration errors are evaluated at the stage of time recording, which is thought to be a stationary stochastic process.

2.8.1 Definition of parameters and co-ordinate system

The measurement model is shown in figure 2.11. Suppose the laser beam originates from a fixed point **F** (laser source) and scans along longitude and latitude directions by *line* and *frame* scanning. We define **F** to be the origin of the measuring co-ordinate system and the y-axis being perpendicular to the frame-scanning plane, denoted by Γ. During the detection process, the detecting plane, denoted by Ψ, scans the space by rotating around a fixed axis which direction is represented by the unit vector $\mathbf{n}_s$. Supposing $\mathbf{n}_s$ and the frame-scanning plane Γ intersect at the point **S**, then the x-axis of the measuring co-ordinate system is defined along FS. Having defined the x- and y- axes, the z- axis follows according to the right-hand rule. The distance between **F** and **S**, denoted by L, is called the baseline length.

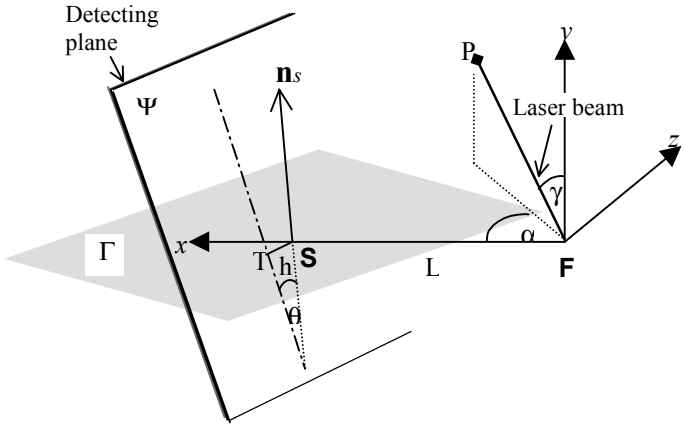


Fig 2.11 model of co-ordinate system

$\mathbf{n}_s$— direction vector of rotation axis of Ψ (the normal vector of plane Γ).
h — distance of **S** to Ψ.
θ — angle between $\mathbf{n}_s$ and Ψ.
L — baseline between **S** and **F**.

It should be noted that in this model the rotation axis $\mathbf{n}_s$ of the detecting plane Ψ, is assumed free-oriented without any constraints with respect to the *y*-axis, the rotation axis of the laser frame scanning. Also, the detecting plane Ψ is assumed to be arbitrarily posed without any mechanical constraints. This freedom in the definition of a co-ordinate system is important because in practice, it is difficult to guarantee some mechanical constraints even there are intentions to do so.

Besides the time parameters, pose parameters are also involved in the representation of a data point with the chosen co-ordinate system. There are four parameters to configure the pose of $\mathbf{n}_s$ and Ψ. Two are used to describe the orientation of $\mathbf{n}_s$ i.e., azimuth ρ_1 and elevation ρ_2 with respect to the *x*-*y*-*z* co-ordinate system. Another two parameters describe the relative pose of the detecting plane Ψ with respect to $\mathbf{n}_s$, i.e., the angle between $\mathbf{n}_s$ and plane Ψ, denoted by θ, and the distance from point **S** to plane Ψ, denoted by 'h' (see figure 2.11). These four parameters are independent of the scanning process.

In summary, the parameters involved in a 3D measurement in the defined co-ordinate system are:

α_0, β_0 and γ_0 — three initial angles of the periodic scanning for α, β and γ.
ρ_1 and ρ_2 — azimuth and elevation angles describing the orientation of the vector $\mathbf{n}_s$ referred to the *x*-*y*-*z* co-ordinates (see figure A.2 in Appendix A).
h — distance between **S** and the detecting plane Ψ.
θ — angle between the detecting plane and $\mathbf{n}_s$.
L — baseline determined by the distance between **S** and **F**.

Once the above eight parameters are determined by a calibration process, for each data point associated to the times t_f, t_l, and t_s, the 3-D co-ordinates can be computed in the defined *x-y-z* co-ordinate system.

The co-ordinate computation is derived from the simple fact that the sensed point is just the intersecting point of the laser beam and the detecting plane. Suppose the point *P* (see figure 2.11) is sensed at the recorded times indicated as t_f, t_l, and t_s. Given t_f and t_l, the line equation of the laser beam is directly available in *x-y-z* co-ordinates from the known α and γ, which are computed from t_f and t_l. Once the equation for the plane Ψ is derived in the same *x-y-z* co-ordinate system, the position of the point *P* expressed in the *x-y-z* co-ordinate system is found by solving these two equations.

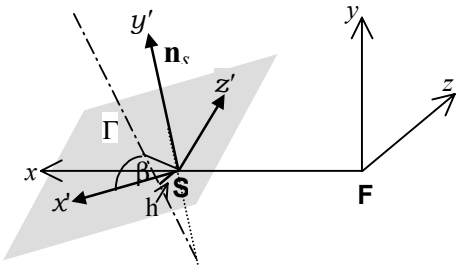


Fig. 2.12. Assistant co-ordinate system *x'-y'-z'* defined in *x-y-z*.

In order to obtain the equation of the detecting plane Ψ in the *x-y-z* co-ordinate system, an assistant co-ordinate system with origin at **S** is defined to describe the pose of Ψ when scanning. This assistant co-ordinate system, denoted as *x'-y'-z'*, is built as described hereafter.

Perpendicular to the rotation axis $\mathbf{n}_s$, and passing through the point **S**, a "scanning-plane" of Ψ is defined, which is denoted as Γ in figure 2.11 (shown in shadow). The *y'*-axis coincides with the vector $\mathbf{n}_s$. The *x'*-axis is defined in the plane Γ along the intersecting line with the plane formed by the *x*- and *y'*-axis. Then the *z'*-axis is derived by applying the "right-hand" rule. This assistant co-ordinate system is shown in figure 2.12.

From figure 2.11 it will become clear that rotation of Ψ can be described by a rotation of the line ST, the distance from **S** to the Ψ. The scanning angle of ST is $\beta = \beta_0 + \omega_s t_s$, where ω_s is the scanning frequency of the detecting plane Ψ and β_0 is the initial angle in each period.

Firstly, the equation of the plane Ψ can be easily represented in the *x'-y'-z'* co-ordinate system. At any time, the normal unit vector of plane Ψ, denoted as **e**, is expressed in *x'-y'-z'* co-ordinates as:

$$\mathbf{e} = [\cos\theta\cos\beta, \quad -\sin\theta, \quad \cos\theta\sin\beta]^T \tag{2.1}$$

So, in *x'-y'-z'* co-ordinates, the equation of the detecting plane Ψ is

$$\mathbf{x}'^T\mathbf{e} = \mathrm{h} \tag{2.2}$$

where h and θ are the parameters defined in figure 2.11.

Secondly, the equation (2.2) is reformulated in the *x-y-z* co-ordinate system by a co-ordinate transformation between *x'-y'-z'* and *x-y-z*, which is described by

$$\mathbf{x} = \mathbf{R}\mathbf{x}' + \mathbf{L} \tag{2.3}$$

The translation vector **L** is the vector pointing from **F** to **S**. The rotation matrix **R** consists of the three base vectors of *x'-y'-z'* viewed in *x-y-z*. According to the definition of *x'-y'-z'*, the three base vectors can be expressed in *x-y-z* co-ordinates, so **R** can be determined.

The equation of the points (e.g. the point P) on the line coinciding with the laser beam, expressed in vector notation in *x-y-z* co-ordinates, is described by

$$\mathbf{x} = K[\sin\gamma\cos\alpha \quad \cos\gamma \quad \sin\gamma\sin\alpha]^T \tag{2.4}$$

where K is a scale factor.

Finally, combining (2.2) with (2.4), the position of the point **x** can be obtained. Details in co-ordinate representation can be found in Appendix A.

2.8.2 Parameter estimation using a planar fitting algorithm

In a calibration process, reference objects are required as the "ground truth" to formulate the equations by which the parameters can be estimated. Without loss of generality, calibration by parameter estimation can be issued as follows.

Given a set of known data S and their measured data S′(**a**), where **a** denotes the vector of measurement parameters, find an estimate $\hat{\mathbf{a}}$ of **a**, satisfying $\hat{\mathbf{a}} = \Theta(S,S',\mathbf{a})$. The cost function Θ is defined according to the estimation criterion.

After the measurement co-ordinate system of the "frequency-difference" laser-scanning system is defined, it is expected that all the intrinsic parameters can be determined within the framework of the modelled co-ordinate system.

Simple and convenient point-matching algorithms for calibration are commonly based on the assumption that correspondence between the reference point and its measured counterpart has been identified. In practice, there exists no kind of illuminating pattern that can guarantee an exact positioning of the target point. Therefore, the measured position may correspond to an ambiguous point, lying in a neighbourhood of the target point. This neighbourhood is determined by the uncertainty in the measured counterpart. Even when the system parameters are exactly known, such an uncertainty still exists.

Following Besl [7], we distinguish between *repeatability* and *accuracy* when quantifying measurement errors in the calibration. *Accuracy* indicates the difference between the measured range and the actual range while *repeatability* indicates the variation of the measured range with respect to a given target. The error of point-matching measurement is in fact evaluated in the sense of *repeatability*. However, the repeatability is less reliable to evaluate the errors caused by the internal uncertainties of the system. Because of the ambiguity in positioning a point target, the known position (the "ground truth" for a given point) might still be different from its "actual" position at the moment of measurement. In contrast, *accuracy* reflects internal causes of errors of the system excluding external influences, so it is more robust to evaluate the performance of a system.

Moreover, when parameters are estimated independently within the system itself, only relative positions of the reference points are known with respect to the defined co-ordinate. Thus more uncertainties are induced in the computation and the reliability of the estimated $\hat{\mathbf{a}}$ will degrade.

In stead of point correspondence, approaches using line- or plane-correspondence for calibration were proposed in [16] and [49]. Methods for self-calibration were also developed for use in mobile applications ([19], [44] and [62]), by which not only the intrinsic parameters but also the co-ordinates of the reference points are estimated, based on multiple images of the same field but acquired at different poses of the camera.

The calibration method proposed for the "frequency-difference" laser-scanning system is in line with previous methods in two aspects, i.e., the correspondence is not point-based and the constraints for correspondence are relaxed. Advantages of this method are:

(1) no a priori knowledge about the position of the calibration object is required. The pose of the calibration object (planes) is estimated simultaneously with the system parameters by a fitting algorithm. This implies that the error of calibration is evaluated in the sense of accuracy.
(2) the errors are evaluated in radian metric, making them independent of the range in triangulation measurement. This guarantees that the calibration fits the assumption of a stationary stochastic model of errors for LS optimisation.

The concept of self-calibration can be illustrated by the following formulation in case of point-correspondence. Suppose a point from an unknown plane is measured as $\mathbf{x}(\mathbf{a})$, where $\mathbf{a}$ denotes the system parameters, then the planarity constraint yields to

$$\mathbf{n}^T\mathbf{x}(\mathbf{a}) = b \tag{2.5}$$

Besides the system parameters, stored in the vector $\mathbf{a}$, also the parameters of the plane, stored in the vector $\mathbf{p}=(\mathbf{n},b)^T$, should be determined by the self-calibration computations. If the number of points in the test plane equals the dimensionality of $\mathbf{a}$ plus $\mathbf{p}$, and the plane is properly posed, then a unique solution of $\mathbf{a}$ and $\mathbf{p}$ is available from the set of non-linear equations.

However, a reliable result can only be obtained in the framework of optimal estimation by using more data points. This can be realised by minimising a defined cost function, as described below. Briefly, our proposed algorithm is to find the parameters of a plane that fits best to a set of measured planar points. To test the "goodness" of fitting, the distance of a point to the plane is used as the measure of error. Assuming the noise $\Delta\mathbf{x}$ to be independent identically distributed (i.i.d), then its distance to the plane $d=\Delta\mathbf{x}^T\mathbf{n}$, where $\mathbf{n}$ denotes the normal vector of the plane, has also the same distribution. For points from the same plane, the function

$$D(\mathbf{a},\mathbf{p}) = \sum_{i=1}^{m} d_i^2 \tag{2.6}$$

can be used to evaluate the variance of noise $\Delta\mathbf{x}$ in statistical sense.

Using a few planes in different poses as the calibration objects, the cost function can be further defined as

$$\Theta(\mathbf{a},\mathbf{p}_k) = \sum_{k=1}^{K} D_k \tag{2.7}$$

where K is the number of calibration planes. By minimising the cost function of (2.7), $\mathbf{a}$ and $\mathbf{p}_k$ (k=1…K) are estimated using the LS criterion.

Because the cost function is based on the effect of the error $\Delta\mathbf{x}$, the solution is obtained by evaluating errors in the sense of *accuracy*, rather than *repeatability*.

2.8.2 Establishing the least-squares criterion in the domain of sensor variables

When applying the algorithm described in section 2.8.2 to calibrate the "frequency-difference" laser scanning sensing system, the result was found unreasonable (in our experiments, the computed 3D co-ordinates of a point exceeded a normal range according to our empirical judgement). The problem was caused by the assumption that the errors measured in Cartesian co-ordinates were additive i.i.d noise in the image domain, which was not correct in reality. In fact, for the triangulation measurement, the range measurement in terms of r_{ij} (see [7]) is indexed by angles. For example, the indices (i, j) correspond to the latitude and longitude angles (α,γ) in the point-beam laser scanning system. After being converted to the *x-y-z* form, the errors represented in *x-y-z* co-ordinates usually do not preserve the properties in the original r_{ij} form. For example, if the error in the angles can be modelled as a stationary stochastic process, then after the non-linear transformation, errors of $\Delta\mathbf{x}$ distributed in *x-y-z* form become a non-stationary pattern. So the LSE criterion with the noise model in the *x-y-z* form can yield unreliable results.

Generally, such kind of error performance is shared by all triangulation measurement systems, in spite of alternatives in sensing patterns. In fact, this is one of the drawbacks that limit the application of triangulation systems for long-distance measurements. To eliminate the influence of the non-linear mapping from the input sensor variables to the output *x-y-z* representation, the measurements errors should be modelled in the domain of the sensor variables, rather than the *x-y-z* form.

The scanning angles are determined by measuring the scanning time. Errors in time recording are thought to fit a stationary stochastic model with constant variance. According to the "cursor-selected" data-sampling approach, the times t_f and t_l, corresponding with the instantaneous pose of the scanning laser beam, are pre-determined, so only the error of t_s in the position of the detecting plane is considered.

Given system parameters $\mathbf{a}$, the co-ordinates of a measured point can be expressed in the form

$$\mathbf{x} = \mathbf{X}(\mathbf{a}, t_l, t_f, t_s) \tag{2.8}$$

In a first-degree approximation, errors in the co-ordinates can be expressed as

$$\Delta\mathbf{x} = \frac{\partial\mathbf{X}}{\partial t_s}\Delta t_s \tag{2.9}$$

Suppose $\mathbf{x}$ is measured from a planar point. Then the error will be

$$\mathbf{n}^T\Delta\mathbf{x} = \mathbf{n}^T\left(\frac{\partial\mathbf{X}}{\partial t_s}\right)\Delta t_s \tag{2.10}$$

where **n** denotes the normal vector of the plane. With

$$w = \left[\mathbf{n}^T \left(\frac{\partial \mathbf{X}}{\partial t_s} \right) \right]^{-1} \quad \text{and} \quad d = \mathbf{n}^T \Delta \mathbf{x} ,$$

Δt_s can be written as

$$\Delta t_s = wd \tag{2.11}$$

Therefore, Δt_s can be interpreted as the "weighted" distance of d in the cost function of (2.10).

Given a set of planar points, the evaluation of the angle error is reformulated as

$$\tilde{D}(\mathbf{a}, \mathbf{p}) = \sum_i w_i^2 d_i^2 \tag{2.12}$$

Assuming the total number of the test plane objects is K, the corrected cost function for calibration is

$$\tilde{\Theta}(\mathbf{a}, \mathbf{p}_k) = \sum_{k=1}^{K} \tilde{D}_k \tag{2.13}$$

Minimising the cost function of (2.13) can be carried out with numerical methods.

This algorithm was applied to calibrate the "frequency-difference" laser-scanning ranging system. The result is shown in table 2.1. The point sets were sampled from a thin wooden plate (size 280×140 mm) at different poses within a range of 0.5~1m along the z-direction. Errors in angle and distance were estimated by the residue of $\Sigma_i(\Delta t_{s,i})^2$ and $\Sigma_i(d_i)^2$, respectively. The gradient-descending numerical method was applied to minimise the cost function of (2.13). This estimation result was found reasonable according to our empirical judgement.

Table 2.1: Result of the calibration of the "space-frequency-difference" laser-scanning system. Parameters were defined in section 2.8.1. Number of sampled points: 500.

α_0	β_0	γ_0	ρ_1	ρ_2	h	θ	L	$\overline{\Delta t_s^2}$ (s)	$\overline{d^2}$ (cm)
2.7615	0.9350	1.7870	0.0784	2.4890	0.2714	0.8767	65.818	3.3143×10^{-6}	1.0630

2.9 Conclusions

This chapter described the principle of the "frequency-difference" laser-scanning system. The measurement is based on a laser point-beam scanning triangulation algorithm. The way of data acquisition differs completely from that with a CCD camera. The sensing approach is effective to overcome the scattering problem in an underwater environment. Reducing

the "scattering volume" improves the signal-noise ratio. This system was designed for machine vision applications in underwater engineering tasks.

A significant element of the sensing technique of this system is the "frequency-difference" scanning principle. This technique offers automatic and high-speed sensing and moreover, has the potential of displaying range information in real-time on a 2D monitor. A real-time data-saving system specifically for this system is still being developed. Currently, the data points are acquired in a static "cursor-positioned" way.

To model and calibrate the measurement system, a system-specific co-ordinate system together with system parameters for 3D computation was defined. A new approach to planar fitting was proposed to estimate the system parameters. The parameters are determined by minimising the cost function and consequently, the measurement errors are evaluated in terms of *accuracy*.

Considering that the transformation from the sensor variables (t_f, t_l and t_s) to the x-y-z form is non-linear, we establish the LSE estimation criterion in the sensing domain (sensor variables) rather than in the image domain (Cartesian co-ordinates). Because the recording of the sensor variables can be treated as a stationary stochastic process, the errors represented in the sensing domain obey the i.i.d noise model closer than those represented in the image domain. Computations with real data confirmed this behaviour.

Chapter 3

Parameter Estimation for Surface Reconstruction Using a Quadric Representation

In a model-based vision system, the collected data and the object model are described by mathematical representations. The choice of such a representation is a critical issue in computer vision. The representations are not only used to calculate various properties of objects in the scene, but also to guide the strategies of recognition, including robustness and efficiency, as well as the segmentation processes.

Given a set of 3-D data obtained from an object surface, the observed shape of the object can be reconstructed by fitting the data to a model representation. This yields a set of model parameters. The surface representations of our interest are the quadric forms. For the quadric representation, we applied surface fitting based on the "bias-corrected" *renormalization* method ([68]), but improved the estimation using a novel approach to obtain a more reliable solution. More generally this approach provides a computational technique to refine the eigenvector solutions of polynomial fitting problems based on the *gradient weighted least-square* criterion.

Some issues concerning surface fitting from range data are discussed in this chapter, including the definition of the cost function for parameter estimation, a solution to the minimisation problem, the robustness of the estimator, and an alternative LSE criterion based on the sensor model.

In the beginning of this chapter, some approaches for 3D object representation are introduced in section 3.1. Focus is on surface primitives describing the class of man-made objects and their quadric representations.

In the first part of section 3.2, the "bias-corrected" renormalization method for surface fitting is introduced. The approach is derived from the popular gradient weighted least-squares criterion ([55], [58] and [68]), upon which the algebraic parameters of quadric fitting can be estimated with eigenvector solution in a re-weighting process ([58]). Since the gradient weighted least-squares estimates are statistically biased ([35] and [68]), the renormalization method is applied to compensate for the perturbation of the matrix and the bias in the iterative computations. In both the re-weighting eigenvector solution and the renormalization method the dependency of the gradient weight, to which a perturbation term is associated, is neglected. In case of low noise level, the perturbation term can be considered small enough to be neglected. When the noise has a significant level, ignorance will degrade the reliability of the cost function. Therefore, although the existing renormalization method yields "unbiased" estimates in the sense of statistics, an inadequate estimation i mplies inaccurate solutions. To correct this problem, we reformulate the eigenvector ap-

proach of the estimation by taking the dependency of the weight into account. In the same notion of "bias-corrected" renormalization, an improved algorithm is proposed to refine the fitting process in the second part of section 3.2.

The improvements of the new renormalization method are illustrated with experiments using synthetic data in section 3.3.

With respect to the robustness of parameter estimation, the method of *regression diagnostics* ([5]) was applied to eliminate outliers in data points. The feasibility of this conventional robust method relies on a good initial estimate, obtained from an eigenvector solution. Ending up in a local minimum of the cost function should be avoided; therefore, finding the global minimum is a critical point in the regression diagnostics approach. This is illustrated in section 3.4.

The criterion used for gradient weighted least-squares fitting (for the bias-corrected renormalization) is based on the assumption that the noise in the data points is additive independent identically distributed (i.i.d). Although such an assumption may not be accurate, it is acceptable because such characterisation offers a (usually good) approximation when an accurate model is not available. However, errors represented in an *x-y-z* form are transformed from the errors in the sensing process. In many cases, it is easier to fit the errors in the "raw" data set obtained in the sensing process (before the transformation) to a stationary stochastic model compared to those in *x-y-z* form (after the transformation). Therefore, a "sensor-driven" fitting algorithm is proposed in the subsequent section 3.5, particularly d e-signed for the data acquisition method used in the laser-scanning ranging system. Using the LSE fitting criterion, a cost function was defined at the level of inputs of the sensor, which fits better to the i.i.d model. Simulation results proved the improvement of the "sensor-driven" algorithm compared with the fitting algorithm in an *x-y-z* form.

Finally, some remarks and conclusions on the new surface fitting approach are given in section 3.6

3.1 Surface representations of 3D objects

Shape descriptions of 3-D objects in range images can be roughly divided into surface-based, volume-based and discontinuity-based descriptions ([2]). Surface-based representations describe the surface geometry and topology; discontinuity-based representations preserve information about the surface changes and volumetric representations refer to the volumes of the objects. The discontinuity-based scheme usually results in 3D curves representing abrupt changes, so only few memory is required. The disadvantage is that in many cases discontinuity information only is not sufficient to describe the curved objects. The volumetric representations, typically superquadrics and generalised cylinders, provide an efficient shape description for surface-bounded objects, with only a few more parameters compared to quadrics. However, they exhibit a problem of non-uniqueness, so this class of representations has been used often for CAD modelling rather than for recognition tasks ([2]). Based on application-specific considerations, we have chosen for surface-based quadric representations for the description of object primitives in this thesis, since our interests are restricted to man-made objects.

Various surface representations have been used for graphics and vision tasks. In general, surfaces can be expressed in an implicit or explicit form. A review of such methods is given in [10]. Briefly, an explicit form of a 3D surface description is a function graph of two independent variables. Most explicit expressions aim at "visible" surface reconstruction with smoothness constraints and a class of spline functions is employed for an optimal approximation. The implicit forms of surface reconstruction focus on the extraction of global properties of these surfaces by parameter estimation. In case of surface representation for recognition tasks, often implicit forms are chosen because of their mathematical tractability.

The implicit expression of a surface can be defined as $f(x,y,z) = 0$. It is noted that given an implicit function, a volume can be defined as $f(x,y,z) \leq 0$. As a matter of fact, according to [10], in that case the distinction between volumetric and surface representation is only semantic. Among the many implicit expressions for surface descriptions, frequently used primitives are super-quadrics, generalised cylinders and implicit polynomials. Super-quadrics and generalised cylinders are also typical for volumetric representations.

Low degree polynomials, i.e., degree 1 or 2, have been very popular to approximate a large scope of manufactured parts, such as planes, cylinders, spheres or quadrics of revolution ([11], [23] and [30]). In summary, implicit polynomial representations require only few parameters for characterising object surfaces; further advantages are low computational cost, insensitivity to noise or to moderate changes in the subset of data and convenient translation of operations in "object space" into natural operations in polynomials ([39]).

Many techniques applying high-degree polynomials to represent more complex objects have been developed in the past decade ([39], [56] and [58]). Particularly, fourth degree polynomials are proposed, being capable to approximate a variety of complex shapes, including super-quadrics. In high-degree polynomial representations, bounded zero sets are emphasised, based on the argument that all data points originate from a bounded object surface ([39] and [56]). Such bounded polynomials in fact yield volumetric representations. However, a man-made object is bounded by a set of simply curved or planar surface patches, so it is easier to characterise them by a surface-specific description rather than by an object-specific description. Surface-based representations with low degree polynomials have some advantages, for the following:

- Representations with more parameters tend to lack uniqueness when dealing with noisy data because there are more degrees of freedom in parameter space to fit the data set. Especially when data points originate from a low-degree parameterised surface patch, higher-degree representations are inherently ill-posed.
- Volumetric representations are sensitive to occlusion ([2] and [63]). On the contrary, low degree polynomials are suited to describe the surface-based data set, assuming the surface of interest is simple-curved. Since they only deal with a "patch" of the bounded object surface, they are not sensitive to occlusion.
- Although the surface-based quadric representations usually only generate "partial" d e-scriptions of a bounded object, mathematical manipulation is easier because of the fewer number of parameters. Therefore, in the top-down recognition scheme, surface features represented at an intermediate level could be a good trade-off between the complexity of data-driven representation and the complexity of model-driven control.

For man-made object recognition from range data, we opt for the surface-based representation in quadric form to describe 3D surfaces. Without loss of generality, an implicit quadric form can be expressed as:

$$f(x,y,z) = a_{11}x^2 + a_{22}y^2 + a_{33}z^2 + a_{12}xy + a_{13}xz + a_{23}yz + v_1x + v_2y + v_3z + k = 0 \qquad (3.1)$$

In addition, the attractivity of the quadric representation emerges also from the fact that any smooth, continuous curve or surface can be locally approximated by first and second degree polynomials. Thus, the quadric approximation is controlled by the definition of local regions.

3.2 Quadric fitting by an improved bias-corrected renormalization method

The parameters of a surface representation are estimated during a fitting process. In general, given a mathematical object representation, parameter estimation faces the following optimisation aspects ([68]):

- definition of a cost function to minimise or maximise (criterion);
- the method to optimise the chosen function (estimation);
- the implementation of the solution.

Quadric (or conic in 2D) fitting is a popular topic in parameter estimation. A tutorial review on conic fitting methods was given in [68].

Given a set of points $\{\mathbf{x}_i\} = \{(x_i, y_i, z_i)\}(i = 1,..,n)$, the coefficients in (3.1) are estimated by minimising a defined cost function. Many alternative fitting methods have been derived from the general least-squares criterion, and the cost functions can be defined in the sense of a distance measure. In this work, the so-called "bias-corrected" *renormalization* method ([68]), which is based on the gradient weighted least-squares optimisation, is applied for quadric fitting. It has the following features:

(1) Optimisation of unbiased estimation provides a possibility to describe the uncertainty of parameters within a statistical model, which is of the major objectives of the work of this thesis.
(2) The variance of noise is recovered simultaneously, which is necessary to evaluate the estimate uncertainty.
(3) The non-linear problem can be solved by an iterative eigenvector method, so finding the global minimum of the cost function is guaranteed. Moreover, a good initial estimate is available so it is possible to early detect outliers using the regression diagnostics ([5]).
(4) Finally, the solution is further optimised, so more reliable estimate can be expected.

In the following sections we discuss successively the gradient weighted least-squares fitting criterion, the algorithms of the bias-corrected renormalization and the improved optimisation method.

3.2.1 Eigenvector approach for gradient weighted least-squares fitting

The bias-corrected renormalization method for surface fitting is based on the gradient weighted least-squares criteria. In ordinary least-squares fitting, the squares of the algebraic distances $\Sigma_i f^2(\mathbf{x}_i)$ are minimised, where $f(\cdot) = 0$ is the implicit quadric form of (3.1). In the gradient weighted least-squares fitting, the squares of the Euclidean distances $\Sigma_i (d_i)^2$ are minimised, where d_i denotes the orthogonal distance between a point $\mathbf{x}_i$ and the surface. Since the expression of d_i is very complicated ([68]), the distance measure can be approximated as describe below ([10]).

Suppose a point $\mathbf{x}$ is measured from the surface $f(\mathbf{x}_0) = 0$, where $\mathbf{x}_0$ denotes the true value. Then, the first-order Taylor expansion is:

$$f(\mathbf{x}) \approx f(\mathbf{x}_0) + \nabla f(\mathbf{x}_0) \cdot \Delta\mathbf{x} ; \tag{3.2}$$

where $\Delta\mathbf{x} = \mathbf{x} - \mathbf{x}_0$ accounts for the noise. The gradient of $f(\mathbf{x}_0)$ can be expressed as:

$$\nabla f(\mathbf{x}_0) = \left\|\nabla f(\mathbf{x}_0)\right\| \mathbf{n}_0 ;$$

where $\mathbf{n}_0$ denotes the unit normal vector at point $\mathbf{x}_0$. Since $f(\mathbf{x}_0) = 0$ and $\mathbf{x}$ is close to $\mathbf{x}_0$, we have:

$$f(\mathbf{x}) \approx \left\|\nabla f(\mathbf{x})\right\| \mathbf{n}_0 \cdot \Delta\mathbf{x} .$$

Therefore, the normal projective distance of $\mathbf{x}$ to the surface (orthogonal distance) can be approximated by

$$d = \mathbf{n}_0 \cdot \Delta\mathbf{x} \approx \frac{f(\mathbf{x})}{\left\|\nabla f(\mathbf{x})\right\|} \tag{3.3}$$

Defining a "weight" w:

$$w = \left\|\nabla f(\mathbf{x})\right\|^{-2} ,$$

the cost function for the gradient weighted least-squares fitting is defined as:

$$\Theta = \sum_{i=1}^{n} w_i f^2(\mathbf{x}_i) \tag{3.4}$$

Orthogonal distances are invariant to Euclidean transformations and it can be proved that the variable d in (3.3) has a constant variance over the surface in first order approximation (assuming that the errors in range data are i.i.d Gaussian noises) ([68]).

Therefore the gradient weighted least-squares fitting is reliable and accepted as an optimal solution of parameter estimation in many applications, including high-degree polynomial representations. Moreover, in many cases the noise term $\Delta\mathbf{x}$ can be treated as being i.i.d Gaussian type, so the distance d is also a Gaussian distributed variable. Therefore the solution by minimising the cost function of (3.4) is optimal in the sense of the Maximum-Likelihood Estimate (MLE).

Minimising (3.4) is in general a non-linear problem and some well-established non-

linear minimisation techniques can be applied. However, for quadric (more generally, polynomials) representations of the function $f(\mathbf{x})$, the eigenvector approach is applicable to obtain an approximated solution ([10], [56] and [58]). This is explained next.

With the expression (3.1), and defining the parameter vector and the measurement vector

$$\mathbf{p} = (a_{11}, a_{22}, a_{33}, a_{12}, a_{13}, a_{23}, v_1, v_2, v_3, k)^T \ , \ \mathbf{M} = (x^2, y^2, z^2, xy, xz, yx, x, y, z, 1)^T \ ,$$

respectively, the cost function (3.4) can be formulated as

$$\Theta = \mathbf{p}\left(\sum_{i=1}^{n} w_i \mathbf{M}_i \mathbf{M}_i^T \right) \mathbf{p} \tag{3.5}$$

Now, let

$$\mathbf{N}_i = \mathbf{M}_i \mathbf{M}_i^T \quad \text{and} \quad \mathbf{N} = \sum_{i=1}^{n} w_i \mathbf{N}_i \ .$$

We have

$$\Theta = \mathbf{p}^T \mathbf{N} \mathbf{p} \ . \tag{3.6}$$

In general, minimisation of the function (3.6) can be solved by the *Lagrange Multiplier* method. Holding the constraint $\|\mathbf{p}\|_2 = 1$, the problem can be transformed into an unconstraint minimisation problem by defining an augmented cost function

$$\Theta = \mathbf{p}^T \mathbf{N} \mathbf{p} - L(\mathbf{p}^T \mathbf{p} - 1) \ , \tag{3.7}$$

where L is the *Lagrange Multiplier*. Minimising the function (3.7) gives

$$\begin{cases} \dfrac{\partial \Theta}{\partial \mathbf{p}} = 0 \\ \dfrac{\partial \Theta}{\partial L} = 0 \end{cases} \tag{3.8}$$

Substituting (3.6) into (3.8) and ignoring the dependency of w_i on $\mathbf{p}$, we have

$$\begin{cases} \mathbf{N}\mathbf{p} = L\mathbf{p} \\ \mathbf{p}^T \mathbf{p} = 1 \end{cases} \tag{3.9}$$

Therefore, the solution of minimising (3.6) is just the eigenvector of $\mathbf{N}$ associated to the smallest eigenvalue L. Because in fact w_i depends on $\mathbf{p}$, the solution is obtained in an iterative re-weighting way: initially, set $w_i = 1 \forall i$ and compute the eigenvector of $\mathbf{N}$ associated to the smallest eigenvalue as the initial estimate of $\mathbf{p}$. At each iteration step, update w_i by the current value of $\mathbf{p}$ and compute the eigenvector of the updated $\mathbf{N}$ to get a new estimate. Finally, a solution is obtained when the iteration shows sufficient convergence.

The advantage of the eigenvector method is that the global minimum is reached automatically without a combinatorial global search. Also, it gives a good initial estimate without any a priori knowledge about the "shape" of the data set, enabling a subsequent outlier test.

3.2.2 The bias-corrected renormalization technique

It has been proved that the estimate from the gradient weighted least-squares fitting is statistically *biased* ([35] and [68]). The bias-corrected renormalization technique was presented earlier by Kanatani ([35]). A simple version of the bias-corrected technique was given in [68]. This technique will be explained next.

Suppose each range point is perturbed by noise $\Delta\mathbf{x}_i$, with $E\{\Delta\mathbf{x}_i\}=0$ and a covariance matrix $Cov\{\Delta\mathbf{x}_i\}=\sigma^2\mathbf{I}\ \forall i$ ($\mathbf{I}$ denotes the unit matrix). Then matrix $\mathbf{N}$ is perturbed too with respect to the ideal matrix $\mathbf{N}_0$. The perturbation matrix $\Delta\mathbf{N} = \mathbf{N}-\mathbf{N}_0$ can be found from the noise terms $\Delta\mathbf{x}_i\forall i$ associated to each element of $\mathbf{N}$, by applying Taylor expansion. If the terms of order higher than 2 are ignored, the expectation of $\Delta\mathbf{N}$ equals:

$$E\{\Delta\mathbf{N}\} = \sum_{i=1}^{n} w_i E\{\Delta\mathbf{N}_i\} = \sigma^2 \sum_{i=1}^{n} w_i \mathbf{B}_i \tag{3.10}$$

where $\mathbf{B}_i$ is a matrix consisting of the coefficients of the second-order expansion. The details of the derivation of $\mathbf{B}_i$ are presented in Appendix B.

Let $\mathbf{p}$ and $\mathbf{p}_0$ be the eigenvectors of $\mathbf{N}$ and $\mathbf{N}_0$, respectively. According to the *perturbation* theory, the perturbation of the eigenvector $\Delta\mathbf{p}=\mathbf{p}-\mathbf{p}_0$ is linear to the perturbation of matrix, i.e., $E\{\Delta\mathbf{p}\}=O(E\{\Delta\mathbf{N}\})$. Since $\mathbf{B}_i \neq 0\forall i$ (see Appendix B), and according to (3.10), we have $E\{\mathbf{N}\}\neq 0$, hence $\mathbf{p}$ is biased.

In order to correct the bias of $\mathbf{p}$, the bias of matrix $\mathbf{N}$ is corrected by defining

$$\overline{\mathbf{N}} = \mathbf{N} - E\{\Delta\mathbf{N}\} = \mathbf{N} - \sigma^2 \sum_{i=1}^{n} w_i \mathbf{B}_i \ . \tag{3.11}$$

So $E\{\Delta\mathbf{N}\} = \mathbf{0}$. By replacing $\mathbf{N}$ with $\overline{\mathbf{N}}$, the eigenvector associated with the smallest eigenvalue is an *unbiased* estimate of the ideal parameter $\mathbf{p}_0$.

Because the variance of the noise is usually unknown, it is estimated in a so-called *renormalization* process, as explained below.

At each step, the constraint of $\mathbf{p}^T\overline{\mathbf{N}}\mathbf{p}$ holds, which argument is based on the fact that $E\{\overline{\mathbf{N}}\} = \mathbf{N}_0$ and $\mathbf{p}_0^T\mathbf{N}_0\mathbf{p}_0 = 0$. If the current smallest eigenvalue $\lambda_{\min}\neq 0$, the value of $c\equiv\sigma^2$ is updated by Δc to satisfy the above constraint. From the expression of $\overline{\mathbf{N}}$, it can be shown that ([68])

$$\Delta c = \frac{\lambda_{\min}}{\mathbf{p}^T\left(\sum_{i=1}^{n} w_i \mathbf{B}_i\right)\mathbf{p}} \ .$$

Setting the initial values $c=0$ and $w_i=1\forall i$, the initial estimate $\mathbf{p}^{(0)}$ is obtained. Then by iterative updating c and w_i in the renormalization process, a solution is obtained if the process converges.

3.2.3 Improvement of the bias-corrected renormalization solution

Although the eigenvector method described in the previous section is powerful in solving

the optimisation problem, the solution is sub-optimal because the dependence of the "weight" w_i on $\mathbf{p}$ in the formation of (3.9) has been ignored. This is illustrated by deriving the complete formulation of the optimisation problem, which takes into account the dependency of the weight w_i on the parameters $\mathbf{p}$.

We apply the *Lagrange Multiplier* method to solve the minimisation problem of (3.6) and take the dependency of w_i on $\mathbf{p}$ into account. From (3.7) and (3.8) we find

$$\begin{cases} \mathbf{N}\mathbf{p} + \frac{1}{2}\sum_{i=1}^{n}\left[(\mathbf{p}^T \mathbf{N}_i \mathbf{p})\left(\frac{\partial w_i}{\partial \mathbf{p}} \right) \right] = L\mathbf{p} \\ \mathbf{p}^T \mathbf{p} = 1 \end{cases} . \tag{3.12}$$

Thus the parameters $\mathbf{p}$ and L should be solved from (3.12).

Clearly, if the second term in the LHS of (3.12) associated to the dependence of w_i on $\mathbf{p}$ is ignored, then (3.12) degenerates to (3.9). However, an exact solution of the optimal minimisation problem of (3.6) should satisfy the equation (3.12), in which the terms associated to $(\partial w_i/\partial \mathbf{p})$ explicitly affect the solution of $\mathbf{p}$, rather than (3.9).

In case of noise absence or low-level noise perturbation in the data set, the quantities of $(\mathbf{p}^T\mathbf{N}_i\mathbf{p})$ mounted to $(\partial w_i/\partial \mathbf{p})\forall i$ are zero or small enough, so difference between solutions of (3.12) and (3.9) can be ignored. In case of a significant noise level, the solution of (3.9) could be unreliable. Since the described bias-corrected renormalization algorithm is based on the formulation of (3.9) but using a "renormalised" matrix $\overline{\mathbf{N}}$, the bias-corrected estimate is also unreliable. However, the bias-corrected renormalisation algorithm can be improved based on the formulation of (3.12).

As we will show, (3.12) can also be converted to an eigenvector problem in the same form as (3.9), but with the matrix $\mathbf{N}$ reformulated by taking into account the term of $(\partial w_i/\partial \mathbf{p})$.

First, we reformulate the second term of the LHS of (3.12) in the form

$$\frac{1}{2}\sum_{i=1}^{n}\left[(\mathbf{p}^T \mathbf{N}_i \mathbf{p})\left(\frac{\partial w_i}{\partial \mathbf{p}} \right) \right] = \frac{1}{2}\sum_{i=1}^{n}\left[\left(\frac{\partial w_i}{\partial \mathbf{p}} \right)\mathbf{p}^T \mathbf{N}_i \right]\mathbf{p} .$$

Suppose the dimensionality of $\mathbf{p}$ is m, and denote the $m \times m$ matrix

$$\mathbf{E} \equiv \frac{1}{2}\sum_{i=1}^{n}\left[\left(\frac{\partial w_i}{\partial \mathbf{p}} \right)\mathbf{p}^T \mathbf{N}_i \right] . \tag{3.13}$$

Then (3.12) can be expressed in the form

$$(\mathbf{N} + \mathbf{E})\mathbf{p} = L\mathbf{p} , \tag{3.14}$$

We define a matrix

$$\widetilde{\mathbf{N}} = \mathbf{N} + \mathbf{E} ,$$

the solution of $\mathbf{p}$ can still be achieved by the re-weighting eigenvector approach as used for (3.9). The matrix $\mathbf{E}$ should also be updated iteratively.

- **Computing the eigenvector of $\widetilde{\mathbf{N}}$**

There exists a difficulty in direct computation of the eigenvector of $\widetilde{\mathbf{N}}$: unlike the matrix **N**, **E** is usually non-symmetric. It can not be guaranteed that the eigenvector is always *real* at each step. To get rid of this problem, we propose an approach to obtain the solution in another way.

Since the matrix **E** is small compared to **N**, it can be treated as a perturbation of **N**. According to perturbation theory, the perturbation of the eigenvector is linearly related to the perturbation of matrix **N**. Let $\mathbf{p}$ and $\widetilde{\mathbf{p}}$ be the eigenvectors of **N** and $\widetilde{\mathbf{N}}$, respectively. Then we have ([53])

$$\Delta\mathbf{p} = \widetilde{\mathbf{p}} - \mathbf{p} = \mathbf{U}\left(\lambda\mathbf{I} - \mathbf{U}^T\mathbf{N}\mathbf{U}\right)^{-1}\mathbf{U}^T\mathbf{E}\mathbf{p} \tag{3.15}$$

Here, **U** is a $m\times(m\text{-}1)$ matrix whose columns are the other m-1 eigenvectors of **N** except for $\mathbf{p}$, and λ is the eigenvalue associated with $\mathbf{p}$.

Therefore, the eigenvector of $\widetilde{\mathbf{N}}$ can be computed as $\widetilde{\mathbf{p}} = \mathbf{p} + \Delta\mathbf{p}$ at each iterative step. To obtain the exact minimiser of (3.6), the solution of (3.9), as used either for gradient weighted least-squares criterion or for bias-corrected renormalization algorithm, should be corrected by $\Delta\mathbf{p}$ at each iterative step. The solution of (3.14) is then achieved when both $\mathbf{p}$ and $\widetilde{\mathbf{p}}$ converge.

Such an improvement of the eigenvector solution can be directly applied for the bias-corrected estimation. In the same notion of [68], in order to get an unbiased estimate, $\widetilde{\mathbf{N}}$ should be further replaced by

$\overline{\widetilde{\mathbf{N}}} = \widetilde{\mathbf{N}} - E\{\Delta\widetilde{\mathbf{N}}\}$.

In the above expression, $E\{\Delta\widetilde{\mathbf{N}}\}$ can be written as

$$E\{\Delta\widetilde{\mathbf{N}}\} = c\sum_{i=1}^{n} w_i\mathbf{B}_i + c\left\{\tfrac{1}{2}\sum_{i=1}^{n}\left[\left(\frac{\partial w_i}{\partial\mathbf{p}}\right)\mathbf{p}^T\mathbf{B}_i\right]\right\}, \tag{3.16}$$

and we have

$$\overline{\widetilde{\mathbf{N}}} = \overline{\mathbf{N}} + \frac{1}{2}\sum_{i=1}^{n}\left[\left(\frac{\partial w_i}{\partial\mathbf{p}}\right)\mathbf{p}^T\left(\mathbf{N}_i - c\mathbf{B}_i\right)\right]. \tag{3.17}$$

Compared with (3.14), we define a new perturbation matrix

$$\overline{\mathbf{E}} = \frac{1}{2}\sum_{i=1}^{n}\left[\left(\frac{\partial w_i}{\partial\mathbf{p}}\right)\mathbf{p}^T\left(\mathbf{N}_i - c\mathbf{B}_i\right)\right]. \tag{3.18}$$

The eigenvector solution can be corrected in the same way as for (3.14), with the substitutions $\overline{\mathbf{N}} \to \mathbf{N}$ and $\overline{\mathbf{E}} \to \mathbf{E}$.

In this way, the implementation of the "bias-corrected renormalization" algorithm is improved by the correction in formulating the cost function of (3.6). While the estimate is unbiased in the sense of statistics, the solution is improved in reliability because of the refinement in the estimation.

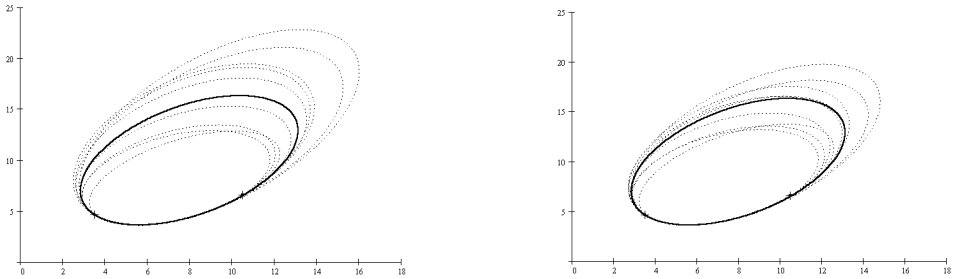
3.3 Verification of the improvement with synthetic data

To show the validity of the proposed procedure, experiments of conic fitting with 2D synthetic data have been carried out.

In these experiments, 2D data points were sampled from a number of synthetic ellipses and Gaussian noise was added to the x and y co-ordinates randomly for each point. The number of points in each set is 80. Given a set of sampled points, both the existing renormalization method as used in [68] and the improved approach proposed in this chapter were applied to obtain a comparison. The results are depicted graphically in figure 3.1, figure 3.2 and figure 3.3. Since the algebraic coefficients describe the conic shape globally, the closed curves of the ellipses with the estimated parameters are plotted to give an insight of the accuracy of the estimates.

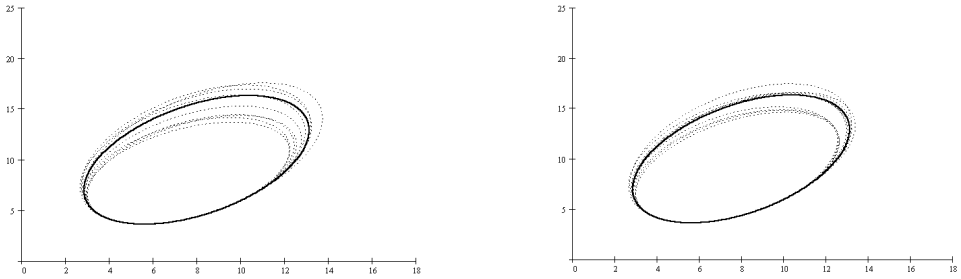
In figure 3.1, the standard deviation of added noise is σ=0.15, which is the same as the unit of the x-y co-ordinate. Graph (a) shows the results using the renormalization method while graph (b) shows the results from the improved approach. The curves in (a) and (b) were plotted pair-wise, i.e., given a set of data points, the fitting results with the two methods were plotted both in (a) and (b), respectively. The primitive used in figure 3.2 is the same as in figure 3.1, but the noise level is σ=0.08. Figure 3.3 shows the results with a different primitive at a noise level of σ=0.15.

From the results we can see that the fittings in the range occupied by the sampled points were good and almost indistinguishable for both methods, but there is an obvious difference between the global descriptions. Even larger variations occurred in the original renormalization method, especially in case of high noise level. This demonstrates clearly the improvement of the new approach.



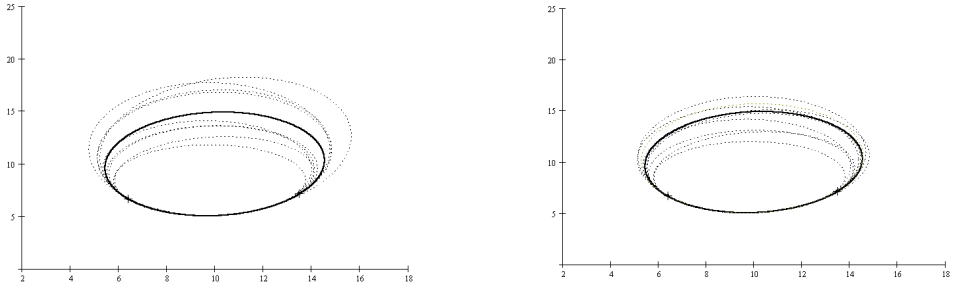
(a) (b)

Fig 3.1 Fitting results with 2D synthetic data at a noise level σ=0.15. For each set of data points, the result using the original renormalization method was plotted in (a), while the result using the new method was plotted in (b). The elliptical primitive for (a) and (b) is shown in solid lines and the estimates are plotted in dotted lines. Data points were samples from the shorter segment marked with the symbol "+".



(a) (b)

Fig 3.2 Results at a noise level of σ=0.08. Data points were generated from the same primitive within the same part as used in figure 3.1. The estimates by the original renormalisation method and the new method are shown in (a) and (b), respectively.



(a) (b)

Fig 3.3 Fitting results at a noise level of σ=0.15. Data points were generated from a different primitive. The estimates by the original renormalisation method and the new method are shown in (a) and (b), respectively. The sampled region is also marked on the curve of the primitive.

It should be noted that the improved method not only yield a higher accuracy in parameter estimation but also performs a better convergence capability. Of all trials at a noise level of σ=0.15 (the number of trials is 100), 8% failed to converge in the original renormalization method, whereas the improved approach results in convergence for all trials. It can be concluded that the term $(\partial w_i/\partial \mathbf{p})$ is associated with a stabilisation function in solving an (ill-posed) non-linear problem.

3.4 Eliminating outliers

Least-squares estimators are vulnerable to outliers ([68]). To get a reliable estimate, robust approaches for parameter estimation have to be conducted. In this thesis, we applied the technique of regression diagnostics ([5]) to detect outliers. This is carried out with the following steps:

(1) Determine an initial estimate by fitting the total data set to a quadric representation and compute the distance between each point and the fitted surface and reject all data points whose values exceed a predetermined threshold.
(2) With the accepted data points, complete the fitting process using the approach as described in section 3.2 (the improved renormalization method).
(3) Re-compute the distances to the estimated surface for all points and reject those having values larger than the threshold.
(4) If there are no updates, output the result of step (2) as the final estimate of the surface fitting, otherwise go back to step 2.

In practice, the threshold for outlier rejection is set in the form $Treshold=r\cdot (d^2)_{med}$, where r is a constant and $(d^2)_{med}$ is the median of square distances between points to the fitting surface.

The initial estimate is critical in applying the regression diagnostics method. Because the initial estimate is given by the eigenvector solution, it is plausible to find a suitable initial fit. Moreover, the standard deviation of the noise is estimated simultaneously with the bias-corrected renormalization algorithm, so the threshold for distance comparison can be reasonably adjusted during the regressive diagnostic process.

3.5 Fitting sensory data based on characterisation of sensing errors

For surface reconstruction problems from range data, the assumption that range data are corrupted by i.i.d noise (either in the three directions of 3D points or only in the depth direction) is valid in most cases. However, when knowledge about the error propagation in data representation is available, the estimation of the parameters can be improved further by optimising the fitting criteria based on the characterisation of errors in the sensing process.

In general, the stochastic behaviour of sensing events, preceding the co-ordinate representation, dictates the pattern of errors occurred in 3D point representations. Due to the non-linear transformation from the input sensor variable to the output formulation in Cartesian co-ordinates, the i.i.d noise modelled at the input variables results in a more complicated pattern of error presence in the *x-y-z* form. In this case, a noise model characterised at the low level of a sensing event could yield more reliable optimal criteria for LSE than the assumption of i.i.d noise in the higher-level *x-y-z* form.

For the range data acquired by the "frequency-difference" laser scanning sensing system, it is possible to reformulate the optimal estimation problem based on characterisation of sensing errors hence surface fitting can be refined. This is explained next.

With the laser-scanning triangulation method, a range measurement is determined by the scanning angles of α, β and γ. According to the "cursor-positioned" data-sampling mechanism, α and γ are pre-determined (by setting the time values), so the distance error is represented by the error in β.

There are many error sources in the sensing process. For example, the finite width of the receiving slit when imaged by the lens forms a narrow sector-shaped sensing space. Instability of the synchronising signal leads to uncertainty in triggering the time counting. The expansion of the laser spot due to laser divergence forms another error source. In general, errors as represented by $\Delta\beta$ can be treated as a stationary stochastic process, and $\Delta\beta$ is further modelled as additive i.i.d noise (assuming the scanning frequencies are stable). Therefore, after non-linear transformation, the errors represented in the *x-y-z* form do not preserve the statistics of i.i.d noise. To get more reliable estimates from the LSE criterion, the cost function for surface fitting should be established on the measure of error $\Delta\beta$, in stead of the distance error$\Delta\mathbf{x}$.

The fitting routine based on modelling noise at the level of sensor input variables can be interpreted as a "sensor-driven" approach. To distinguish, the routine based on modelling noise represented in the *x-y-z* form, i.e., the output of the sensor, is called the "image-driven" approach hereafter.

Given the system parameters, the co-ordinates of a measured point can be expressed in the form $\mathbf{x}=\mathbf{X}(\alpha, \beta, \gamma)$, In first-degree approximation, errors in the co-ordinates can be expressed as

$$\Delta\mathbf{x} = \frac{\partial \mathbf{X}}{\partial \beta}\Delta\beta \ .$$

According to the gradient weighted least-squares fitting criteria, the distance of point $\hat{\mathbf{x}}$ to the surface is approximated with

$$d = \mathbf{n}^T \Delta\mathbf{x} = \mathbf{n}^T \left(\frac{\partial \mathbf{X}}{\partial \beta} \right) \Delta\beta \ ; \tag{3.19}$$

where $\mathbf{n}$ denotes the normal vector of the surface.
Defining

$$r = \left[\mathbf{n}^T \left(\frac{\partial \mathbf{X}}{\partial \beta} \right) \right]^{-1} ,$$

we have

$$\Delta\beta \approx rd \ . \tag{3.20}$$

Consequently $\Delta\beta$ can be interpreted as the "weighted" distance of d in expression of (3.3). Combining with (3.3), the cost function of (3.4) can be re-written as

$$\overline{\Theta} = \sum_{i=1}^{n} r_i^2 d_i^2 = \sum_{i=1}^{n} r_i^2 w_i f^2(\hat{\mathbf{x}}_i) = \sum_{i=1}^{n} \overline{w}_i f^2(\hat{\mathbf{x}}_i) \ . \tag{3.21}$$

Here $\overline{w}_i = r_i^2 w_i$. Now the optimisation is achieved by minimising the cost function (3.21).

It can be seen that the described "bias-corrected" renormalization algorithm and the improved approach can be applied straightforwardly to the minimisation problem of (3.21), just by substituting the "weight" w_i with $\overline{w}_i$. The quantity of $\mathbf{B}_i$ in (3.10) should be reformulated by the second order expansion of $\Delta\beta$ and the noise level of c is interpreted as the square of the standard deviation of $\Delta\beta$.

To compare the sensor-driven method and the image-driven method used in section 3.2, synthetic data were used to obtain the fitting results. The data points were generated from a synthetic spherical surface by simulating the "frequency-difference" laser-scanning sensing process. Noise was added to the variable β recorded for each point. Table 3.1 shows some results for different noise levels.

For comparison, we calculate the distance between an estimate and the true parameters in the 10 dimensional parameter space as the measure of "goodness". From these results, we can see that both approaches had little differences at low level noise perturbation of β. At large noise levels, the sensor-driven method yields better solutions than the image-driven method, while its computation was more time-consuming.

From this comparison with simulated data, the image-driven way seems to perform well at moderate noise perturbation, even when the noise was nonlinearly transformed into *x-y-z* co-ordinates. Since it is relatively easy to implement, it is still recommended as a recommendable approach in solving quadric fitting problems. However, when the noise significantly influences the accuracy of the fitting solution, the sensor-driven method should be considered to refine the estimate.

Table 3.1 Comparison of fitting results with synthetic data using the image-driven and the sensor-driven method. 400 points were sampled from a spherical surface patch in the simulated laser-scanning process. Noise is added to angle β (quantity of radian) for each point. In each row the second line with symbol "*" is the result of the sensor-driven method.

Noise level	Fitting results (coefficients in (3.1)) ($\times 10^{-6}$)	Iteration number
σ=0.0002	(136, 138, 140, 0, 0, -6, -8155, 389, -22349, 999717)	5
	* (136, 138, 140, 0, -1, -5, -8152, 311, -22341, 999717)	5
σ=0.0005	(132, 137, 143, 0, -1, -15, -7880, 1012, -22560, 999714)	7
	* (133, 137, 142, 1, -1, -12, -7890, 781, -22527, 999715)	16
σ=0.001	(124, 135, 147, 0, -1, -33, -7367, 2259, -22972, 999706)	8
	* (127, 135, 145, 2, -2, -4, -7468, 1599, -22837, 999710)	20
True values	(139, 139, 139, 0, 0, 0, -8331, 0, -22216, 999718)	

It should be pointed out that the simulated results were obtained using the improved algorithm. When trying the existing renormalization approach, the image-driven method more

often failed in convergence than the sensor-driven method at high noise level. This also proves the merit of the sensor-driven approach.

3.6 Conclusions

This chapter discussed parameter estimation in surface fitting. Based on the "bias-corrected" renormalization algorithm, we proposed a new approach to improve the accuracy of solutions in the renormalisation process. Experiments on 2D synthetic data revealed a significant improvement in performance, compared with the original renormalization method.

The fitting criterion in the sense of LSE is applied to minimise the variation of noise in the measurements. Without any a priori knowledge about data formation in *x*-*y*-*z* coordinates, usually it is assumed that the i.i.d noise is added to each point, either in three dimensions or only in depth direction. However, once the error of measurement in the sensing process is characterised, the fitting criterion can be optimised on the level of sensor input variables and the estimation can be refined. With respect to the "frequency-difference" laser scanning sensing system described in chapter 2, this chapter contributes to solving the fitting problem alternatively, using a "sensor-driven" approach. Compared with the "image-driven" approach, the "sensor-driven" approach yielded better estimates in simulation experiments.

It should be pointed out that although all discussions are limited to the quadric representation, the algorithms described in this chapter can be extended to representations of higher order polynomials.

Finally, it should be noted that surface reconstruction could be an ill-posed problem, especially when the inputs contain inadequate information to determine the dimensionality of the representations. For example, when a surface patch is too "flat" or when the noise level is high, the surface fitting routine might fail to converge. Although some other non-linear minimisation techniques could help to find the global minimum, this chapter does not present a systematic investigation of solving ill-posed problems in surface fitting.

Chapter 4

Modelling of Uncertainties in Invariants and Pose Estimation Based on Quadric Representations

The goal of a model-based vision system is to identify and to localise objects of interest in a scene, which is generally achieved through interpretation of the collected data using the a priori knowledge in a model base. Following the module of object representation with the range data, the matching between models and the scene is implemented on the basis of features correspondence. To approach the feature descriptions, invariants, which are extracted from object representations, have gained in importance and are widely used for model-based recognition. Since surface-based primitive representation is the focus of this thesis, the discussion of feature description and classification will be limited to quadric primitives in this chapter.

In general, invariants are object-centred shape descriptors, which do not change under some kinds of transformations, especially, to make them viewpoint independent. The use of invariants enables us to perform the correspondence by direct comparison between the scene descriptions and models without external factors. Based on the quadric representations of object surfaces, which have been discussed in chapter 3 as the parametric descriptions for the given range data, this chapter first discusses the invariants descriptions of quadric primitives for surface-based recognition, incorporated with recovery of pose transformation. Then the uncertainty in invariants extraction, which is of our main interests of this chapter, is investigated and a model to describe the uncertainties is proposed.

One of the key issues in object recognition is to cope with the inherent uncertainty problem in measurements. Because of noise presence in data acquisition, errors in parameter estimation of the object representation are inevitable, causing uncertainties in the computed invariants. When using invariants for feature correspondence, the uncertainties in the estimates of the invariants will consequently cause ambiguity in the decisions resulting from the matching process. In a statistical framework, the matching criteria are usually based on the metric of likelihood defined on the object features. There are some approaches for feature matching, which are applicable ([6], [60] and [61]). However, to establish the systematic Bayesian-based principle in feature classification, modelling of uncertainties in feature estimation with *explicit* formulations is ultimately necessary. Concerning the surface-based quadric representations described in chapter 3, where the algorithm of surface fitting is specified, the uncertainties in parameter estimation and extraction of invariants, as

well as in pose representation, are formulated in this chapter with a covariance representation.

Generally, the model-based surface primitive recognition scheme, proposed in this thesis, is established on the estimates of invariants of quadric representation and an explicit representation of modelling the uncertainties of the estimates, i.e., the covariance matrices. This enables us to use an analytical probabilistic density function (pdf) to realise the optimal recognition of surface primitives. In other words, the estimate plus the covariance matrix comprise the *optimal descriptions* of an estimate, both of which are extracted from measurements, comprising a given set of range data. In the work of this thesis, as will be described in section 4.5, such optimal descriptions in a model-base are always assumed to be obtained from measurements. Based on the optimal descriptions, we also propose an approach to optimise such a modelling process.

The acquisition of the "optimal description" from range data of a quadric primitive is briefly shown in figure 4.1. An overview of the proposed optimal recognition scheme is depicted in figure 4.2.

There has been research published concerning the formulation of uncertainties in object descriptions. Waite and Ferrie ([63]) discussed the non-uniqueness problem in volumetric representations. They characterised the ambiguity of parameter estimates in terms of the "ellipsoid of confidence", quantifying the level of acceptability of a model and which information can be used to plan a new direction of view that minimise the ambiguity of subsequent interpretation. As argued by the authors, the representation was limited to partially communicate nonuniqueness at a single minimum in parameter space. The work of [55] dealt with object in high order polynomial representations. They used the asymptotic form of estimates to describe the posterior distribution of the parameters in polynomial fitting, and a criterion of minimum-error-probability for recognition was proposed. However, the reality of their probabilistic model, especially the derivation of covariance information, relies on the assumption that a large number of data are involved in fitting. Moreover, the knowledge of noise perturbation in data acquisition was not adequate for the description of the uncertainties in parameter estimation.

Based on the quadric representations of surfaces, we derive the formulation of uncertainties in the parameter estimation and the extraction of the invariants within a framework

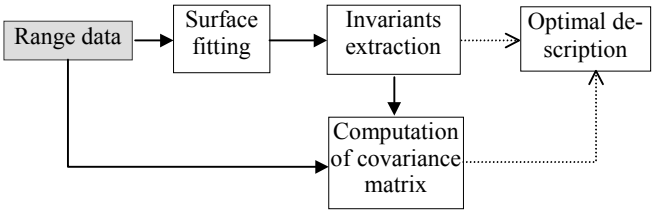


Fig. 4.1. Acquisition of the optimal description for quadric surface primitive.

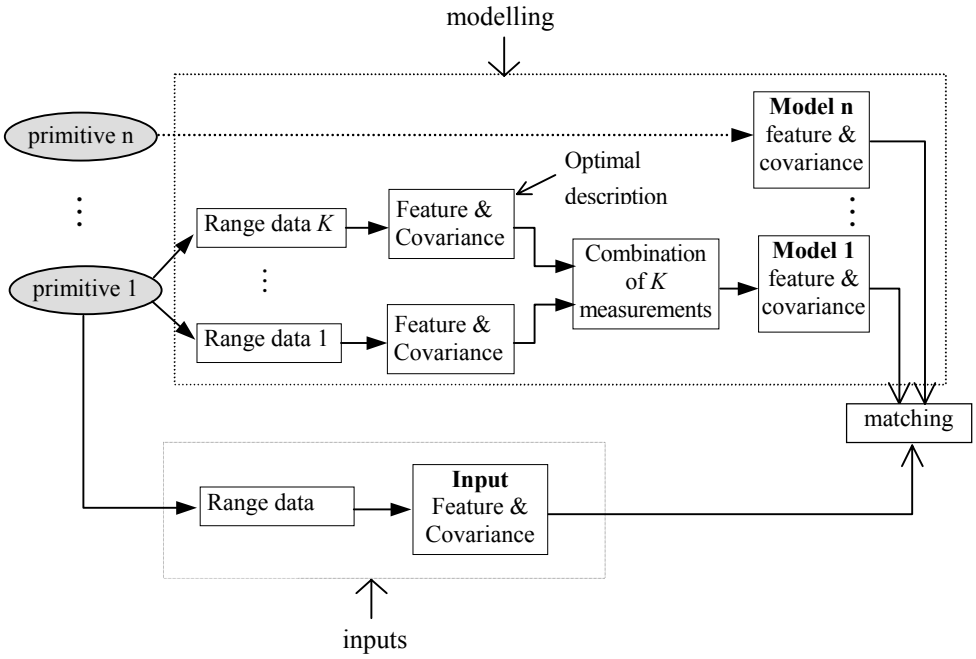


Fig 4.2. The recognition scheme of surface primitives.

of statistical perturbation. The linear approximation is applied in error analysis for parameter estimates, as it was used in [63]. However, linear approximation in our case is thought to be reliable for the estimates to possess the property of being statistically "unbiased" in surface fitting. Because the variance of noise in data points has been obtained in estimation, the linearisation of error propagation from data acquisition to parameter representation results in explicit formulation of covariance descriptions. For the concerning applications, we believe that the major statistical behaviour of estimates is contributed by the first and second order moments (expectation and covariance). Therefore, a normal pdf (probabilistic density function) is proposed to describe the estimates of parameters and the invariants.

Although our discussion about the feature-based optimal classification is limited to the surface primitives, the principles is expected to apply for more complicated objects by utilising the knowledge of pose transformation. Because the surface-based representation only partially communicates the shape information of a bounded object, spatial relations between surface primitives are expected to communicate the topological structures of a complex-constructed object. In other words, pose representation can be employed in object description on the scheme of *relational structure* (RS) representation. Towards this recognition scheme, discussions about the relational pose between two surface primitives and its uncertainty in estimation are also included in this chapter.

This chapter is organised as follows. In the first section, the invariants for object recognition are introduced. Concerning the quadric representation used in this work, invariants of a quadric surface are defined as the algebraic coefficients in the normalised standard frame, and stacked into a vector to describe the surface feature. In the normalisation process, the rotation and translation of the standard frame with respect to the world co-ordinate frame can be derived simultaneously.

Section 4.2 discusses the error analysis in parameter estimation and extraction of invariants, focused on the approach of formulating the covariance matrix. In section 4.3, a probabilistic model is proposed, describing the uncertainty in the estimates of the surface features as a normal pdf. The likelihood function, conditioned on the knowledge of the feature of object in model base, is derived in an analytical form. According to the probabilistic model, the sensor-based modelling scheme can be implemented within an optimal framework to establish the model base. This is described in section 4.4. Section 4.5 presents experimental results using synthetic data for surface primitive classification by feature correspondence. It demonstrates that the Bayesian-based matching criterion with the *statistical distance*, which is defined on the formulated likelihood function, is much more reliable than the one based on the measure of the conventional Euclidean distance where only the first order statistics (the estimates) are utilised. The pose representation is discussed in section 4.6. Based on the definition of pose transformation, the relational pose is further specified. Similar to the analysis of invariants, the covariance of the pose estimate is also discussed. The conclusions and some discussions are given in section 4.7.

4.1 Extraction of invariants

4.1.1 Invariants for recognition

An invariant is always defined in the context of a particular transformation. As shape descriptors for 3D objects, the invariants associated with geometric properties of the objects are often used. They are invariant to rotation and translation transformation. Examples are the lengths of the principal axes of an ellipse or the curvatures at surface points. Extensive studies on invariants theory and applications can be found in [24], [25], [38], [37] and [57].

According to the descriptive properties in object representations, invariants can be further distinguished as global or local. Global invariants are defined at object level, describing the shape of object as a whole ([51]), such as moment invariants, Fourier descriptors or algebraic invariants of implicit polynomials (e.g., the principal lengths of an ellipse). Local invariants are defined at each point of shape. Typical are the differential invariants (e.g., the curvature at a point on an ellipse). An advantage of local invariants is that they are less sensitive to occlusion. Local invariants usually require the computation of higher order derivatives, so they are vulnerable to noise perturbation. The advantage of global invariants is that they need neither extraction of local features nor the computation of derivatives, so they are less sensitive to noise perturbation. But, because they utilise the knowledge of the whole shape, they are susceptible to partial occlusion.

Just as the implicit polynomial representations, global invariants are formalised as functions of polynomial coefficients ([25] and [57]). Since the algebraic coefficients are estimated through fitting of data to a specific representation, the influences of occlusion on global invariants rely on the feature-representation. In general, the more parameters are required in representations, the more freedom in the shape descriptions in case of occlusion. For example, the problem of non-uniqueness in volumetric representations (see chapter 3) is mainly caused by partial occlusion ([63]). However, in case of low-degree polynomial surface representations, i.e., the quadrics used in this work, occlusion does not significantly influence the invariants at the level of surface description. This is because the surface representation neglects the "volumetric" property of the object, i.e. the surface-based shape descriptors relate only parts of the entity of the whole bounded object (to the quadrics). Therefore, the invariants we deal with in this section are "global" in the viewpoint of primitive representations at low level, but "local" in the viewpoint of being components of the scene descriptions at the higher level recognition phase, as used in the modelling strategy of section 4.6. In the notion of [43] and [64], such invariants could be regarded as "semi-global".

Below, the invariants to describe the quadric representation are defined and extracted through linear transformation.

4.1.2 Invariants extraction by normalisation

To uniquely determine the geometric properties of a quadric representation, the dimensionality of the invariants must be equal to the minimum number of the parameters of the object-centred shape description. According to algebra, multivariate quadrics can be converted by linear transformations into their *standard forms*, which have the minimum number of parameters required for uniqueness of the representation. For 3D surface representation in quadrics, the linear transformation can be implemented by a rotation and translation of the Cartesian co-ordinate systems. The co-ordinate system associated with the standard form is called the "standard co-ordinate system" or "standard frame". The standard form of quadrics uniquely and completely determines the shape of the surface. Naturally, as shape descriptor, the invariants can be explicitly expressed as the coefficients of the standard form. For a quadric representation, the extraction of invariants is the process of formulating the standard frame through a rotation and translation transformation. We call this process the *normalisation* of the surface representation.

The advantage of defining the invariants based on the standard form is in the simultaneous pose-extraction of a quadric surface as its co-ordinate transformation between the standard frame and the world frame. In this viewpoint, the normalisation in fact operates as an action on parameters to map them into a manifold in which the invariant geometry and the spatial properties are explicitly separated.

Without loss of generality, the standard form of a quadrics can be formulated in a "standard frame" x_s-y_s-z_s as

$$x_s^2 + \gamma_1 y_s^2 + \gamma_2 z_s^2 + c = 0 \tag{4.1}$$

In the above expression, distinctions of different quadric types are embedded within the selection of the coefficients. For example, an ellipsoid corresponds to the positive set

$\{\gamma_1,\gamma_2\}$ and negative c. A hyperbola corresponds to the set $\{\gamma_1,\gamma_2\}$ with at least one of the invariants being negative. The last case degenerates to a cone if c=0. A cylinder corresponds to the case that one of the invariants of the set $\{\gamma_1,\gamma_2\}$ is zero. We have imposed a scale constraint, normalising the implicit function of (4.1) such that the coefficient of the term x^2 becomes equal to 1, assuming that the coefficient of x^2 has the maximum absolute value among the three quadric terms.

Since the algebraic coefficients in the standard form of (4.1) are independent of its pose, they are used as invariants to comprise a "feature vector"

$$\mathbf{g} \equiv [\gamma_1, \gamma_2, c]^T \tag{4.2}$$

The feature vector **g** uniquely determines the geometric properties of a quadric surface.
The general quadric form of (3.1) represented in world co-ordinates can be rewritten in a form

$$\mathbf{x}^T\mathbf{A}\mathbf{x} + \mathbf{v}^T\mathbf{x} + k = 0\,, \tag{4.3}$$

where

$$\mathbf{A} = \begin{bmatrix} a_{11} & \frac{1}{2}a_{12} & \frac{1}{2}a_{13} \\ \frac{1}{2}a_{12} & a_{22} & \frac{1}{2}a_{23} \\ \frac{1}{2}a_{13} & \frac{1}{2}a_{23} & a_{33} \end{bmatrix};\quad \mathbf{v} = [v_1, v_2, v_3]^T\,.$$

According to the definition of (4.2), extraction of invariants equals formulating the standard form of (4.1) through co-ordinate transformation, i.e., through the normalisation process. The normalisation is implemented with two steps which are detailed next.

First a rotation transformation **x** = **Rx'** is applied to diagonalise the matrix $\mathbf{A}$. This yields

$$\mathbf{x}'^T\mathbf{A}'\mathbf{x}' + \mathbf{v}'^T\mathbf{x}' + k = 0\,, \tag{4.4}$$

with $\mathbf{A}'=\mathbf{R}^T\mathbf{A}\mathbf{R}$ a diagonal matrix which diagonal components are denoted as λ_1, λ_2, λ_3 and with $\mathbf{v}'=\mathbf{R}^T\mathbf{v}$.

Then we apply the translation transformation to let the origin of the co-ordinate system to be at the standard position. Let

$$\mathbf{x}' = \mathbf{x}_s + \mathbf{t}'\,,$$

where **t'** denotes the translation vector from **x'** to $\mathbf{x}_s$. Then we have

$$\mathbf{x}_s^T\mathbf{A}'\mathbf{x}_s + (2\mathbf{t}'^T\mathbf{A}' + \mathbf{v}'^T)\mathbf{x}_s + \mathbf{t}'^T\mathbf{A}'\mathbf{t}' + \mathbf{v}'^T\mathbf{t}' + k = 0 \tag{4.5}$$

Compared with (4.1), the second term in LHS (left hand side) of (4.5) is zero. Thus the translation can be determined from

$$2\mathbf{t}'^T\mathbf{A}' + \mathbf{v}'^T = 0\,, \tag{4.6}$$

and so we get

$$\mathbf{t}' = -\frac{1}{2}\mathbf{A}'^{-1}\mathbf{v}' . \tag{4.7}$$

The translation from $\mathbf{x}$ to $\mathbf{x}_s$, denoted as $\mathbf{t}$, is derived by

$$\mathbf{t} = \mathbf{R}\mathbf{t}' = -\frac{1}{2}\mathbf{A}^{-1}\mathbf{v} \tag{4.8}$$

Finally, to convert (4.5) into the expression (4.1), we get the coefficients in the standard form, i.e., the invariants, as below.

$$\gamma_1 = \frac{\lambda_2}{\lambda_1}; \gamma_2 = \frac{\lambda_3}{\lambda_1}; c = \frac{k'}{\lambda_1} \tag{4.9}$$

where

$$k' = \mathbf{t}'^T\mathbf{A}'\mathbf{t}' + \mathbf{v}'^T\mathbf{t}' + k . \tag{4.10}$$

Depending on the applied notation, the finally obtained expression is the one of (4.1).

4.2 The covariance matrices to describe the uncertainties in estimated parameter and feature representation

As stated in the previous section, the invariants and the feature vector defined by (4.2) are computed from the quadric parameters $\mathbf{A}$, $\mathbf{v}$ and k of (4.3) representing the quadric in world co-ordinates. However, because of noise in the acquired range data, uncertainties propagate into the estimates of the parameters $\mathbf{A}$, $\mathbf{v}$ and k during surface fitting. Consequently the uncertainties will propagate into the computation of the invariants. Since the surface recognition is based on the matching of the invariants between the model base and the inputs of the observations, uncertainties in parameter estimation and feature representation must be considered in the recognition strategy. To optimise the recognition scheme, knowledge about the statistical attributes of the estimates of the feature is required. Essentially, the second order moment of the estimates, i.e., the covariance matrix, can be used to characterise the uncertainties of the estimates.

Given the estimate of the quadric parameters, obtained from surface fitting, this section deals with the formulation of the uncertainties in the estimated parameters and the extracted invariants. In the perturbation framework for error analysis, the statistics of uncertainties are explicitly expressed by the covariance matrices. To represent the uncertainties in the invariants, we start with the uncertainties in the estimated parameters. This step is followed by the formulation of the covariance matrix of the feature vector, which is derived from the covariance matrix of the quadric parameters.

In reality, the performance of uncertainties in parameter estimation could be very complicated and it is impossible to perfectly fit an analytical mathematical model. However, in many cases, the major statistical characteristics of a random variable (or a multivariate random vector) can be described by its first and second order moments. The first order mo-

ment, i.e., the expectation of the estimate of surface parameters, is associated with surface fitting which is characterised by the bias-corrected optimisation. The second order moment, i.e., the covariance, can be employed to describe the uncertainties of the estimate. Without further refinement in statistical analysis, we use the covariance model, including the optimal estimate and the covariance formulation, to represent the overall statistical properties of the estimated parameters, obtained from surface fitting, and the invariants, extracted from these parameters.

Because the mapping of perturbations in data points (measurement noise) into uncertainties in the descriptions of invariants, is non-linear, it is not practical to try to formulate exact expressions for the errors in the estimates in terms of noise in data points. Therefore, a linear approximation is applied in the error analysis. Moreover, taking into account that the improved bias-corrected optimisation is used for parameter estimation, it is expected that the linear approximation used in the following discussion is more reliable than that used in [55] and [63]. Obviously, the assumption of linearity is acceptable in case of low noise level. For significant noise levels, non-linearity of the perturbation might violate the conditions and consequently degrade the reliability of the model. According to the experiments with synthetic data, reliability of the proposed approach for surface classification has been verified with additive noise of moderate level. More generally, the proposed model can be treated as a trade-off between the tractability in mathematics and the complexity in practice of modelling uncertainties in representations.

4.2.1 Perturbation of the invariants in measurement

The errors in the estimated invariants can be approximated by linearisation of their dependency on the errors of the estimated parameters obtained by surface fitting. The linear approximation is based on the assumption that both the errors of the estimates obtained from surface fitting and the errors in the derived invariants can be treated as bounded perturbations that are almost linearly related to the noise in the data points.

First of all, a vector representing the surface parameters is defined, by which the uncertainties in the parameter estimation can be properly formalised. The general expression of a quadric representation was given by the implicit form of (3.1). Since multiplying of (3.1) with any non-zero constant results in the same quadric representation, a scale constraint on $\mathbf{p}$, which consists of 10 coefficients (chapter 3), is introduced. Considering the inherent constraint $\|\mathbf{p}\|_2^2 = 1$ among the coefficients, the dimensionality of the parameter space reduces to 9. Because the scale constraint for (3.1) does not change the invariants expressed by (4.1), for simplicity in parameter representation, we choose k=1 as the constraint in (3.1) and (4.3). This is realised by multiplying a scalar to all of the components of the solution $\mathbf{p}$ in surface fitting. Therefore, we define the parameter vector $\boldsymbol{\theta}$ as

$$\boldsymbol{\theta} = [a_{11}, a_{22}, a_{33}, a_{12}, a_{13}, a_{23}, v_1, v_2, v_3]^T .$$

In the following, the parameters of $\mathbf{A}$ and $\mathbf{v}$ in (4.3) are assumed to be represented by the vector $\boldsymbol{\theta}$.

Defining $\hat{\mathbf{A}}, \hat{\mathbf{v}}$ as the estimates of the algebraic parameters obtained from surface fitting of (4.3). To extract the invariants by normalisation, three eigenvalues, denoted as $\hat{\lambda}_1, \hat{\lambda}_2, \hat{\lambda}_3$, are first derived from the symmetric matrix $\hat{\mathbf{A}}$, which incorporates the derivation of the rotation matrix **R**. Then the normalised variables from (4.9)

$$\hat{\gamma}_1 = \frac{\hat{\lambda}_2}{\hat{\lambda}_1}; \hat{\gamma}_2 = \frac{\hat{\lambda}_3}{\hat{\lambda}_1}$$

will serve as the invariants represented in the standard form. Because of the deviation of $\hat{\mathbf{A}}$ from its ideal values, the derived eigenvalues also deviate from their true values, which yields the errors in invariants computation. Denoting

$$\Delta\lambda_i = \hat{\lambda}_i - \lambda_i, \forall i \in \{1,2,3\}$$

as the error in the estimated eigenvalue with respect to its ideal value, within linear approximation, the errors of the invariants expressed in (4.9) can be expressed as

$$\Delta\gamma_1 = \frac{\Delta\lambda_2}{\lambda_1} - \frac{\lambda_2\Delta\lambda_1}{\lambda_1^2}; \Delta\gamma_2 = \frac{\Delta\lambda_3}{\lambda_1} - \frac{\lambda_3\Delta\lambda_1}{\lambda_1^2}; \Delta c = \frac{\Delta k'}{\lambda_1} - \frac{k'\Delta\lambda_1}{\lambda_1^2} \tag{4.11}$$

To discuss the statistics of the error in the feature vector $\Delta\mathbf{g} = [\Delta\gamma_1, \Delta\gamma_2, \Delta c]^T$, we start with the properties of the perturbation of eigenvalues. Based on the *perturbation* theory and the assumption that $\|\Delta\mathbf{A}\|$ is small enough so that we can take the linear approximation. The perturbation $\Delta\lambda$ of an eigenvalue λ can be approximated, as described in [53], by

$$\Delta\lambda \approx \mathbf{e}^T \Delta\mathbf{A}\mathbf{e} \tag{4.12}$$

where **e** is the eigenvector of **A** associated with the true λ.

A convenient way to express the perturbation of the eigenvalues in terms of parameters is to select a special co-ordinate system that is formed by using the three eigenvectors of **A** as the basis vectors of a new co-ordinate system. From the process of invariant extraction by normalisation, described in the subsection above, it is easy to see that such a co-ordinate system is obtained by a rotation transformation with $\mathbf{x}'=\mathbf{R}^T\mathbf{x}$. Hereafter, this transformed co-ordinate **x'** is denoted as the "rotation-normalised" co-ordinate, which is always available because the corresponding eigenvectors and the rotation transformation can be simultaneously computed. To avoid confusion, the co-ordinate $\mathbf{x}_s$ in the final standard form is denoted as the "normalised' co-ordinate.

In "rotation-normalised" co-ordinates, we have

$$\mathbf{e}_1 = [1,0,0]^T, \mathbf{e}_2 = [0,1,0]^T, \mathbf{e}_3 = [0,0,1]^T.$$

Then straight forward we have

$$\Delta\lambda_i = \Delta a'_{ii} \quad \forall i \in \{1,2,3\},$$

with $\Delta a'_{ii} = \hat{a}'_{ii} - a'_{ii}$

The $\hat{a}'_{ii}$ and a'_{ii} are respectively the estimate and the true value of the i^{th} diagonal element of the transformed matrix **A'**. Also from (4.11), the component Δc can also be expressed in terms of $\Delta a'_{ii}$ and $\Delta v'_i$ with i=1,2,3 under the same conditions.

Strictly speaking, the "rotation-normalised" co-ordinate is defined by an ideal transformation for the eigenvectors $\{\mathbf{e}_i\} \forall i \in \{1,2,3\}$, which are associated to the ideal formation of **A**. Because only $\hat{\mathbf{A}}$, consisting of estimates of the parameters, is known and used for computation of **R**, errors are unavoidably induced. However, due to the unbiased property of the parameter estimates, an optimisation exists such that $\hat{\mathbf{A}}$ is unbiased, and as a result **R** and **x'** too.

Without trivialities in formality from (4.11), the error of the feature vector $\hat{\mathbf{g}}$ can be briefly expressed by

$$\Delta \mathbf{g} = \mathbf{M} \Delta \boldsymbol{\theta}' \tag{4.13}$$

where $\Delta \boldsymbol{\theta}'$ is the error in the estimate of the surface parameters assuming obtained in the rotation-normalised co-ordinate and **M** is a 3×9 matrix associated to the coefficients of (4.11). Details of the derivatives of (4.13) and the expression of **M** can be found in Appendix B of this thesis.

In the following, the covariance matrix of $\hat{\boldsymbol{\theta}}$, describing the uncertainties of estimates in surface fitting, is generally formulated under the world co-ordinate system. The formulation with respect to $\hat{\boldsymbol{\theta}}'$ can be easily obtained just by applying rotation to those data points in the expression of covariance matrix with respect to $\hat{\boldsymbol{\theta}}$. Subsequently, the covariance matrix of the feature vector $\hat{\mathbf{g}}$ can be derived from (4.13).

4.2.2 Formulation of the covariance matrix

As mentioned at the beginning of 4.2, the parameters in surface fitting are estimated with a reliable "bias-corrected" method. Since parameter transformation is linear between different co-ordinates, the parameter vector $\hat{\boldsymbol{\theta}}'$, estimated in the rotation-normalised co-ordinate system, can be treated as unbiased. According to (4.13), $\hat{\mathbf{g}}$ is also unbiased, i.e., we have $E[\hat{\mathbf{g}}]=\mathbf{g}$.

To derive the covariance matrix of $\hat{\mathbf{g}}$, we have to start from formulating the covariance matrix of the estimated surface parameter vector $\hat{\boldsymbol{\theta}}$ in the world co-ordinate system. In surface fitting, the estimate $\hat{\boldsymbol{\theta}}$ is obtained by minimising the cost function of (3.4), i.e.,

$$\frac{\partial}{\partial \boldsymbol{\theta}} \left(\sum_{i=1}^{n} d_i^2 \right) = 0 .$$

Substituting the expression of d_i^2 with (3.3), and applying a linear approximation in the dependency of $\Delta \boldsymbol{\theta}$ on $\Delta \mathbf{x}_i \forall i$, we can obtain the covariance matrix of $\hat{\boldsymbol{\theta}}$, defined as

$$\Psi = E\left[\Delta \boldsymbol{\theta} \Delta \boldsymbol{\theta}^T \right],$$

expressed in terms of $\{\mathbf{x}_i\}$ and the estimate $\hat{\boldsymbol{\theta}}$, as well as the variance of noise in data points.

The derivation of the covariance matrix $\boldsymbol{\Psi}$ and its formulation is described in Appendix C.1. The final expression of $\boldsymbol{\Psi}$ is the form of (4.14).

$$\boldsymbol{\Psi} = \sigma^2 \left(\sum_{i=1}^{n} w_i \mathbf{m}_i \mathbf{m}_i^T \right)^{-1} \tag{4.14}$$

where $\mathbf{m}_i = \left[x_i^2, y_i^2, z_i^2, x_i y_i, x_i z_i, y_i z_i, x_i, y_i, z_i \right]^T$.

Supposing that the parameter $\hat{\boldsymbol{\theta}}$ is estimated in the "rotation-normalised" co-ordinate, then its covariance matrix can be expressed by substituting $\{\mathbf{x}_i\}$ and $\hat{\boldsymbol{\theta}}$ with $\{\mathbf{x}'_i\}$ and $\hat{\boldsymbol{\theta}}'$, i.e., the values after rotation transformation from world co-ordinate to the rotation-normalised co-ordinate.

Denoting $\mathbf{G}$ the covariance matrix of $\hat{\mathbf{g}}$, then according to (4.13), we have

$$\mathbf{G} = \mathbf{M} \boldsymbol{\Psi} \mathbf{M}^T \tag{4.15}$$

So it is just required to compute the matrix $\boldsymbol{\Psi}$ in rotation-normalised co-ordinates. Since the transformation between world co-ordinates and rotation-normalised co-ordinates can be simultaneously derived through the computation of $\mathbf{R}$, the computation of $\mathbf{G}$ can also be derived simultaneously. In addition, the variance of the measurement noise σ^2 is estimated simultaneously with surface fitting.

From observation of the expression (4.14), we can find that the covariance matrix $\boldsymbol{\Psi}$ relates to the measure of *condition number* of the coefficient matrix in the sense of LSE solution of parameter $\boldsymbol{\theta}$. When treating the weight $\{w_i\}$ independent of $\boldsymbol{\theta}$, the solution of $\boldsymbol{\theta}$ can be obtained as the LSE solution of the following set of n linear equations $\mathbf{F}\boldsymbol{\theta}=\mathbf{b}$ with

$$\mathbf{F} = \left[\sqrt{w_1}\mathbf{m}_1, \sqrt{w_2}\mathbf{m}_2, .., \sqrt{w_n}\mathbf{m}_n \right]^T$$

an $n \times 9$ matrix and

$$\mathbf{b} = \left[-\sqrt{w_1}, -\sqrt{w_2}, .., -\sqrt{w_n} \right]^T$$

an $n \times 1$ vector.

To solve this over-determined equation, the linear LSE approach can be applied. Neglecting the details of solution, we just notice that the condition number of the matrix $\mathbf{F}$, defined as $cond(\mathbf{F}) = \|\mathbf{F}\|_2 \left\|\mathbf{F}^{-1}\right\|_2$, is associated to the above covariance matrix $\boldsymbol{\Psi}$. Since

$$\mathbf{F}^T\mathbf{F} = \sum_{i=1}^{n} w_i \mathbf{m}_i \mathbf{m}_i^T \ ,$$

there is

$$cond(\mathbf{F}) = \sqrt{\rho_{\max} / \rho_{\min}} \ ,$$

here ρ_{max} and ρ_{min} are the maximum and minimum eigenvalues of $\mathbf{\Psi}$.

Because the condition number is a measure for the "stability" of the LSE solution, the uncertainty in the estimated parameters (the parameter vector $\hat{\boldsymbol{\theta}}$) by the covariance matrix $\mathbf{\Psi}$ can be further viewed.

In summary, the optimal feature representation can be implemented with the following steps:

1. Carry out surface fitting with the given range data in the world co-ordinates, as described in chapter 3.
2. Extract invariants by normalisation to form the feature vector $\mathbf{g}$, incorporating retrieval of rotation and translation, as described in (4.4) to (4.10).
3. Compute the covariance matrix Ψ in the rotation-normalised co-ordinate system by substituting the measurement of $\hat{\boldsymbol{\theta}}$ and $\{\mathbf{x}_i\}$ in the world co-ordinates with $\hat{\boldsymbol{\theta}}'$ and $\{\mathbf{x}'_i\}$ in the rotation-normalised co-ordinates.
4. Extract the covariance matrix $\mathbf{G}$ for $\mathbf{g}$ from $\mathbf{\Psi}$ with (4.14) and (4.15).

Once the covariance matrix is formulated, the statistical performances of the estimate of feature $\mathbf{g}$ are then further represented using a probabilistic model proposed in the next section.

4.3 Modelling the uncertainties of estimates with normal pdf

In the sense of statistical pattern recognition, the Bayesian rule is applicable if the pdf (probabilistic density function) of the estimates is known. Once an analytical representation of the pdf is available, all of the statistical attributes of the estimates can be derived from the pdf. Therefore, given the estimates and the analytical pdf, implementation of optimal recognition system is straightforward, without the requirement of large samples stored in the model base.

In this section, an analytical representation based on a normal pdf is proposed to model the estimate of feature vector $\hat{\mathbf{g}}$. This representation is suggested from a few aspects:

(1) As noted in the covariance formulation, the approximation of a linear dependency of the perturbation of parameters on the noise of data points is supported by the improved "bias-corrected" parameter estimation. Since the noise $\{\Delta\mathbf{x}_i\}$ is treated as Gaussian noise, the perturbation $\Delta\boldsymbol{\theta}$ of the parameters stacked in the parameter vector $\boldsymbol{\theta}$ is also approximately normally distributed.

(2) The gradient weighted LSE can be interpreted ([55]) in terms of *maximum likelihood estimation* (MLE), whose asymptotic distribution is normal.

(3) In general, the covariance and the expectation reveal the major statistical behaviour of an estimate. Thus the normal pdf is a tractable model since it describes all statistics of the estimate by the expectation and the covariance. Therefore, at moderate noise level, the normal pdf can be used as an analytical probabilistic representation for the estimate $\hat{\mathbf{g}}$.

Denoting with $\mathbf{g}$ the "true" parameters describing the object and with $\hat{\mathbf{g}}$ an estimate obtained from the measurements, the normal pdf of $\hat{\mathbf{g}}$ is expressed by

$$f(\hat{\mathbf{g}}|\mathbf{g}) = C\exp\left\{-\tfrac{1}{2}(\hat{\mathbf{g}}-\mathbf{g})^T\mathbf{G}^{-1}(\hat{\mathbf{g}}-\mathbf{g})\right\} \tag{4.16}$$

Here, the covariance matrix $\mathbf{G}$ is formulated by (4.15) and C is the normalised constant.

In many cases, the true parameters $\mathbf{g}$ can not be obtained exactly. As will be illustrated in the next section, we consider the establishment of a model base being sensor-based, i.e. the models used in the recognition phase are also obtained from measurements. Considered in this way, the matching problem is in fact a correspondence problem between two observations. Therefore, what we concern in recognition is a probabilistic representation that two estimates come from the same class, irrespective of what are the "true" parameters of that class. Below we will see that such a probabilistic representation can be formulated by a conditional joint pdf of two observed features and further a "statistical distance" that can be derived to characterise the similarity of the two features.

Suppose that $\hat{\mathbf{g}}_1$ and $\hat{\mathbf{g}}_2$ are two estimates of the same surface feature noted by the same class m and that they are measured independently. Then the joint pdf of $\hat{\mathbf{g}}_1$ and $\hat{\mathbf{g}}_2$ is of the form

$$f(\hat{\mathbf{g}}_1,\hat{\mathbf{g}}_2|\mathbf{g}_m) = f(\hat{\mathbf{g}}_1|\mathbf{g}_m)f(\hat{\mathbf{g}}_2|\mathbf{g}_m) \tag{4.17}$$

where $\mathbf{g}_m$ denotes the true feature of the m^{th} model.

Since $\mathbf{g}_m$ is unknown, we need to know the joint pdf of $\hat{\mathbf{g}}_1$ and $\hat{\mathbf{g}}_2$ only conditioned on the assumption that both are from the same model, in spite of the true value $\mathbf{g}_m$. Under such constraint, the joint pdf can be expressed by

$$f(\hat{\mathbf{g}}_1,\hat{\mathbf{g}}_2|H) = \int f(\hat{\mathbf{g}}_1,\hat{\mathbf{g}}_2|\mathbf{g}_m)d\mathbf{g}_m = \int f(\hat{\mathbf{g}}_1|\mathbf{g}_m)f(\hat{\mathbf{g}}_2|\mathbf{g}_m)d\mathbf{g}_m \tag{4.18}$$

where the constraint is

H : $\hat{\mathbf{g}}_1$ and $\hat{\mathbf{g}}_2$ belong to the same model

Substituting (4.16) into (4.18) and assuming that the covariance matrices both for $\hat{\mathbf{g}}_1$ and $\hat{\mathbf{g}}_2$ are constant, the integration of (4.18) yields the following expression

$$f(\hat{\mathbf{g}}_1,\hat{\mathbf{g}}_2|H) = C\exp\{-\tfrac{1}{2}(\hat{\mathbf{g}}_1-\hat{\mathbf{g}}_2)^T\mathbf{G}_{1,2}^{-1}(\hat{\mathbf{g}}_1-\hat{\mathbf{g}}_2)\} \tag{4.19}$$

where C is again the normalised constant and with

$$\mathbf{G}_{1,2} = \mathbf{G}_1 + \mathbf{G}_2 \tag{4.20}$$

In (4.20), $\mathbf{G}_1$ and $\mathbf{G}_2$ are the covariance matrices for $\hat{\mathbf{g}}_1$ and $\hat{\mathbf{g}}_2$ that are independently obtained from the measurements. The joint pdf in (4.19) is also called the *likelihood* function.

Formulations (4.19) and (4.20) enable us to apply the principle of Bayesian recognition for feature classification. The recognition can be stated as: given observations which classes are predefined, classify a new observation to one of the classes. This can be interpreted as a classical *hypotheses test* problem. Supposing $\hat{\mathbf{g}}_{m1}$ and $\hat{\mathbf{g}}_{m2}$ are two observations from the

known class m_1 and m_2, respectively. Given a new observation $\hat{\mathbf{g}}_s$, if we are going to classify $\hat{\mathbf{g}}_s$ to either m_1 or m_2, the hypotheses are stated as:

$$H_0: \quad E\{\hat{\mathbf{g}}_s\} = E\{\hat{\mathbf{g}}_{m1}\},\ H_1: \quad E\{\hat{\mathbf{g}}_s\} = E\{\hat{\mathbf{g}}_{m2}\},$$

According to the Bayesian rule, assuming that the risks for misclassification are equal, then the classification is determined by

$$\text{if } r(\hat{\mathbf{g}}_s) = \frac{f(\hat{\mathbf{g}}_s, \hat{\mathbf{g}}_{m1} | H_0)}{f(\hat{\mathbf{g}}_s, \hat{\mathbf{g}}_{m2} | H_1)} \geq 1, \text{ classify } \hat{\mathbf{g}}_s \text{ to } m_1; \text{ otherwise, classify } \hat{\mathbf{g}}_s \text{ to } m_2;$$

where the joint pdf is formulated by (4.19). The ratio r(·) is also called *likelihood ratio*. Equally we can use the measure of

$$d_{sm}^2 = \tfrac{1}{2}\left(\hat{\mathbf{g}}_s - \hat{\mathbf{g}}_m\right)^T \mathbf{G}_{s,m}^{-1}\left(\hat{\mathbf{g}}_s - \hat{\mathbf{g}}_m\right) \tag{4.21}$$

for comparison, i.e., if $(d_{s,m1})^2 \geq (d_{s,m2})^2$, classify $\hat{\mathbf{g}}_s$ to m_1; otherwise, classify $\hat{\mathbf{g}}_s$ to m_2.

In case of multiple classes, supposing $\hat{\mathbf{g}}_l$, with $l \in \{1,\ldots,\mathrm{L}\}$ is the observed feature of the l^{th} class of L modelled primitives. Given a measurement $\hat{\mathbf{g}}_s$ from the scene, it is classified to i^{th} class if

$$d_{si}^2 = \min_{l \in \{1,..,L\}} \{d_{sl}^2\} \tag{4.22}$$

Without confusion, we call the $d_{ij}^2 \forall (i,j)$ defined in (4.21) the "statistical distance" between two measurements $\hat{\mathbf{g}}_i$ and $\hat{\mathbf{g}}_j$. It is noted that if the modelling process is ideal, i.e., the "true" model feature $\mathbf{g}_j$ is available and consequently the covariance matrix $\mathbf{G}_j$ is zero, then d_{ij}^2 is equivalent to the well-known *Mahalanobis* distance.

4.4 Optimisation in sensor-based modelling process

In model-based recognition, a modelling process is required to establish the reference knowledge about the objects. As illustrated in figure. 4.1, the models concerned in the work of this thesis are described with the estimates of features and the covariance matrices with respect to defined classes of surface primitives.

Generally, two modelling approaches can be distinguished: sensor-based and through a CAD/CAM system. The sensor-based approach uses actual objects to build model descriptions by measurements. The approach with CAD/CAM system uses a set of predefined prototypes to construct the CAD model of an object. The former modelling process is usually noise corrupted while the latter is noise free. Without comparison between these two schemes, we focus on the sensor-based modelling approach.

In the sensor-based modelling process, we suppose that more than one measurement is carried out for each primitive which class has been predefined. $\{\mathbf{g}_{m,k}, \mathbf{G}_{m,k}\}\ \forall k \in \{1,\ldots,K\}$ are the K observations for the primitive m. Since the "true" value for class m is unknown, we need to find a "best" value for model m from the K observations. Based on the prob-

abilistic representation, i.e., the analytical pdf of the estimates as introduced in section 4.3, such a "best" description can be obtained as an optimal estimate.

Assuming that these observations are independent with each other, the joint pdf of the measurements is expressed as

$$f(\hat{\mathbf{g}}_{m,1},\ldots,\hat{\mathbf{g}}_{m,K}|\mathbf{g}_m)=\prod_{k=1}^{K} f(\hat{\mathbf{g}}_{m,k}|\mathbf{g}_m)\,, \tag{4.23}$$

where $\mathbf{g}_m$ represents the "true" feature of the primitive m.

To integrate the information of all the measurements to estimate the parameter $\mathbf{g}_m$, the *Maximum Likelihood Estimate* (MLE) can be applied because the pdf's of feature $\hat{\mathbf{g}}_{mk}$ $\forall k\in\{1,\ldots,K\}$ have been explicitly formulated. Thus the feature stored in the model base, denoted $\overline{\mathbf{g}}_m$, is the one that maximises the joint pdf $f(\hat{\mathbf{g}}_{m,1},\ldots,\hat{\mathbf{g}}_{m,K}|\mathbf{g}_m)$. Substituting $f(\hat{\mathbf{g}}_{m,k}|\mathbf{g}_m)$ in (4.23) with the normal pdf of (4.16) and let

$$\frac{\partial}{\partial \mathbf{g}} f(\hat{\mathbf{g}}_{m,1},\ldots,\hat{\mathbf{g}}_{m,K}|\mathbf{g}_m)\bigg|_{\mathbf{g}_m=\overline{\mathbf{g}}_m}=0\,,$$

we have

$$\overline{\mathbf{g}}_m=\left(\sum_{k=1}^{K}\mathbf{G}_{m,k}^{-1}\right)^{-1}\left(\sum_{k=1}^{K}\mathbf{G}_{m,k}^{-1}\hat{\mathbf{g}}_{m,k}\right) \tag{4.24}$$

The result of (4.24) gives an optimal estimate of the feature of a model given a number of measurements in the modelling process.

The covariance matrix of the estimate $\overline{\mathbf{g}}_m$, denoted as $\mathbf{G}_m$, is computed with the expression

$$\overline{\mathbf{G}}_m=E\{(\overline{\mathbf{g}}_m-E(\overline{\mathbf{g}}_m))(\overline{\mathbf{g}}_m-E(\overline{\mathbf{g}}_m))^T\}$$

Because each $\hat{\mathbf{g}}_{mk}$ $\forall k$ is obtained independently, the correlation between $\hat{\mathbf{g}}_{mi}$ and $\hat{\mathbf{g}}_{mj}$ $\forall\{i,j|$ $i\neq j\}$ is thought to be zero. Then $\overline{\mathbf{G}}_m$ can be derived with (4.24) as

$$\overline{\mathbf{G}}_m=\left(\sum_{k=1}^{K}\mathbf{G}_{m,k}^{-1}\right)^{-1} \tag{4.25}$$

Finally, the pair ($\overline{\mathbf{g}}_m$, $\overline{\mathbf{G}}_m$) constructs an optimal description for the model m.

4.5 Optimal classification of surface primitives using synthetic data

Based on the optimal description of surface primitives, the feature-based classification can be implemented with the Bayesian criterion, using the probabilistic model of the normal

pdf. This section describes the results of experiments of classification using synthetic data. Applications for real range images are described in chapter 6. To demonstrate the optimality of the proposed probabilistic model, the classification results based on the conventional Euclidean distance, which only relies on the estimates of the feature itself without characterisation of the covariance, is also reported for a comparison.

The 3-D data points were generated from synthetic quadric surfaces. Four ideal ellipsoid prototypes labelled with $l \in \{1,2,3,4\}$ were selected as the four classes for classification. The surface primitives were generated from the four prototypes within a 2×2 window at *x*-*y* plane, corrupted with additive Gaussian noise. The sampling distance at both *x* and *y* directions is 0.1. The number of used points is 400 for all trials. In the following, all the measurement values have been normalised in an arbitrary unit.

Assuming that the sensor-based modelling process was applied, the model base was established from measurements with the same conditions used for getting inputs of the scenes. Hereafter, the "model" refers to the estimate of feature in the model base obtained from the noise-corrupted measurement and the "prototype" refers to the ideal value.

Given the surface primitive of 3-D data points sampled from each prototype, Gaussian noise was added to the three co-ordinates of each data point. Then the optimal description for the primitive, i.e., the estimate of feature vector and the covariance matrix, was obtained through the procedure as shown in figure 4.1.

First, each model of the four classes was constructed only from one randomly selected set of measurements. It means that such a model is less "accurate" than what is obtained in a number of observations. The feature and the covariance matrix for class *m* is noted as $\{\mathbf{g}_{m,l}, \mathbf{G}_{m,l}\}\ \forall l \in \{1,2,3,4\}$. The features of the ideal prototypes and their generated models are listed below. The standard deviation of the added noise is σ=0.02 for each model.

class	*1*	*2*	*3*	*4*
Prototypes	$[0.667,0.333,-1.667]^T$	$[0.727,0.455,-1.819]^T$	$[0.8,0.6,-2]^T$	$[0.889,0.778,-2.222]^T$
Models	$[0.605,0.306,-1.425]^T$	$[0.671,0.416,-1.589]^T$	$[0.777,0.586,-1.932]^T$	$[0.837,0.719,-2.039]^T$

Then trials of inputs as the observations of the surface primitives were repeatedly generated from the four prototypes, with added Gaussian noise. The feature of each input scene, denoted $\mathbf{g}_s$, and the associated covariance matrix $\mathbf{G}_s$ were obtained in the same way as in the modelling process.

Applying the Bayesian recognition algorithm stated in (4.21), the classification was carried out using the following criterion:

Classify $\mathbf{g}_s$ to class *i*, if it satisfies the condition:

$$\left(\mathbf{g}_s - \mathbf{g}_{m,i}\right)^T \mathbf{G}_{s,i}^{-1} \left(\mathbf{g}_s - \mathbf{g}_{m,i}\right) \le \left(\mathbf{g}_s - \mathbf{g}_{m,j}\right)^T \mathbf{G}_{s,j}^{-1} \left(\mathbf{g}_s - \mathbf{g}_{m,j}\right), \forall j \ne i, \quad \text{with} \quad i, j \in \{1,2,3,4\}$$

where $\mathbf{G}_{s,i} = \mathbf{G}_s + \mathbf{G}_{m,i}$.

To test the reliability of the approach, the inputs comprised 50 trials for each prototype. Because the classification is determined by the "statistical distance" defined in (4.21), a

comparison of the classification results respectively based on the statistical distance, denoted as $(d_{std})^2$ and the Euclidean distance, denoted as $(d_{euc})^2$, is given.

The classification results are listed in table 4.1 and table 4.2 for noise levels of σ=0.05 and σ=0.02. All inputs generated from the four prototypes were classified to one of the labels $l \in \{1,2,3,4\}$ and the "miss-classifications" were counted, i.e., the kind of errors that a trial is classified as "not matching" to its actual model is considered and the ratios were computed. The experimental results strongly demonstrate what was already expected, that the Euclidean distance based classification could be unreliable if the surfaces "shapes" are similar, as is the case involved in the experiment. In contrast, using the proposed optimal description via the probabilistic model, the classification based on the statistical distance yields much better results.

Intuitively, the capability of a recognition system to "identify" different objects depends on the measure of "closeness" between two observations for the same object. In such a notion, the statistical distance and Euclidean distance were used to measure the "closeness" between a model and its observations in experiments. A number of inputs were generated from the prototype of label l=2 with 100 trials (with noise level of σ=0.05) and the distances between each trial and the fixed model $\mathbf{g}_{m,2}$ were computed in terms of "statistical distance" and Euclidean distance. The histograms of the distance distribution within these trials are shown in figure 4.3 and figure 4.4. In order to give a comparison of their capabilities of measure of "closeness", the distance measure was normalised by a factor, which was selected as the *median* among the distances to model $\mathbf{g}_{m,1}$ of all trials. Therefore, the quantities along x-axes in figure 4.3 and figure 4.4 indicate relative distances.

From these results we see that the dispersion in the case of the statistical distance is smaller then that of the Euclidean distance. It implies that the statistical distance yields a better "identification" for different classes. In other words, since the distances have been normalised, the histograms reflect the distributions of likelihood ratio for two kinds of likelihood, $(d_{std})^2$ and $(d_{euc})^2$. Comparing the results of figure 4.3 with figure 4.3, we also conclude that using the statistical distance $(d_{std})^2$ in stead of $(d_{euc})^2$ is more "likely" to correctly classify an observation.

Given a group of trials generated from one primitive, the mean values of the statistical distance $(d_{std})^2$ and the Euclidean distance $(d_{euc})^2$ to each model are listed in table 4.3 and table 4.4, respectively. Comparing the (relative) distance of the corresponding model and other models, we can further see that the statistical distance performs better identifications than that of the Euclidean distance.

4.6 Pose description with surface representations

In general, recovery of pose information of objects is included in the model-based recognition. The success of matching between an observed object and its established model always implies a proper interpretation of the rotation and translation transformation between the measuring co-ordinate system and the model co-ordinate system. The importance of pose representation can be viewed from two aspects:

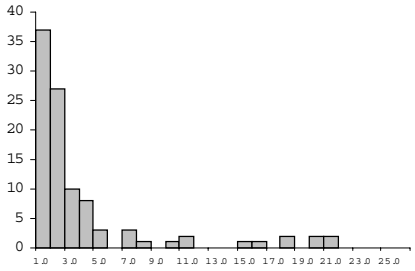


Fig. 4.3. Histogram of distance distribution in terms of statistical distances between $\mathbf{g}_{m,2}$ and the trials generated from the same prototype.

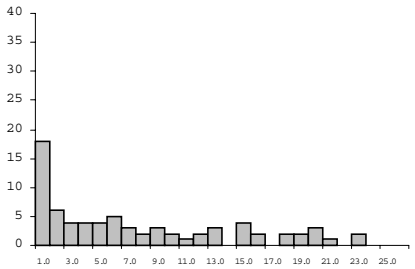


Fig.4.4 Histogram of distance distribution in terms of Euclidean distances between $\mathbf{g}_{m,2}$ and the trials generated from the same prototype.

Table 4.1. Error ratio of classification with statistical distance and Euclidean distance. Inputs were generated from the 4 prototypes with 50 trials. The noise level is σ=0.05.

correspondence	*1*	*2*	*3*	*4*
Using $(d_{std})^2$	0.1	0.14	0.1	0
Using $(d_{euc})^2$	0.54	0.66	0.5	0.34

Table 4.2. Error ratio of classification with statistical distance and Euclidean distance. Inputs were generated from the 4 prototypes with 50 trials. The noise level is σ=0.02.

correspondence	*1*	*2*	*3*	*4*
Using $(d_{std})^2$	0	0	0	0
Using $(d_{euc})^2$	0.26	0.78	0.52	0.24

Table 4.3. Mean value of statistical distances $(d_{std})^2$ between $\mathbf{g}_{s,i}$, $i\in\{1,\ldots,4\}$ and $\mathbf{g}_{m,j}$ $j\in\{1,\ldots,4\}$. The noise level is σ=0.05.

	$\mathbf{g}_{m,1}$	$\mathbf{g}_{m,2}$	$\mathbf{g}_{m,3}$	$\mathbf{g}_{m,4}$
$\mathbf{g}_{s,1}$	1.42	10.84	44.84	76.09
$\mathbf{g}_{s,2}$	12.45	1.66	13.72	38.60
$\mathbf{g}_{s,3}$	29.31	10.87	1.01	8.29
$\mathbf{g}_{s,4}$	64.71	34.91	9.62	1.14

Table 4.4. Mean value of Euclidean distances $(d_{euc})^2$ between $\mathbf{g}_{s,i}$, $i\in\{1,\ldots,4\}$ and $\mathbf{g}_{m,j}$, $j\in\{1,\ldots,4\}$. The noise level is σ=0.05.

	$\mathbf{g}_{m,1}$	$\mathbf{g}_{m,2}$	$\mathbf{g}_{m,3}$	$\mathbf{g}_{m,4}$
$\mathbf{g}_{s,1}$	0.31	0.24	0.33	0.48
$\mathbf{g}_{s,2}$	0.35	0.22	0.21	0.31
$\mathbf{g}_{s,3}$	0.56	0.32	0.10	0.11
$\mathbf{g}_{s,4}$	0.91	0.59	0.22	0.15

(1) In most applications, localisation of objects is required for robotic manipulations and is a task of machine vision;
(2) Pose description provides a geometric configuration of the spatial relations between two surface patches, which is often necessary for scene interpretation.

It should be noted that the latter argument can be utilised to approach the object representations. As mentioned in chapter 3, quadric representations are surface-based and neglect the volumetric properties of objects. Assuming that a bounded object can be decomposed into several quadric surface primitives, the *relational pose* transformations between these spatial correlated surface patches describe the topological structures of the object. When treating each surface patch as a "node" and the spatial relation between a pair of connecting surfaces as the "connection", then the object can be interpreted with a symbolic Attributed Relational Graph (ARG). Such kind of a relational graph belongs to the so-called *Relational Structure* (RS) representations ([17]) for object modelling. In this notion, the quadric representations can be consequently extended towards "volumetric" object representations. While the invariant feature of each single surface primitive describes the *unary* properties of the node, the descriptions of the relational pose between adjacent surface primitives can be utilised as the *binary* features to specify the joints between these nodes. The RS representation of object model is illustrated in figure 4.5.

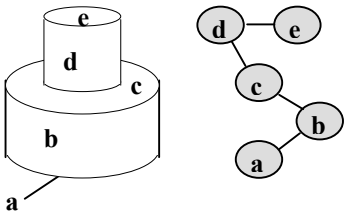


Fig. 4.5. The surface-based RS representation for object model.

This section discusses pose representation of quadric surfaces. Based on the pose representation of a quadric surface with respect to world co-ordinates, the relational pose between two surfaces is further formulated by defining a feature vector for pose description. Because the pose of a surface patch is recovered from the formulation of the quadrics, errors in surface fitting also lead to uncertainties in pose estimation. In the context of section 4.2 and 4.3, uncertainties in pose estimation are also described with a covariance model, by which the pose representation, in the sense of feature description, is optimised.

4.6.1 Formulation of pose representation

The pose of an object is usually described by rotation and translation transformations with respect to a pre-defined Cartesian co-ordinate system (the world co-ordinate system). As described in section 4.2, the advantage of applying quadric representations for the modelling of object surfaces is in the simultaneous recovery of the pose of a surface patch from the normalisation process for extraction of invariants, which is expressed by the rotation matrix $\mathbf{R}$ and the translation vector $\mathbf{t}$. Denoting with x_s-y_s-z_s the standard co-ordinate system and with $x_w y_w z_w$ the world co-ordinate system, the pose transformation is expressed by

$$\mathbf{x}_w = \mathbf{R}\mathbf{x}_s + \mathbf{t} = \mathbf{R}(\mathbf{x}_s + \mathbf{t}'), \tag{4.26}$$

where the vectors $\mathbf{x}_s$ and $\mathbf{x}_w$ denote the 3D co-ordinates with respect to the frames x_s-y_s-z_s and x_w-y_w-z_w, respectively. Because $\mathbf{R}$ consists of the three basic vectors of the normalised standard frame, there are only three degrees of freedom in the formation of $\mathbf{R}$. Thus the maximum number of degrees of freedom in pose description is six. To formulate the pose representation, the parameters to describe the rotation must be specified. Together with the translation vector $\mathbf{t}$, the pose parameters can be defined in a vector form.

Naturally, the rotation parameters can be selected as the three rotation angles about the (world) co-ordinate axes x_w-y_w-z_w, denoted as α, β and γ. Thus the rotation matrix $\mathbf{R}$ is expressed by

$$\mathbf{R} = \begin{bmatrix} \cos\alpha & -\sin\alpha & 0 \\ \sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \cos\beta & 0 & \sin\beta \\ 0 & 1 & 0 \\ -\sin\beta & 0 & \cos\beta \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\gamma & -\sin\gamma \\ 0 & \sin\gamma & \cos\gamma \end{bmatrix},$$

Here the angles are defined in the range $0 \le \alpha, \beta, \gamma \le 2\pi$. To distinguish the rotation and the translation, we define a vector $\boldsymbol{\mu}=[\alpha, \beta, \gamma]^T$ to represent the rotation transformation. For convenience of computation, we define the translation vector as $\mathbf{t}'=\mathbf{R}^T\mathbf{t}$, instead of $\mathbf{t}$ in (4.26).

Consequently, the pose parameters represented in the world co-ordinate can be defined as

$$\boldsymbol{\tau}_w = [\alpha, \beta, \gamma, t'_x, t'_y, t'_z]^T \tag{4.27}$$

In translation representation, a special case is formed by cylindrical surface because the origin of the standard co-ordinate system of a cylindrical form is non-fixed. According to the normalisation process for the extraction of invariants, a cylindrical surface is ideally expressed in the standard form as $\lambda_x x^2+\lambda_y y^2+1=0$. So the diagonal element λ_z of the matrix $\mathbf{A}'$ in (4.4), associated to the term of z is zero. According to the normalisation process stated by (4.3)-(4.10), $\mathbf{t}'$ is computed by $2\mathbf{t}'\mathbf{A}'+\mathbf{v}'^T=0$, so the value of t'_z can be arbitrarily selected. As a result, there are only two deterministic parameters to represent the parameter vector $\mathbf{t}'$. In other words, the 3D point represented by vector $\mathbf{t}'$ degenerates to a straight line. In such a case, the pose vector $\boldsymbol{\tau}_w$ also degenerates to

$$\boldsymbol{\tau}_{wc} = [\alpha, \beta, \gamma, t'_x, t'_y]^T \tag{4.28}$$

In reality, the eigenvalue λ_z can not be ideally zero, because of the uncertainties occurring in the parameter estimation by surface fitting, even when the object is in fact a cylinder. However, the special case should be distinguished because too small (absolute) value of λ_z not only yields an unreliable solution of t'_z, but also might cause numerical difficulties in computation since the domain of translation is infinite. In reality, identification of a cylinder is possible by comparing λ_z with a threshold. If it is lower than the threshold, it can be treated as a perturbation from zero and the cylindrical constraint is imposed. Alternatively, in using invariants for model-based recognition, success of matching to a model can also provide the a priori knowledge of the surface type of the observed instance.

To suffice the descriptions of man-made objects, we have to deal with planar surfaces. This simple surface type is not included in the discussions of surface representation of

chapter 3. Because either curved or planar surfaces are extracted in the segmentation process, the approach of planar fitting is described in the next chapter, which is related to the segmentation algorithm. In reality, because a planar representation does not possess unary features for shape identification, only the pose of a plane can be utilised to describe the spatial relations between it and other surfaces. For compactness, we conduct the pose definition of a planar surface in this section as a supplement.

Suppose a plane is expressed by the implicit form $\mathbf{n}^T\mathbf{x}$-k=0, where $\mathbf{n}=[n_x,n_y,n_z]^T$ is the unit vector representing the surface normal and k is the distance between the origin of the co-ordinate system to the plane. Considering that there are two degrees of freedom in $\mathbf{n}$, we simply define a pose vector $\boldsymbol{\rho}_w$ with respect to the world co-ordinate system by

$$\boldsymbol{\rho}_w = \left[\bar{n}_x, \bar{n}_y, \bar{k}\right]^T \tag{4.29}$$

with

$$\bar{n}_x = \frac{n_x}{n_z}; \quad \bar{n}_y = \frac{n_y}{n_z}; \quad \bar{k} = \frac{k}{n_z}.$$

4.6.2 Representation of relational pose between two surfaces

The relational pose representation is used to parameterise the spatial constraint between two surfaces. It is defined as the pose transformation between the standard frames of two quadric surface representations. To approach the RS representation of a 3D object, the relational pose can be employed as the *binary* feature description, while the invariants of a single surface is the *unary* feature description.

Let $(\mathbf{R}_a,\mathbf{t}'_a)$ and $(\mathbf{R}_b,\mathbf{t}'_b)$ denote the pose transformations between the world co-ordinate frame and the standard frame of two surfaces a and b, respectively. Then:

$$\mathbf{x}_w = \mathbf{R}_a(\mathbf{x}_a + \mathbf{t}'_a)\,;\ \mathbf{x}_w = \mathbf{R}_b(\mathbf{x}_b + \mathbf{t}'_b)\,. \tag{4.30}$$

When using the standard frame of a as the reference co-ordinate system, the pose transformation between a and b can be derived from (4.30) as

$$\mathbf{x}_a = \mathbf{R}_a^T\mathbf{R}_b(\mathbf{x}_b + \mathbf{t}'_b) - \mathbf{t}'_a \tag{4.31}$$

Now defining

$$\mathbf{R}_{ab} = \mathbf{R}_a^T\mathbf{R}_b \ \text{ and } \ \mathbf{t}'_{ab} = \mathbf{t}'_b - \mathbf{R}_{ab}^T\mathbf{t}'_a\,, \tag{4.32}$$

we obtain

$$\mathbf{x}_a = \mathbf{R}_{ab}(\mathbf{x}_b + \mathbf{t}'_{ab})\,. \tag{4.33}$$

Thus (4.33) is of the same structure as the one in (4.26). Therefore, the relational pose can be defined with the vector $\boldsymbol{\tau}$ in the same form as (4.27), but the rotation angles and translation components are derived from $\mathbf{R}_{ab}$ and $\mathbf{t}'_{ab}$. To avoid confusion, we use the notation $\boldsymbol{\tau}$ to

represent the relational pose and $\boldsymbol{\tau}_w$ to represent the pose of a single surface with respect to the world co-ordinate system. Obviously, $\boldsymbol{\tau}_w$ can be treated as a special case when the standard frame of surface *a* is identical to the world co-ordinate system.

Analogous to the derivation of (4.28), the relational pose in case of cylindrical expression of one of the two surfaces or both is defined specifically. Assuming that *b* is expressed with a cylindrical form, the relational pose is consequently defined by (4.28), where the parameters are derived from $\mathbf{R}_{ab}$ and $\mathbf{t}'_{ab}$ in (4.33). Another degenerated, but perhaps frequently encountered, case occurs when both *a* and *b* are cylindrical surfaces. Since both $t'_{a,z}$ and $t'_{b,z}$ are underdetermined in that case, from (4.32) and (4.33), we see that $\mathbf{t}'_{ab}$ can not be derived as a point vector, but can only be configured with a single constraint when eliminating the two unknown parameters. To be clear, we express $\mathbf{t}'_{ab}=[t'_x,t'_y,t'_z]^T$ in (4.32) with the following equations:

$$\begin{bmatrix} t'_x \\ t'_y \\ t'_z \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} \begin{bmatrix} t'_{bx} \\ t'_{by} \\ t'_{bz} \end{bmatrix} - \begin{bmatrix} t'_{ax} \\ t'_{ay} \\ t'_{az} \end{bmatrix} \tag{4.34}$$

Treating t'_{az} and t'_{bz} as two variables, (4.34) states a surface representation in the space of $\mathbf{t}'$ in an explicit form. By eliminating the two variables t'_{az} and t'_{bz}, (4.34) can be converted into an implicit form as

$$\mathbf{u}^T\mathbf{t}' = l\,, \tag{4.35}$$

where $\mathbf{u}$ is a coefficient vector consisting of only rotation parameters; l is a constant determined by both the rotation parameters and the known (t'_{ax}, t'_{ay}, t'_{bx}, t'_{by}). It means that in case of two cylindrical surfaces, their relational translation $\mathbf{t}'$ is only configured by a plane equation. Consequently, we define a one-dimensional parameter $t_l=l$ to represent the mapping from $\mathbf{t}'$ to this planar constraint. In fact, t_l corresponds to the geometric distance between two straight lines, i.e., the central axes of these two cylinders.

Hence the relational pose in the case of two cylindrical representations is defined as

$$\boldsymbol{\tau}_{cc} = [\alpha,\beta,\gamma,t_l]^T \tag{4.36}$$

To deal with planar surfaces, the relational pose should be defined specifically. Suppose that one of the two surfaces, e.g., surface *b*, is planar, the relational pose is just the expression (4.29) under the standard co-ordinate system of *a*, i.e.,

$$\boldsymbol{\rho}_a = [\bar{n}_{ax},\bar{n}_{ay},\bar{k}_a]^T \tag{4.37}$$

The components in (4.37) can be derived from the normal vector of plane *b* and the rotation matrix of surface *a* with respect to the world co-ordinate system. Detailed expressions of (4.37) are formulated in Appendix C.2.

A further degenerated case is the one where surface *a* is cylindrical and *b* is planar. In this case, the parameter k_a is unknown, thus the definition of (4.37) degenerates to be

$$\boldsymbol{\rho}_{ac} = [\bar{n}_{ax},\bar{n}_{ay}]^T \tag{4.38}$$

If both *a* and *b* are planar surfaces, their relational pose is only described by the angle between the two normal vectors, i.e., we define

$$\boldsymbol{\rho}_{nn} = \mathbf{n}_a^T \mathbf{n}_b \,, \tag{4.39}$$

to represent the relational pose of the two planes.

4.6.3 Descriptions of uncertainties in pose estimation

Following the principle applied to the formulation of the uncertainties of the estimated invariants, we can also use a covariance model to describe the uncertainties in the pose estimation. We first cope with the uncertainty problem of pose estimation for a single quadric surface in the world co-ordinate system. Then by applying linear approximation, the covariance expression with respect to the relational pose of two surfaces will be derived.

Suppose that $\hat{\boldsymbol{\tau}}_w$ is the estimate of the pose parameters of a single surface in quadric fitting, and $\boldsymbol{\tau}_w$ is the ideal parameter vector. According to (4.27), the error of $\hat{\boldsymbol{\tau}}_w$ is expressed by

$$\Delta\boldsymbol{\tau}_w = \left[\Delta\alpha, \Delta\beta, \Delta\gamma, \Delta t'_x, \Delta t'_y, \Delta' t_z\right]^T .$$

Because the components of $\hat{\boldsymbol{\tau}}_w$ are derived from the estimates of surface parameters $\hat{\boldsymbol{\theta}}$, $\Delta\boldsymbol{\tau}_w$ is uniquely determined by the errors of $\hat{\boldsymbol{\theta}}$. However, it is not convenient to formulate directly the expression of $\Delta\boldsymbol{\tau}_w$ through the dependency on $\Delta\boldsymbol{\theta}$, because of its non-linearity. In stead, we derive the formulation of $\Delta\boldsymbol{\tau}_w$ in the same way as used for the analysis of $\Delta\boldsymbol{\theta}$.

It should be noticed that in the normalisation process the invariants, i.e. the definition of feature $\mathbf{g}$ with respect to the equation of (4.1), are defined as the algebraic coefficients represented in the transformed "standard frame", and the transformation, i.e., the definition of $\boldsymbol{\tau}_w$, is represented by the pose parameters. This process can be interpreted as a one-to-one mapping between the quadric parameters $\boldsymbol{\theta}$ observed in the world co-ordinate system and the feature $\mathbf{g}$ together with the pose parameters $\boldsymbol{\tau}_w$, i.e., $\boldsymbol{\theta} \Leftrightarrow \mathbf{g} \cup \boldsymbol{\tau}_w$. Therefore, the quadric surface can be equivalently represented by the parameter vector defined as $\boldsymbol{\eta}=(\mathbf{g}^T, \boldsymbol{\tau}_w{}^T)^T$. While the estimate $\hat{\boldsymbol{\theta}}$ is obtained through surface fitting, $\hat{\boldsymbol{\eta}}$ minimises the cost function at the same time. Alternatively, the estimation is converted as

$$\hat{\boldsymbol{\eta}} = \arg\min\left\{\sum_{i=1}^{n} d_{s,i}^2\right\}$$

with

$$d_{s,i}^2 = w_{s,i} f_s^2(\mathbf{x}_{s,i}(\boldsymbol{\tau}_w), \mathbf{g})$$

defined in the standard co-ordinate system with the standard form

$$f_s(\mathbf{x}_s, \mathbf{g}) = x_s^2 + \gamma_1 y_s^2 + \gamma_2 z_s^2 + c = 0 \tag{4.40}$$

Therefore, we can apply the same analysis used in section 4.2 to derive the covariance matrix of $\hat{\boldsymbol{\theta}}$, to obtain the linear approximation in dependency of $\Delta\boldsymbol{\eta}$ on $\{\Delta\mathbf{x}_i\}\forall i$. Then the covariance matrix $\boldsymbol{\Phi} = E[\Delta\boldsymbol{\eta}\Delta\boldsymbol{\eta}^T]$ is formulated. Since only the pose parameters are of concern, the covariance matrix of $\hat{\boldsymbol{\tau}}_w$, denoted $\mathbf{T}_w$, can be directly derived with the low-right sub-matrix of $\boldsymbol{\Phi}$.

To obtain the covariance matrix of the relational pose estimate $\hat{\boldsymbol{\tau}}$ with respect to surface *a* and *b*, we have to utilise the definition of the relational pose expressed in (4.32). By further applying the linear approximation, the error $\Delta\boldsymbol{\tau}$ can be expressed in terms of ($\Delta\mu_a$, $\Delta\mu_b$, $\Delta t'_a$, $\Delta t'_b$). Then the covariance matrix of $\hat{\boldsymbol{\tau}}$, denoted as $\mathbf{T}$, can be derived from the covariance matrices $\mathbf{T}_{wa}$ and $\mathbf{T}_{wb}$. To avoid trivialities in mathematics, the derivation of the covariance matrix of the relational pose is neglected in this chapter. More details of the discussions, including the degenerated situations for pose representation, are presented in Appendix C.3.

Once the uncertainties of pose estimates are explicitly expressed with a covariance matrix, we think that the main statistical performances of the pose estimate $\hat{\boldsymbol{\tau}}$ are configured by its covariance. Therefore, we can apply the probabilistic model of the normal pdf to describe the pose estimates, either for a single surface measurement or the relational pose measurement. Especially, the explicit pdf model for description of the relational pose estimate enables us to deal with the *binary* feature in object representation and recognition in an optimal framework as used for the *unary* feature of invariants of a single surface representation. Consequently, the RS representations of objects and ARG's-based matching scheme are systematically optimised.

4.7 Conclusions

In this chapter, we discussed the feature representation of quadric primitives, which are used as shape descriptors for surface-based recognition. In order to establish a systematic optimal recognition framework, we cope with the uncertainty problem in parameter estimation with a covariance representation, by which an optimal description of a quadric primitive becomes available as the combination of the estimate and its covariance matrix. Consequently, an analytical probabilistic representation with a normal pdf is proposed to characterise the statistical attributes of the estimates and enabling the application of the Bayesian criterion for optimal classification of surface primitives. The major points in our investigation are summarised below.

- **Extraction of invariants of quadric primitives**

The feature vector of quadric surfaces was defined from the invariants of the implicit representation of quadrics. The invariants are presented as the standard form of quadrics, which describe the global geometric properties of the surface. A merit of using such feature definition is that the pose information can be recovered simultaneously through the extraction

of invariants, i.e. the normalisation process via transformation of rotation and translation from the viewer-centred frame to the standard frame.

- **Describing the uncertainties in parameter/feature estimation with the covariance representation**

Because of noise presence in data acquisition, errors exist in parameter estimation by surface fitting and uncertainties are conducted in feature extraction. This chapter aimed at modelling the uncertainties with an explicit representation, i.e., derivation of the covariance matrix of an estimate. To do that, we started from the formulation of the surface fitting problem and applied the linear approximation in the expression of errors of estimates. Under the assumption of Gaussian noise in range data, the computation of covariance matrix of the estimate in surface fitting, as well as the estimate of feature vector, were explicitly formulated.

Based on the optimal description of the features of surface primitives, we proposed a probabilistic model, i.e. the analytical normal pdf, to characterise the statistical attributes of the estimates. It should be noted that such an analytical model relies on the assumption of linear dependency of uncertainties of estimates and noise. However, it was also argued to be supported by the optimal characteristics of the estimates in surface fitting, i.e. the refined "bias-corrected" solution.

- **Optimisation in establishing the model base**

Assuming that the model base in terms of feature descriptions was established from measurements, the modelling process can be optimised as the outcome of applying the probabilistic model for feature estimates. Given artificially trials of measurements generated in the modelling process, the optimal descriptions in the model base were obtained under the optimal criterion of MLE.

- **Pose/relational pose representations**

The last part of this chapter discussed the problem of pose representation. More of our interests, the relational pose between two surface primitives was characterised in an explicit expression. By the normalisation process for extraction of invariants, pose parameters, represented under the viewer-centred co-ordinate system, were first derived. Then the relational pose, describing the geometric connectivity between two surface primitives, was formulated. The aim of formulating the relational pose is to extend the surface-based primitive description to represent more complicated objects through an object representation scheme of RS (relational structure) representation. By regarding the features of surface primitives as the "unary" features and the relational pose as the "binary" features, the object recognition can be implemented with the scheme of ARG (symbolic Attributed Relational Graph). Because of the simplicity of quadrics in mathematical manipulations, it is expected to arrive at a sophisticated object recognition strategy as a trade-off between the complexity of representation in low and intermediate level and the complexity of control in higher level. Furthermore, the approach to represent the uncertainties in pose estimates was also discussed.

As an application of modelling of uncertainties in surface parameter estimation, segmentation of a scene from range data can be implemented based on an optimal description of the surface primitive, using the region-based approach. Because the success of segmentation in region-based approaches rely on the reliability of parameter estimation of surface regions, the optimal description of parameter estimates enables us to carry out the primitive clustering algorithm within an optimal framework. This is discussed in chapter 5.

Chapter 5

Segmentation of Range Images Based on the Optimal Description of Surface Primitives

This chapter discusses the segmentation of range images, which is an essential low-level task for image analysis and object recognition. Once having obtained the range image with the ranging system, segmentation is accomplished to build up partitioned primitives of the scene. The primitives are then classified by matching with the models of primitive descriptions in the subsequent recognition stage. Since the object recognition scheme in this thesis is obtained through surface-based representations, the purpose of segmentation in this chapter is to partition the input raw range data into a group of regions each of which is described with a quadric (or planar) representation.

Generally, segmentation of (greylevel or range) images can be defined as a process of partitioning a given image into a set of meaningful regions. A meaningful region represents a region in which all pixels (or voxels) possess similar properties. Then symbolic descriptions can be made of the partitioned regions, to be used by higher level modules. Critical in segmentation is the definition of "similarity" or "homogeneity" of partitioned regions. It is often impossible to define the homogeneous property purely from a strict mathematical context, because the requirements to the result of segmentation are usually guided by the subsequent recognition module in machine vision, or by the visual interpretation of human beings. For 2D greylevel images, the homogeneity criterion is based on segments of "visually pleasing regions", i.e., segments of the image are in accordance with our perceptual experience about the pattern of illumination. But such intuitive visual judgement is meaningless for 3D data. In fact, in segmentation of range images, the partitioning principles not only depend on the nature of the input image, but also on the kind of symbolic representations of the objects, because any segmentation task must be put in perspective with the final objective of a vision system.

According to the surface-based object representations discussed in the previous chapter, this chapter applies to a region-based algorithm in segmentation. A homogenous region is thought to be a piece of the surface that fits either a planar or a quadric representation. Thus the output of the segmentation is a group of surface primitives that are ready to pass through the subsequent process of object recognition using surface-based representations.

The segmentation approach proposed in this chapter consists of three modules. First, outlier detection is applied to the input range image by testing the local properties of the data. In this process, data points near the boundaries or edges of objects, as well as the

sparse distant points due to noise, are extracted and marked as outliers. These data points thus are excluded from the region-growing process of the second module. Second, the whole range image is partitioned into small regions with a grid pattern. Then these over-segmented regions are merged in an iterative region-growing scheme. In this process, the merging criterion is based on an optimal estimation-based clustering algorithm in parameter space. Finally, based on the results of the segmentation of the second module, a point-based refinement is applied in the third module, which yields the labelled final segments and unlabeled points. This approach contains the assumptions that the objects involved in the scene are piecewise smooth and that each segment of the surfaces can be described by either a quadric or a plane. Being based on the optimal description of the surface primitives, the segmentation is approached within an optimal framework.

This chapter is organised as follows. Section 5.1 briefly reviews related work on segmentation of range images. Restricting to the class of man-made objects only, in section 5.2 an approach for region-based segmentation for such objects is proposed, based on an optimal criterion for homogeneity. Considering that planar surfaces are frequently encountered in man-made objects, the segmentation algorithm deals with planar surface estimation separately. Section 5.3 discusses the method of separation and parameter estimation of planar surface regions. Finally, discussions and conclusions are given in section 5.4.

5.1 Background of segmentation of range images

For a recognition system, segmentation is a very important and perhaps the most difficult task. One should note that in this work, pixels are the greylevel representations of 3D distance measurements, which will have its consequences for the development of a segmentation algorithm.

Typical segmentation difficulties are described in [33]. According to the assessment of the state-of-the-art in planar range image segmentation, it is asserted here that this problem is “not solved”. There is no universal solution of segmentation for all instances of images, because of the possible complexities of the scene, neither for 2D greylevel images nor for 3D range images. An important reason is the lack of a unified metric to assess the segmented outputs. On the other hand, segmentation is often followed by higher level processing, so the algorithm of segmentation should usually be implemented taking into account the available a priori information and guided by the strategy of description and recognition of a vision system.

Conventional methods of range image segmentation can be classified into two approaches: region-based and edge-based. In region-based methods, pixels having similar properties are grouped together and finally the images are partitioned into a set of “homogenous” regions. This is often implemented by the region growing technique, in which a seed of region is expanded by checking neighbours with a defined homogeneity criterion through an iterative process ([46], [31], [8], [59] and [32]). A homogenous region is usually described by some uniformity properties such as planar, quadric, convex or concave, etc., based on the assumption that the object surface is piecewise smooth. One of the representative works was described by Besl and Jain ([8]). In their method of variable-order fitting,

any piecewise smooth surface is partitioned into a few fundamental primitives based on differential geometry, then the simply described surfaces pass through a fitting stage using variable-order polynomials.

Another widely used segmentation scheme is based on edge-based methods. The edge-based methods obtain the segments from edge descriptions of images. Some researchers combine both region-based and edge-based techniques for segmentation of range images ([47], [65] and [9]). In the past decade, robust estimation techniques have been applied for segmentation of range images, which can be categorised as primitive extraction through robust estimation ([66] and [13]).

In region-based segmentation of range images, homogeneity of a region needs to be defined. Generally this is based on exploration of the properties of each pixel and its neighbours, i.e., it is assumed that range image data embed surface coherence ([8]). The definition of homogeneity properties of a region can vary depending on different surface descriptions. In many cases, criteria such as "the same sign of curvature" of a smooth surface could yield segments being intractable for high level description because of the shape variations of smooth surfaces. However, when surface-based object representation is utilised in intermediate and high level vision tasks, homogeneous regions can be defined by parametric primitives, such as the quadric description as used in this thesis. Using parameterised models, a homogenous region is thought to "fit well" the parametric description whereas two regions are thought to be homogenous if they can be expressed as two observations of the same surface primitive. Commonly used parametric models in segmentations are piecewise polynomials ([8], [58] and [52]).

Since region-based approaches are guided by the detection of similarities of data subsets, techniques for clustering of data subsets are required. Basically, a homogeneity test determines the process of clustering. Many segmentation algorithms employ some threshold parameters for the purpose of clustering, which are usually set by training to account for the noise in the data and errors in the estimation of homogeneity properties ([33]). Naturally, for parameterised models, homogeneity between two sub-regions is inferred from parametric descriptions. However, the reliability of the homogeneity test by estimation-based algorithms could suffer from the lack of explicit representation of uncertainties of region estimates. In [42], a Bayesian segmentation methodology is proposed, using parametric models. In that approach, instead of direct parameter estimation, a so-called "probability of homogeneity" is derived from the a posteriori probability conditioned by two joint observations in the spatial domain. Thus the uncertainties in parameterisation of region primitives are considered in the homogeneity test in an implicit way. Although that methodology provides an optimal test of homogeneity between regions using parametric models (this requires integration among the multidimensional parameter manifold), it is still unknown how a homogeneous region "fits" the used model, because no parameters are estimated in that approach. However, once the parameter uncertainties are known explicitly, an estimation-based optimal clustering algorithm can be implemented, with equal optimality of the approach of [42]. Moreover, the "degree of fitting" of a homogeneous region to a parametric model is represented as well.

Following the surface-based object representation used in this thesis, we apply the region-based segmentation algorithm using the parametric model; this is described in section

5.2. Because the uncertainties in surface estimation have been represented with the approach in chapter 4, an optimal estimation-based clustering criterion in parameter space can be established.

5.2 The algorithm of the optimal region-based segmentation

As stated previously, the major objective of the vision system concerned in this thesis is to recognise and localise man-made objects, where in most cases the objects can be represented mathematically by quadrics. Naturally, the parametric primitives to characterise the region homogeneity are selected as quadric or planar descriptions. The representations for surface primitives in the proposed approach are the same as in [36], i.e., second-order polynomials. But the segmentation algorithm proposed here is different from that of [36] which was in fact based on edge extraction (although the surface representations were employed). As we will see below, the approach proposed in this chapter is driven by surface representations (and surface data points). The key point of the method is the *optimisation* in region-based parameter estimation. A flowchart of the segmentation approach is depicted in figure 5.1.

The new method consists of three modules. In the first stage, outliers (including edge or boundary points and distant error points) are extracted. This is achieved by fitting the data points in a small sampling window to a planar description and testing the goodness of fit. The central point of the window whose planar fitting is recognised as "bad" is marked as an outlier. By moving the sampling window over the whole range image, centred at each pixel each time, we extract the outliers of the input range image. In the second module, the whole image is initially partitioned into small regions with a rectangular grid pattern. According to [52], these initial grid regions are noted as "surface elements", or *sels*. These sels then pass through the merging process, according to the criterion of homogeneity. The merging process starts with a seed of subregions consisting of a few neighbouring sels. Such a group of neighbouring sels is called a *patch*. When the growing of a patch stops, a new seed patch is selected and the process repeats until no more seeds can be found. At the end of the second module, the merged regions form several isolated segments with different labels in the range image. Such a segment is denoted as *dominant region*. In the last module, a process of point-based refinement is applied, with which the points near the boundary of dominant regions are re-classified according to a measure of the distance between the point and the estimated surface.

In the process of computing the seed patch, a planarity test is applied according to the result of the quadric fitting of the seed patch. If the seed turns out to be planar, then it is represented as such and the updated region is estimated using the planar representation in the subsequent merging process. Otherwise, the seed and its updated region are estimated through a quadric representation.

Finally, the outcomes of the segmentation include the labelled regions, either in quadric or planar representations, and the unlabeled points.

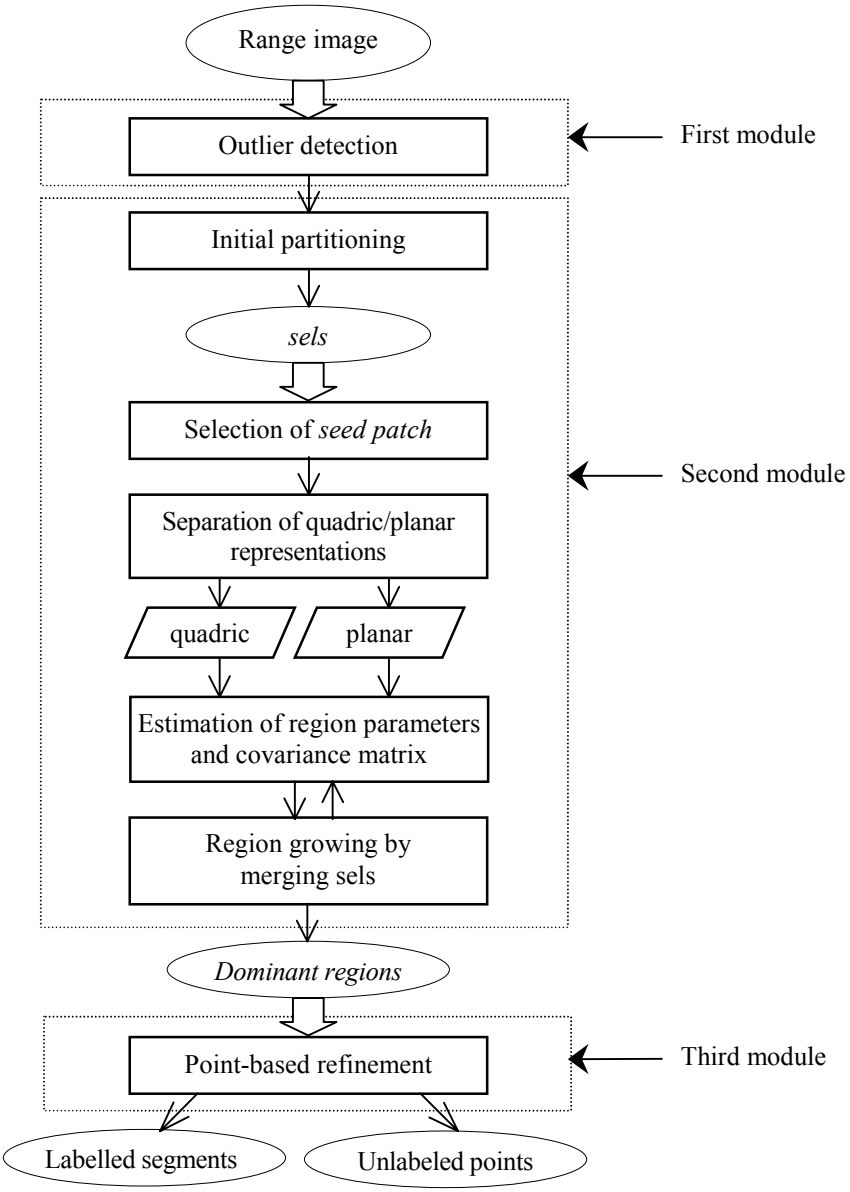


Fig 5.1 Flowchart of the segmentation algorithm

5.2.1 Outlier detection

To hold the condition of piecewise smoothness for surface fitting, boundary or edge points of the objects, as well as the distant data points, are detected and excluded from surface fitting. To do that, we model the points within a local region as a set of 3D random data points and the outliers are detected according to the pattern of their 3D distributions.

Given the raw input range data, a small rectangular window is used as the mask to test the local property of each point. The size of the window is pre-defined, according to the sampling resolution of the range data as well as the size of the objects in consideration. As an example, the window size used in our experiments is 5×5 (point^2). The window is centred at each point and all points covered by the window are considered. Because these points are locally distributed, they are hypothesised coming from a planar patch of the smooth object surface and the result is subjected to a planarity test. Obviously, if the window covers the boundary and partially embeds different object surfaces or different objects, or contains any local irregularity of an object, this region will fail to pass the planar fitting test. Then the centre point of the window is considered as an outlier.

Treating the points sampled within the window as 3D random points, we can use the algorithm of principal-component extraction to describe the spatial variance of these points for surface fitting. Supposing $\{\mathbf{x}_i\}$ $\{i = 1,..,m\}$ are the points of the window, the matrix

$$\mathbf{Z} = \frac{1}{m-1}\sum_{i=1}^{m}(\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T \text{ , where } \bar{\mathbf{x}} = \frac{1}{m}\sum_{i=1}^{m}\mathbf{x}_i ,$$

is an unbiased estimate of the covariance of the set $\{\mathbf{x}_i\}$. The three eigenvectors of $\mathbf{Z}$ are called the *principal components* and the eigenvalues represent the variance at these principal directions. Naturally, if $\{\mathbf{x}_i\}\forall i$ belongs to a plane in the spatial domain, the smallest eigenvalue is ideally zero and its corresponding eigenvector is just the normal vector of that plane. Therefore, the smallest eigenvalue of $\mathbf{Z}$, denoted $c_{\min}$, can be used as a measure to characterise the uniformity of these points. If $c_{\min}$ exceeds a threshold, then this patch is thought to be inhomogeneous and the centre point is classified as outlier.

For different range data, the predefined threshold, denoted T_{win}, should vary according to different noise levels. Commonly in robust estimation, the threshold is set by

$$T_{win} = r_{win} \cdot \bar{c}_{\min} ,$$

where $\bar{c}_{\min}$ is the median of $c_{\min}$ for all of the windows in the range data, and r_{win} is a chosen constant. If $c_{\min} \geq T_{win}$, the centre point is classified as outlier, otherwise it is a normal point.

At the same time, another criterion is used to test the irregularity of the centre point. If the difference in depth values between the centre point and its nearest neighbouring point exceeds a pre-defined threshold, this centre point is thought to be outlier. The threshold is chosen as $T_{dept} = r_{depth} \cdot \bar{c}_{\min}$, where $\bar{c}_{\min}$ is the same as used above, and r_{depth} is a selected constant.

Moving the window over the entire range image, centred at each point, the outliers are extracted using the above two criteria. Examples of the detected outliers are shown in fig-

ure 6.6 (b) in our experiments (chapter 6), where the detected outliers are displayed highlighted (white points).

Although there are more refined methods for the outlier detection, such as those used for edge-based segmentation algorithms, the outlier detection in the first module is not critical for the proposed region-based segmentation approach. As we can see in the examples of experiments described in chapter 6, sometimes it was difficult to find all the edge points near the connection part of two surfaces in case of a "roof" edge. The segmentation still succeeded in the second module, i.e. the region-based clustering process. The first module provides a preliminary description of the "inhomogeneity" of the scene. More important, a successful outlier detection reduces the risk of falsely merging two inhomogeneous regions, as well as the computational burden of surface fitting in the region-growing process. The extracted outliers will be re-evaluated in the third module.

5.2.2 Estimation-based region-growing scheme

After detection of outliers, the range image passes through the estimation-based region-growing stage. Initially, the range image is manually partitioned into a set of small subregions using a rectangular grid pattern. Then the region-growing process starts with this over-segmented small partitions, i.e., the "surface elements" or *sels*. Selecting the size of a sel is done manually. On one hand, it should be small enough so that points from a smooth surface lying within the sel are thought to be homogeneous and planar descriptive. On the other hand, a too small size of the sel might cause difficulties in parameter estimation at the beginning of growing because of insufficient shape information in a very small patch. Moreover, the smaller the sels, the more iterations the updating process takes and the more computational time will be required. In practice, the size of the sel can be set about the same as that of the window used in outlier detection.

If the number of outliers in a sel exceeds a threshold, it is marked as "ambiguous", meaning that it may cover part of the object boundary or local irregularities. Others are termed as "ordinary" sels. Only the ordinary sels participate in the region-growing process. For each of these ordinary sels, the local variance of the noise is estimated by applying the computing algorithm used in outlier detection, i.e., the variance of noise is estimated with the value of $c_{\min}$.

At the beginning of the region growing, a subregion should be selected as a seed; this subregion is noted as *patch*. The patch is an array of neighbouring sels with a selected number. In our approach, the type of array is selected in notion of a second order *neighbourhood system*, treating the grid-pattern sels as "pixels" (see figure 5.2). So the maximum size of a patch is nine sels and the minimum size is set to six, which is a subset neighbouring with each other.

Starting with such a seed patch, the region grows by merging it with the ordinary sels that are neighbours of the current region and the region is updated in an iterative way. This process is controlled by the criterion of the homogeneity test, which can be interpreted as the clustering principle of the sels. The clustering principle is based on the *hypothesis test* implemented in parameter space. The uncertainties in parameter estimation have been ex-

plicitly represented, as discussed in chapter 4. The test criterion is established based on the Bayesian principle as worked out in this chapter.

Denote the current region at the k^{th} iteration as $R^{(k)}$. For a neighbouring sel S, the merged region is represented as

$$R^{(k+1)} = R^{(k)} \bigcup S .$$

Suppose ($\boldsymbol{\theta}_k$, $\boldsymbol{\Psi}_k$) are the estimates of parameters for the current region $R^{(k)}$, where $\boldsymbol{\theta}$ denotes the surface parameter vector and $\boldsymbol{\Psi}$ denotes the covariance matrix. After merging the sel S, the estimates for $R^{(k+1)}$ are ($\boldsymbol{\theta}_{k+1}$, $\boldsymbol{\Psi}_{k+1}$). Assuming that $R^{(k)}$ and S belong to the same surface that can be represented with their primitive representation, then the estimates of the surface parameters $\boldsymbol{\theta}_k$ and $\boldsymbol{\theta}_{k+1}$ are two observations of the same primitive.

Assuming that the expectations of the two estimates are $\overline{\boldsymbol{\theta}}_k$ and $\overline{\boldsymbol{\theta}}_{k+1}$, respectively, the hypotheses are stated as:

H_0 : $\quad \overline{\boldsymbol{\theta}}_k = \overline{\boldsymbol{\theta}}_{k+1}$ $\qquad$ ($R^{(k)}$ and S are homogeneous)

H_1 : $\quad \overline{\boldsymbol{\theta}}_k \neq \overline{\boldsymbol{\theta}}_{k+1}$ $\qquad$ ($R^{(k)}$ and S are inhomogeneous)

Applying the covariance model proposed in chapter 4, both estimates $\boldsymbol{\theta}_k$ and $\boldsymbol{\theta}_{k+1}$ are described with a normal pdf. We define

$$\boldsymbol{\varepsilon}_{k,k+1} = \boldsymbol{\theta}_k - \boldsymbol{\theta}_{k+1} .$$

When hypothesis H_0 holds, the same way as used to derive (4.19), $\boldsymbol{\varepsilon}_{k,k+1}$ is also expressed with the normal pdf as

$$f(\boldsymbol{\varepsilon}_{k,k+1}) = C \exp(-\tfrac{1}{2}\boldsymbol{\varepsilon}^T_{k,k+1}\boldsymbol{\Psi}^{-1}_{k,k+1}\boldsymbol{\varepsilon}_{k,k+1}) \; ; \; (\boldsymbol{\Psi}_{k,k+1} = \boldsymbol{\Psi}_k + \boldsymbol{\Psi}_{k+1}),$$

where C is the normalisation constant. Therefore, the "statistical distance"

$$d^2_{k,k+1} = \tfrac{1}{2}\boldsymbol{\varepsilon}^T_{k,k+1}\boldsymbol{\Psi}^{-1}_{k,k+1}\boldsymbol{\varepsilon}_{k,k+1} \tag{5.1}$$

can be used to test the hypothesis. If $d^2_{k,k+1} \geq T_d$, where T_d is a selected threshold, then H_0 is rejected and H_1 is accepted, meaning $R^{(k)}$ and S are inhomogeneous. Otherwise, they are thought to be homogeneous and merging of S with $R^{(k)}$ is acceptable. The threshold T_d is associated with the so-called *significance level*, denoted α, which can be expressed as T_d = t(α). The significance level α is just the probability of type-1 error, that is

$$P\{\text{rejecting } H_0 \,|\, H_0 \text{ is true}\} = \alpha .$$

Theoretically, the value of α can be chosen according to the design of the probability of the type-1 error and to the distinction of different object surfaces. However, as the covariance computation was derived only approximately, so in reality, the threshold T_d should be determined empirically.

Once the sel passes the test and is merged, the region $R^{(k+1)}$ is represented with the updated description ($\boldsymbol{\theta}_{k+1}$, $\boldsymbol{\Psi}_{k+1}$) and the next neighbouring sel is tested. If all of the neighbouring sels of the current region fail to pass the test, the growing process of the selected

seed stops. From the remaining ordinary sels, a new seed patch is selected and the growing process as described above repeats. The whole process is finished when no more seed patches can be found. At the end of this module, the input range image is segmented into several fundamental regions each of which consists of a set of sels, which are noted as the *dominant regions*. The rest of the range image includes "ambiguous" sels and some ordinary sels, which number of neighbourhood is less than the minimum requirement of a patch. Examples of the dominant regions after merging sels are shown in figure 6.6(c) of chapter 6.

5.2.3 New features of the estimation-based merging algorithm

In this section, we address some features of the region-growing process. .

5.2.3.1 Where to look next

In general, any segmentation implemented with the region-growing principle meets two fundamental questions:

(1) When does the growing stop?
(2) In which order should the candidate sels be chosen for merging with the patch?

The answer of the first question is in fact the criterion of the homogeneity test for region-based approaches. The second question, however, has not yet fully treated until now. An intuitive answer to the second question is that the points adjacent to the current region with the "best" output from the homogeneity test, are given highest priority for merging. For example, the sel that generates the smallest value of $(d_{k,k+1})^2$ could be taken into $R^{(k+1)}$ firstly. However, such a priority only comes from the "closeness" in the homogeneity test, but the "reliability" of the estimate of the homogeneous region itself is neglected. In other words, merging the "best homogeneous" points can not guarantee to yield the "best estimate" of the updated region $R^{(k+1)}$. This is because the uncertainties in parameter estimation not only depend on the noise level of the data points involved in the surface fitting, but also depend on the spatial locations of these points. This can be illustrated with the heuristic example of figure 5.3. Suppose that the current region $R^{(k)}$ is a rectangular portion of a cylindrical surface which length extends along the axis of the cylinder, and that both of the sels S_1 and S_2 pass the homogeneity test. If merging S_1 into $R^{(k)}$, usually the estimate of $R^{(k+1)}$ is less reliable than that of merging S_2 into $R^{(k)}$. Intuitively, the merging of S_1 results in an extension of the length of $R^{(k)}$ along the axis, so $R^{(k+1)}$ is still "flat" somehow (assuming its width is narrow compared with the radius of the cylinder) so the parameter estimation of $R^{(k+1)}$ is unstable. On the contrary, merging of S_2 will embed more information of the "curve" of the surface in estimating $R^{(k+1)}$, resulting in less uncertainty (assuming the noise level remains the same). Therefore, although S_1 could result in a smaller distance measure between $R^{(k)}$ and $R^{(k+1)}$, the sel S_2 should have higher priority for merging in view of the *reliability* of surface representation.

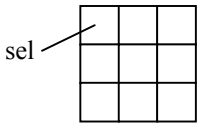


Fig 5.2 The array of sels consisting of the maximum-sized *patch*.

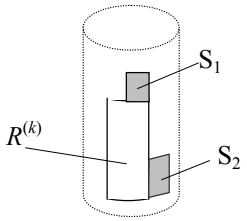


Fig 5.3 selection of the candidate sel for the next merging

In fact, from the expression of the statistical distance of (5.1), it can be seen that a large value of the covariance matrix $\mathbf{\Psi}_{k+1}$ results in a small value of $(d_{k,k+1})^2$. However, this is not equivalent to improve the reliability of the representation of the current region $R^{(k+1)}$.

According to the above argument, in the proposed approach, the merging priority in the region-growing process is determined by the *measure of uncertainties* in the estimate of the updated region. For that purpose, the *measure of uncertainties* is defined by the quantity

$$\textit{measure of uncertainty} \equiv \left\|\mathbf{\Psi}\right\|_2 .$$

Therefore, among all the neighbouring sels that pass the homogeneity test, the one that generates the smallest $\left\|\mathbf{\Psi}_{k+1}\right\|_2$ is merged first. In order to reduce the burden of parameter estimation for each candidate sel, in the experiments the value of $\left\|\mathbf{\Psi}_{k+1}\right\|_2$ is estimated only by approximation. In stead of re-computing the region parameters after merging each sel, the current estimate $\boldsymbol{\theta}_k$ is used for the computation of $\left\|\mathbf{\Psi}_{k+1}\right\|_2$, just by adding data points of each sel into the current group of points in $R^{(k)}$. Thus the sel with the smallest $\left\|\mathbf{\Psi}_{k+1}\right\|_2$ has the highest priority of passing through the homogeneity test. If it fails to pass the homogeneity test, the next is the second in the queue ordered by the value of $\left\|\mathbf{\Psi}_{k+1}\right\|_2$ from the smallest to the largest.

5.2.3.2 Selection of the seed patch

Although there is more than one candidate for the patch to be selected as the seed of growing, an optimal choice is still based on the reliability of representation in surface estimation. Following the same principle as discussed above, the patch with the smallest measure of uncertainty $\left\|\mathbf{\Psi}_{patch}\right\|_2$ is selected as the seed. Here $\mathbf{\Psi}_{patch}$ is the covariance matrix estimated in fitting data points in the patch to the quadric representation. It should be noticed that the selection of the proper seed patch is important in the region-growing approach. Since the estimates of parameters for the homogeneous region are updated while the patch expands, an improper choice of the seed patch could lead to an unreliable result in surface representation.

Proceeding the same way as for selecting the queued sels, the selection of the seed for the next region is also queued according to the value of $\left\|\boldsymbol{\Psi}_{patch}\right\|_2$ in the search of the rest candidates.

5.2.3.3 The computing process

The overall region-growing approach consists of the following steps:

1. Partition the range image into sels, and classify these sels as ambiguous or ordinary.
2. Find all possible patches that occur, by merging a set of neighbouring sels and estimate the quadric parameters together with their covariance matrices, by applying the surface fitting method described in chapter 3. From these patches, select the seed patch that has the smallest value of *measure of uncertainty*.
3. Starting from the seed patch, the region is updated by merging its neighbouring sels that pass the homogeneity test. The smaller the *measure of uncertainty* in merging the sel, the higher priority in the queue for merging. The region parameters are updated after each merge.
4. If all neighbouring sels of the current region fail to be merged, the growing process stops and the region is labelled.
5. For the remaining sels, repeat step 2 to 4 to generate newly labelled regions. The whole process ends if no more patch can be found.

At the end of the second module we obtain a set of labelled regions, i.e., the *dominant regions* that are formed with sels. But in reality, the partitions of the range image, representing different clusters in parameter space for surface representations, will not be exactly separated by the boundaries of the sels in the spatial domain. Points near the boundaries of these segments should be treated in a refined process, i.e., these sels should be split at point level and further classified. This is implemented in the third module, as described below.

5.2.4 Point-based refinement

There are two kinds of sels that should be considered to pass through the refinement process: merged and non-merged. The merged sels are those at the boundary between labelled *dominant regions*. In this situation, points in these boundary sels are re-classified as either their original segment or one of their neighbouring segments. The non-merged sels adjacent to a segment are split down to points and these points are further classified as surface points of the known segment or non-surface points. In both situations, the criterion for classification is based on the spatial "closeness" of a point to the specified surface. Typically, we can use the geometric distance between a point and a surface as the measure of such spatial closeness.

As worked out in chapter 3, estimation of surface parameters is based on the gradient weighted least-squares criterion. The solution is derived by minimising the sum of squares of distances from points to the estimated surface. Assuming i.i.d Gaussian noise added to each point, we have shown that the signed distance, i.e., the projection at the normal direction of the surface, can also be approximated as Gaussian noise with expectation of zero

and variance of σ^2 (σ is the standard deviation of noise in the data points). Denoting $d(\boldsymbol{\theta})$ the distance from point $\mathbf{x}$ to the surface which parameters are noted by $\boldsymbol{\theta}$, its pdf can be expressed by

$$f(d|\boldsymbol{\theta}) = Cexp\left(-\frac{d^2(\boldsymbol{\theta})}{2\sigma^2}\right).$$

According to the discussion in chapter 3, the distance $d(\boldsymbol{\theta})$ is computed by

$$d(\boldsymbol{\theta}) = \frac{h(\mathbf{x},\boldsymbol{\theta})}{\|\nabla h\|};$$

where $h(\cdot) = 0$ is the implicit quadric representation of the surface.

Therefore, identification of a surface point is still based on the hypothesis test algorithm. Given a significance level α_0, if

$$\frac{d^2(\boldsymbol{\theta})}{2\sigma^2} \le t(\alpha_0),$$

it is identified as a surface point with respect to surface of $\boldsymbol{\theta}$. Otherwise, it is identified as a non-surface point. Therefore, for the case of non-merged sels, a threshold is selected to set $t(\alpha_0)$. This threshold can be fixed according to the noise level of the whole range image. Considering that the noise estimates might be different for different segments, we select the threshold of $d^2(\boldsymbol{\theta})$ to be $T = r_{\text{point}} \cdot c_{\text{seg}}$, where c_{seg} is the estimate of noise variance for the concerned segmentation and r_{point} is a constant.

If there is more than one surface for which the criterion satisfies, then it becomes an optimal classification problem, in which case the point is classified as belonging to the surface to which the distance is the smallest. The above principle is also applied for the adjacent merged sels.

It should be pointed out that in this process, the outliers that are detected in the first module also participate in the classification. This is because some of the detected outliers are in fact surface points, but located near the boundaries or edges of the object. If the outlier is caused by a local irregularity, its distance to the surface of the segment will be large and it will be classified as a non-surface point. For simplicity, during the computation of the squared distance $d^2(\boldsymbol{\theta})$, the surface parameters $\boldsymbol{\theta}$ keep the values of the estimates from the last version in the region-growing process. After reclassifying those points, the surface parameters of the refined segment are re-computed.

Finally, the outcomes of the third module are the labelled final segments and the unlabeled points. In figure 6.6(d), the final segmentation results in our experiments are given.

5.3 Separation of planar regions

Most of the objects in the scene of interest are bounded by simple surface primitives: curved and planar surfaces. Therefore, the quadric description is expected to fit most of the

surface segments that are tractable for the purpose of recognition. For the planar surfaces, however, they are not suitable to be described by quadrics. Although a planar surface can also be represented by higher order polynomials, the representation may suffer from non-uniqueness. Supposing a planar surface is described by a plane equation in the form:

$$ax + by + cz + k = 0\,,$$

then if a quadric representation fits this plane, it must have the following form:

$$(dx + ey + fz + g)(ax + by + cz + k) = 0\,. \tag{5.2}$$

Obviously, any (non-zero) value of the parameters $\{d, e, f, g\}$ represents the same planar surface. In other words, quadric representations for planar surfaces are under-determined. Thus when applying quadric fitting for planar surface, theoretically the solutions are non-unique. In case of noise presence, there can be a single solution associated to the minimum, but the solution is heavily unstable because it is an intrinsically ill-posed problem.

To avoid this problem, in the second module, where the region parameters are estimated, we apply a routine of planarity test to identify the planar surfaces. Then the planar surfaces are estimated using the planar representation instead of quadric representation and their region-growing proceeds with planar representations. In the next section a method for planarity testing is introduced first, then the parameter estimation in planar representation, together with the description of uncertainties, are discussed.

5.3.1 Planarity test

According to the procedure for quadric fitting stated in chapter 3, the first step is to obtain an initial estimate of the quadric parameters using the eigenvector solution. If we transfer expression (5.2) into the general form of a quadric representation, i.e., in the form of the parameter vector $\mathbf{p} = [a_{11}, a_{22}, a_{33}, a_{12}, a_{13}, a_{23}, v_1, v_2, v_3, k]^T$, $\mathbf{p}$ does not uniquely represent a plane, due to the freedom in selecting $\{d, e, f, g\}$, as shown before. To find out when this situation will occur, a test method should be developed.

Suppose that $\{\mathbf{x}_i\}$ (i=1,..,n) are points, sampled from a planar surface and they are used to fit a quadric presentation

$$f(\cdot) = \mathbf{p}^T\mathbf{m} = 0\,, \text{ where } \mathbf{m} = \left[x^2, y^2, z^2, xy, xz, yz, x, y, z, 1\right]^T\,,$$

if using the fitting criterion

$$\mathbf{p} = \arg\min\{\sum_{i=1}^{n} f_i^2\}\,,$$

the solution of $\mathbf{p}$ is the eigenvector of the matrix

$$\mathbf{M} = \sum_{i=1}^{n} \mathbf{m}_i\mathbf{m}_i^T\,,$$

associated with the smallest eigenvalue. In the noise-free ideal case, the smallest eigenvalue of **M** is zero. However, since the solution of **p** is not unique, there are more than one eigenvectors with zero eigenvalue. Considering the normalisation requirement for the algebraic coefficients, there are four eigenvectors with zero eigenvalues. Therefore, when ordering the ten eigenvalues as $\lambda_1 \geq \lambda_2 \geq .. \geq \lambda_{10}$, ideally $\lambda_7 = .. = \lambda_{10} = 0$.

So, we can identify a set of planar points in quadric representation just by checking the eigenvalue of λ_7. Of course in reality, the value of λ_7 is non-zero because of noise, but it can be expected that for very "flat" surfaces, this eigenvalue is small. Therefore, we can simply use a threshold T_{plane} to test the planar points. The criterion is

$$\text{If } (\frac{\lambda_7}{n}) \leq T_{\text{plane}} \quad (n \text{ is the number of points}) \tag{5.3}$$

the points belong to a planar surface, otherwise a quadric surface.

In the second module, the seed patch is searched from the candidate patches before the region-growing process. In this searching stage, each patch is subjected to the planarity test. Of all planar patches that are found, a seed is selected and the growing is performed in planar representation.

5.3.2 Homogeneity test in planar representation

The homogeneity measure used for merging planar regions is also based on a parametric representation. By modelling the uncertainties of the estimated planar parameters, the criterion for the homogeneity test can thus be established within an optimal framework.

Normally a planar surface can be represented by the implicit form:

$$\mathbf{n}^T \mathbf{x} + b = 0\,, \tag{5.4}$$

where $\mathbf{n} = [n_x\ ,\ n_y\ ,\ n_z\ ]^T$ is the unit normal vector of the plane. The parameter estimation with respect to the representation of (5.4) is described in Appendix D.

Since the sels are spatially connected, only the normal vector **n** of a plane is of concern. To convert the parameter as unconstrained, we define a vector, denoted $\boldsymbol{\theta}_n$, as the planar representation:

$$\boldsymbol{\theta}_n = \left[n_x / n_z\ , n_y / n_z\right]^T . \tag{5.5}$$

Applying the method of modelling uncertainties in the quadric estimate, as described in chapter 4, we can derive the covariance matrix of the vector $\boldsymbol{\theta}_n$.

Defining a general surface parameter

$$\boldsymbol{\theta}_{nb} = \left[\boldsymbol{\theta}_n^T , b/n_z\right]^T , \tag{5.6}$$

the estimate of $\boldsymbol{\theta}_{nb}$ can be treated as the solution of minimising the cost function

$$\Theta = \boldsymbol{\theta}_{nb}^T \mathbf{M}_p \boldsymbol{\theta}_{nb}\,;$$

where $\mathbf{M}_p = \sum \mathbf{m}_p \mathbf{m}_p^T$; $\mathbf{m}_p = [x, y, 1]^T$.

In the same way as the analysis in section C.1.1 of Appendix C.1, we can derive the covariance matrix of $\boldsymbol{\theta}_{nb}$ with the expression

$$\boldsymbol{\Psi}_{nb} = \sigma^2 \mathbf{M}_p^{-1} \tag{5.7}$$

Consequently, the covariance matrix for the parameter vector $\boldsymbol{\theta}_n$, denoted $\boldsymbol{\Psi}_n$, is obtained as the top-left 2×2 matrix of $\boldsymbol{\Psi}_{nb}$.

Using the covariance model for the planar representation, classification of planar estimates can also be implemented within an optimal framework. The "statistical distance" between the estimate $\boldsymbol{\theta}_{n,i}$ of the i^{th} region and the estimate $\boldsymbol{\theta}_{n,j}$ of the j^{th} region is then defined as

$$d_{i,j}^2 = (\boldsymbol{\theta}_{n,i} - \boldsymbol{\theta}_{n,j})^T \boldsymbol{\Psi}_{n,ij}^{-1} (\boldsymbol{\theta}_{n,i} - \boldsymbol{\theta}_{n,j}) \quad (\boldsymbol{\Psi}_{n,ij} = \boldsymbol{\Psi}_{n,i} + \boldsymbol{\Psi}_{n,j}). \tag{5.8}$$

According to the discussion above, the growing process in the planar representation is performed in a similar way as used for the quadric representation.

5.4 Conclusions

This chapter presented a region-based algorithm for range image segmentation. The feasibility of using the region-based technique arises from two aspects: the assumption of piecewise smoothness of the object surfaces and modelling of the uncertainties in the estimates of the region parameters, which has been described in chapter 4. The preference of using a region-based method instead of an edge-based method is inherently driven by the strategy of surface-based object representation and recognition, where the output of segmentation is represented with surface parameters.

The proposed approach consists of three modules. In the first stage, a planar approximation is applied to the local set of data points within a sampling window and the outliers are detected. This can be regarded as a pre-processing stage. In the second stage, a region-growing algorithm is applied by iterative merging of the initially over-segmented elementary regions (the sels). In this stage, region parameters are estimated either using a quadric representation or a planar representation, associated with the proposed geometric model. The features of the estimation-based clustering algorithm are:

(1) The distance between two clusters in parameter space is measured with the "statistical distance", so the clustering is performed in an optimal framework.
(2) The reliability of region estimates is specified. By defining the *measure of uncertainties*, the selection of the seed for region growing is determined in view of "being the most reliable in representation". Furthermore, the growing process always runs towards the most stable state in representation, where the candidate sels to be merged are ordered according to the *measure of uncertainties*.

In the last stage, the fundamental segments resulting from the region growing stage are further refined through point-based classification.

The planar regions and quadric regions are treated separately in this approach. To distinguish between planar and quadric regions, a method of planarity test was proposed. It should be pointed out that in case of noise presence, there is no exact criterion to define the "planarity" of a surface. In fact, the purpose of separating the planar representation from the quadric representation is to avoid the possible ill-posed problem when a very "flat" surface is represented in quadric form. As we can see from the experimental results in chapter 6, in reality, the quadric representation can be used to describe most of the surface segments, even though some of them are expected to be "planar" from our visual judgement.

Chapter 6

Experimental Results with Real Range Data

The surface-based recognition algorithm introduced in chapter 4 and the region-based segmentation approach introduced in chapter 5 were applied to experiments with real range data. Most of the range images were acquired with the range imaging system developed at the UT (a system description is given in [18]). Some images were obtained from the "frequency-difference" laser scanning sensing being developed at the OUQ (this system was introduced in chapter 2 of this thesis). The former system generates dense range images within a smaller measurement area of the scene and the latter generates scattered 3D data within a larger area of the scene. In the approaches proposed in this thesis, both types are considered as composed of scattered 3D data.

In this chapter, the experimental results of the surface-based recognition algorithm are given first. All samples came from man-made objects, whose surfaces are assumed to be modelled by quadric representations. Figure 6.1 shows the objects used in experiments in a greylevel image and figure 6.2 shows the range images of the objects, as obtained from the ranging system of [18]. Figure 6.3 shows the greylevel pictures of the objects that were used to obtain the scattered 3D data points with the laser scanning ranging system. For each of these objects, a quadric representation has been used to describe the visible surfaces. Range data of these objects as acquired with the ranging systems act as the inputs for the recognition. The model base was established by applying the sensor-based modelling scheme as proposed in section 4.4. The results of this recognition experiment are described in section 6.1. The experiments on scene segmentation with the range images acquired from the ranging system of [18] are detailed in section 6.2. Finally, brief conclusions are given in section 6.3.

6.1 Surface-based recognition in quadric representations

In chapter 4, the optimal classification algorithm of quadric surfaces with range data has been described. The uncertainties of the estimated parameter are represented by a covariance matrix that is obtained simultaneously in feature extraction. Then a probabilistic model with normal pdf is applied to describe the statistical properties of the estimated features. Therefore the classification can be implemented within a Bayesian framework, by which the decision rule is established on the measure of "statistical distance" (section 4.3).

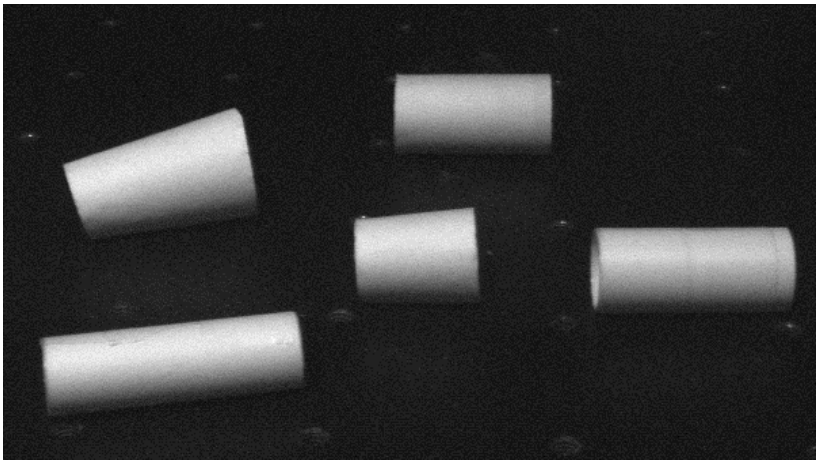


Fig 6.1Greylevel pictures of the objects corresponding to the range images of (a) ~ (e).

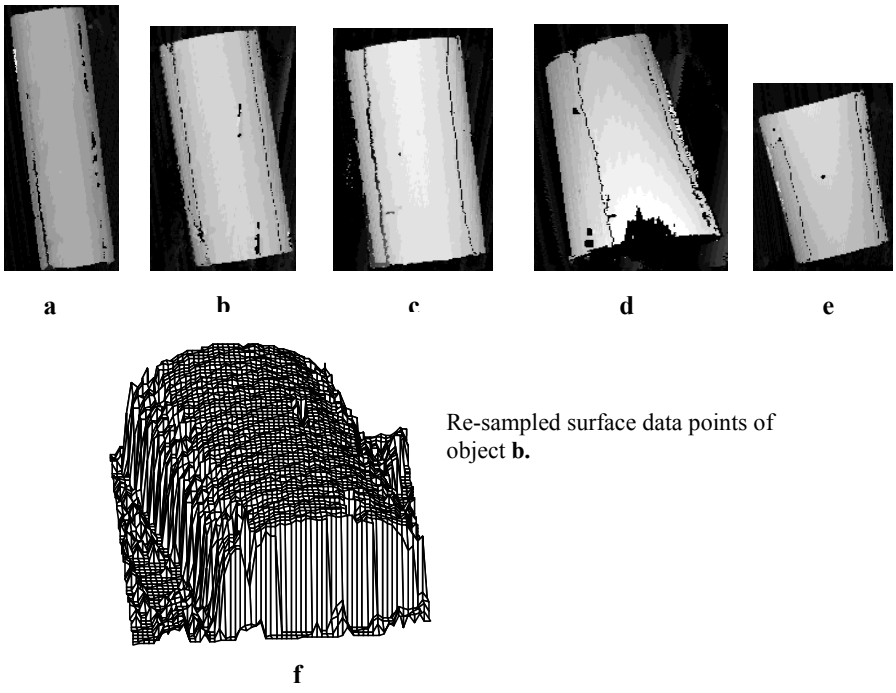


Fig 6.2 Range images of the objects used in the experiments. The images were obtained with the range system of the UT. The re-sampled data points from image **b** are plotted in **f**.

1 **2** **3** **4**

Fig 6.3 The objects used to obtain 3D data with the frequency-difference laser scanning system that was developed at the OUQ (The pictures displayed here are greylevel images made by a camera).

The purpose of the experiments is to test the optimal descriptions for quadric feature representation, as proposed in chapter 4. The experiments follow the recognition scheme as illustrated in figure 4.2, according to the algorithms of modelling and recognition described in chapter 4.

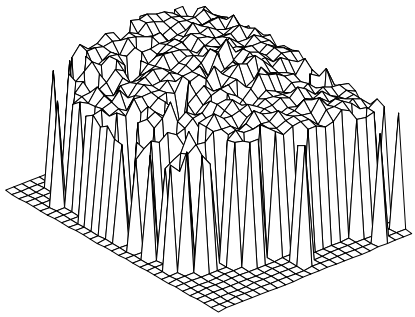


Fig 6.4 The 3D surface data points obtained with the laser-scanning sensing system for the object **3** in figure 6.3

For each object in figure 6.2, range images have been obtained with different poses of the object. A range image contains 3D data points of both the visible surface of the object and the flat background, see the examples displayed in figure 6.2. Since the scene was composed of (part of) a quadric surface and a piece of the planar background only, the surface points were simply extracted just by thresholding the depth value. These surface data points in the dense range images were re-sampled with a lower resolution. The purpose of this re-sampling are: (1) to reduce the computational burden, because the dense data set is often redundant for low degree (quadric or planar) surface representations, as is the case in our experiments; (2) to obtain a better approximation of the measurement error model (as i.i.d noise), because the noise in neighbouring pixels of a dense range image is possibly strongly correlated.

For the range images of (a)~(d) in figure 6.2, the re-sampling interval in horizontal direction is two pixels and that in vertical direction is three pixels. For the image of (e), the intervals in both horizontal and vertical directions are two pixels.

After re-sampling the range images, the model base was established using a number of trials of each of the objects.

First, the surface fitting process as introduced in chapter 3 was applied to obtain the surface parameters. The next step is to derive the invariants of the quadric representation, to build up the feature vector. The computation of the covariance matrix is performed simultaneously. The feature vector and the covariance matrix stored in the model base were obtained with the MLE algorithm for each of the objects, according to (4.24) and (4.25). The number of images to construct each object model was about 10.

After the modelling process, more range images were obtained and re-sampled to act as inputs for recognition. The feature vector and covariance matrix were extracted for each input range data. In the recognition stage, the statistical distances between the input feature vector and each of the model vectors are computed according to (4.21). Denote the input feature as s and each model as $l\ \forall l \in \{1,..,L\}$. So, the input s is classified to model l' if $d_{sl'}^2 = \min\{d_{sl}^2 | \forall l\}$, where d_{sl}^2 is the distance between s and the model l.

- ***Using range images from the UT ranging system***

The five objects in figure 6.2, whose surfaces are either cylinders or cones, were used as primitives for the purpose of classification. In order to test the discriminatory power of the statistical distance as described in chapter 4, the classification results using the statistical distance are compared to the results using Euclidean distance. Table 6.1 lists two sets of feature vectors in the model base describing the objects in figure 6.2. The first, denoted as $\bar{\mathbf{g}}_m$, is obtained by MLE with the joint pdfs of a set of measurements, according to the proposed optimal modelling process. They are printed in bold. The other set is obtained just as the mean of a set of measurements, and is denoted as $\bar{\mathbf{g}}_n$. Obviously, in case only first order statistics of the measurements is available, $\bar{\mathbf{g}}_n$ is the optimal estimate of the expectation of the feature vector. However, when higher order statistics of the feature are recovered, i.e., the pdf, $\bar{\mathbf{g}}_m$ is the optimal estimate.

The number of samples measured as inputs for each object in figure 6.2 is between 17 and 20. The samples were obtained for different poses within a limited field of view. The estimated variance of the noise in the range images is within a range of 0.01~0.08 (mm^2). Table 6.2 lists the classification results. The models are given row wise, the inputs column wise. The cells contain the number of inputs that were classified to the model in the first column; empty cells indicate no hits. The upper number corresponds to the classification results based on the statistical distance; the lower one (in round brackets) is the result based on the Euclidean distance.

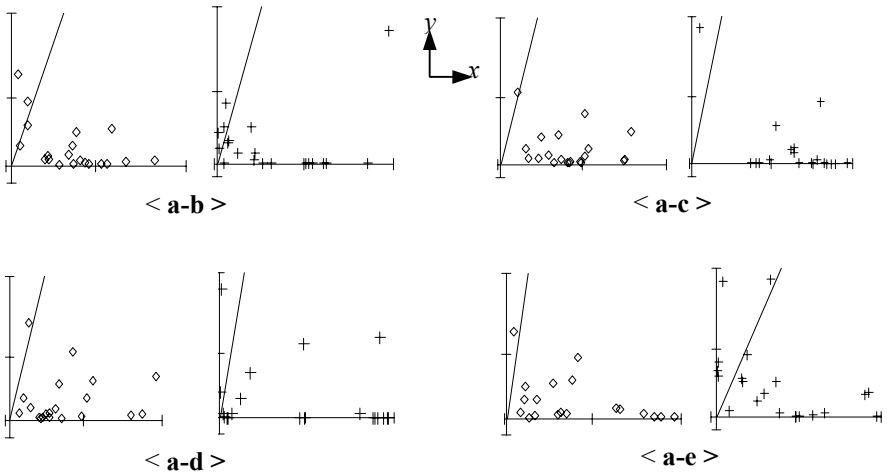
Figure 6.5 shows an overview of the relative distance measures for the trials from class **a** and **e,** plotted as distance distribution patterns. For a given input, the distance to the actual model (for instance **a** in the upper left part of the figure) and the distance to one of other models (for instance **b** in the upper left) are displayed using a two-dimensional point (a diamond symbol) in the x-y plane. The y co-ordinate represents the distance to its actual model and x represents the distance to the other. So for each input four such diagrams are given, corresponding to the actual model and the four other models. The discriminatory line

Table 6.1 Feature vectors obtained in the modelling process for the objects in figure 6.1. The bold lines refer to MLE and the second lines refer to means.

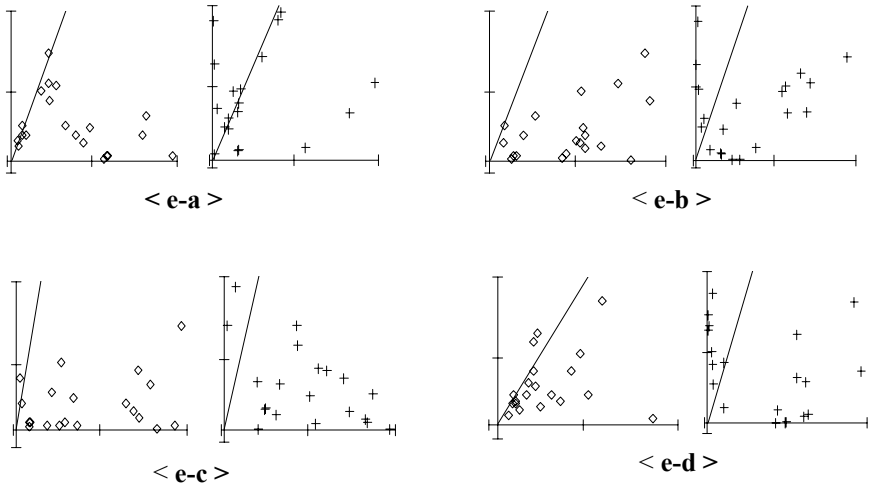
model	Feature vector $\overline{\mathbf{g}}_m$ ($\overline{\mathbf{g}}_n$)
a	[**0.3904 0.00004162 -28.42**]^T
	[0.3956 –0.0002171 –28.98]^T
b	[**0.4618 0.0004876 -49.03**]^T
	[0.4607 0.0006075 –54.81]^T
c	[**0.5452 -0.001241 –85.99**]^T
	[0.5393 –0.001184 –84.08]^T
d	[**0.3715 -0.001938 12.02**]^T
	[0.3779 -0.001834 5.431]^T
e	[**0.3628 -0.0008914 -9.567**]^T
	[0.3989 -0.002049 -27.91]^T

Table 6.2 Classification results: the number of trials in the associated columns to be classified to the model in the associated rows. Classification results using the Euclidean distance are shown in round brackets

		Input [number of trials]				
Model		**a** [*20*]	**b** [*18*]	**c** [*17*]	**d** [*18*]	**e** [*20*]
	a	18 (9)	1 (3)		1 (1)	5 (4)
	b	2 (4)	17 (14)	1 (1)		(4)
	c	(2)	(1)	16 (16)	1 (1)	1
	d	(1)			16 (13)	3 (8)
	e	(4)			(3)	11 (4)



(1).All trials are from class **a**. Distances to model **a** and to models **b, c, d, e** are displayed separately.



(2).All trials are from class **e**. Distances to model **e** and to models **a**, **b, c, d** are displayed separately.

Fig 6.5 Measurement of distances. For any input, the *y* co-ordinate represents the distance to the actual model, and the *x* co-ordinate represents the distance to another model. In (1), all trials are from class **a**, and in (2) all trials are from class **e**. The statistical distances are marked by a diamond symbol and the Euclidean distances are marked by a "+" symbol.

$y = x$ is also plotted in each diagram. Obviously, a correct classification for an input occurs only if all the positions of this input in the four diagrams are below the discriminatory line: this means that the distance to the actual model is smaller than all of the others. For the purpose of comparison, the Euclidean distances are plotted too, as plus-symbol. For the purpose of visibility, the x and y co-ordinates in these diagrams have been differently scaled.

As expected, the classification results in Table 6.2 illustrate that the classification errors in case of the statistical distance are less than when the Euclidean distance is used. This proves that the statistical distance has a better discriminatory power than the conventional Euclidean distance. From the results, we also conclude that more errors occurred in classifying object **e**. This is explained by the fact that the surface data points of **e** were sampled over a smaller region, so the uncertainty in parameter estimation is consequently larger than for the other objects. Therefore, it makes sense taking into account such uncertainties in feature estimates, to reduce ambiguities in the classification, as is illustrated by comparing the outcomes in table 6.2. The overall results in this experiment show that the statistical distance measure results in a better *class separability* than the Euclidean distance measure. This is the essential advantage of using the statistical distance in classification.

It should be noted that in general the computed statistical distance between an input and its actual model is larger than what could be predicted by the normal pdf model. This is due not only to the approximations made in the computation of the covariance matrix, but also to the fact that the noise in a real measurement has not the ideal Gaussian i.i.d property. Non-linear errors in the ranging process result in shape deformation of the object in the representations. However, according to the experimental results, together with the results from the synthetic data as described in chapter 4, it can be concluded that the proposed covariance model is a good approximation to optimise quadrics-based surface representations and the associated object recognition.

- ***Using range data from the OUQ laser-scanning system***

The 3D data obtained with the frequency-difference laser scanning system, which has been introduced in chapter 2, were also subjected to the recognition process. Since the implementation of this sensing system is not yet finished, the number of samples obtained for experiments was limited. The four objects used in this experiment, as shown in figure 6.3, are four cups of which two have the shape of a cylinder and two have a conical shape. As an illustration, figure 6.4 shows an instance of the scattered 3D data points obtained by this system, generated from object **3** in figure 6.3.

The feature vectors of the models and the inputs are listed in Table 6.3, together with the estimates of the variance in noise. The classification results are listed in Table 6.4. Again, the left column contains the models and the top row indicates the inputs. The cells contain the values of the distance measures with respect to all models. At the upper lines, results are given based on the statistical distance; those in brackets refer to Euclidean distances. The smallest value in each row is printed in bold, indicating that the input in the corresponding column is classified to the model in the corresponding row. Alternatively,

where the Euclidean distance is used as the criterion, the classification results are marked by "+", corresponding to the smallest value in the column.

Table 6.3 Feature vectors and noise variance obtained from the objects in figure 6.3 with the frequency-difference laser scanning system.

model	Feature vector	$\sigma^2(cm^2)$
m1	$[-0.7429\ \ 0.2011\ \ -0.8934]^T$	0.9552
m2	$[0.7532\ \ -0.1169\ \ -10.68\ ]^T$	0.4172
m3	$[-0.6213\ \ -0.2433\ \ 4.762]^T$	0.6029
m4	$[0.3687\ \ 0.1469\ \ -7.916\ ]^T$	0.4404
input	Feature vector	$\sigma^2(cm^2)$
s1	$[-0.3785\ \ 0.1322\ \ -0.0844]^T$	0.8134
s2	$[0.7068\ \ -0.01314\ \ -10.21]^T$	0.3586
s3	$[-0.5785\ \ -0.2256\ \ 4.509]^T$	0.5508
s4	$[0.4083\ \ 0.1463\ \ -9.748]^T$	0.3318

Table 6.4 classification results for the inputs and models shown in table 6.3. The smallest statistical distance between an input to all models (in one column) is shown in bold type. Figures in brackets are Euclidean distances; the smallest value in one column is marked with "+".

	s1	**s2**	**s3**	**s4**
m1	**215.2** +(0.7920)	1562 (88.97)	304.9 (29.41)	1390 (79.73)
m2	1630 (113.3)	**11.07** +(0.2207)	1487 (232.1)	230.8 +(1.733)
m3	414.1 (23.68)	1373 (226.0)	**2.636** +(0.06556)	1127 (211.8)
m4	1412 (61.89)	209.5 (5.406)	1076 (155.4)	**5.293** (3.358)

From these results, all inputs were correctly classified when using the statistical distance. However, an error occurred using the Euclidean distance, notably object **s4** was misclassified to object **m2**. This is not surprisingly, because the surface shapes of these two

objects are very close. Although one is (part of) a cylinder (label 4) and the other is a cone (label 2), the estimated surface parameters in feature space are closer to each other than to object 1 and 3, the latter having larger diameters. However, when considering the uncertainties in the parameter estimation, the *class separability* can be improved because the probabilistic representation provides an optimal measure of the separability.

6.2 Segmentation with region-based approach

In section 6.1, the objects involved in the experiments were considered as isolated quadric primitives (in figure 6.2 and figure 6.3), and used for examining the optimal classification algorithm based on modelling of uncertainties in feature extraction. A practical scene may consist of a set of objects, or one object could be described with a combination of primitives. Prior to the recognition task, segmentation is to partition the input image into a set of meaningful regions that can be described with primitives.

In view of the surface-based quadric representations used in this work, the objects in the images need be partitioned as to describe them as a set of quadric surface primitives. The range images used in these experiments are shown as (a) in figure 6.6. All have been taken from the range system of the UT.

As described in chapter 5, the segmentation algorithm used in this work is based on a region-based approach. The main stream of processing consists of three modules: outliers detection, region growing by merging the *sels* and finally point-based refinement. For these experiments, the range images have been re-sampled at a lower resolution. In the following, we will always refer to points from re-sampled images.

In the first module, a small rectangular window moves stepwise over the whole image area, centred at a data point at any instant. Points covered by this window are assumed to belong to a planar patch; the mean squared distance from data points in the window to the estimated patch surface, denoted $\hat{\sigma}^2$, is calculated. If the result satisfies $\hat{\sigma}^2 \geq r_{win}\bar{\sigma}^2$, then the centre point of the window is marked as outlier. Here $\bar{\sigma}^2$ is the median of all $\hat{\sigma}^2$ and r_{win} is a pre-defined coefficient. In the experiments, the size of the moving window is 5×5 (point^2) and $r_{win} = 25$. Furthermore, if the depth value of the centre point differs to one of its neighbouring points (4-connective) over a threshold T_{depth}, that centre point is marked as outlier. Empirically, we selected $T_{depth} = 1$(mm) in the experiments.

In the second module, the input range image, of which the outliers have been marked, is initially partitioned into a group of grid areas, i.e., the *sels* (as shown in figure 6.6b). The size of each sel in the experiments is 5×5 (point^2). If a sel contains a number of outliers exceeding a pre-defined fraction, i.e., 20% of the total points in a sel, this sel is marked as *ambiguous* and will not participate in the merging process. All others are considered to be ordinary sels. Of these ordinary sels, a subregion consisting of a few neighbouring sels is searched as the seed for region growing, which is termed as the *seed patch*. The number of sels in a patch is limited between 6 and 9. For each candidate patch, the quadric parameter is estimated by surface fitting and the covariance matrix $\mathbf{\Psi}$ is computed simultaneously. Applying the *measure of uncertainty* in the quadric representation defined in chapter 5, i.e.,

$\|\boldsymbol{\Psi}\|_2$, the patch that has the minimum value of *measure of uncertainty* is selected as the seed. In the region-growing process, the seed patch expands successively by merging its (8-connective) neighbouring sels. The criterion for merging is

$$d_s^2(\mathbf{p}, \boldsymbol{\Psi}, \mathbf{p}', \boldsymbol{\Psi}') \le T_d ,$$

where d_s^2 is the statistical distance between the regions before and after merging, ($\mathbf{p}$,$\boldsymbol{\Psi}$) and ($\mathbf{p}'$,$\boldsymbol{\Psi}'$) are the quadric parameters and covariance matrices before merging and after merging and T_d is the selected threshold. During this merging process, the quadric parameters are successively updated with the current merged data points.

Another criterion to control the merging process is that if there are enough points in the sel whose distances to the current surface exceed a certain threshold (these points are regarded as non-fitted points), the current merging is rejected. The threshold of distance is $d^2 \ge r_d d_{med}^2$, where d_{med}^2 is the median of the squared distances within the sel, while the constant used in the experiment is $r_d = 25$. The threshold of the number of non-fitted points was $T_n = 20\%$ of the total number of the sel.

According to the growing strategy as proposed in chapter 5, the sel that is predicted to have the smallest value of *measure of uncertainty* has the highest priority to enter the test of the merging criterion. Then the second candidate is the one having the smallest *measure of uncertainty* of the remainder, and so on. If all neighbouring sels fail to pass the merging criterion (evidently the ambiguous sels fail for merging) or have been merged previously, the region growing stops. A new seed is searched from the remaining sels and the above process is repeated. The whole merging process is finished if no more seed can be find in the range data.

It is noted that the bias-corrected renormalization approach for surface fitting often fails to result in convergence at the start of region growing, because of noise and the limited size of the seed patch. In stead, we applied the gradient weighted least-squares criterion for surface fitting. The noise variance used in the covariance computation was estimated from planar fitting. After fitting data points of a sel to a planar representation, the mean squared distance of those points to the fitted plane is obtained. Then averaging the mean squared distances among the merged sels results in an estimate of the noise variance with respect to the current patch. As stated in chapter 3, the estimate using the gradient weighted least-squares criterion is biased. However, because the sel is small and is merged one by one, the bias in the estimated parameters before and after merging does not significantly change. Approximately, in the computation of the statistical distance, both biases are cancelled out with each other. In all cases, the marked outliers in sels were excluded from surface fitting.

An important parameter in the merging process is the statistical distance threshold T_d. Usually, a large value of T_d may cause under-segmentation, i.e., inhomogeneous sels are merged into one primitive representation, whereas a small value of T_d could cause over-segmentation, i.e., homogeneous sels fail to be merged. In practice, the value of T_d should be selected empirically. However, in the view of hypothesis test, the value of T_d determines the "significance level" in the test. So the selection of T_d makes sense of optimality in the definition of homogeneity.

To set the threshold automatically, a proper value of T_d can be found by referring to the histogram of the statistical distances between all the candidate seed patches and their neighbouring sels. In the process of finding the seed patch, the distance d_s^2 for each candidate patch (between that before and after merging and its neighbouring sel) that has the lowest *measure of uncertainty* was recorded and the histogram of d_s^2 is calculated. We found experimentally to set T_d around the last peak of the histogram, which is thought to be the bound of the "permitted" distances between two homogeneous regions. The selected values of T_d and the histograms for the instances of (1) ~ (4) are depicted in figure 6.7.

As stated in chapter 5, in order to avoid the ill-pose problem in surface fitting, we distinguish between the planar surfaces and the curved surfaces at the start of a merging routine. In searching the seed, a planarity test is carried out for all candidate patches. In the experiments, the threshold for the planarity test in (5.3) of chapter 5 was set to $T_{plane} = 0.0005$. Once a candidate patch was found to be planar, it grows by merging with its neighbouring sels in the planar representation and stops if all neighbouring sels either failed to pass the merging criterion or have been merged to other labels.

After region growing, the homogeneous sels are grouped together and labelled. These labelled sels, parameterised either with quadric or planar representations, comprise the *dominant regions* of the scene.

In the third module a point-based refinement is applied to modify these dominant regions into the final segments. First, points in the boundary sels of two connected different regions are re-classified to the one having the smallest distance between this point and the surface. Second, all of the neighbouring sels of these merged regions, whether ambiguous or ordinary, are split at point level. If the distance of a point to the labelled surface is less than a certain threshold, it is merged. Experimentally, this threshold was selected to be $T_{pointdist} = 0.8$(mm). After that, points whose distance exceeds this threshold and those in the sels that are not neighbours of the labelled region are marked as "unlabeled".

Figure 6.6, (1)~(4) shows four instances of the input range images and their segmentation results. The scene includes object(s) in a background (a flat table in the system set-up). The greylevel of the image is associated to the depth value of a pixel to the background. As an example, in scene (1), picture (a) is the original input range image. This image is re-sampled with a resolution of 2 pixels/point). Picture (b) shows the result of outlier detection (the first module), in which the detected outliers are displayed highlighted (white points). It is noted that the point size has been doubled to keep the same size of the original image. The whole image of (b) has been initially segmented with a grid of sels. Picture (c) shows the result of region growing by the second module. Four separate groups of sels have been merged and form the dominant regions (displayed in different greylevels). Finally, picture (d) shows the result of refinement by the third module.

For the inputs of (2) and (4) the re-sampling resolution was also 2 pixels/point, but for (3) it is 3 pixels/point.

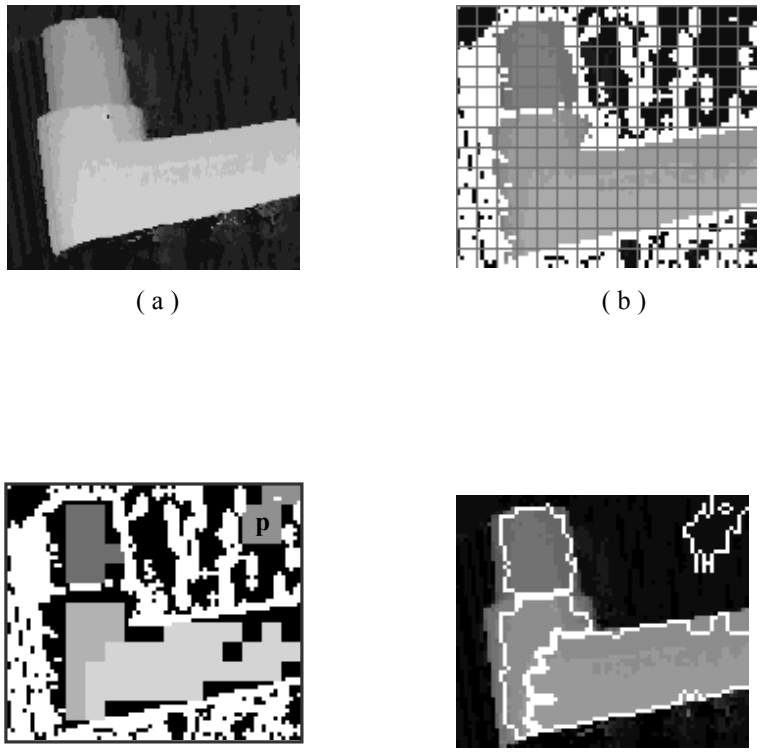
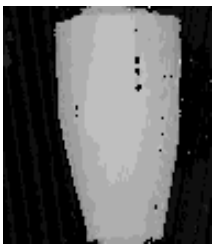


Fig 6.6 Segmentation results with real range images (1 ~ 4) obtained with the ranging system of the UT. Picture (a) is the original range image (depth value is mapped to greylevel). (b) shows the result of outlier detection (the bright pixels), where the initial grid of sels is also displayed. (c) shows the result of merging sels. The labelled segments i.e., the dominant regions, are shown with different greylevels (outliers are also displayed with higher brightness). (d) shows the final segmentation result after the refinement step.

Of the dominant regions in picture (c), the planar regions have been marked with letter “p”, others are regarded as quadric regions.



(a)

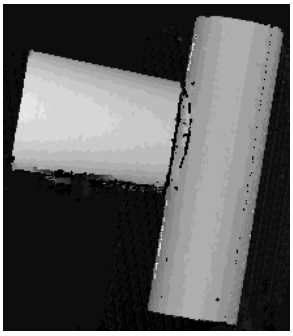
(b)

(c)

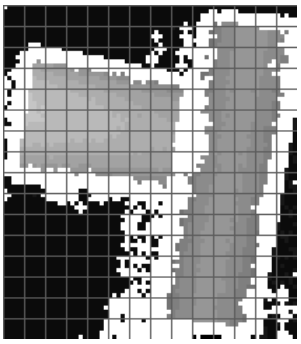
(d)

< 2 >

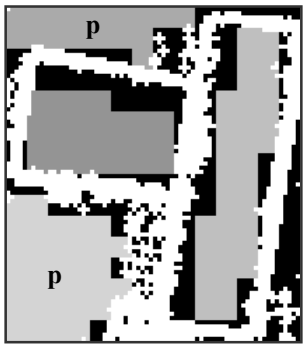
Fig 6.6



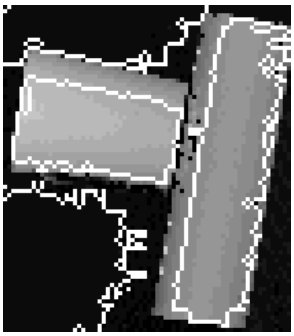
(a)



(b)



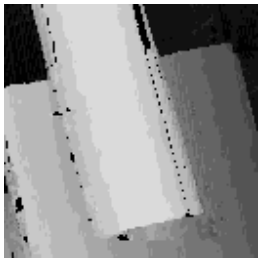
(c)



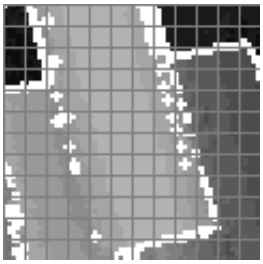
(d)

< 3 >

Fig 6.6



(a)



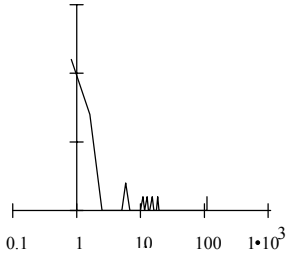
(b)

(c)

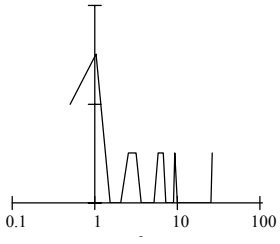
(d)

< 4 >

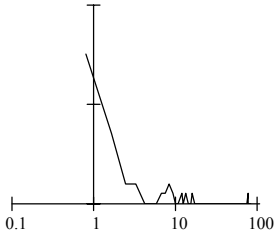
Fig 6.6



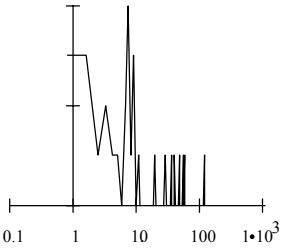
Result of scene <1> in figure 6.6. The used threshold T_d =30.



Result of scene <2> in figure 6.6. The used threshold T_d =15.



Result of scene <3> in figure 6.6. The used threshold T_d =90



Result of scene <4> in figure 6.6. The used threshold T_d =80

Fig 6.7 The histogram of d_s^2 in the process of searching the seed patch. The x co-ordinate is on a log-scale. The values of T_d as used in the experiments are also given.

It should be pointed out that in the second module of the segmentation, some regions that were thought to be planar returned as quadric and have grown in the quadric representation. A typical example is a part of the background. This is due to noise in the input range images, leading to errors in the shape description of the planar patch. Of course, if the threshold T_{plane} in the separation of planarity increases, more regions could be recognised as being planar, but the risk of mis-recognising a quadric patch as planar also increases, because a small patch could easily fit to a planar representation. In order to avoid an over-segmented result, we prefer reducing the risk of returning a curved patch to a planar description. Therefore, T_{plane} was set to a small value. In fact, the reason to separate planar and quadric representations is to avoid the possible ill-posed problem when fitting low-dimensional surface data to a higher dimensional representation. As long as the ill-posed

problem is tolerated, a low threshold value for T_{plane} is recommended. In these segmentation results, the planar segments were marked with "p" and others were in quadric representation.

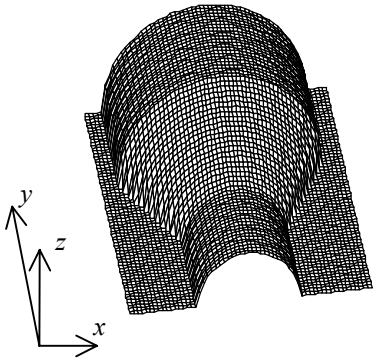
From the experiments, it was noticed that the boundary points of a curved surface were usually difficult to be merged. This might be due to inaccurate range data acquisition at the boundaries (for example, the results of <3> and <4> in figure 6.6). Such results seem "imperfect" according to our visual judgement. However, as pointed out in the description of the region-based algorithm in chapter 5, what we concerned and expected from the segmentation is the reliability of surface representation. A "visually pleasing" judgement for range image segments should be discarded.

The segmentation algorithm was also applied to synthetic data. In figure 6.8 (a), the 3D scattered data points generated from a synthetic object surface and a planar background are depicted. Gaussian noise has been added to the three co-ordinates of each point. With the same unit, the radius of the larger cylinder portion equals 20 and the length of the whole object (along y axis) equals 56. The standard deviation of the noise is $\sigma = 0.02$. The resampling interval at x and y direction is 0.8.

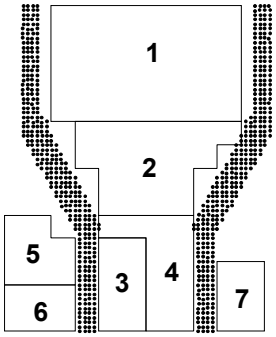
The same segmentation approach was applied as for the images in figure 6.6. Also the parameters have been selected the same, except for the threshold of the statistical distance in the merging criterion, which is set to $T_d = 100$, and the size of the sel, now being 4×4. The segmentation result of the *dominant regions* is illustrated in figure (b) with a perspective view at the z direction. The histogram of d_s^2 is depicted in (c). Finally, the seven segmented regions are shown in (d) after applying the process of point-based refinement (in exploded view along the z direction, for the purpose of visualisation). All those regions returned as quadrics, although in fact the regions "5", "6" and "7" are planar. Other points are marked as unlabeled.

A notable result of the segmentation is that the bottom cylinder portion in figure (b) was over-segmented into two regions, i.e., labels "3"' and "4". This was caused by noise in the data points. However, according to the final estimate of the surface parameters for each segment, it can be found that one of the two segments of this over-segmented cylinder surface, i.e., label "3", has well recovered the true parameters of this cylinder representation. But the other gives an unexpected result. The recovered surface parameters of the object surface regions and the actual surface parameters are listed in Table (6.5).

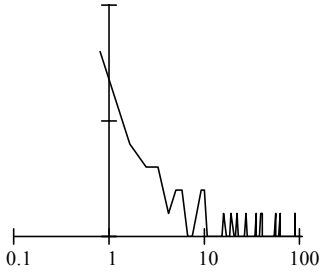
It can be seen that the regions with label "1", "2" and "3" in fact recover their original surfaces. From this example, we can also conclude that for range image segmentation, the reliability of surface representation is most critical and should be focused in the development of segmentation approach, rather than a result of "visually pleasing" outcome.



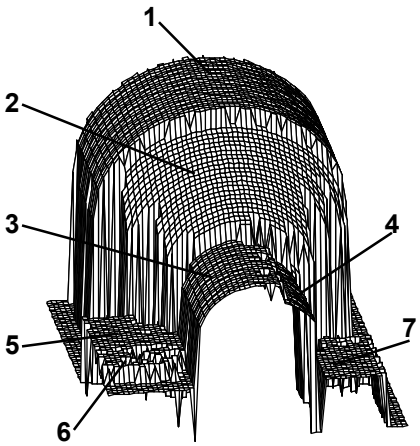
(a) The scene consisting of synthetic data. The radius of the larger cylinder is 20. The sampling interval is 0.8. Gaussian noise of standard deviation σ=0.02 has been added to the three co-ordinates for each points (in the picture, noises are only added to *z* co-ordinates for the purpose of plotting).



(b) The perspective view of the dominant regions by merging sels. The size of the sel is 4×4. The seven regions are labelled with figures 1 ~ 7.



(c) The Result of the histogram of d_s^2 . The used parameter T_d = 100



(d) The final result of refinement. The figured regions correspond to the labels in (b). In order to view different segments, labelled regions have been detached at *z* direction.

Fig 6.8 Segmentation result with synthetic data.

Table 6.5. Estimated parameters of the segmented surfaces of the object in figure 6.8. Their actual values are also given.

	Estimate of surface parameters	Actual values
1	8.2×10^{-5}, 0, 8.3×10^{-5}, 0, 0, 0, -0.016464, -0.000037, -0.008264	8.3×10^{-5}, 0, 8.3×10^{-5}, 0, 0, 0, -0.016529, 0, -0.008264
2	1.06×10^{-4}, -3.4×10^{-5}, 1.02×10^{-4}, 0, 1×10^{-6}, 4×10^{-6}, -0.021221, 0.006540, -0.010630	1.06×10^{-4},-3.3×10^{-5},1.06×10^{-4}, 0,0,0, -0.021246, 0.006557 ,-0.010623
3	8.1×10^{-5}, 0, 8.2×10^{-5}, 0, -2×10^{-6}, -1×10^{-6}, -0.016155, -0.000032, -0.007954	8.1×10^{-5}, 0, 8.1×10^{-5}, 0, 0, 0, -0.01613, 0, -0.008064
4	3.2×10^{-5}, 0, 5.5×10^{-5}, -2×10^{-6}, 8.9×10^{-5}, -6×10^{-6}, -0.011376, 0.000846, -0.015185	

6.3 Conclusions

This chapter presents the experimental results of optimal classification of quadric primitives and region-based segmentation of range images, both being based on the approach of modelling uncertainties in quadric representation as proposed in chapter 4.

In section 6.2, the algorithm of the Bayesian classification through feature extraction of quadric representation has been tested. The range data used in the experiments were sampled from object surfaces which were assumed quadric primitives. The results demonstrated that the classification using the statistical distance as the measure of likelihood for feature correspondence is more reliable than the conventional Euclidean distance measure. It should be noted that the parameter estimation was carried out by applying the improved approach of quadric fitting proposed in chapter 3. So it was expected that the improved solution of surface parameter estimates could increase the reliability of the likelihood measure even for Euclidean distances. However, in order to establish the optimal framework of object classification, modelling of uncertainties in feature representation is ultimately required. Because of the complexity of noise in real measurements and errors in parameter estimation, the proposed covariance model is only an approximate model for feature representation. However, as a trade-off between the accuracy of classification and the complexity of mathematics, the experimental results presented in this chapter, together with the results of synthetic data in chapter 4, suggested that the proposed approach is practical in optimal classification for surface-based quadric representations.

The approach of range image segmentation proposed in chapter 5 was applied for real range images, as well as synthetic data. The experiments in section 6.2 were concerned with the cases where the scene in a range image contains object(s) consisting of quadric or planar surfaces and a planar background only. The method used in this work, which is discussed in chapter 5, can be regarded as the region-based approach. But the uncertainties in

parameter estimation have been explicitly included in the surface description. Consequently, the definition of surface homogeneity has been optimised by applying the proposed probabilistic model.

As mentioned in chapter 5, one of the difficulties in range image segmentation is finding proper criteria for the assessment of a segmentation result. The experimental results showed that the proposed method is practical, i.e., it gives the expected results. However, as argued in the discussion of chapter 5, since the criterion for region growing is estimation-based, the assessment of segmentation of range images should also be representation-based, rather than a "visually pleasing" judgement. This argument might be emphasised when we keep in mind that usually the segmentation module is followed by a module of primitive recognition at a higher level in a recognition system. Because of noise in range image acquisition and the possible deviations of a physical surface to the mathematical model, the segmentation results could be different to what we expected or assumed. As shown in experiments, data points near the boundary of the curved surface (for example, the case of <3> in figure 6.6) were difficult to be merged. A notable example is the segmentation result with synthetic data, shown in figure 6.8. Although the small cylinder was falsely partitioned into two regions (label "3" and "4"), the surface parameters were still well recovered by one of them (label "3"). Therefore, we can argue that such a result of segmentation is false for our perceptual geometry, but correct for parameterised representation. From this point of view, the segmentation method proposed in this chapter seems more close to the notion of "primitive extraction". Therefore the reliability of homogeneity test was in fact the focus of the experiments described above.

According to the results presented here, we expect that the proposed approach based on modelling of uncertainties in parameter estimation have improved the reliability of region estimation. However, a comparison with other methods required.

A susceptible question in using this segmentation method is the determination of the parameter T_d, the threshold of statistical distance in the homogeneity test. Too small or too large values could cause either over segmentation or under segmentation. Theoretically, the selection of T_d controls the quantitative definition of homogeneity, or the *class separability*, in the sense of statistical classification. However, it is not realistic to fix this parameter for all kinds of range images. A solution to determine this parameter is to consider the number of final clusters in a given scene as a priori information. This relates to research work on the specification of the number of clusters called *cluster validation* ([34] and [67]), in which the number of clusters is determined by the maximum entropy principle. In this experiment, it was found to set the parameter T_d just empirically, simply by referring to the histogram obtained in the process of searching the seed patch from the candidates. Moreover, according to the experimental results, small changes of the value of T_d (for example, variation within ±20%) did not significantly change the segmentation results. This also illustrates that the statistical distance yields a reliable measure for *class separability*.

Chapter 7

Conclusions and Prospects

7.1 Conclusions

This thesis is about surface-based representations for the recognition of man-made objects from range data. Quadric primitives are thought to be appropriate for modelling the class of man-made objects, where in most cases the objects are reconstructed from simply shaped surface patches. In a top-down recognition strategy, representations of objects and models are usually established at a low level or intermediate level of the vision system, meaning that the outcomes of a feature description are highly data-driven. Because of the error presence in the raw range data, uncertainties inherently exist in representations and could further plague the robustness of higher level matching schemes. The theme of this thesis is thus to improve the reliability of primitive representations, being restricted to the class of quadric representations in our work.

The research of this thesis has been subdivided in a few topics: (1) Extension of the "frequency-difference" laser scanner, from a visual inspection system to a 3D measurement system. (2) Parameter estimation with quadric primitive representation, where a new computing algorithm for quadric fitting was proposed to improve the reliability of the estimates. (3) A study on the uncertainties of estimates of quadric representations, where a covariance model was proposed to explicitly represent the uncertainties in feature space. (4) Range image segmentation, where the covariance model of surface estimates was applied to optimise the region-based estimation and the representation of surface homogeneity.

- **3D measurements with the "frequency-difference" laser scanning sensing system**

Firstly, chapter 2 gave an introduction of the "frequency-difference" laser scanning sensing system, by which this thesis was initially motivated. Its particular scanning technique, i.e., the "frequency-difference" scanning, and the extension of the sensing modality, requires the design of special hardware and software. A new planar fitting algorithm has been proposed, that allows the accurate estimation of metric parameters in a 3D co-ordinate system. The measurement error is defined with respect to the sensor variables, rather than to the general *x-y-z* form, because they are closely related to the model of the independent identically distributed (i.i.d) noise. In this chapter we denoted this technique the "sensor-driven" manner. By a more detailed analysis, we also illustrated that estimates of parameters using the planar fitting algorithm are more reliable in terms of *accuracy* than those using a method based on point correspondence.

- **Parameter estimation with quadric primitive representation**

In chapter 3, a new approach for parameter estimation through surface fitting with range data has been described. Quadric fitting is a popular issue in surface reconstruction and a variety of estimation methods are applicable. We focussed on the bias-corrected renormalization approach because unbiasedness is a desired feature in the sense of optimisation. More important, we can incorporate the estimates of surface parameters with a popular statistical model, i.e., the normal pdf, for which only the statistics of expectation and covariance are required. However, in spite of having optimised the estimation, the solution of that approach is still deficient. We have proved that such a deficiency arises from the inexact formulation of a gradient-weighted least squares criterion in the eigenvector implementation, although it is commonly applied in alternative methods for quadric and conic fitting. To obtain better solutions with the renormalization approach, as well as for alternatives based on the general gradient-weighted least squares criterion, we proposed an approach to improve the reliability of the solutions in eigenvector implementations. Experimental results of conic fitting using synthetic data have confirmed the improvement.

We also studied the parameter estimation method incorporated in the sensing mechanism. Specific to the "frequency-difference" laser-scanning sensing technique, a sensor-driven approach for surface fitting has been proposed. Since the measurement errors are characterised based on a refined analysis of the sensing process, the estimates can also be improved. Simple simulations have verified the improvement (although slightly). A further indication for this improvement was obtained from the 3D measurement system (described in chapter 2), where the planar fitting utilised the sensor-driven algorithm (in contrast, the image-driven routine generated unexpected results). Considering that most of the 3D sensing modalities are application-specific, the sensor-driven strategy can facilitate undoubtedly the optimisation of the solutions in many low-level processing problems.

- **Uncertainties in estimates of quadric representation**

Chapter 4 focused on the subject of modelling the uncertainties in parameter estimation and feature extraction with an explicit form. First, the quantitative feature of a quadric representation is defined by algebraic invariants. For the implicit quadric form, the invariants can be explicitly expressed by the coefficients of the standard form of quadrics, for which a standard co-ordinate system is specified. Due to the simple structure of quadrics, feature extraction is implemented through a transformation between the sensor-specific representation (in the world co-ordinate system) and the feature-specific representation (in the standard co-ordinate system). The pose parameters (rotation and translation) are derived simultaneously.

Starting from the surface fitting routine, second order statistics of the estimates of quadric parameters, i.e., the covariance, were investigated. Within the perturbation framework and under the assumption of i.i.d Gaussian noise to model the errors in the raw range data, the covariance matrix of the estimate, as well as the feature vector, were formulated. Next, the parametric representation was established on an optimal framework —estimate plus covariance, which has been termed as the covariance model in this thesis. Therefore, the statistical behaviour of the representation can be formulated with an explicit form of a normal

pdf. This enables us to implement the routine of feature classification within the Bayesian principle.

It has been noticed that such a probabilistic model is an approximation. The basic assumption is the low-level i.i.d Gaussian noise to model the measurement errors in range data. Although such an assumption seems popular in image processing, errors in real measurements are much more complicated. This can explain the fact that in the experiments of optimal classification of surface primitives, the results when using synthetic data (described in chapter 4) were more optimistic than that with real range images (described in chapter 6). However, according to the experimental results of this thesis, the reliability of surface classification using the proposed optimal representation (based on the "statistical distance") was indeed improved, compared with that of a conventional representation (only the first-order statistics used, i.e., based on the Euclidean distance). For example, in the experiments with real range images, for the used trials, the lowest ratio of "correct classification" by the statistical distance was 55%, while that by the Euclidean distance was 20%. Of course, we can expect that further characterisation of measurement errors, particularly associated to certain sensing modalities, will improve the reliability of the covariance computation and, perhaps, refinement of the formulation of the probabilistic model.

Another notable topic discussed in chapter 4 is a pose description for surface representations. In most 3D vision systems, localisation of objects is ultimately required. Utilising quadric representations, pose parameters can be simultaneously derived via feature extraction at a surface-based primitive level. More attention has been paid to the formulation of "relational pose" parameters, by which the topological geometry between surface primitives of a solid object is specified. As will be mentioned in the next subsection, quadric primitives and relational pose can be combined to provide a generic modelling strategy for 3D object recognition.

- **Region-based range image segmentation through an optimal estimation-based implementation**

Chapter 5 discussed the low-level task of range image segmentation. Region-based segmentation requires the representation of a region, no matter how the surface region is parameterised. In a general sense, robustness of segmentation relies on the reliability of region representations. From this viewpoint, we regard the proposed region-based segmentation approach being optimised at the representation level.

Most existing approaches for range image segmentation, either edge-based or region-based, aim at application-independent interpretations of the scene. Some generic properties about the surface geometry, in particular surface-smoothness, are emphasised for the definition of surface homogeneity. Such general characteristics are used in low-level vision tasks because the data-driven strategy requires an application-independent process. However, because the outcomes of the segmentation are closely interrelated to the representation scheme, the role of surface representation in controlling segmentation requires sufficient attention. Such an idea has been expressed in our approach through a few points:

(1) the region-growing routine is estimation-based, i.e., the homogeneity is defined in the domain of surface representation using a parametric description.

(2) the reliability of representation has been considered in the homogeneity test, i.e., the "statistical distance" was used so the measure was in fact taken in probability space.

(3) the region-growing process was guided by the "measure of uncertainty", leading the process towards the reliable representation of the segments.

Therefore, the strategy of region-based segmentation used in our approach can also be regarded as "representation-based". In other words, we are more interested in how reliable the representations of these segments are, rather than how the scene is partitioned. This has been further illustrated with the experiments in chapter 6. Compared with other presented approaches, it seems that our method is more close to the spirit of "primitive extraction".

It should be pointed out that there is still space for improvement. Except for the possible refinement of the covariance model in quadric representation, the selection of the threshold in the merging criterion used in our method is basically a heuristic criterion. To set the selection of the threshold within a more mathematically rigorous solution, an optimistic scheme is to tackle this problem with the topic of *cluster validation*, with which the number of final clusters (the segments) is determined through other clustering principles based on a stochastic modelling procedure. Thus the threshold to distinguish different clusters can be inferred from the solution of cluster validation. This will be one of the topics in future research.

7.2 Prospects of 3D object recognition with quadric primitives

The work of this thesis was limited to a surface-based representation scheme, which was largely at low and intermediate level. Since the goal of a vision system is ultimately the interpretation of the scene, the consequence of this work falls in the high-level recognition phase.

For that, a natural extension of the research of this thesis is to utilise the optimal representation of quadric primitives to facilitate a generic object recognition scheme.

One of the general methodologies to model 3D objects is the surface-primitive modelling strategy, i.e., a set of surfaces and a graph describing their connectivity is sufficient to represent complex solid objects (the idea of so-called *boundary representation*, or B-rep). A typical example is the Delaunay triangulation (polyhedral patches) technique to describe arbitrary shaped objects ([48]). In the high level stage, the objects are usually modelled in a *relational structure* (RS) scheme and the techniques of matching of *attributed relational graphs* (ARG's) can be used in the recognition routine ([21]).

Compared with the polyhedral patches, quadric surfaces can generate representations for more complicated objects. In the ARG modelling scheme, the features of a quadric representation, as formulated in chapter 4 of this thesis, describe the *intrinsic* attributes of the graph, while the relational poses (also defined in chapter 4) describe the *extrinsic* attributes of the graph. Then the recognition can be implemented using graph matching techniques.
In image analysis, an optimal and systematic scheme is the Markov Random Field (MRF) model. Using the stochastic model for scene description, the knowledge-based matching is

formulated by the *maximum a posteriori probability* estimation. The MRF model is widely used in 2D image analysis (mostly for low-level tasks such as image segmentation) and extension to higher-level recognition receives much attention. Based on the studies of this thesis, optimisation in quadric primitive representation enables us to apply an explicit probabilistic model for object (scene) description within an ARG scheme. Till now, the subject of object recognition or scene interpretation using MRF model at high level has not yet well studied, especially in the field of 3D object recognition. However, scene interpretation based on primitive representations using the MRF model (or alternative optimal approaches) seems an attractive and optimistic solution for 3D vision, which could be a research subject in future work.

References

[1] Aloimonos, Y. , editor, Active Perception. Lawrence Erlbaum Associates, Hillsdale, NJ, 1993.

[2] Arman, F. and Aggarwal, J.K., "Model-based object recognition in dense-range images-a review", ACM Computing Survey, vol.25, no.1, pp. 5-43, 1993.

[3] Bajcsy, R., Active perception, Proceedings of the IEEE, vol. 76, no. 8, pp. 996-1005, 1988.

[4] Barrow, H. G and Tenenbaum, J. M., "Recovering intrinsic scene characteristics from images", Computer Vision Systems, Hanson, A. R. and Riseman, E.M., editors, Academic Press, New York, 1978.

[5] Belsley, D.A., Kuh, E. and Welsch, R. E., Regression Diagnostics, John Wiley, New York, 1980.

[6] Besl. P. J. and McFay, N. D., "A method for registration of 3-D shapes", IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 14, no. 2, pp.239-255, 1992.

[7] Besl, P. J, "Active, optical range imaging sensors", Machine Vision and Applications, vol. 1, pp. 127-152, 1988.

[8] Besl, P. J. and Jain, R. C., "Segmentation through variable-order surface fitting", IEEE Trans. Pattern Analysis and Machine Intelligence, vol.10, no.2, pp. 167-192, 1988.

[9] Bhandarkar, S. M. and Siebert, A., "Integrating edge and surface information for range image segmentation", Pattern Recognition, vol.25, no.9, pp.947-692, 1992.

[10] Bolle, R. M. and Vemuri, B. C., "On three-dimensional surface reconstruction methods", IEEE Trans. Pattern Analysis and Machine Intelligence, vol.13, no.1, pp.1-13, 1991.

[11] Bolle, R.M., Kjeldsen, R. and Sabbah, D., "Primitive shape extraction from depth maps", Proceedings of IEEE Workshop Comp. Vision (1987), pp. 324-326.

[12] Bolles R.C, et al, "Projector camera range sensing of three-dimensional data", Machine Intelligence Research Applied to Industrial Automation, SRI International (1983), pp. 29-43.

[13] Boyer, K.L., Mirza, M.J. and Ganguly, G., "The robust sequential estimator: a general approach and its application to surface organization in range data", IEEE Trans. Pattern Analysis and Machine Intelligence, vol.16, no.10, pp. 987-1001, 1994.

[14] Cha, Y.Y. and Gweon, D.G., "A calibration and range-data extraction algorithm for omnidirectional laser range finder of a free-range mobile robot", Mechatronics, vol. 6 no. 6, pp.665-689, 1996.

[15] Champleboux, G, et al, "Accurate calibration of cameras and range imaging sensors: the NPBS method", Proceedings IEEE Int. Conf. Robotics and Automation (1992), pp.1552-1557.

[16] Chen, C. H and Kak, A.C., "Modeling and calibration of a structured light scanner for 3D robot vision", Proceedings IEEE Conf. Robotics and Automation (1987), pp. 807-815.

[17] Cheng J.K. and Huang, T.S., "Image registration by matching relational structures", Pattern Recognition, vol. 17, no. 1, pp.149-159, 1984.

[18] Dop, E.R. van, Multi-sensor object recognition: the case of electronics recycling, Ph.D. thesis, University of Twente, 1999.

[19] Du, F and Bradly, M, "Self-calibration of the intrinsic parameters of cameras for active vision system", Proceedings IEEE Conf. Computer Vision and Pattern Recognition, (1993), pp.477-482.

[20] Dijck, H.A.L. van, "Object recognition with stereo vision and geometric hashing", Ph.D. thesis, University of Twente, 1999.

[21] Fan, T.J., Medioni, G. and Nevatia, R., "Recognizing 3-D objects using surface descriptions. IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 11, no. 11, pp. 1140-1157, 1989.

[22] Faugeras, O. D, Luing, O.T. and Maybank, S.J., "Camera self-calibration: theory and experiments", Proceedings Europe Conf. Computer Vision (1992), pp. 321-334.

[23] Faugeras, O. D., Herbert, M. and Pauchon, E., "Segmentation of range data into planar and quadratic patches", Proc. 3rd Computer Vision and Pattern Recognition Conf., (1983), pp. 8-13.

[24] Forsyth, D. et al, "Invariant descriptors for 3-D object recognition and pose", IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 13, no. 10, pp.971-991, 1991.

[25] Grace J. and Young, A., The algebra of Invariants, Cambridge University Press, 1903.

[26] Grimson, W. E. L. and Lozano-Perez, T., "Model-based recognition and localization from sparse range or tactile data", The International Journal of Robotics Research, vol. 3, no. 3, pp. 3-35, 1984.

[27] Grimson, W. E. L. and Lozano-Perez, T., "Localizing overlapping parts by searching the interpretation tree", IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 9, no. 4, pp. 469-482, 1987.

[28] Grimson, W. E. L., "The combinatorics of object recognition in cluttered environments using constrained search", Artificial Intelligence, vol. 44, pp. 121-165, 1990.

[29] Gu, J, "A study on frequency-difference laser scanning principle and underwater 3D vision sensing technique", Ph.D. thesis, Zhe Jiang University, 1990.

[30] Hall, E. L. et al, "Measuring curved surfaces for robot vision", IEEE Computer, pp.42-54, 1982.

[31] Han, J.H., Volz, R.A. and Mudge, T.M., "Range image segmentation and surface parameter extraction for 3-D object recognition of industrial parts", Proceedings of IEEE Int. Conf. on Robotics and Automation (1987), vol. 1, pp. 380-386.

[32] Hoffman, R. and Jain, A.K., "Segmentation and classification of range images", IEEE Trans. Pattern Analysis and Machine Intelligence, vol.9, no.5, pp.608-620, 1987.

[33] Hoover, A., et al, "An experimental comparison of range image segmentation algorithms", IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 18, no. 7, pp. 673-689, 1996.

[34] Jain, A. K., Cluster analysis, Handbook of Pattern Recognition and Image Processing, T. Y. Young and K. S. Fu Ed., New York: Academic, 1986.
[35] Kanatani, K.,"Statistical bias of conic fitting and renormalization", IEEE Trans. Pattern Analysis and Machine Intelligence, vol.16, no.3, pp.320-326, 1994.
[36] Kaveti, S., Khwang, E. and Wang, H., "Second-order implicit polynomials for segmentation of range images", Pattern Recognition, vol.29, no.6, pp. 937-949, 1996.
[37] Keren, D., "Some new invariants in computer vision", IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 16, no.7, pp.1143-1149, 1994.
[38] Keren, D., Subrahmonias, J. and Cooper, D., "Robust object recognition based on implicit algebraic curves and surfaces", Proc. IEEE Conf. on computer vision and pattern recognition, (Champaign, June, 1992), pp.791-794.
[39] Keren, D., Cooper, D. and Subrahmonia, J., "Describing complicated objects by implicit polynomials", IEEE Trans. Pattern Analysis and Machine Intelligence, vol.16, no.1, pp.1-13, 1994.
[40] Lamdan, Y., and Wolfson, H.J., "Geometric hashing: a general and efficient model-based recognition scheme", Proc. IEEE International Conference on Computer Vision (1988), pp. 238-249.
[41] Landy, M.S., Maloney, L.T. and Pavel, M., editors. Exploratory Vision: The active eye. Springer Verlag, New York, 1996.
[42] LaValle, S.M. and Hutchinson, S.A., "A Bayesian segmentation methodology for parametric image models", IEEE Trans. Pattern Analysis and Machine Intelligence, vol.17, no.2, pp. 211-217, 1995.
[43] Li, S. Z., "Invariant representation, matching and pose estimation of 3D space curves under similarity transformations", Pattern Recognition, vol. 30, no. 3, pp.447-458, 1997,
[44] Ma, S.D, "A self-calibration technique for active vision system", IEEE Trans. Robotics and Automation, vol. 12, pp. 114-120, 1996.
[45] Marr, D. Vision-a computational investigation into the human representation and processing of visual information. Freeman, San Francisco, 1982.
[46] Oshima, M. and Shirai, Y., "Object recognition using three-dimensional information", IEEE Trans. Pattern Analysis and Machine Intelligence, vol.5, no.4, pp. 353-361, 1983.
[47] Pavlidis T. and Liow, Y., "Integrating region growing and edge detection", IEEE Trans. Pattern Analysis and Machine Intelligence, vol.12, no.3, pp. 225-233, 1990.
[48] Preparata, F. P., and Shamos, M. L., Computational Geometry — An introduction, Springer Verlag, Berlin, 1985.
[49] Reid, I, "Projective calibration of a laser-stripe range finder", Image and Vision Computing, vol. 14, pp. 659-666, 1996.
[50] Requicha, A.A., "Representation for rigid solids: theory, methods and systems", ACM Compt. Surveys, vol. 12, no. 4, pp. 437-464, 1980.
[51] Rivlin, E. and Weiss, I., "Local invariants for recognition", IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 17, no. 3, pp.226-238, 1995.
[52] Sabata, B., Arman, F. and Aggarwal, J. K., "Segmentation of 3D range images using pyramidal data structures", CVGIP: Image Understanding, vol. 57, no. 3, pp.373-387, 1993.

[53] Stewart, G.W., Introduction to matrix computations, Academic Press, New York, 1973.
[54] Stockman, G., "Object recognition and localization via pose clustering", Computer Vision, Graphics and Image Processing, vol. 40, pp. 361-387, 1987.
[55] Subrahmonias, J., Cooper, D. and Keren, D., "Practical reliable Bayesian recognition of 2D and 3D objects using implicit polynomials and algebraic invariants", IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 18, no. 5, pp.505-519, 1996.
[56] Taubin, G., et al, "Parameterized families of polynomials for bounded algebraic curve and surface fitting", IEEE Trans. Pattern Analysis and Machine Intelligence, vol.16, no.3, pp.287-303, 1994.
[57] Taubin, G. and Cooper, D., "3D object recognition and positioning with algebraic invariants and covariants", Symbolic and Numerical Computations Towards Integration, B. R. Donald, D. Kapur, and J. Mundy eds., Academic Press, pp. 147-182, 1992.
[58] Taubin, G., "Estimation of planar curves, surfaces, and nonplanar space curves defined by implicit equations with applications to edge and range image segmentation", IEEE Trans. Pattern Analysis and Machine Intelligence, vol.13, no.11, pp.1115-1138, 1991.
[59] Taylor, R.W., Savini, M. and Reeves, A. P., "Fast segmentation of range imagery into planar regions", CVGIP, vol.45, pp.42-60, 1989.
[60] Ullman, S., "Aligning pictorial descriptions: an approach to object recognition", Cognition, vol. 32, pp.193-254, 1989.
[61] Umasuthan, M. and Wallace, A. M., "Model indexing and object recognition using 3D viewpoint invariance", Pattern Recognition, vol. 30, no. 9, pp.1415-1434, 1997.
[62] Wei, G.Q, Arbter, K. and Hirzinger, G., "Active self-calibration of robotic eyes and hand-eye relationships with model identification", IEEE Trans. Robotics and Automation, vol. 14, no. 1, pp. 158-166, 1998.
[63] Whaite, P. and Ferrie, F.P., "From uncertainty to visual exploration", IEEE Trans. Pattern Analysis and Machine Intelligence, vol.13, no.10, pp.1038-1049, 1991.
[64] Wolfson, H.J., "On curve matching", IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 12, no. 5, pp.483-489, 1990.
[65] Yokoya, N. and Levine, M. D., "Range image segmentation based on differential geometry: a hybrid approach", IEEE Trans. Pattern Analysis and Machine Intelligence, vol.11, no.6, pp.643-649, 1989.
[66] Yu, X., Bui, T.D. and Krzyzak, A., "Robust estimation for range images segmentation and reconstruction", IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 16, no. 5, pp. 530-538, 1994.
[67] Zhang, J. and Modestino, J. W., "A model-fitting approach to cluster validation with application to stochastic model-based image segmentation", IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 12, no. 10, pp. 1009-1017, 1990.
[68] Zhang, Z., "Parameter estimation techniques: a tutorial with application to conic fitting", Image and Vision Computing, vol.15, pp.59-76, 1997.
[69] Zheng, G, et al, "Laser difference frequency scanning 3-D vision sensing system for underwater robot", Proceedings Intervention ROV'91 (1991), pp. 90-96.
[70] Zheng, G and Zheng, B., "A new laser 3D image information display method for ROV", Proceedings Intervention ROV'92 (1992), pp. 183-188.

List of Symbols

Chapter 1:

r_{ij}	a regularly organised range image, where r indicates the range measurement associated to a specified co-ordinate system and (i, j) is a pair of integers as the indices.

Chapter 2:

S_0	baseline of laser scanning triangulation
P	point of the object being sensed
F	laser emitter and the origin of the measuring co-ordinate system
x-*y*-*z*	the measuring co-ordinate system; its origin is **F**, and the *y*-axis is perpendicular to the frame-scanning plane Γ
S	detector
L	baseline length = distance between **F** and **S**.
α	*frame*-scanning angle of laser scanning triangulation $\alpha=\alpha_0+\omega_f t_f$
β	*slit*-scanning angle of laser scanning triangulation $\beta=\beta_0+\omega_s t_s$
γ	*line*-scanning angle of laser scanning triangulation $\gamma=\gamma_0+\omega_l t_l$
α_0	initial angle of the periodic scanning for α
β_0	initial angle of the periodic scanning for β
γ_0	initial angle of the periodic scanning for γ
t_f	time with respect to starting time of *frame*-scanning at which the point P is scanned
t_s	time with respect to starting time of *slit*-scanning at which the point P is scanned
t_l	time with respect to starting time of *line*-scanning at which the point P is scanned
ω_f	*frame*-scanning frequency
ω_s	*slit*-scanning frequency
R_{sf}	*frequency-difference* coefficient defined as the quotient of ω_s and ω_f.
Ψ	detecting plane
Γ	frame-scanning plane
$\mathbf{n}_s$	unit vector indicating the direction of the axis about which the detecting plane Ψ rotates
ρ_1	azimuth angle of $\mathbf{n}_s$ with respect to the *x*-*y*-*z* co-ordinate system
ρ_2	elevation angle of $\mathbf{n}_s$ with respect to the *x*-*y*-*z* co-ordinate system
θ	angle between $\mathbf{n}_s$ and plane Ψ
h	the distance of point **S** to plane Ψ

x'-y'-z'	assistant co-ordinate system connected to the *frame-scanning* plane Γ
e	unit normal vector of the plane Ψ expressed in *x'-y'-z'* co-ordinates
L	translation vector pointing from **F** to **S** with length L
S	set of known data
a	parameters of the measurement system
S′ (**a**)	measured set of data of S
â	estimate of **a**
Θ	cost function defined according the estimation criterion

Chapter 3:

$\mathbf{x}_i$	the i^{th} data point in *x-y-z* co-ordinates in vector form
$f(x,y,z)$	implicit form to describe a surface
$\mathbf{x}_0$	the true value of the point on the surface $f(\mathbf{x}_0)=0$
$\Delta\mathbf{x}$	$= \mathbf{x}-\mathbf{x}_0$ represents the noise
n	the normal vector of the surface at **x**
d	the distance of a data point **x** to the surface measured in the direction of **n**
w_i	weight $\equiv \left\|\nabla f(\mathbf{x}_i)\right\|^{-2}$
p	parameter vector consisting of the algebraic coefficients of quadrics
$\mathbf{p}_0$	the ideal parameter vector
M	measurement vector defined as $[x^2, y^2, z^2, xy, xz, yz, x, y, z, 1]^T$ associated with a data point **x**
$\mathbf{N}_i$	the matrix formed by the vector product $\mathbf{M}_i\mathbf{M}_i^T$ with $\mathbf{M}_i$ the measurement vector for the i^{th} data point $\mathbf{x}_i$
N	the weighted sum of matrices $\equiv \Sigma_{i=1\ldots n}\, w_i\mathbf{N}_i$
$\Delta\mathbf{N}$	perturbation matrix defined as $\mathbf{N}-\mathbf{N}_0$
L	Lagrange multiplier
σ^2	the variance of the Gaussian noise $\Delta\mathbf{x}_i\ \forall i$
$\mathbf{B}_i$	a matrix consisting of the coefficients of the second-order expansion with respect to $\Delta\mathbf{x}_i$ of each of the elements of $\mathbf{N}_i$
$\overline{\mathbf{N}}$	matrix defined as $\mathbf{N}-E[\Delta\mathbf{N}]$
$\mathbf{p}^{(0)}$	initial estimate of **p**
c	$\equiv\sigma^2$, being estimated in the renormalisation process
E	matrix introduced in the derivation of the *improved bias-corrected renormalisation solution* defined as $\mathbf{E} \equiv \frac{1}{2}\sum_{i=1}^{n}\left[\left(\frac{\partial w_i}{\partial \mathbf{p}}\right)\mathbf{p}^T\mathbf{N}_i\right]$
$\widetilde{\mathbf{N}}$	matrix defined as $\widetilde{\mathbf{N}} = \mathbf{N} + \mathbf{E}$
$\widetilde{\mathbf{p}}$	eigenvector of $\widetilde{\mathbf{N}}$
$\overline{\widetilde{\mathbf{N}}}$	matrix defined as $\widetilde{\mathbf{N}} - E[\Delta\widetilde{\mathbf{N}}]$

Chapter 4:

x_s-y_s-z_s	standard co-ordinate system or frame

$\{\gamma_1,\gamma_2,c\}$	the set of invariants, determining the type of quadric
$\mathbf{g}$	the feature vector defined as $[\gamma_1, \gamma_2, c]^T$
$\mathbf{A}$, $\mathbf{v}$	the symmetric 3×3 matrix $\mathbf{A}$ and the 3×1 vector $\mathbf{v}$ describing a quadric in world co-ordinates as $\mathbf{x}^T\mathbf{A}\mathbf{x}+\mathbf{v}^T\mathbf{x}+k=0$
$\mathbf{R}$	the rotation matrix of the rotation transformation between Cartesian co-ordinates
$\mathbf{A}'$, $\mathbf{v}'$	the surface parameters represented in the rotation-transformed co-ordinate system
$\mathbf{t}'$	the translation vector of the transformation from the rotated co-ordinate system to the standard co-ordinate system
$\mathbf{t}$	the translation vector of the transformation from the world co-ordinate system to the standard co-ordinate system
$\boldsymbol{\theta}$	the unconstrained parameter vector as $[a_{11}, a_{22}, a_{33}, a_{12}, a_{13}, a_{23}, v_1, v_2, v_3]^T$, while setting k=1 in quadric expression
$\mathbf{x}'$	the "rotation-normalised" co-ordinate system, in which the matrix $\mathbf{A}'$ is diagonal
$\boldsymbol{\theta}'$	estimate of the surface parameters assuming that they are obtained in the rotation-normalised co-ordinate system
$\mathbf{M}$	a 3×9 coefficient matrix (see appendix B) relating the perturbations $\Delta\mathbf{g}$ and $\Delta\boldsymbol{\theta}'$ by $\Delta\mathbf{g}=\mathbf{M}\Delta\boldsymbol{\theta}'$
$\boldsymbol{\Psi}$	covariance matrix of $\boldsymbol{\theta}$
$\mathbf{G}$	covariance matrix of $\mathbf{g}$
$f(\hat{\mathbf{g}}\|\mathbf{g})$	the normal pdf of the estimate $\hat{\mathbf{g}}$ with respect to the ideal $\mathbf{g}$
$f(\hat{\mathbf{g}}_1, \hat{\mathbf{g}}_2\|\mathbf{g}_m)$	the joint pdf of $\hat{\mathbf{g}}_1$ and $\hat{\mathbf{g}}_2$, as two observations of the ideal $\mathbf{g}_m$
$f(\hat{\mathbf{g}}_1, \hat{\mathbf{g}}_2\|H)$	the joint pdf of $\hat{\mathbf{g}}_1$ and $\hat{\mathbf{g}}_2$, conditioned on the hypothesis H
$\mathbf{G}_{ij}$	matrix defined as $\mathbf{G}_{ij} = \mathbf{G}_i+\mathbf{G}_j$
$\mathbf{G}_i$	covariance matrix of $\hat{\mathbf{g}}_i$
$\hat{\mathbf{g}}_m$	observation from the known class m
$\hat{\mathbf{g}}_s$	observation from the scene
$(d_{s,m})^2$	the "statistical distance" defined as $\frac{1}{2}(\hat{\mathbf{g}}_s - \hat{\mathbf{g}}_m)^T\mathbf{G}_{s,m}^{-1}(\hat{\mathbf{g}}_s - \hat{\mathbf{g}}_m)$
$\overline{\mathbf{g}}_m$	the feature stored in the model base for the m^{th} primitive
$\overline{\mathbf{G}}_m$	the covariance matrix stored in the model base for the m^{th} primitive
$\boldsymbol{\mu}$	vector defined as $[\alpha, \beta, \gamma]^T$ containing the rotation angles of the rotation matrix $\mathbf{R}$
$\boldsymbol{\tau}_w$	the vector representing the pose parameters of a surface under the world co-ordinate system, defined as $[\alpha, \beta, \gamma, t'_x, t'_y, t'_z]^T$
$\boldsymbol{\tau}_{wc}$	pose vector of a cylindrical surface
$\boldsymbol{\rho}_w$	pose vector of a plane
$\mathbf{R}_{ab}$	rotation between the two standard co-ordinate systems of surfaces a and b
$\mathbf{t}'_{ab}$	translation between the two standard co-ordinate systems of surfaces a and b
$\boldsymbol{\tau}$	vector representing the "relational pose" of two surfaces
$\boldsymbol{\tau}_{cc}$	relational pose vector of two cylindrical surfaces

$\boldsymbol{\rho}_a$	vector representing the relational pose of planar surface *b* with respect to surface *a*
$\boldsymbol{\rho}_{ac}$	vector representing the relational pose of planar surface *b* with respect to cylindrical surface *a*
$\boldsymbol{\rho}_{nn}$	vector representing the relational pose of two planar surfaces
$\boldsymbol{\eta}$	vector of surface parameters consisting of feature vector and the pose vector
$\mathbf{T}_w$	covariance matrix of $\boldsymbol{\tau}_w$
$\boldsymbol{\Phi}$	covariance matrix of $\boldsymbol{\eta}$

Chapter 5:

$\mathbf{Z}$	the estimate of the spatial covariance of a set of points $\{\mathbf{x}_i\}$ in a window
$c_{\min}$	the smallest eigenvalue of $\mathbf{Z}$
T_{win}	the threshold of $c_{\min}$ to detect outliers
T_{depth}	the threshold of the depth difference between the centre point of a window- and its nearest point to detect outliers
$R^{(k)}$	the merged region at k^{th} step
$(\boldsymbol{\theta}_k , \boldsymbol{\Psi}_k)$	the estimate of the surface parameter vector and the covariance matrix of the region $R^{(k)}$
$\boldsymbol{\varepsilon}_{k,k+1}$	$\equiv \boldsymbol{\theta}_k - \boldsymbol{\theta}_{k+1}$
$d^2_{k,k+1}$	the statistical distance between the regions before and after merging a sel ($R^{(k)}$ and $R^{(k+1)}$)
T_d	the threshold of $d^2_{k,k+1}$ for homogeneity test
$\left\|\boldsymbol{\Psi}_k\right\|_2$	the *measure of uncertainty* of the region $R^{(k)}$
c_{seg}	estimate of the variance of noise in the segmented region
T_{plane}	the threshold for planarity test
$\boldsymbol{\theta}_{nb}$	vector representing the parameters of a planar surface
$\boldsymbol{\theta}_n$	vector representing the orientation of a planar surface (part of $\boldsymbol{\theta}_{nb}$)
$\boldsymbol{\Psi}_{nb}$	covariance matrix of $\boldsymbol{\theta}_{nb}$
$\boldsymbol{\Psi}_n$	covariance matrix of $\boldsymbol{\theta}_n$

Appendix A

Computation of 3D Co-ordinates of the Range Data Acquired with the "Frequency-Difference" Laser Scanning Sensing System

After defining the measurement parameters in chapter 2, we can formulate the three-dimensional co-ordinates of the sensed point under the defined Cartesian system. To illustrate the principles of deriving the 3D formulations, we combine Fig 2.8 and Fig 2.9 to show the co-ordinate definitions in Fig A.1.

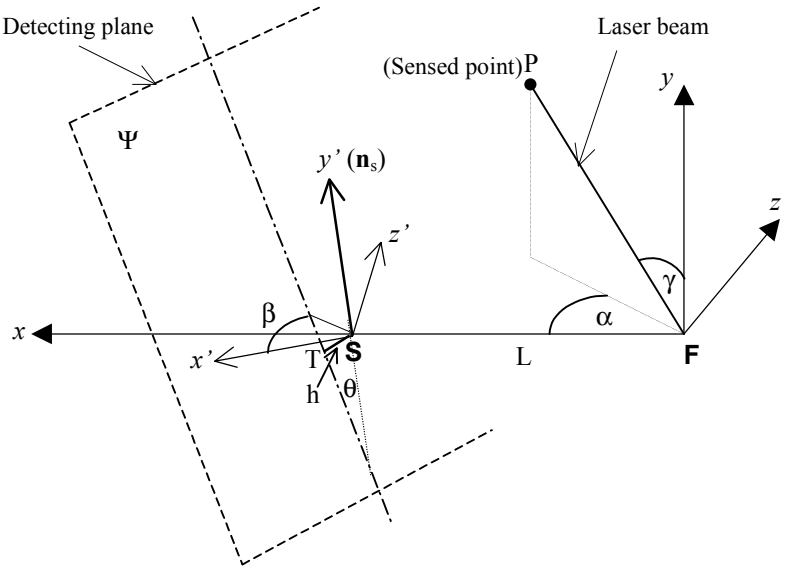


Fig A.1 The measurement co-ordinate system.

Notations of the co-ordinate system for 3D measurement are explained below:

F — Origin of the scanning laser beam.
S — **I**ntersecting point of the rotation axis $\mathbf{n}_s$ and the frame scanning plane (the *x*-*z* plane), from which the *x*-axis is defined.
L — The distance between **F** and **S**, i.e., the baseline.
$\mathbf{n}_s$ — The direction of rotation axis of the detecting plane. The detecting slit runs around this central axis in representation of β.
h — The distance between point S and the detecting plane Ψ.
θ — The angle between $\mathbf{n}_s$ and the plane Ψ.

Obviously, a sensed point P is just the intersection of the scanning laser beam and the detecting plane Ψ. Therefore, if we express the plane equation of Ψ and the line equation of the laser beam under the defined *x*-*y*-*z* co-ordinate system, then the solution of these two equations give rise the formulation of the 3-D co-ordinates of point P.

It is easy to write down the expression of the line equation of the laser beam under the *x*-*y*-*z* co-ordinate system. We can express it in a vector form as

$$\mathbf{x} = K\begin{bmatrix}\cos\alpha\sin\gamma \\ \cos\gamma \\ \sin\alpha\sin\gamma\end{bmatrix} = K\mathbf{q}, \tag{A.1}$$

where K is the scale factor and $\mathbf{q}$ is the unit vector of the line. To express the plane equation under the *x*-*y*-*z* co-ordinate system, we need to utilise the assistance of the system parameters.

First, we define an assistant co-ordinate system *x'*-*y'*-*z'*, which is originated at point **S**. The *y'*- axis is defined on $\mathbf{n}_s$. The *x'*- axis is defined that it is orthogonal to *y'* and lies on the plane consisting of *y'*- and *x*- axes. Then *z'*- axis is defined as the cross production of *x'* and *y'* (the right-hand rule).

Under the *x'*-*y'*-*z'* co-ordinate system, it is easy to express the plane equation of Ψ as

$$\mathbf{x}'^{T}\mathbf{e} = h, \tag{A.2}$$

where $\mathbf{e}$ is the unit normal vector of Ψ, which can be formulated as

$$\mathbf{e} = \begin{bmatrix}\cos\beta\cos\theta \\ -\sin\theta \\ \sin\beta\cos\theta\end{bmatrix} \tag{A.3}$$

Second, we use the transformation $\mathbf{x} = \mathbf{R}\mathbf{x}' + \mathbf{L}$ to represent the express of (A.2) under the *x*-*y*-*z* co-ordinate system:

$$(\mathbf{x} - \mathbf{L})^{T}\mathbf{R}\mathbf{e} = h. \tag{A.4}$$

Substituting (A.1) into (A.4), we get:

$$K = \frac{h + (\mathbf{R}^T \mathbf{L})^T \mathbf{e}}{(\mathbf{R}^T \mathbf{q})^T \mathbf{e}}. \tag{A.5}$$

Then the 3-D co-ordinates of point P can be expressed according to (A.1) and (A.5).

Now we further derive the expressions of (A.5) in form of the system parameters. Denoting the three basic vectors of the co-ordinate system *x'-y'-z'* as ($\mathbf{v}_x$, $\mathbf{v}_y$, $\mathbf{v}_z$), the rotation matrix **R** can be expressed as

$$\mathbf{R} = \begin{bmatrix} \mathbf{v}_x & \mathbf{v}_y & \mathbf{v}_z \end{bmatrix}.$$

According to the definition of the co-ordinate system *x'-y'-z'* , we know that $\mathbf{v}_y = \mathbf{n}_s$, therefore,

$$\mathbf{v}_y = \begin{bmatrix} n_x \\ n_y \\ n_z \end{bmatrix} = \begin{bmatrix} \sin\rho_1 \cos\rho_2 \\ \cos\rho_1 \\ \sin\rho_1 \sin\rho_2 \end{bmatrix}, \tag{A.6}$$

where ρ_1 and ρ_2 are the azimuth and elevation angles of $\mathbf{n}_s$ (see Fig A.2).

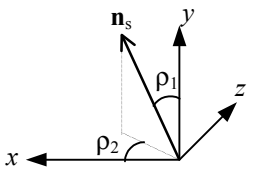


Fig A.2. Azimuth and elevation angles of $\mathbf{n}_s$.

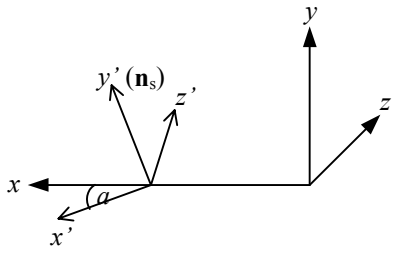


Fig A.3 definition of the angle (*a)*. Noted that *x*, *y'* and *x'* are coplanar.

Because the axes of $\mathbf{n}_s$, *x* and *x'* are co-planar, we can derive the vector $\mathbf{v}_x$ from the following expression:

$$\mathbf{i} = \cos(a)\mathbf{v}_x + \sin(a)\mathbf{n}_s , \tag{A.7}$$

where **i** is the unit vector of the axis *x*, *a* is the angle between *x* and *x'* (see Fig A.3). Therefor,

$$\mathbf{v}_x = \frac{\mathbf{i} - \sin(a)\mathbf{n}_s}{\cos(a)} = \frac{1}{\cos(a)} \begin{bmatrix} 1 - n_x^2 \\ -n_x n_y \\ -n_x n_z \end{bmatrix}. \tag{A.8}$$

Here we have used the relations of

$$\mathbf{i} = [1,0,0]^T \; ; \; \sin(a) = n_x .$$

Finally the vector $\mathbf{v}_z$ can be derived from $\mathbf{v}_z = \mathbf{v}_x \times \mathbf{v}_y$. According to (A.8), we get

$$\mathbf{v}_z = \frac{1}{\cos(a)}(\mathbf{i} \times \mathbf{n}_s) = \frac{1}{\cos(a)} \begin{bmatrix} 0 \\ -n_z \\ n_y \end{bmatrix} \tag{A.9}$$

Now the three basic vectors ($\mathbf{v}_x$, $\mathbf{v}_y$, $\mathbf{v}_z$) have been formulated in terms of the parameters ρ_1 and ρ_2 through (A.6), (A.8) and (A.9).

Denoting $\mathbf{q}' = \mathbf{R}^T\mathbf{q}$, we can obtain the following results:

$$\begin{aligned} q'_x &= \tfrac{1}{\cos(a)}(\sin\gamma\cos\alpha - \sin(a)q'_y) \\ q'_y &= \tfrac{1}{\cos(a)}(\sin\rho_1\cos\rho_2\sin\gamma\cos\alpha + \cos\rho_1\cos\gamma + \sin\rho_1\sin\rho_2\sin\gamma\sin\alpha) \\ q'_z &= \tfrac{1}{\cos(a)}(-\sin\rho_1\sin\rho_2\cos\gamma + \cos\rho_1\sin\gamma\sin\alpha) \\ \sin(a) &= \sin\rho_1\cos\rho_2; \quad \cos(a) = \sqrt{1-\sin^2(a)} \end{aligned} \tag{A.10}$$

Denoting $\mathbf{L}' = \mathbf{R}^T\mathbf{L}$, we can obtain

$$\begin{aligned} L'_x &= \tfrac{1}{\cos(a)}(L - \sin(a)L'_y) \\ L'_y &= L\sin\rho_1\cos\rho_2 \\ L'_z &= 0 \end{aligned} \tag{A.11}$$

Expanding (A.5) in the form of

$$K = \frac{h + L'_x e_x + L'_y e_y}{q'_x e_x + q'_y e_y + q'_z e_z}, \tag{A.12}$$

and combining (A.3), (A.10) and (A.11), the parameter K can be further expressed with the system parameters.

Finally, the 3-D co-ordinates are represented in the form

$$\begin{cases} x = K\sin\gamma\cos\alpha \\ y = K\cos\gamma \\ z = K\sin\gamma\sin\alpha \end{cases}, \tag{A.13}$$

where K is computed from (A.12).

Appendix B

Expression of the Matrix $\mathbf{B}_i$ Used in (3.10)

To derive the expression of the matrix $\mathbf{B}_i$, first we write down the matrix $\mathbf{N}_i$ in (3.5) with its expression

$$\mathbf{N}_i = \begin{bmatrix} x_i^4 & x_i^2y_i^2 & x_i^2z_i^2 & x_i^3y_i & x_i^3z_i & x_i^2y_iz_i & x_i^3 & x_i^2y_i & x_i^2z_i & x_i^2 \\ x_i^2y_i^2 & y_i^4 & y_i^2z_i^2 & y_i^3x_i & y_i^2x_iz_i & y_i^3z_i & y_i^2x_i & y_i^3 & y_i^2z_i & y_i^2 \\ x_i^2z_i^2 & y_i^2z_i^2 & z_i^4 & z_i^2x_iy_i & z_i^3x_i & z_i^3y_i & z_i^2x_i & z_i^2y_i & z_i^3 & z_i^2 \\ x_i^3y_i & y_i^3x_i & z_i^2x_iy_i & x_i^2y_i^2 & x_i^2y_iz_i & y_i^2x_iz_i & x_i^2y_i & y_i^2x_i & x_iy_iz_i & x_iy_i \\ x_i^3z_i & y_i^2x_iz_i & z_i^3x & x_i^2y_iz_i & x_i^2z_i^2 & z_i^2x_iy_i & x_i^2z_i & x_iy_iz_i & z_i^2x_i & x_iz_i \\ x_i^2y_iz_i & y_i^3z_i & z_i^3y_i & y_i^2x_iz_i & z_i^2x_iy_i & y_i^2z_i^2 & x_iy_iz_i & y_i^2z_i & z_i^2y_i & y_iz_i \\ x_i^3 & y_i^2x_i & z_i^2x_i & x_i^2y_i & x_i^2z_i & x_iy_iz_i & x_i^2 & x_iy_i & x_iz_i & x_i \\ x_i^2y_i & y_i^3 & z_i^2y_i & y_i^2x_i & x_iy_iz_i & y_i^2z_i & x_iy_i & y_i^2 & y_iz_i & y_i \\ x_i^2z_i & y_i^2z_i & z_i^3 & x_iy_iz_i & z_i^2x_i & z_i^2y_i & x_iz_i & y_iz_i & z_i^2 & z_i \\ x_i^2 & y_i^2 & z_i^2 & x_iy_i & x_iz_i & y_iz_i & x_i & y_i & z_i & 1 \end{bmatrix} \tag{B.1}$$

Because of noise presence in $\{\mathbf{x}_i\}_{i=1}^n$, the matrix $\mathbf{N}_i$ is perturbed from its ideal value. According to [68], we carry out the Taylor expansion to each of the elements of matrix $\mathbf{N}_i$ in term of Δx_i, Δy_i and Δz_i,. Under the assumption of i.i.d Gaussian noise, i.e.,

$$E\{\Delta\mathbf{x}_i\} = \mathbf{0}; \quad E\{\Delta\mathbf{x}_i\Delta\mathbf{x}_i^T\} = \sigma^2\mathbf{I},$$

we get the expectation of $\Delta\mathbf{N}_i$ in the form

$$E\{\Delta\mathbf{N}_i\} = \sigma^2 \begin{bmatrix} 6x_i^2 & x_i^2+y_i^2 & x_i^2+z_i^2 & 3x_iy_i & 3x_iz_i & y_iz_i & 3x_i & y_i & z_i & 1 \\ x_i^2+y_i^2 & 6y_i^2 & y_i^2+z_i^2 & 3x_iy_i & x_iz_i & 3y_iz_i & x_i & 3y_i & z_i & 1 \\ x_i^2+z_i^2 & y_i^2+z_i^2 & 6z_i^2 & x_iy_i & 3x_iz_i & 3y_iz_i & x_i & y_i & 3z_i & 1 \\ 3x_iy_i & 3x_iy_i & x_iy_i & x_i^2+y_i^2 & y_iz_i & x_iz_i & y_i & x_i & 0 & 0 \\ 3x_iz_i & x_iz_i & 3x_iz_i & y_iz_i & x_i^2+z_i^2 & x_iy_i & z_i & 0 & x_i & 0 \\ y_iz_i & 3y_iz_i & 3y_iz_i & x_iz_i & x_iy_i & y_i^2+z_i^2 & 0 & z_i & y_i & 0 \\ 3x_i & x_i & x_i & y_i & z_i & 0 & 1 & 0 & 0 & 0 \\ y_i & 3y_i & y_i & x_i & 0 & z_i & 0 & 1 & 0 & 0 \\ z_i & z_i & 3z_i & 0 & x_i & y_i & 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} = c\mathbf{B}_i \tag{B.2}$$

The matrix $\mathbf{B}_i$ in (3.10) is thus expressed by (B.2).

Appendix C.1

Formulation of the Covariance Matrices of Surface Parameters

Within the perturbation framework, the covariance matrices of the estimated surface parameters and the derived feature vector can be formulated by applying the linear approximation.

C.1.1 Covariance matrix of surface parameters

In chapter 4, we defined the parameter vector of a quadric surface as

$$\boldsymbol{\theta} = [a_{11}, a_{22}, a_{33}, a_{12}, a_{13}, a_{23}, v_1, v_2, v_3]^T .$$

The cost function for surface fitting is

$$\Theta = \sum_{i=1}^{n} d_i^2 ,$$

where $d_i \forall i$=1,..,n is the geometric distance between the point $\mathbf{x}_i$ to the surface.

Denoting $\boldsymbol{\theta}_0$ the true parameters in case of noise-free estimation and $\mathbf{x}_{i0}$ is the ideal measurement of $\mathbf{x}_i$, we can take the Taylor expansion of d_i^2 up to second order approximation:

$$\begin{aligned} d_i^2 = {} & d^2(\mathbf{x}_{i0}, \boldsymbol{\theta}_0) + \Delta\theta^T \frac{\partial d^2(\mathbf{x}_{i0}, \boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} + \Delta\mathbf{x}_i^T \frac{\partial d^2(\mathbf{x}_{i0}, \boldsymbol{\theta}_0)}{\partial \mathbf{x}_i} + \Delta\boldsymbol{\theta}^T \frac{\partial^2 [d^2(\mathbf{x}_{i0}, \boldsymbol{\theta}_0)]}{\partial \boldsymbol{\theta} \partial \mathbf{x}_i} \Delta\mathbf{x}_i \\ & + \Delta\mathbf{x}_i^T \frac{\partial^2 [d^2(\mathbf{x}_{i0}, \boldsymbol{\theta}_0)]}{\partial \mathbf{x}_i^2} \Delta\mathbf{x}_i + \Delta\boldsymbol{\theta}^T \frac{\partial^2 [d^2(\mathbf{x}_{i0}, \boldsymbol{\theta}_0)]}{\partial \boldsymbol{\theta}^2} \Delta\boldsymbol{\theta} \end{aligned} \tag{C.1}$$

where $\Delta\mathbf{x}_i = \mathbf{x}_i - \mathbf{x}_{i0}$ is the noise perturbation and $\Delta\boldsymbol{\theta} = \boldsymbol{\theta} - \boldsymbol{\theta}_0$ is the error in the estimated parameter $\boldsymbol{\theta}$.

Because in case of noise-free measurement, $\boldsymbol{\theta}_0$ is just the estimate by minimising the cost function, there are

$$d^2(\mathbf{x}_{0,i}, \boldsymbol{\theta}_0) = 0; \quad \frac{\partial d^2(\mathbf{x}_{0,i}, \boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} = 0 \ \forall i .$$

Therefore, taking the operation of $\partial\Theta/\partial\boldsymbol{\theta} = 0$ and substituting (C.1) with the expression of d_i^2, we get

$$\Delta\boldsymbol{\theta} = -\mathbf{M}_{\theta\theta}^{-1}\sum_{i=1}^{n}\mathbf{M}_{i,x\theta}\Delta\mathbf{x}_i\,, \tag{C.2}$$

where

$$\mathbf{M}_{\theta\theta} = \sum_{i=1}^{n}\mathbf{M}_{i,\theta\theta}\ ;\ \mathbf{M}_{i,\theta\theta} = \frac{\partial^2}{\partial\boldsymbol{\theta}^2}d^2\left(\boldsymbol{\theta}_0, \mathbf{x}_{i0}\right);\ \mathbf{M}_{i,x\theta} = \frac{\partial^2}{\partial\boldsymbol{\theta}\partial\mathbf{x}_i}d^2\left(\boldsymbol{\theta}_0, \mathbf{x}_{i0}\right). \tag{C.3}$$

Assuming that $\Delta\mathbf{x}_i\ \forall i$ is i.i.d Gaussian noise, there are

$$E\{\Delta\mathbf{x}_i\Delta\mathbf{x}_i^T\} = \sigma^2\mathbf{I}\ \ \forall i;\quad E\{\Delta\mathbf{x}_i\Delta\mathbf{x}_j^T\} = 0\ \ \forall i \neq j\,,$$

thus the covariance matrix of $\boldsymbol{\theta}$ is formulated as

$$\Psi = E\left[\Delta\boldsymbol{\theta}\Delta\boldsymbol{\theta}^T\right] = \sigma^2\mathbf{M}_{\theta\theta}^{-1}\left(\sum_{i=1}^{n}\mathbf{M}_{i,x\theta}\mathbf{M}_{i,x\theta}^T\right)\mathbf{M}_{\theta\theta}^{-T} \tag{C.4}$$

The expression of (C.4) can be further simplified. Because $d(\mathbf{x}_{i0}, \boldsymbol{\theta}_0) = 0\ \forall i$, from the expression of (C.3), we have

$$\mathbf{M}_{i,x\theta} = 2\left(\frac{\partial d_i}{\partial\boldsymbol{\theta}}\right)\left(\frac{\partial d_i}{\partial\mathbf{x}_i}\right)^T;\ \mathbf{M}_{i,\theta\theta} = 2\left(\frac{\partial d}{\partial\boldsymbol{\theta}}\right)\left(\frac{\partial d}{\partial\boldsymbol{\theta}}\right)^T. \tag{C.5}$$

In (C.4), we can rewrite the expression

$$\mathbf{M}_{i,x\theta}\mathbf{M}_{i,x\theta}^T = 4\left(\frac{\partial d_i}{\partial\boldsymbol{\theta}}\right)\left(\frac{\partial d_i}{\partial\mathbf{x}_i}\right)^T\left(\frac{\partial d_i}{\partial\mathbf{x}_i}\right)\left(\frac{\partial d_i}{\partial\boldsymbol{\theta}}\right)^T. \tag{C.6}$$

Call in mind that $d_i = \dfrac{f_i}{\left\|\nabla f_i\right\|}$; where $f_i(\cdot) = 0\forall i$ is the implicit form of surface equation, we have

$$\frac{\partial d_i}{\partial\mathbf{x}_i} = \frac{1}{\left\|\nabla f_i\right\|}\left(\frac{\partial f_i}{\partial\mathbf{x}_i}\right) = \frac{\nabla f_i}{\left\|\nabla f_i\right\|};$$

therefore, $\left(\dfrac{\partial d_i}{\partial\mathbf{x}_i}\right)^T\left(\dfrac{\partial d_i}{\partial\mathbf{x}_i}\right) = 1$.

Then (C.6) is simplified as

$$\mathbf{M}_{i,x\theta}\mathbf{M}_{i,x\theta}^T = 4\left(\frac{\partial d_i}{\partial\boldsymbol{\theta}}\right)\left(\frac{\partial d_i}{\partial\boldsymbol{\theta}}\right)^T \tag{C.7}$$

Combing (C.4), (C.5) and (C.7), we obtain

$$\mathbf{\Psi} = \sigma^2 \left[\sum_{i=1}^{n} \left(\frac{\partial d_i}{\partial \boldsymbol{\theta}} \right) \left(\frac{\partial d_i}{\partial \boldsymbol{\theta}} \right)^T \right]^{-1} \tag{C.8}$$

Since

$$d_i = \frac{1}{\|\nabla f_i\|} \boldsymbol{\theta}^T \mathbf{m}_i ;$$

where $\mathbf{m}_i = \left[x_i^2, y_i^2, z_i^2, x_i y_i, x_i z_i, y_i z_i, x_i, y_i, z_i \right]^T$, (C.8) is finally expressed as

$$\mathbf{\Psi} = \sigma^2 \left(\sum_{i=1}^{n} w_i \mathbf{m}_i \mathbf{m}_i^T \right)^{-1}, \tag{C.9}$$

where $w_i = \|\nabla f_i\|^{-2}$, which is the same of (4.14).

C.1.2 Covariance matrix of the feature vector

The feature vector of quadric representation consists of the invariants defined by (4.9). Perturbations of the invariants are expressed by (4.11). For further illustration, we rewrite (4.11) below.

$$\Delta\gamma_1 = \frac{\Delta\lambda_2}{\lambda_1} - \frac{\lambda_2 \Delta\lambda_1}{\lambda_1^2} \;;\; \Delta\gamma_2 = \frac{\Delta\lambda_3}{\lambda_1} - \frac{\lambda_3 \Delta\lambda_1}{\lambda_1^2} ;\; \Delta c = \frac{\Delta k'}{\lambda_1} - \frac{k' \Delta\lambda_1}{\lambda_1^2} . \tag{C.10}$$

In (C.10), λ_i $i \in \{1,2,3\}$ are the eigenvalues of $\mathbf{A}$ in the quadric form of (4.3), and k' is expressed by (4.10).

As discussed in section 4.2 of chapter 4, the perturbation of the eigenvalue in (C.10) can be approximately expressed by

$$\Delta\lambda_i = \Delta a'_{ii} \quad \forall i \in \{1,2,3\}, \tag{C.11}$$

where $\Delta a'_{ii}$ are the errors of parameters estimated in the rotation-normalised co-ordinate system. In the transformed rotation-normalised co-ordinate system, the errors of estimates of parameter $\boldsymbol{\theta}'$ can also be computed using the covariance matrix (C.9), just transforming the measurement of $\mathbf{x}_i$ $\forall i$ into the representation of $\mathbf{x}'_i$ $\forall i$. Therefore, the covariance matrix of the feature vector can be derived from the relation of (C.10) and the covariance matrix of $\boldsymbol{\theta}'$.

To do that, we express the dependency of the errors of invariants on the errors of elements of $\boldsymbol{\theta}'$ in an explicit form. From (C.10), we have

$$\Delta\gamma_1 = \frac{1}{\lambda_1} \Delta a'_{22} - \frac{\lambda_2}{\lambda_1^2} \Delta a'_{11} ; \tag{C.12.a}$$

$$\Delta\gamma_2 = \frac{1}{\lambda_1}\Delta a'_{33} - \frac{\lambda_3}{\lambda_1^2}\Delta a'_{11}; \tag{C.12.b}$$

$$\Delta c = \frac{1}{\lambda_1}\Delta k' - \frac{k'}{\lambda_1^2}\Delta a'_{11}; \tag{C.12.c}$$

From (4.7) and (4.10), we get

$$k' = -\tfrac{1}{4}\left(\frac{v_1'^2}{\lambda_1} + \frac{v_2'^2}{\lambda_2} + \frac{v_3'^2}{\lambda_3}\right) + k_0 . \tag{C.13}$$

Note that according to the scale constraints in definition of parameter $\boldsymbol{\theta}$, in fact $k_0 = 1$. Also applying the linear approximation to (C.13), we have

$$\Delta k' = -\tfrac{1}{4}\left(\frac{2v'_1}{\lambda_1}\Delta v'_1 + \frac{2v'_2}{\lambda_2}\Delta v'_2 + \frac{2v'_3}{\lambda_3}\Delta v'_3 - \frac{v_1'^2}{\lambda_1^2}\Delta a'_{11} - \frac{v_2'^2}{\lambda_2^2}\Delta a'_{22} - \frac{v_3'^2}{\lambda_3^2}\Delta a'_{33}\right). \tag{C.14}$$

From

$$\Delta\mathbf{g} = [\Delta\gamma_1, \Delta\gamma_2, \Delta c]^T;$$

$$\Delta\boldsymbol{\theta}' = [\Delta a'_{11}, \Delta a'_{22}, \Delta a'_{33}, \Delta a'_{12}, \Delta a'_{13}, \Delta a'_{23}, \Delta v'_1, \Delta v'_2, \Delta v'_3]^T,$$

From the dependencies of (C.12.a) ~ (C.12.c) and combining with (C.13) and (C.14), we get

$$\Delta\mathbf{g} = \mathbf{M}\Delta\boldsymbol{\theta}', \tag{C.15}$$

with

$$\mathbf{M} = \begin{bmatrix} \dfrac{-\lambda_2}{\lambda_1^2} & \dfrac{1}{\lambda_1} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \dfrac{-\lambda_3}{\lambda_1^2} & 0 & \dfrac{1}{\lambda_1} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \dfrac{v_1'^2}{4\lambda_1^3} & -\dfrac{k'}{\lambda_1^2} & \dfrac{v_2'^2}{4\lambda_1\lambda_2^2} & \dfrac{v_3'^2}{4\lambda_1\lambda_3^2} & 0 & 0 & 0 & -\dfrac{v'_1}{2\lambda_1^2} & -\dfrac{v'_2}{2\lambda_1\lambda_2} & -\dfrac{v'_3}{2\lambda_1\lambda_3} \end{bmatrix}$$

Therefore, the covariance matrix of $\mathbf{g}$ can be formulated as

$$\mathbf{G} = \mathbf{M}\boldsymbol{\Psi}\mathbf{M}^T, \tag{C.16}$$

where $\boldsymbol{\Psi}$ is the covariance matrix of the parameters $\boldsymbol{\theta}'$.

Appendix C.2

The Relational Pose in Degenerated Cases

In section 4.6.2, we defined the relational pose between two surface primitives, denoted as *a* and *b*. Generally, the relational pose parameters are defined by three "relational" rotation angles and a "relational" translation vector, which are derived from (4.32). But for some degenerated cases, the expressions of the relational pose are derived specifically.

In case of one of the two surfaces being planar, the relational pose was defined specifically. Without loss of generality, supposing surface *b* to be planar, the relational pose is thus defined by (4.37), i.e., the pose parameters of plane *b* viewed in the standard frame of *a*. It is

$$\boldsymbol{\rho}_a = \left[\bar{n}_{ax}, \bar{n}_{ay}, \bar{k}_a\right]^T . \tag{C.17}$$

To compute the components of $\boldsymbol{\rho}_a$, we first need to express the plane equation under the co-ordinate system of *a*, which is in the form

$$\mathbf{n}_a^T \mathbf{x}_a - k_a = 0 . \tag{C.18}$$

Supposing in the world co-ordinate system, the plane of *b* is expressed as

$$\mathbf{n}^T \mathbf{x}_w - k = 0 , \tag{C.19}$$

According to (4.30), there is

$$\mathbf{n}_a = \mathbf{R}_a^T \mathbf{n} \quad \text{and} \quad k_a = k - \mathbf{n}_a^T \mathbf{t}'_a . \tag{C.20}$$

Expressing the matrix $\mathbf{R}_a$ in the form $\mathbf{R}_a = [\mathbf{v}_{ax}, \mathbf{v}_{ay}, \mathbf{v}_{az}]$, therefore,

$$n_{ax} = \mathbf{v}_{ax}^T \mathbf{n};\ n_{ay} = \mathbf{v}_{ay}^T \mathbf{n};\ n_{az} = \mathbf{v}_{az}^T \mathbf{n};\ . \tag{C.21}$$

therefore, $\boldsymbol{\rho}_a$ in (C.17) can be expressed from (C.18) ~ (C.21) with

$$\bar{n}_{ax} = \frac{n_{ax}}{n_{az}}; \bar{n}_{ay} = \frac{n_{ay}}{n_{az}}; \bar{k}_a = \frac{k_a}{n_{az}} . \tag{C.22}$$

Other degenerated cases in our concerns have been discussed in section 4.6.2 of chapter 4.

Appendix C.3

Formulation of the covariance matrix of pose parameters

C.3.1 Formulation of the covariance matrix of pose parameters

As stated in section 4.6.3, the covariance matrix of the pose parameters is also derived from the formulation of surface fitting. In stead of the algebraic parameters $\boldsymbol{\theta}$, we utilise a new representation $\boldsymbol{\eta}= [\mathbf{g}^T, \boldsymbol{\tau}^T_w]^T$, where $\mathbf{g}$ is the feature vector and $\boldsymbol{\tau}_w$ is the pose vector, to formulate the minimisation problem. Then the fitting is stated as

$$\hat{\boldsymbol{\eta}} = \arg\min\left\{\sum_{i=1}^{n} d_{s,i}^2\right\},$$

where

$$d_{s,i}^2 = w_{s,i} f_s^2(\mathbf{x}_{s,i}(\boldsymbol{\tau}_w), \mathbf{g})$$

is defined in the standard co-ordinate system with the standard form

$$f_s(\mathbf{x}_s, \mathbf{g}) = x_s^2 + \gamma_1 y_s^2 + \gamma_2 z_s^2 + c = 0\,. \tag{C.23}$$

Here

$$\mathbf{x}_s = \mathbf{R}^T(\boldsymbol{\eta})\mathbf{x} - \mathbf{t}'(\boldsymbol{\eta})\,. \tag{C.24}$$

Now, applying the same principle used for derivation of (C.2), we get

$$\Delta\boldsymbol{\eta} = -\mathbf{M}_{\eta\eta}^{-1} \sum_{i=1}^{n} \mathbf{M}_{i,x\eta} \Delta\mathbf{x}_i\,, \tag{C.25}$$

where

$$\mathbf{M}_{\eta\eta} = \sum_{i=1}^{n} \mathbf{M}_{i,\eta\eta}\,,\ \mathbf{M}_{i,x\eta} = \frac{\partial^2}{\partial\boldsymbol{\eta}\partial\mathbf{x}} d^2(\boldsymbol{\eta}_0, \mathbf{x}_{i0}),\ \mathbf{M}_{i,\eta\eta} = \frac{\partial^2}{\partial\boldsymbol{\eta}^2} d^2(\boldsymbol{\eta}_0, \mathbf{x}_{i0})$$

In the computations of the above matrices, the transformation between $\mathbf{x}$ in the world system and $\mathbf{x}_s$ in the standard system is used. Then the covariance matrix of $\boldsymbol{\eta}$ can be expressed by

$$\Phi = E\left[\Delta\boldsymbol{\eta}\Delta\boldsymbol{\eta}^T\right] = \sigma^2\mathbf{M}_{\eta\eta}^{-1}\left(\sum_{i=1}^{n}\mathbf{M}_{i,x\eta}\mathbf{M}_{i,x\eta}^T\right)\mathbf{M}_{\eta\eta}^{-T} \tag{C.26}$$

Then the covariance matrix of $\boldsymbol{\tau}_w$, denoted $\mathbf{T}_w$, can be directly derived with the low-right sub-matrix of Φ.

C.3.2 Formulation of the covariance matrix of the relational pose parameters

According to (4.32), the three rotation angles associated to $\mathbf{R}_{ab}$, denoted α, β and γ, are derived from the rotation angles of $\mathbf{R}_a$ and $\mathbf{R}_b$ associated to the standard frames of a and b, which are independently obtained in the world co-ordinate system. In form of vector $\boldsymbol{\mu}$ to represent the three rotation angles, there is

$$\Delta\mathbf{R}_{ab}(\boldsymbol{\mu}) = \Delta\mathbf{R}_a^T(\boldsymbol{\mu}_a)\mathbf{R}_b(\boldsymbol{\mu}_b) + \mathbf{R}_a^T(\boldsymbol{\mu}_a)\Delta\mathbf{R}_b(\boldsymbol{\mu}_b)\,. \tag{C.27}$$

When substituting the expression of $\mathbf{R}$ with respect to α, β and γ into (C.27) and using linear approximation, we get

$$\mathbf{D}_{ab}\Delta\boldsymbol{\mu} = \mathbf{D}_a\Delta\boldsymbol{\mu}_a + \mathbf{D}_b\Delta\boldsymbol{\mu}_b \tag{C.28}$$

where $\mathbf{D}_{ab}$, $\mathbf{D}_a$ and $\mathbf{D}_b$ are 3×3 matrices derived from the differential operations to those components of $\mathbf{R}_{ab}$, $\mathbf{R}_a$ and $\mathbf{R}_b$ respectively. Then the errors of relational rotation parameters can be expressed as

$$\Delta\boldsymbol{\mu} = \mathbf{D}_{ab}^{-1}\mathbf{D}_a\Delta\boldsymbol{\mu}_a + \mathbf{D}_{ab}^{-1}\mathbf{D}_b\Delta\boldsymbol{\mu}_b \tag{C.29}$$

The errors of relational translation parameters can also be derived from (4.32) in the form

$$\Delta\mathbf{t}' = \Delta\mathbf{t}'_b - \Delta\mathbf{R}(\boldsymbol{\mu})\mathbf{t}'_a - \mathbf{R}_{ab}^T\Delta\mathbf{t}'_a \tag{C.30}$$

The expression of (C.30) can be further formulated as

$$\Delta\mathbf{t}' = \Delta\mathbf{t}'_b - \mathbf{D}_t\Delta\boldsymbol{\mu} - \mathbf{R}_{ab}^T\Delta\mathbf{t}'_a\,, \tag{C.31}$$

where $\mathbf{D}_t$ is also a 3×3 matrix. Combined with (C.29) and (C.31), finally the errors of the relational pose can be expressed in the form

$$\Delta\boldsymbol{\tau} = \mathbf{M}_a\Delta\boldsymbol{\tau}_{wa} + \mathbf{M}_b\Delta\boldsymbol{\tau}_{wb} \tag{C.32}$$

where $\mathbf{M}_a$ and $\mathbf{M}_b$ are 6×6 coefficient matrices that can be computed out with the estimated pose parameters. Therefore, the covariance matrix of $\hat{\boldsymbol{\tau}}$ can be derived as

$$\mathbf{T} = E\{\Delta\boldsymbol{\tau}\Delta\boldsymbol{\tau}^T\} = \mathbf{M}_a\mathbf{T}_{wa}\mathbf{M}_a^T + \mathbf{M}_b\mathbf{T}_{wb}\mathbf{M}_b^T, \tag{C.33}$$

where $\mathbf{T}_{wa}$ and $\mathbf{T}_{wb}$ are the covariance matrices of pose parameters for surfaces *a* and *b*, respectively. Note that in formulation of (C.33), the pose estimates of *a* and *b* are thought to be non-correlated.

At last, we discuss the cases that planar surfaces involve in the relational poses. The errors of estimate $\hat{\boldsymbol{\rho}}_a$ in (C.17) can be expressed from (C.19) and (C.20) as

$$\Delta n_{ai} = \Delta\mathbf{v}_{ai}^T\mathbf{n} + \mathbf{v}_{ai}^T\Delta\mathbf{n}; i \in \{x, y, z\};\ \Delta k_a = \Delta k - \Delta\mathbf{n}_a^T\mathbf{t}'_a - \mathbf{n}_a^T\Delta\mathbf{t}'_a \tag{C.34}$$

Without trivialities in mathematics, the covariance matrix of $\hat{\boldsymbol{\rho}}_a$ can be expressed in the form

$$\mathbf{T}_\rho = \mathbf{m}_1\mathbf{T}_{wa}\mathbf{m}_1^T + \mathbf{m}_2\mathbf{Y}\mathbf{m}_2^T \tag{C.35}$$

where $\mathbf{T}_{wa}$ is again the covariance matrix of pose estimate of *a* in the world co-ordinate system; $\mathbf{m}_1$ is a 3×6 coefficient matrix. The notation $\mathbf{Y}$ is the covariance matrix of the estimate of planar parameter $\hat{\boldsymbol{\rho}}_w$ in the world co-ordinate system, which can be obtained in planar fitting; $\mathbf{m}_2$ is a 3×3 coefficient matrix.

The same way in analysis of (C.34), in case of surface a being cylindrical, covariance matrix for $\hat{\boldsymbol{\rho}}_{ac}$ is expressed as

$$\mathbf{T}_{\rho c} = \mathbf{m}_{1c}\mathbf{T}_{wa}\mathbf{m}_{1c}^T + \mathbf{m}_{2c}\mathbf{Y}\mathbf{m}_{2c}^T \tag{C.36}$$

As for two planar surfaces, the pose parameter is defined as a one-dimensional variable and its variance can be obtained directly from the definition of (4.39) and the computation of (C.34).

Appendix D

Bias-corrected Parameter Estimation of a Planar Representation

A planar surface can be represented by the implicit form:

$$\mathbf{n}^T\mathbf{x} - b = 0 \tag{D.1}$$

where $\mathbf{n}$ is the unit normal vector of the plane. In general, the parameters $\mathbf{n}$ and b are estimated by fitting data points to the expression of (D.1). To do that, we first derive the vector $\mathbf{n}$ with the method of principal component extraction, as used in the variance estimation in outlier detection of chapter 5.

Given a set of points $\{\mathbf{x}_i\}$ $\{i = 1,..,m\}$, the normal vector $\mathbf{n}$ is estimated by the eigenvector of the matrix

$$\mathbf{Z} = \sum_{i=1}^{m}(\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T \text{ , where } \bar{\mathbf{x}} = \tfrac{1}{m}\sum_{i=1}^{m}\mathbf{x}_i \text{ ,} \tag{D.2}$$

associated to the smallest eigenvalue. Then the parameter b is estimated by

$$\hat{b} = \hat{\mathbf{n}}^T\bar{\mathbf{x}} \text{ .} \tag{D.3}$$

Analogous to the analysis of [68] in chapter 5, we can obtain the "bias-corrected" estimate of the normal vector $\hat{\mathbf{n}}$. Let

$$\mathbf{Z}_i = (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T = \begin{bmatrix} (x_i - \bar{x})^2 & (x_i - \bar{x})(y_i - \bar{y}) & (x_i - \bar{x}_i)(z_i - \bar{z}) \\ (x_i - \bar{x})(y_i - \bar{y}) & (y_i - \bar{y})^2 & (y_i - \bar{y})(z_i - \bar{z}) \\ (x_i - \bar{x})(z_i - \bar{z}) & (y_i - \bar{y})(z_i - \bar{z}) & (z_i - \bar{z})^2 \end{bmatrix}$$

Suppose point $\mathbf{x}_i$ is perturbed with noise $\Delta\mathbf{x}_i$, then $\mathbf{Z}_i$ is perturbed with $\Delta\mathbf{Z}_i$. When applying the Taylor expansion to the components of $\mathbf{Z}_i$ to derive $\Delta\mathbf{Z}_i$, only the diagonal elements are associated to the second order term of $\Delta\mathbf{x}_i$. For the first diagonal component, there is

$$\Delta(x_i - \bar{x})^2 = 2(x_i - \bar{x})(\Delta x_i - \Delta\bar{x}) + (\Delta x_i - \Delta\bar{x})^2 \text{ .} \tag{D.4}$$

Under the assumption of i.i.d Gaussian noise, there is

$$E\{\Delta x_i\} = 0;\ E\{\Delta x_i^2\} = \sigma^2; \forall i \text{ and } E\{\Delta x_i \Delta x_j\} = 0 \forall i \neq j$$

where σ is the standard deviation of noise. From simple computations, we obtain

$$E\{\Delta(x_i - \bar{x})^2\} = \frac{m-1}{m}\sigma^2 \tag{D.5}$$

So the expectations of $\Delta\mathbf{Z}$ is

$$E\{\Delta\mathbf{Z}\} = \sum_{i=1}^{m} E\{\Delta\mathbf{Z}_i\} = (m-1)\sigma^2\mathbf{I}\,, \tag{D.6}$$

where $\mathbf{I}$ is the unit matrix. Defining

$$\overline{\mathbf{Z}} = \mathbf{Z} - E\{\Delta\mathbf{Z}\}\,, \tag{D.7}$$

then the eigenvector of $\overline{\mathbf{Z}}$ associated to the smallest eigenvalue is the "bias-corrected" e s-timate of $\mathbf{n}$. The parameter b is estimated from (D.3). The variance of noise σ^2 is estimated from the smallest eigenvalue of the matrix $\mathbf{Z}$.