Universitat de Girona

3D HAND-HELD SENSOR FOR LARGE SURFACE REGISTRATION

Carles MATABOSCH GERONÈS

ISBN: 978-84-691-2386-7
Dipòsit legal: GI-207-2008

Department of Electronics, Computer Science and Automatic Control

PhD Thesis

Hand-held 3D-scanner for large surface registration

Thesis presented by Carles Matabosch Geronès,
to obtain the degree of:
PhD in Computer Engineering.

Supervisors:
Dr. Joaquim Salvi, Universitat de Girona.
Dr. David Fofi, Université de Bourgogne

Girona, April 2007

Agraïments

Sempre és difícil agrair en poques paraules a tota la gent que ha col·laborat d'una manera o altra en l'elaboració d'aquesta tesi. M'agradaria especialment començar per la principal persona que ha fet que això sigui possible. Tot començà durant el cinqué durs d'EI, on el Dr. Joaquim Salvi era el professor de *Robòtica Industrial*. Des d'un bon principi ja em vaig deixar seduir pel món de la robòtica i un cop acabat el meu projecte final de carrera vaig anar a parlar amb ell per endinsar-me en aquest món. Els elements més importants de la visió per ordinador, les matemàtiques i la geometria sempre m'havien atret, bé, això fins que en Quim em va "embolicar" amb les equacions de *Kruppa*. També em va donar la màxima confiança fent-me secretari del congrés MCMC celebrat a Girona. Però, pel que realment estic més agraït és que sempre ha estat disposat a ajudar-me, corregint publicacions, o, a vegades, reescrivint-les de nou. També agrair-li la possibilitat que m'ha brindat, no únicament per formar-me com a doctor, sinó com a persona.

Des d'un bon principi, en Quim va insistir que el millor era un doctorat europeu, espero que no fos per lliurar-se de mi durant uns mesos. Però encara que fos així, l'hi he d'agrair que em deixes en molt bones mans, àvec le Sieur Dr. David Fofi. L'experiencia a França no va començar gaire bé, a l'arribar el meu francés es limitava a Paris i Roland Garros, i a sobre, em vaig trobar tirat a la gare TGV esperant un no-se-qui que no va arribar mai. Però, exceptuant aquest desafortunat inici, la resta va ser immillorable. I tot això, va ser possible gràcies a en David, que sempre va estar disposat a donar-me un cop de mà (o tots dos), especialment per anar a fer unes *bieres*. No gens menys important fou la resta de membres de Le2i, Fabrice, Thierry, Alexandra,.. i especialment a en Pierre i la Noe. En Pierre sempre va estar disposat a obrir-me la seva casa (o la dels seus avis) per fer més amena l'estada a Le Creusot (o le Corbeau, com era anomenada aquesta població francesa pels membres de la residència, Samy and cia). I que dir de la Noe, que deixant a part les diferències entre un català i una madrilenya, es convertí en una gran amiga, amb el que vaig compartir pràcticament totes les experiencies viscudes en aquesta petita població francesa.

També m'agradaria agrair a tots els membres del grup VICOROB. El grup és bastant gran, i, malauradament, no puc anomenar a tothom, ja que serien més llargs els agraïments que la tesi en si. Pero gracies a tots, des dels membres fundadors del grup, Batlle and cia,

als primers doctorants per marcar el camí a seguir, els *submarinus*, i no només per les festes que organitzen, al laboratori de visió, els membres del PAS, secretàries, i darrerament a la nova fornada de doctorants, la majoria dels quals vinguts des de fora, els que m'han ensenyat, entre d'altres coses, que el meu anglès necessita millorar. M'agradaria especialment destacar els altres membres de percepció 3D, Xavier Armangué, Josep Forest, Jordi Pagés (Per cert, algún dia haurem de continuar amb el "buscamines", has estat realment un deixeble excel·lent!), Radu Orghidan i darrerament, la Bet Batlle. Tot a canviat molt des del principi on era jo qui esta sempre "molestant" a l'eminència Mangui, on ara, faig el que puc per ajudar a la Bet, passant per una etapa intermitja de col·laboració mútua entre tots. I que dir de l'Arnau, amb les seves porres i les seves brometes, però, el millor, la seva darrera ensaimada!

Un capítol a part per en Forest, per donar-me l'oportunitat d'aplicar el coneixement adquirit a l'empresa AQSense. També destacar als diferents membres de la mateixa, Jordi, Ramon, Lluís, Xevi, Carles i Enric. Aquesta oportunitat m'ha ensenyat que el món real és molt diferent a la recerca. Gràcies Jordi per les impagables lliçons de programació i a totes per les converses sobre *aprimaments*, *llançaments*, i *obusos*.

També m'agradaria recordar els meus amics de la Bisbal, que tot i no tenir massa idea de visió, les festes dels dissabtes a la nit m'han permés desfogar-me quan em calia, i perquè no, afogar les penes en alcohol.

Finalment agrair a la meva familia tot el suport rebut, no únicament aquests darrers 4 anys, sinó al llarg de tota la meva vida. Especialment destacar els meus pares i el meu germà Xavier que són els que han fet tot el possible per formar-me com a persona.

Acknowledgments

I would like to dedicate this thesis to all those people who have helped me during this period of time. Thanks to them, these years have been full of good and interesting moments, and the bad moments were easy to overcome in the company of good friends.

My first words are dedicated to Joaquim Salvi, who is "guilty" of this work. Dr. Salvi opened the doors of research to me when I had just finished my degree.

I would like to thank the University of Girona, not only for the pre-doctoral fellowship but also for allowing me to improve my knowledge during these years.

Special thanks also go to David Fofi, who helped me during both of my research stays in France, in 2004 and later in 2005. This help was significant, especially at the beginning, when my French was limited to "Paris" and "Roland Garros". I would also like to thank his institution, the Le2i of University of Burgundy, for providing me with the resources required to develop part of my work.

I am most grateful to all the members of the VICOROB group in Girona, which was always prepared to help me and, also, to organize good parties. Thanks also to Josep Forest, who let me join AQSense where I am still working in 3D.

Additional thanks to all the people I met in France, not only people from the Le2i, but also students in the residences, staff of the university and other people of Le Creusot. Special thanks to Noe, Pierre, Thierry, Monique, Alexandra, Samy, Olaf and Mathias.

This thesis is dedicated to my family for always being beside me.

Hand-held 3D-scanner for large surface registration

Resum

La visió per computador és un camp complex i ampli que preten simular el comportament humà per tal de percebre i interpretar una escena. Tradicionalment, aquesta percepció ens ha permés tenir una representació bidimensional del món. Per contra, la complexa visió humana és capaç d'obtenir informació de la profunditat de l'escena. És per aquest motiu, que durant els darrers anys s'han presentat diferents càmeres 3D, és a dir, sistemes de visió capaços d'obtenir informació de la profunditat de l'escena. No obstant, aquestes càmeres només són capaces d'obtenir una representació parcial de l'escena. Per contra, la ment humana és capaç d'extrapolar i d'imaginar-se la totalitat de l'objecte. Basicament la ment aprofita informació de l'objecte (simetries) o observacions anteriors del mateix objecte o objectes similars per tal de sustituir aquelles parts de l'objecte que estan fora de l'avast de l'ull humà.

Un dels primers pasos per tal d'imitar el comportament humà consisteix en l'obtenció d'un model únic de l'escena a partir d'un conjunt d'adquisicions parcials de la mateixa. L'idea principal és la d'alinear les diferents adquisicions i obtenir una representació única de l'objecte. Per tal d'alinear dues o més superficies, cal minimitzar la distància entre elles.

L'objectiu d'aquesta tesi és l'estudi de les diferents tècniques per alinear vistes 3D. Aquest estudi ens ha permés detectar els principals problemes de les tècniques existents, aportant una solució novedosa i contribuint resolent algunes de les mancances detectades especialment en l'alineament de vistes a temps real. Per tal d'adquirir les esmentades vistes, s'ha dissenyat un sensor 3D manual que ens permet fer adquisicions tridimensionals amb total llibertat de moviments. Així mateix, s'han estudiat les tècniques de minimització global per tal de reduir els efectes de la propagació de l'error.

Hand-held 3D-scanner for large surface registration

Abstract

Computer vision is a huge and complex field that aims to simulate the human vision in order to perceive and interpret an object or scene. This perception task traditionally provides a bidimensional representation of the world. Human vision, on the other hand, is more complex and is able to retrieve information from the depths of a scene.

During the last years, 3D cameras that can gather information about the depth of a scene have been commercialized. However, they can only obtain a partial representation of the scene while the human mind is able to imagine and extrapolate all the information from the object. Basically, the human mind gets information from the object (symmetries) or uses past observations of the same or similar objects to replace non-observed parts of the scene.

One of the first steps in imitating the human behavior consists of obtaining a single representation of the scene from a set of partial acquisitions. In order to fuse this set of views, all of them have to be aligned.

Therefore, the main idea is to align different acquisitions and obtain a single representation of the object.

The goal of this thesis is to study the different techniques used to register 3D acquisitions. This study detects the main drawbacks of the existing techniques, presents a new classification and provides significant solutions for some perceived shortcomings, especially in 3D real time registration. A 3D hand-held sensor has been designed to acquire these views without any motion restriction and global minimization techniques have been studied to decrease the error propagation effects.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The acquisition of scenes and objects constitutes the framework of this thesis. This chapter presents a general overview of the topics included in the thesis as well as its motivation and objectives. In addition, an outline of the thesis is presented at the end of the chapter.

1.1 Computer Vision

Computer vision is concerned with the theory and technology for building artificial systems that obtain information from images or multi-dimensional data. Roughly, computer vision tries to emulate the visual perception of the human beings from the first stage of light detection to the complex task of understanding what is being perceived. In this case, light is detected by cameras, while the high level tasks of image understanding are processed by computer algorithms. The goal of achieving a comprehensive perception of the world through computer vision is still far from being attained. The main reason is that the way the human brain works is still fairly unknown, and computer vision community is still dealing with low-level problems related to vision, like color and texture perception, and intermediate-level problems, like motion detection, depth and shape acquisition and object recognition. Another reason why it is hard to emulate the human perception is that humans take advantage of active perception capabilities to optimize perceptive tasks. For example, we can converge or diverge our eyes, move our head, change our point of view, etc. The use of these capabilities to optimize perception tasks is another important

research field in computer vision known as active vision [Aloimonos *et al.*, 1987].

The history of computer vision began in the 1960s when some artificial intelligence experts proposed a student's summer project. The goal of this project was to enable a computer to see like humans. Although this task is still incomplete, computer vision has emerged as a discipline in itself with strong connections to mathematics and computer science and marginal connections to physics, the psychology of perception and the neurosciences.

Although the way biological vision works is still unknown, practical and theoretical successes have been obtained from the beginning. To single out one example, the possibility of guiding vehicles on regular roads or on rough terrain using computer vision technology was demonstrated many years ago. This complex task involves a set of computer vision topics: a) image acquisition; b) segmentation and interpretation to detect the road; and c) camera modeling to obtain metric distance from pixels combined with stereovision to get depth information.

The acquisition of an image consists of a bidimensional digitalization of the observed scene. A camera produces a grid of samples of the light received from the scene from different directions. The position within the grid where a scene point is imaged is determined by the perspective transformation. The amount of light recorded by the sensor from a certain scene point depends on the type of lighting, the reflection characteristics and orientation of the surface being imaged, and the location and spectral sensitivity of the sensor.

The segmentation of the image consists in partitioning it into several constituent components. Interpretation is defined as the extraction of qualitative and quantitative information about the shape, location, structure, function, quality, condition and the relationship of and between objects, etc.

The last step is part of geometric computer vision, which describes the change in appearance of the objets as a function of their shape and the camera's parameters. In order to model this behavior, *intrinsic* and *extrinsic* parameters are studied. Intrinsic parameters model the physical elements of the camera to transform metric units to pixels, and vice versa. Extrinsic parameters determine the position and orientation of the camera in the world.

A multitude of applications can be obtained from the acquired images. There are lots of applications in very different fields, such as medical imaging, pose estimation, map

building, 3D modeling, etc. Some of these applications require computation of the third dimension, known as 3D imaging.

1.2 3D Imaging

The problem of inferring 3D information from an object/scene has a long history. Basically, existing techniques can be divided into two groups, *contact* and *non-contact* techniques.

Nowadays, the use of contact techniques has been reduced due to the problems involved in touching the object/scene. Objects can be deformed during the operation, producing errors on the estimation of the shape. This group of techniques can be divided into *destructive* techniques, like *slicing*, or *non-destructive* techniques, like the articulated arms of *CMM*[1].

Non-contact techniques have some advantages. Shape can be obtained in the presence of delicate objects, hot environments, deformable objects, big scenes, etc. This group of techniques can be divided into two main groups, *reflective* and *transmissive* techniques.

Transmissive techniques are based on computer tomography. This method, based in X-rays, is widely used in medical imaging. Other transmissive scanners are based on time-of-flight lasers, which compute the distance to a surface by timing the round-trip time of a pulse of light. A laser is used to emit a pulse of light and the amount of time that passes before the reflected light is seen by a detector is timed. Since the speed of light is a known, the round-trip time determines the travel distance of the light, which is twice the distance between the scanner and the surface. This kind of range finder is used in large object acquisition and sometimes in airborne applications.

Reflective techniques can be classified into optical and non-optical. Non-optical techniques such as sonars or microwave radars are mainly used in robotics. In addition, optical techniques can be divided into two types, active sensors and passive sensors.

Active sensors emit some kind of radiation and detect its reflection to probe the presence of an object or environment. The possible types of radiation used include DLP[2], laser triangulation and interferometry.

Passive scanners do not emit any kind of radiation themselves, but instead rely on

[1] Coordinate-measuring machines
[2] Digital Light Projectors

detecting the radiation reflected by objects. Most scanners of this type detect visible light because it is readily available ambient radiation. Other types of radiation, such as infrared, could also be used. Passive methods can be very cheap because in most cases they do not need special hardware. Most common examples are shape-from-X, where X represents the method used to determine the shape, that is motion, stereo, shading, silhouette and texture, among others. All these techniques are based on the use of several images of the object/scene, and are known as multiple-view geometry.

The first research directly related to multiple-view geometry resulted in a paper by the Austrian mathematician Kruppa [Kruppa, 1913]. This paper demonstrated that two views of 5 points are enough to acquire the relationship between views and the 3D location of the points up to a finite number of solutions. Although the first techniques using the term computer vision in 3D imaging appeared in 1970s, the origin of the modern treatment is attributed to the introduction of epipolar geometry by Longuet-Higgins [Longuet-Higgins, 1981].

Epipolar geometry has been widely studied in the last decades, producing a wealth of knowledge in this field, and has been extensively used in camera calibration, 3D acquisition and correspondence problem simplification.

Despite multi-view geometry being widely used in computer vision, it presents some drawbacks when it is used in 3D imaging. The first problem is the correspondence problem. In other words, determining relationships from pixels of different views is not a trivial step. The second important problem is the resolution of the acquisition. These techniques usually work with a small number of points, so that dense reconstructions are difficult to obtain.

In order to overcome both drawbacks, active sensors are commonly used when dense reconstructions are required. Based primarily on laser or coded structured light, several commercial sensors are available nowadays. This kind of sensor is basically used to get 3D models of objects or scenes. However, modeling is not only a three-dimensional acquisition of the object/scene but a complex problem composed of several steps, which are briefly illustrated in Fig. 1.1. First, some techniques can be applied to determine the best position of the camera with the aim of reducing the number of views of a given object/scene. Sometimes, it is also used to acquire images from incomplete objects or scenes, especially in map building [Wong *et al.*, 1999; Stoev and Strasser, 2002]. Second, 3D acquisition involves obtaining the object/scene structure or depth, and the methods for that are

briefly commented in Chapter 2. Depth information can basically be delivered in two different representations known as range maps and clouded points.

A range map is a bidimensional representation of the object/scene, where the intensity of each pixel of the image is directly related to the depth. Clouded points are a set of unorganized 3D points. Depending on the application, one partial acquisition is not enough to represent the object/scene. In such a situation, several acquisitions of the object/scene must be made. In the general case, the pose of the sensor is unknown. Hence, the motion from several viewpoints must be estimated to align all the acquisitions in a single coordinate frame. The techniques that estimate these Euclidean transformations are known as registration techniques (see Chapter 2).

When all acquisitions are aligned, a set of unorganized 3D points is obtained. However, this is not a continuous representation of the surface. Moreover, redundant data is commonly present due to the overlapping between partial views. The integration step attains a continuous surface by triangulating the 3D points, generating a parameterized surface and creating occupancy grids, meshing partial triangulations, etc. [Curless and Levoy, 1996; Soucy and Laurendeau, 1995; Turk and Levoy, 1996; Hilton and Illingworth, 2000; Peng *et al.*, 2002].

Finally, when the surface is obtained, texture can be stuck to the structure to get a more realistic representation of the object/scene.

1.3 Context and motivations

This thesis has been developed within the framework of several Spanish government research projects. The thesis has been funded by a scholarship from the University of Girona.

Most of the thesis has been developed within the VICOROB[3] group of the University of Girona composed of 36 members including researchers and PhD students. The research areas of the group are underwater robotics and vision, mobile robotics, 3D perception and image analysis. The research activities are currently supported by several national projects, like the development of autonomous underwater vehicles, monitoring the deep sea floor on the mid-Atlantic ridge or mammographic image analysis based on content,

[3]Computer Vision and Robotics Group. http://vicorob.udg.es

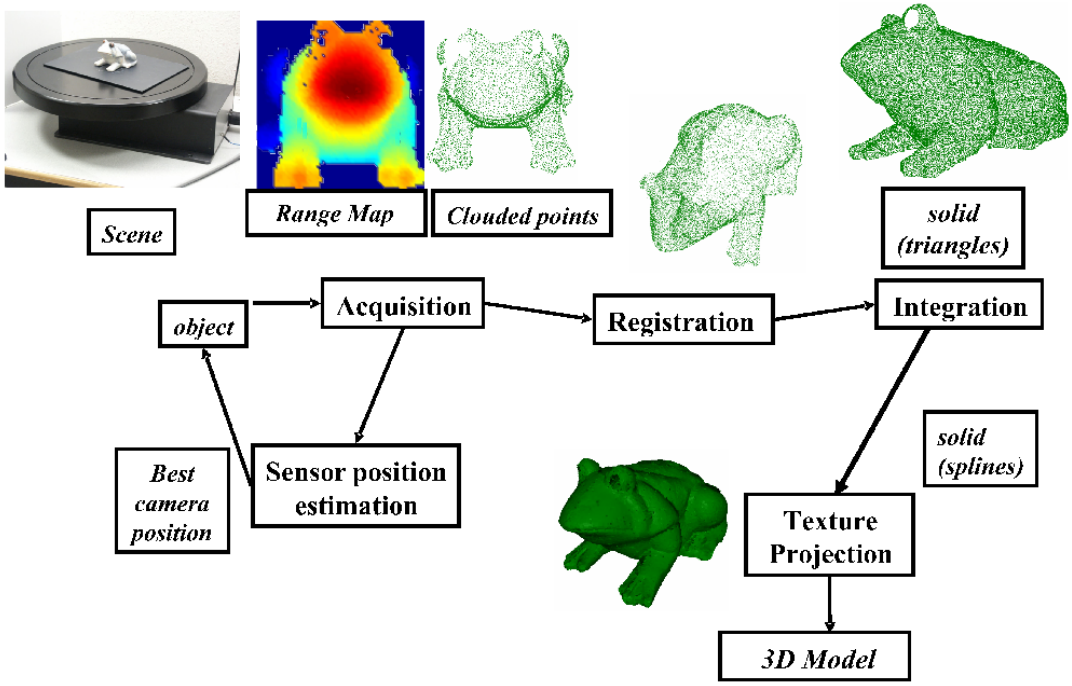


Figure 1.1: Steps in 3D modeling

and computer vision applied to visual surveillance.

Some members of the group are working in 3D imaging. The main topic is acquisition, and several acquisition prototypes have been made based on stereovision, coded structured light, omnidirectional systems and laser triangulation. Specifically, Dr. Forest presented a thesis titled *New methods for triangulation-based shape acquisition using laser scanners*. This thesis presented the common techniques of shape acquisition based on laser triangulation. The author also developed a 3D scanner based on a scanning laser. The setup of this scanner is presented in Fig. 1.2, while an acquisition sample is shown in Fig. 1.3.

Taking advantage of their knowledge, Forest and Salvi, together with other partners, founded AQSense S.L., a spin-off specialized in 3D imaging. Nowadays, the company is focused on laser segmentation, acquisition, and more complex tasks like visual inspection, all of them in real time.

Registration is an interesting topic and is actually the natural evolution of 3D acquisition to explore techniques to achieve complete acquisitions of complex objects/scenes. In visual inspection tasks, the object/scene must be aligned to the model before comparing them with many industrial applications and interest by AQSense. Additionally, the

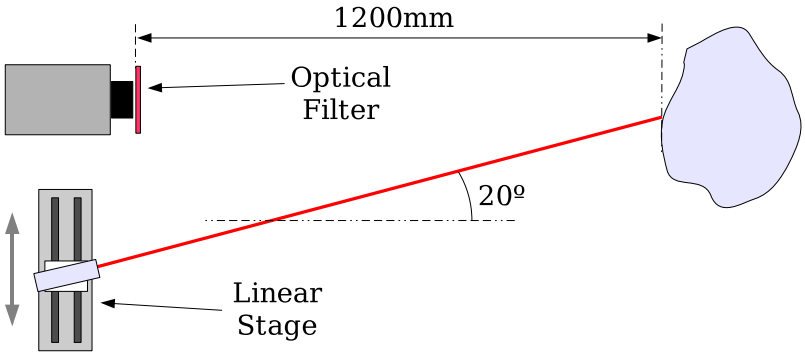


Figure 1.2: Set up of Forest's scanner

(a) (b)

Figure 1.3: Results of the Forests' 3D scanner: a) 3D acquisition; b) image of the acquired object

possibility of developing a hand-held scanner is also interesting as it would increase the number of applications of the company. Unlike Forest's prototype, the new one should be small, and it should be able to move freely to scan complex objects/scenes.

Apart from the company requirements, the VICOROB group is working on some projects related to 3D imaging, and part of the work developed in this thesis has contributed to the following Spanish projects:

- The MCYT[4] project TAP[5] 1999-0443-C05-01 from 31/12/99 up to 31/12/02. The aim of this project was the design, implementation and accuracy evaluation of mobile robots fitted with distributed control, sensing and a communicating network. A computer vision-based system was developed to provide the robots with the ability to explore an unknown environment and build a dynamic map. This project was part of a larger project coordinated by the Polytechnic University of Valencia (UPV) involving both the Polytechnic University of Catalonia (UPC) and the University of Girona (UdG).

- The MCYT project TIC[6] 2003-08106-C02-02 from 01/12/03 to 30/11/06. The aim of the overall project is the design and development of FPGA-based applications with fault tolerance applied to active vision-based surveillance tasks in large scenarios like airports and train stations. Some of the tasks involved are the automatic detection of dangerous situations or suspicious behaviors, and people tracking. The project is again developed in collaboration with the UPV.

Although the research areas of the projects comprise topics that are familiar to the group, collaboration with similar groups seemed to be an interesting way to share knowledge and experience. That was the original idea behing the collaboration with the Le2i[7] group at the University of Burgundy, Le Creusot, France, which is part of the CNRS[8]. One of the main lines of research of the group is 3D acquisition and analysis. Apart from their valuable knowledge, the superior equipment in their laboratory allowed us to set up the experiments we could not perform in our lab. Our first collaboration with them consisted of a 4 month stay in 2004 to study the fundamentals of registration under the

[4]Ministerio de Ciencia y Tecnología
[5]Tecnologías Avanzadas de Producción
[6]Tecnologías de la Información y de las Comunicaciones
[7]Laboratoire d'Electronique, Informatique et Image. http://vision.u-bourgogne.fr/le2i/index.htm
[8]Centre National de la Recherche Scientifique

supervision of Dr. Fofi. At the end of the stay, and due to the good results coming from it, both research groups decided to continue the collaboration. We also asked Dr. Fofi to join as a co-director of this thesis.

A second stay of four months was planned in 2005. The goal was to take advantage of their 3D acquisition knowledge and also to obtain accuracy evaluation of our registration algorithms by coupling our brand new hand-held scanner to mechanically controlled structures. As a result of this collaboration, some articles have been joinly published and the relationship continues to grow through other research stays.

1.4 Objectives

The goal of this thesis consists of the depth estimation of real objects and scenes, which is also related to the main goals of some of the projects described in the previous sections. Bearing in mind that 3D acquisition is one of the topics we work with in our group, the aim of this work is to complement this acquisition with a set of techniques designed to achieve complete acquisition from a set of partial acquisitions.

3D acquisition is a very important area whose goal is to acquire not only the images of the object/scene but also the third dimension of each point in the image. Because some occlusions can be present, and due to the limitations of the field of view of the camera, some parts of the object/scene cannot be reconstructed in a single acquisition.

The alignment of several acquisitions lets us obtain complete objects/scenes. However, this is not a trivial step. These partial acquisitions are represented in different coordinate frames and the Euclidean motion that relates them is generally unknown. Bringing about this transformation is the first goal of our thesis, a state-of-the art analysis of the main techniques is realized to judge the pros and cons of each method.

The second and more complex task is the use of previous algorithms to develop a system that lets us acquire information about large surfaces without using complex mechanical systems. Specifically, the idea is to develop a portable scanner to scan objects/scenes which are not easy to acquire with commercial scanners. Due to the large-scale object/scene dimensions, huge numbers of acquisitions are required. In this case, error propagation can be an important impeding to good alignments. To solve this problem, global minimization algorithms will be studied.

Two steps are required to perform these tasks. First, a global registration algorithm is proposed. Second, a prototype of a 3D hand-held scanner is designed to acquire consecutive clouded points of an object/scene, which are, a posteriori, aligned by means of the proposed registration algorithm.

Specifically, the objectives of this thesis are:

- Perform a state-of-the-art analysis on pair-wise registration techniques and a study of multi-view alignment
- Classify the surveyed techniques analyzing their pros and cons
- Propose a new method that solves some of the existing drawbacks especially in on-line registration
- Design a new hand-held scanner for large-surface measurement
- Perform experimental procedures to analyze accuracy in comparison with other, existing methods.

1.5 Thesis outline

3D acquisition is a very important topic in computer vision. The primary goal is to obtain depth information from an observed scene. This information is useful in many applications. Hence, Chapter 2 begins with a brief overview of the most common techniques to acquire the third dimension. Due to occlusions and shadows, but mainly the limited field of the scanner, only a part of the object/scene can be acquired. This is also common in human vision, as people are not able to determine the full shape of a viewed object if it is partly occluded. However, if the object/scene is observed from different positions, the brain is able to align all partial observations to get the full shape of the object/scene. In computer vision, the techniques that align partial acquisitions are known as *registration* techniques. The state-of-the-art of most relevant techniques is also presented in Chapter 2. This survey determines which techniques are the most suitable under different situations and presents quantitative evaluation of the registration errors of some implemented techniques.

Most techniques presented are based on the alignment of two different acquisitions. However, in real applications, more than two views are required to obtain a complete

acquisition. When more than two views are registered, there is another problem related with the propagation of registration errors through all the views. Some authors have proposed several algorithms to solve this problem. In Chapter 3, an overview of these techniques is presented, and then a new method is detailed. The proposed technique is based on the simultaneous registration of a set of views to decrease the propagation errors. The technique is compared with a similar approach to test accuracies in both synthetic and real data.

The proposed technique lets us robustly register a set of 3D acquisitions of the same object/scene. This technique can be used in combination with a hand-held scanner to reconstruct 3D surfaces without costly and complex mechanics. A first prototype of a hand-held scanner is presented in Chapter 4. The laser triangulation principle is used to get the third dimension and a single shot is enough to achieve an acquisition. Thus, the scanner can be held by hand and used in the presence of moving objects. Experiments to determine the acquisition accuracy of the hand-held scanner are explained in this chapter.

Some experiments are done to integrate all the steps presented in preceding chapters. A sample of objects is acquired by our hand-held scanner and then registered to align all the partial acquisitions. In order to determine the accuracy of the multi-view registration, the pose of the scanner in each acquisition must be known. Hence, the scanner is coupled to the end-effector of a robot arm, which leads to the eye-to-hand calibration problem, a brief summary of which is commented, detailing our calibration steps. Experimental results are presented in Chapter 5.

Finally, Chapter 6 presents the conclusions of the thesis, including a list of related published articles and conference contributions. Further work derived from the results and some additional perspectives are also discussed.

Chapter 2

State-of-the-art in Surface Alignment

During recent years several commercial sensors have been introduced to obtain three-dimensional acquisitions. However, most of them can only attain partial acquisitions and certain techniques are required to align several acquisitions of the same object to get a full reconstruction of it. Range image registration techniques are used in Computer Vision to obtain the motion between sets of points. It is based on the computation of the motion that best fits two (or more) sets of clouded points. This chapter presents an overview of the existing techniques, as well as a new classification of them. We have employed a set of representative techniques in this field and some comparative results are presented. The techniques presented in this chapter are discussed and compared taking into account their 3D registration performance.

2.1 Techniques to acquire the depth information

Surface acquisition is one of the most important topics in visual perception. Without it, acquiring the third dimension is impossible. In order to attain this depth perception, various systems have attempted to imitate the human vision system.

These methods are classified into what Woodham [Woodham, 1978] refers to as direct

methods and indirect methods. Direct methods are those that try to measure distance ranges directly, for example, pulsed laser based systems, where the depth information is the only information available. Indirect methods are those that attempt to determine distance by measuring parameters calculated from images of the illuminated object. Several direct and indirect methods commonly refer to these techniques as shape from X, where X is one of a number of options resulting from the spread of such technologies in the last few years.

Shape from X techniques can be divided into four main groups:

- *Techniques based on modifying the intrinsic camera parameters, i.e. depth from focus/defocus and depth from zooming,* consisting of the acquisition of several images of the scene from the same point of view by changing the camera parameters. By using depth from focus/defocus, the camera parameters can be dynamically changed during the surface estimation process [Favaro and Soatto, 2002]. Depth from zooming involves the use of multiple images taken with a single camera coupled with a motorized zoom.

- *Techniques based on considering an additional source of light projected onto the scene, i.e. shape from photometric stereo and shape from structured light.* Photometric stereo considers several radiance maps of the measuring surface captured by a single camera and a set of known light sources. The use of at least three radiance maps determines a single position and orientation for every imaged point [Solomon and Ikeuchi, 1996]. The structured light technique is based on the projection of a known pattern of light onto the measuring surface, such as points, lines, stripes or grids. 3D information of the scene is obtained by analyzing the deformations of the projected pattern when it is imaged by the camera [Salvi *et al.*, 2004].

- *Techniques based on considering additional surface information, i.e. shape from shading, shape from texture and shape from geometric Constraints.* Shape from shading uses the pattern of shading in a single image to infer the shape of the surface. Often, the parameter of the reflectance map is unknown. In this case we have to estimate the albedo and the illuminant direction. From the reflection map and by assuming local surface smoothness, we can estimation local surface normal, which can be integrated to give local surface shape [Gibbins, 1994]. The basic principle behind shape from texture is the distortion of the individual texels. Their variation across the image gives an estimation of the shape of the observed surface.

The shape reconstruction exploits perspective distortion which makes objects farther from the camera appear smaller, and foreshortening distortion which makes objects not parallel to the image plane seem shorter. Assuming that the normals are dense enough and the surface is smooth, these distortions can be used to reconstruct the surface shape [Chantler, 1994]. Finally, shape from geometric constraints considers the problem of obtaining 3D reconstruction from 2D points localized in a single image. Planarity, colinearity, known angles and other geometric properties provided by the "user" are taken into account to remove ambiguities from the scene and, if possible, obtain a single reconstruction [Grossmann, 2002].

- *Techniques merely based on multiple views, such as shape from stereo and shape from motion.* Shape from stereo is based on solving the correspondence problem between two or more views of a given surface taken from different locations [Armangué and Salvi, 2003]. Each image point determines an optical ray which intersects with the others in space in order to compute the 3D surface point. Shape from motion exploits the relative motion between camera and scene [Matabosch *et al.*, 2003][Armangué *et al.*, 2003]. Similar to the stereo technique, this process can be divided into the subprocesses of finding correspondences from consecutive frames and reconstructing of the scene. The differences between consecutive frames are, on average, much smaller than those of typical stereo pairs because the image sequences are sampled at higher rates. Motion computation can be obtained by estimating optical flow and using differential epipolar constraint.

Stereovision is one of the most important topics in Computer Vision since it allows the three dimensional position of an object point to be obtained from its projective points in the image planes [Faugeras, 1993]. The main problem in stereovision is the well known *correspondence problem*, or determining the correspondences between two images. There are geometrical constraints, based on epipolar geometry, that help to find the correspondences. However, this information is not enough, and additional information is required. Therefore, stereovision can only by applied in textured scenes, where the color and texture information is used to establish correspondences.

Active stereovision aims solve this problem. The term *active* is not used because the visual systems can interact with the scenes but rather because it is based on the illumination of the scene with structured light. Previous structured light techniques were based on projecting simple primitives like a single dot or a single line of light, usually

provided by lasers. The advantage of projecting such structured light primitives is that the correspondence problem of the illuminated points in the images is directly solved. Nevertheless, the number of correspondences per image is very small. In order to increase the number of correspondences, structured light patterns like arrays of dots, stripes, grids or concentric circles [Robinson *et al.*, 2004] were introduced. However, in this technique, known as uncoded structured light, the correspondence problem is not directly solved and the indexing problem must be solved. The indexing problem consists of identifying the relationship between pattern projections and light acquisition, which provoked the emergence of coded structured light [Batlle *et al.*, 1998]. In this case, the projected patterns are coded so that each element of the pattern can be unambiguously identified in the images. Therefore, the aim of coded structured light is to robustly obtain a large set of correspondences per image independently of the appearance of the object being illuminated and the ambient lighting conditions.

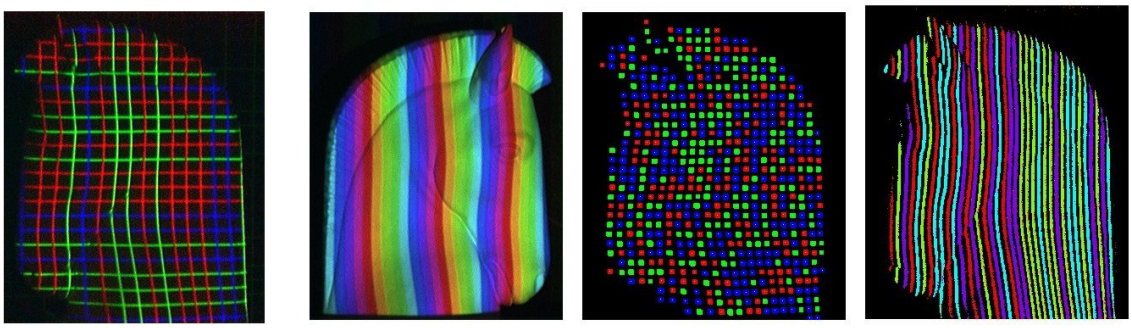


Figure 2.1: Some examples of coded structured light patterns

In order to get good acquisitions, several colors must be projected to the scene (see Fig. 2.1). However, the use of colors may be a problem depending on the texture of the scene and also on the presence of uncontrolled ambient light. On the other hand, if only one color is used, several uncoded patterns must be projected to code each point of the scene (see Fig. 2.2). Therefore, moving scenes can not be acquired.

In summary, although case coded structured light is generally preferred to uncoded structured light, in one-shoot pattern applications or in colored scenes it is better to use of uncoded structured light [Matabosch *et al.*, 2005*b*].

Another type of range finders are based on laser projections. This technique consists of the projecting a laser profile onto the scene. Generally, a plane is projected onto the surface, producing a 3D curve in the space. As only a single plane is projected, motion

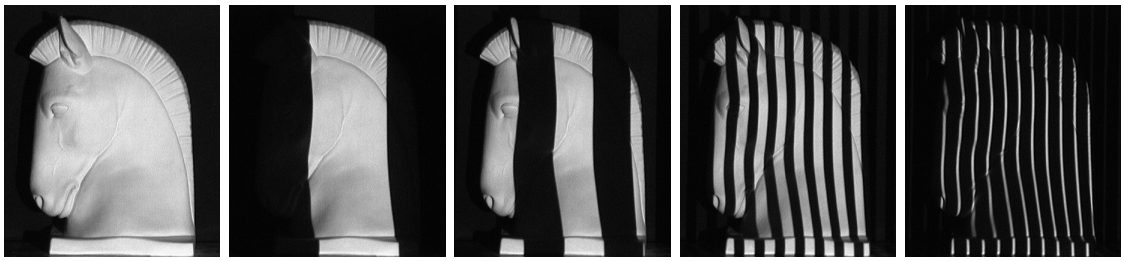


Figure 2.2: Sequence of patterns to code the scene

must be added to scan the surface. Most typical is the use of a rotation mirror to change the laser direction, the use of translation tables to displace the laser position, or the use of rotation tables to change the orientation of the object (see Fig. 2.3).

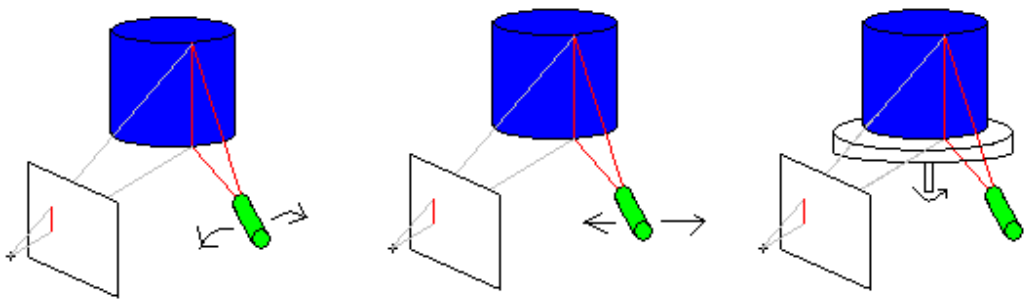


Figure 2.3: Examples of laser projection triangulation principle

These kind of sensors are often used due to the large resolution of the acquired surface. Furthermore, the accuracy is generally much better with respect to coded structured light. On the other hand, it is obvious that only static scenes can be acquired because movement or vibration produces a misalignment in the acquisition. Moreover, the precision of the mechanical systems also affects the final result. Additionally, a control system is required to synchronize motion and acquisition.

There are also one-shot patterns based on grids of dots, parallel planes, coaxial circles, etc. These patterns can also be considered as uncoded structured light. In this case, one pattern is enough, however, the indexing problem must still be solved and, also, the resolution is limited.

Although different types of range finders are introduced, most of them can only obtain partial information of the scene. This is due to several reasons such as occlusions, shadows and, especially, the limited point-of-view of the sensor. In order to get a full model, several

views acquired from different poses of the sensor are required. The fusion of all of these views gives us the complete information of the scene. As the pose of the camera in each acquisition is generally unknown, aligning partial acquisitions is not trivial. There are several techniques to fuse a set of views:

- *Mosaicing* technique based on homographies to integrate multiple images in to a continuous one, increasing the field of view of the camera. This technique can only obtain planar reconstructions.
- *Simultaneous localization And mapping* technique used by robots and autonomous vehicles to build up a map within an unknown environment while at the same time keeping track of their current position. They usually use odometry information to get an estimation of the motion, which is refined by means of computer vision.
- *Photogrammetry* measurement technology in which the three-dimensional coordinates of points on an object are determined by measurements made in two or more photographic images taken from different positions. Algorithms for photogrammetry typically express the problem as that of minimizing the sum of the squares of a set of errors. The minimization is itself often performed using the bundle adjustment.
- *Registration:* This technique aligns two or more data sets by minimizing the distance between them. There are several variants, 2D/2D, 2D/3D and 3D/3D among others.

The term *range image registration* is defined as the set of techniques that represents a set of range images to a common coordinate frame. *Range image* is an *mxn* grid of distances that describe a surface in Cartesian or cylindrical coordinates. This concept is based mainly on the fact that most acquisitions are based on structured light. In these systems, for a set of pixels on the image, distances between the sensor and the scene are obtained. Applying several steps, a surface can be obtained from this range image. First of all, 3D clouded points can be generated. Then, a triangulation is required to get a continuous surface. Some additional steps can be applied to smooth the surface, remove false triangles, etc. Finally, texture can be applied to the surface.

Although the term range image only refers to a bidimensional representation of the distance between the sensor and a set of points of the scenes, it can be used as the information given by 3D sensors. Therefore, the term *range image registration* represents the set of techniques that aligns several partial clouded points of a common scene obtained

by a range sensor, which can be based on coded structured light, laser projection, time-of-flight, etc.

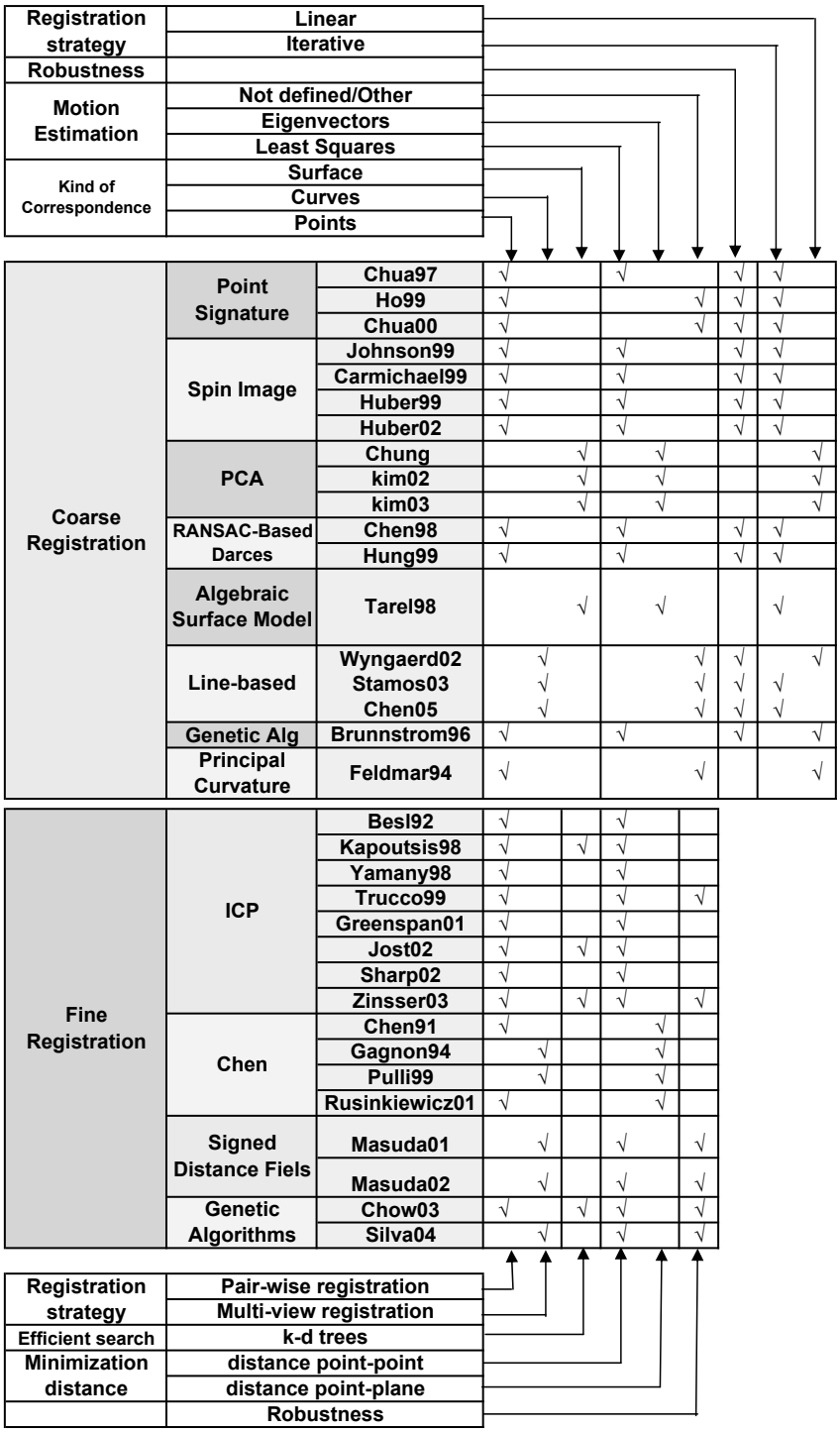
Initially, range image registration was developed to be used in surface-to-model registration. A partial acquisition is registered with the complete model of the scene. Applications of this kind of registration are visual inspection, recognition tasks, etc. Some years later, the surface-to-surface registration appears. In this case, the goal is to obtain the motion between two partial surfaces of the same scene. The main problem with this type of registration is the presence of points on the first surface that are not present on the second surface. Some modifications must be made to remove false correspondences. The main application of this kind of registration is the computation of 3D surfaces by aligning a set of partial acquisitions.

The chapter is structured as follows: first, the classification is presented in section 2.2. Second, in section 2.3, coarse registration techniques are explained. In section 2.4, fine registration techniques are presented. In section 2.5, the experimental results obtained with a set of implemented techniques are presented. Section 2.6 concludes with a discussion of advantages and drawbacks.

2.2 Classification of registration techniques

The goal of registration is to find the Euclidean motion between a set of range images of a given object taken from different positions in order to represent them all with respect to a reference frame. The proposed techniques differ as to whether initial information is required, so that a rough registration can only be estimated without an initial guess. If an estimated motion between views is available, a fine registration can then be computed. The classification of the surveyed methods is revealed in Table 2.1.

Table 2.1: Classification of Registration methods



In coarse registration, the main goal is to compute an initial estimation of the rigid motion between two clouds of 3D points using correspondences between both surfaces, as explained in Section 2.3. These methods can be classified in terms of a) the kind of correspondences used; b) the method used to compute the motion; c) the robustness of the method; and d) the registration strategy (see Table 2.1). In general, the most common correspondence method used is point-to-point, such as the *point signature* method [Chua, 1997] and the method of *spin-image* [Johnson, 1997]. However, there are other methods that align lines, like methods of *bitangent curves* [Wyngaerd, 2002] and others that match the surfaces directly, like the *algebraic surface model* [Tarel *et al.*, 1998]. Another important aspect of coarse registration is the way of computing the motion when correspondences are found. Robustness in the presence of noise is another important property, because there are usually no corresponding regions between views. Most methods are robust, looking for the best combination of correspondences [Chua, 1997; Chen *et al.*, 1998; Johnson and Hebert, 1999]. Other methods may converge to a local solution [Feldmar and Ayache, 1994], and in theory this fact increases the speed of the method but the solution is not always the best, and in some cases it is far from the right solution. In general, coarse registration methods are iterative, usually maximizing the rate of overlapping points. However, a few provides linear solutions, like the methods based on *Principal Component Analysis* [Kim *et al.*, 2003] or the *Algebraic Surface model* [Tarel *et al.*, 1998].

In fine registration, the goal is to obtain the most accurate solution as possible. These methods use an initial estimation of the motion to first represent all range images with respect to a reference system, and then refine the transformation matrix by minimizing the distances between temporal correspondences, known as closest points. Table 2.1 also classifies fine registration methods in terms of: a) the registration strategy; b) the use of an efficient search method, such as k-d trees in order to speed up the algorithm; c) the way of computing the minimization distance, either point-to-point or point-to-plane; d) the way of computing the motion in each iteration; and e) the robustness of the method.

The registration strategy can differ according to whether all range views of the object are registered at the same time (*multi-view registration*) or the method registers only a pair of range images in every execution (*pair-wise registration*). Moreover, fine registration methods need a lot of processing time to decide which is the closest point. In order to deal with this problem, several proposals to increase the searching speed have been presented, such as the use of k-d trees to alleviate the problem of searching neighbors.

Another important parameter is the distance to minimize. Most methods use the distance between point correspondences, while others use the distance between a given point in the first range image and the corresponding tangent plane in the second. The problem of point-to-point distance is that the correspondence of a given point in the first view may not exist in the second view because of the limited number of points acquired by the sensor, especially on low resolution surfaces. To address this problem, some authors use the point-to-plane distance. In this case, a tangent plane in the second view is computed at the position pointed by the given point in the first view. The distance between the point in the first view and that tangent plane in the second is the minimization distance. Theoretically, point-to-plane converges in less iterations than point-to-point.

Finally, robust methods can cope with noise and false correspondences due to the presence of non-overlapping regions. In real images, the robustness is very important, especially when only a small part of the first view has a correspondence in the second, that is in the presence of a reduced overlapping region.

2.3 Coarse Registration techniques

Coarse registration methods search for an initial estimation of the motion between pairs of consecutive 3D views leading to the complete registration of the surface. In order to compute this motion, distances between correspondences in different views are minimized. Features from both surfaces are usually extracted with the aim of matching them to obtain the set of correspondences, whereas other techniques find such correspondences without any feature extraction but with some Euclidean invariants. The most common correspondences are points, curves and surfaces.

In some situations, coarse registration techniques can be classified on *shape features* or *matching methods*. The first group searches for characteristics of points, using usually neighborhood information, in order to search for correspondences. Examples of this group are *Point Signature*, *Spin Image*, etc. Matching methods are based on the process of matching points from both surfaces, as *RANSAC-based Darces* or *Genetic Algorithm*. In some situations both techniques can be combined to find correspondences, as Brunnström [Brunnström and Stoddart, 1996], who used the normal vectors at every point to define the fitness function of the genetic algorithm. On the other hand, techniques of both groups can be used independently as *RANSAC-based Darces* which do not use fea-

tures in the matching process or *Point Signature* that when points are characterized only a comparison between features from both surfaces is required to detect correspondences.

2.3.1 Point Signature

Point Signature is a point descriptor introduced by Chua [Chua, 1997] and used to search for correspondences. Given a point p, the curve of the surface that intersects with a sphere of radius r centered to p gives the contour of points (C). These points are then represented in a new coordinate frame centered at p. The orientation axes are given by the normal vector (n_1) at p, a reference vector (n_2) and the vector obtained by the cross-product. All points on C are projected to the tangent plane giving a curve C'. The vector n_2 is computed as the unit vector from p to a point on C' which gives the largest distance. Thus, every point on C can be characterized by: a) the signed distance between its own correspondence in C'; and b) a clockwise rotation angle θ from the reference vector n_2. Depending on the resolution, different $\Delta\theta$s are chosen. Then, the point signature can be expressed as a set of distances in each θ from 0^o to 360^o. Finally point signatures from two views are compared to determine potential correspondences. The matching process is very fast and efficient.

The main drawback of the algorithm is the process to compute the Point signature. The intersection of a sphere to the surface is not very easy, especially when the surface is represented as a cloud of points or a triangulated surface. In this situation interpolation is required, incrementing the computing time and decrementing the quality of the point signature. Moreover the computation of the reference vector is very sensible to noise, and errors in this computation effects the Point Signature descriptor obtained considerably.

2.3.2 Spin Image

Spin image is a 2D image characterization of a point belonging to a surface [Johnson, 1997]. Like point signature, spin image was initially proposed for image recognition. However, it has been used in several registration applications since then.

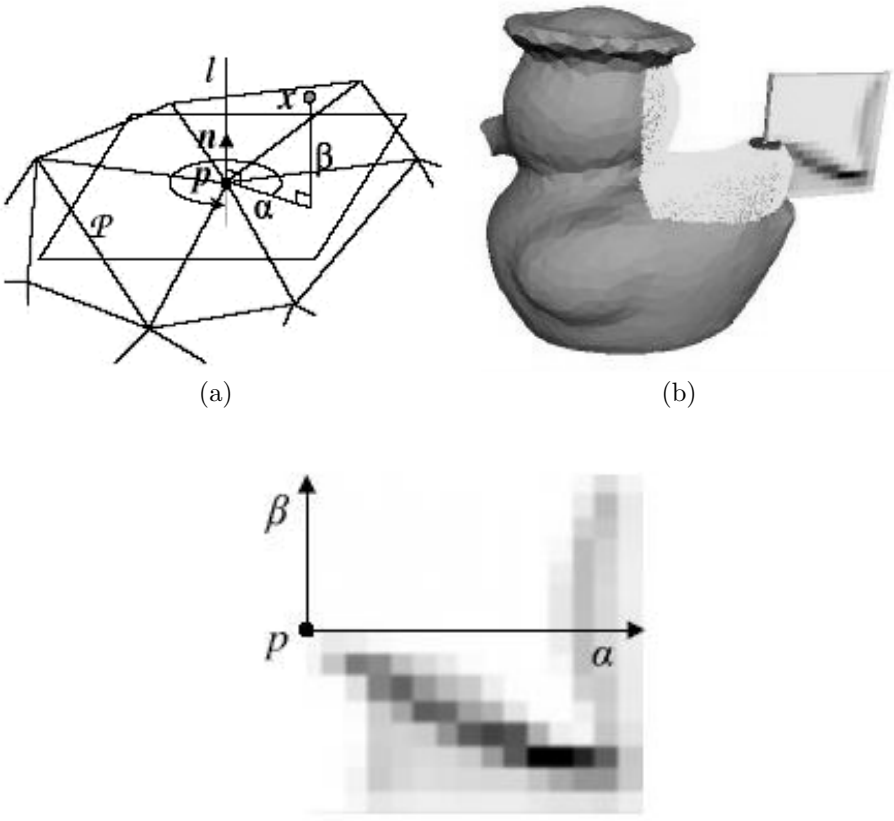
Consider a given point at which a tangent plane is computed by using the position of its neighboring points. Then, a region around the given point is considered in which two distances are computed to determine the spin image: a) the distance α between each

point to the normal vector defined by the tangent plane; and b) the distance β between this point to the tangent plane; obtaining:

$$\alpha = \sqrt{\|x - p\|^2 - (n(x - p))^2} \tag{2.1}$$

$$\beta = n(x - p) \tag{2.2}$$

where p is the given point, n is the normal vector at this point, and x is the set of neighboring points used to generate the spin image. Using these distances, a table is generated representing α on the x-axis and β on the y-axis. Each cell of this table contains the number of points that belong to the corresponding region. In order to choose the size of the table that determines the resolution of the image, the double length of the triangle mesh is selected. One example of the spin image is shown in Fig. 2.4.



(a) (b)

(c)

Figure 2.4: Example of Spin Image construction. a) Representation of the axis of the spin image . b) Visualization of the process to obtain the spin image and c) Final result of the spin image

Some spin images are computed in the first view and then, for each one, the best correspondences are searched for in the second view. When the point correspondences are found, outliers are removed by using the mean and the standard deviation of the residual as a threshold. The rigid transformation is finally computed from the best correspondence found.

The main problem of this method is that the spin image strongly depends on the resolution of the method. In order to solve this problem, Carmichael [Carmichael *et al.*, 1999] proposed the *face-based spin image* in which a set of points are interpolated inside every triangular mesh with the aim of uniforming the number of points in every spin image computation. In addition, other approaches have been presented to solve the problem of false mesh triangles given by surface boundaries and occlusions [Huber and Hebert, 1999]. In this case, the method is used as a filter to remove such false triangles before registration.

Finally, using the variants of Spin Image, good results can be found in Range Image Registration. The Spin Image feature is very robust, except in case of symmetries or repeated regions in the object. However this is a problem present in most part of coarse registration techniques.

2.3.3 Principal Component Analysis

This method is based on using the direction of the main axis of the volume given by the cloud of points of the range image to align the sequence of range images between them. If the overlapping region is large enough, both main axes should be almost coincident and related to a rigid motion so that registration may succeed. Therefore, this transformation matrix is found to be the one that aligns both axes by only applying a simple product (see eq. 2.5). This method is very fast with respect to others that identify point or curve correspondences. However, the overlapping region must be a very important part of the view in order to obtain good results. Chung [Chung *et al.*, 1998] proposed a registration algorithm using the direction vectors of a cloud of points (a similar approach was used by Kim [Kim *et al.*, 2003]). The method involves calculating the covariance matrix of each range image as follows:

$$Cov = \frac{1}{N}\sum_{i=0}^{N-1}(p_i - \overline{p})(p_i - \overline{p})^T \tag{2.3}$$

where N is the number of points, $\overline{p}$ is the center of mass of the cloud of points, and p_i is the ith point of the surface . Then, the direction U_i of the main axis can be computed by singular value decomposition:

$$Cov_i = U_i D_i U_i^T; \tag{2.4}$$

The rotation is determined by the product of the eigenvector matrices:

$$R = U_1 U_2^{-1} \tag{2.5}$$

and the translation is determined by the distance between the centers of mass of both clouds of points, expressed with respect to the same axis:

$$t = \overline{\mu}_2 - R\overline{\mu}_1 \tag{2.6}$$

Principal component analysis is very fast. However, it can only be used with effectiveness when there is a sufficient number of points. In addiction, this method obtains accurate solutions when most part of the points are common. Results are less accurate when the overlapping region constitutes a smaller part of the image. In practice, a 50% overlapping of the region is critical. However, the solution obtained can be used as an initial guess in a further fine registration. The main problem of principal component analysis is its limitation in coping with surfaces that contain symmetries. Thus, if the eigenvalues obtained representing two axes are similar, the order of these axes can change in the matrix U_i, and the final result obtained is completely different from the correct solution. Although PCA provides a fast solution, in most cases this one is far from the expected.

2.3.4 RANSAC-Based Darces

This method is based on finding the best three point correspondences between two range images to obtain an estimation of the Euclidean motion. Three points are the minimum required to compute the motion between both surfaces if no other information is used [Chen *et al.*, 1998]. As will be commented in section 2.3.8, Feldmar used only a single point but

also considered the normal vector and the principal curvature to obtain enough information to compute the rigid motion [Feldmar and Ayache, 1994].

Three points (primary, secondary and auxiliary) in the first view are characterized by the three distances between them (d_{ps},d_{pa} and d_{sa}). Each point in the second view is hypothesized to be the correspondence of the primary point (p'). Next, the secondary point is searched for among the points located at a distance d_{ps} from p'. If there are not any points in that position, another primary point is tested. Otherwise, a third point in the second view that satisfies the distances defined in the triplet is searched. Once a triplet is identified, the rigid transformation between both points can be determined. This search is repeated for every satisfied triplet between both views and a set of potential Euclidean motions is obtained. The correct transformation is the one that obtains the largest number of corresponding points between both views.

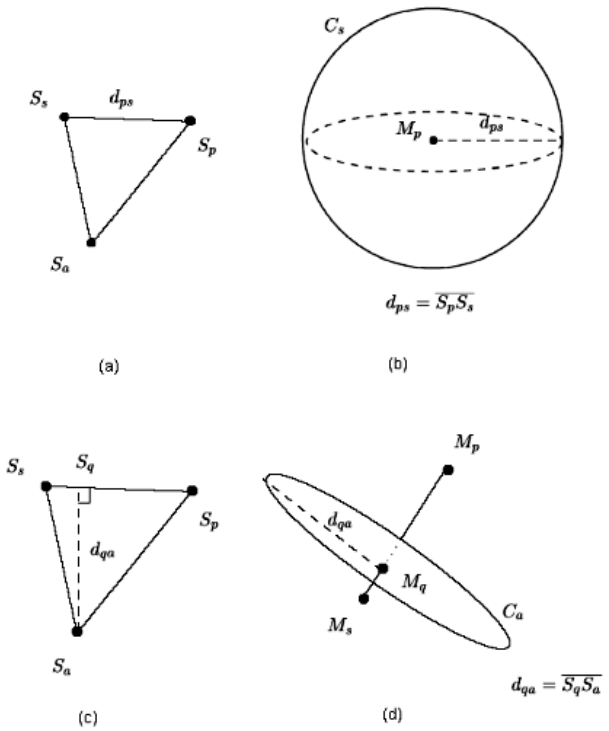


Figure 2.5: Method of search of points: a,b)search of the secondary point in the surface of a sphere of radius d_{ps}; c,d)search of the auxiliary point in the circle obtained by the intersection of the two spheres of radius d_{pa} and d_{sa}

A modification of this method focused on decreasing the computing time related to the search of correspondences was proposed [Chen *et al.*, 1999]. The results obtained were very good because of its robustness even in the presence of outliers. However, it can only be used when the number of points in each view is relatively small. Theoretically it is a good method. However, the precision depends on the resolution of the surface and the time increases considerably with the number of points, so that it can only be used in applications where time is not critical.

2.3.5 Algebraic Surface Model

Tarel [Tarel *et al.*, 1998] proposed a method to estimate the motion between surfaces represented as a polynomial model. First, two implicit polynomial models are determined from all the points of both range images using *3L Fitting*, a linear algorithm based on Least Squares. In general, the algorithms used to obtain a model are iterative, and require a lot of processing to compute the polynomial function. However, the linear algorithm does not require so much computational time and offers better repeatability compared to other implicit polynomial fitting methods.

This method is based on obtaining a function of the distance between the polynomial model and the points, where these distances are nearly zero. In order to improve the accuracy of this method, fictional points are added to the range image located at distances of +c and -c from the surface.

As this method does not need points or curve correspondences, the computation time is faster compared to others. However, a normal vector at each point is required to estimate the model, which it is not easy to compute when only points are available. If the range scanner gives this information, the computing time decreases considerably. The principal drawback of this method is the requirement that a large part of both images must belong to the overlapping region. The author reports good results with less than 15% of non-overlapping region[1], which is quite unusual in range image registration.

[1]This means that more than 85% of overlapping is required

2.3.6 Line-based algorithm

Some authors proposed to use lines to find pairs of correspondences. Examples are the straight line-based method proposed by Stamos [Stamos and Leordeanu, 2003] and the curved line-based method proposed by Wyngaerd [Wyngaerd, 2002].

The former is based on the extraction of straight segments directly in the range images which are further registered with the aim of computing the motion between the different views. The algorithm is applied to large and structured environments such as buildings in which planar regions and straight lines can be easily found. The segmentation algorithm determines a set of border lines and their corresponding planes. First, a robust algorithm is used to efficiently search pairs of lines based on line length and plane area. Then, the rotation and translation among potential pairs is computed. Finally, the one that maximizes the number of planes is taken as the solution.

Some years later, the same authors changed the approach used in computing the motion between straight lines [Chen and Stamos, 2005]. As most part of lines in a structured environment is contained in the three planes of a coordinate system, they proposed to compute first the three main directions of every view. Hence, 24 combinations arise to potential align both views. Then, the rotation matrix is computed for every combination and, finally, the one that maximizes the number of diagonal elements is selected as the rotation solution. The rest of rotation matrices are kept because final results are supervised by an operator. Translation vectors are computed as the one that connect midpoints of two pair of segments. The more repeated vector is selected to become the solution. Finally, the registration is refined by using an ICP-based method.

The algorithm obtains good results even considering that it is classified into coarse registration. The main drawback is the difficulty to segment the straight segments as well as the supervisor required to check the final results given by the method. Both drawbacks decrease the number of applications but the method has performed very well in the registration of buildings.

The general case of line-based matching is the consideration of curved lines in order to register free-form surfaces

Vanden Wyngaerd [Wyngaerd, 2002] proposed a rough estimation of motion by matching bitangent curves (see Fig. 2.6). A bitangent curve is a pair of curves composed by the union of bitangent points, which are simultaneously defined as a pair of points tangent to

the same plane. The bitangent curves are found by means of a search in the dual space.

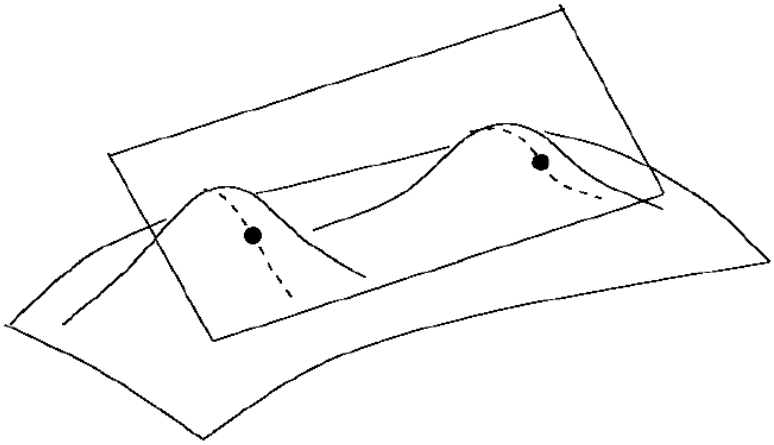


Figure 2.6: Exemple of a pair of bitangent points

The main idea is that all bitangent points are coincident in the dual space. In order to do the search, it is necessary to represent the four parameters of any plane using only three components or coordinates. So, the normal vectors at each point of the range image are computed and their norms are set to one. Using these vectors and the coordinates of their points, it is easy to compute the four parameters of the plane (a, b, c and d) tangent to that point.

Since the norms of the normal vectors are set to one, it is possible to represent this vector using just two parameters. The author used a and b to parameterize the normal vector. In theory, it is possible to construct the dual space using a, b and d. However, it is necessary to normalize the parameter d between -1 and +1 to scale the values.

Once all the bitangent curves present in a range image are extracted from the dual space, the matching between these curves with the curves in the next range image starts. In this way, an invariant description of a pair of bitangent curves is used with the goal of matching only the most representative curves, i.e. the 15 longest ones. The invariant used is defined as the set of distances between bitangent points.

In order to increase efficiency, the curve is divided into segments of equal length. Once a correspondence is found, four corresponding points, that is the two end-points of both bitangent segments, are obtained. With these four correspondences, the Euclidean transformation can be computed, and then the error can be analyzed by transforming all the points with respect to the reference system. The matching of bitangent segments

that correspond to the minimum error is selected as the best one among all the potential matches.

Compared to other methods in which the correspondence is based on points, this method has the advantage that the range image is previously transformed into the dual space before the search for possible matches starts. This transformation decreases the computing time and increases the robustness. However, depending on the shape of the object, the number of bitangent points can be insufficient to ensure good results.

2.3.7 Genetic Algorithm

Brunnström [Brunnström and Stoddart, 1996] used a genetic algorithm to solve the problem of searching for correspondences between two range images. The interest in this method is centered on defining the vector that contains the n index of correspondences between both range images, where the size of the vector is set to n, i.e. the number of points in the second range image (the image that is matched with respect to the first). Genetic algorithms require a fitness function to measure the quality of each potential solution. In order to determine this fitness function, four invariants between the two pairs of correspondences are used:

$$\|\overrightarrow{v}_{ij}\| = \|\overrightarrow{r}_j - \overrightarrow{r}_i\| \tag{2.7}$$

$$cos(\theta_{ij}) = \frac{\overrightarrow{n_j} \cdot \overrightarrow{v_{ij}}}{\|\overrightarrow{n}_j\| \cdot \|\overrightarrow{v}_{ij}\|} \tag{2.8}$$

$$cos(\theta_{ji}) = \frac{\overrightarrow{n_i} \cdot \overrightarrow{v_{ji}}}{\|\overrightarrow{n}_i\| \cdot \|\overrightarrow{v}_{ji}\|} \tag{2.9}$$

$$cos(\beta_{ji}) = \frac{(\overrightarrow{n}_j \times \overrightarrow{v}_{ij}) \cdot (\overrightarrow{n}_i \times \overrightarrow{v}_{ij})}{\|\overrightarrow{n}_i\| \cdot \|\overrightarrow{n}_j\| \cdot \|\overrightarrow{v}_{ij}\|^2} \tag{2.10}$$

where r_i and r_j are the position of two points belonging to the same surface and n_i and n_j are the normal vectors at both points, respectively.

Using these invariants, the quality of the correspondences is computed analyzing the distance error and the error in the normal parameters as follows:

$$q_d(\alpha_i, \alpha_j) = e^{-\frac{[\|\overrightarrow{v}_{\alpha_i\alpha_j}\| - \|\overrightarrow{v}_{ij}\|]^2}{2\sigma^2}} \tag{2.11}$$

$$q_n(\alpha_i, \alpha_j) = e^{\frac{(\theta_{\alpha_i,\alpha_j} - \theta_{ij})^2 + (\theta_{\alpha_j,\alpha_i} - \theta_{ji})^2 + (\beta_{\alpha_i,\alpha_j} - \beta_{ij})^2}{2\mu^2}} \tag{2.12}$$

$$Q(\alpha_i, \alpha_j) = q_d(\alpha_i, \alpha_j) q_n(\alpha_i, \alpha_j) \tag{2.13}$$

where σ and μ are experimental parameters that must be estimated; i and j are the indexes that determine two points in the second range image; and α_i and α_j represent the indexes of two other points in the first range image. Then, the quality of a correspondence α_i can be computed as the sum of the qualities of every pair of correspondences between α_i and the rest of the points.

$$Q(\alpha_i) = \sum_{j \neq i} Q(\alpha_i, \alpha_j) \tag{2.14}$$

The previous function indicates the quality of a pair of correspondences, while the fitness function indicates the global matching quality, which is expressed as a function of $Q(\alpha_i)$ as follows,

$$f(\overrightarrow{\alpha}) = e^{\sqrt{\sum_i Q(\alpha_i)}} \tag{2.15}$$

When the fitness function is defined, the cross-over and mutation probabilities are fixed to characterize the algorithm. The mutation is not very important when searching for correspondences because nearby indexes do not imply nearby points in the space. Therefore, the author set the probability of mutation at 1%, with a crossover of 90%.

Finally, the stopping criteria, which it is not a very well-studied problem in genetic algorithms, must be defined. Three different approaches were presented: a) setting a % of good correspondences (computed starting from the closest points); b) supervising the fitness function so that it does not increase after a certain number of iterations; and c) counting iterations until a certain number. When the algorithm finishes, the Euclidean motion might be computed because the chromosome that maximises the fitness function contains the point correspondences. However, some correspondences in the chromosome might be wrong, which means that these bad correspondences must be previously removed in order to guarantee the computation of a good Euclidean Motion, using SVD, for instance. So, only the 30% of the correspondences that maximize $Q(\alpha_i)$ are used in the computation. As in most genetic approaches, the results obtained are quite good but the computing time is expensive, specially in the presence of a large number of points, where exists lots of potential correspondences. As Ransac-based DARCES algorithm, it is not appropriate when time is critical.

2.3.8 Principal curvature

Feldmar [Feldmar and Ayache, 1994] proposed that use of the differential of points characterized by the principal curvature in the matching process. This method only needs a single correspondence to compute the Euclidean motion between two range images. It characterizes a point by its principal curvatures (k_1, k_2). Principal curvatures are the maximum and the minimum curvature of the surface at a point. Additionally, the normal vector and the principal direction corresponding to the principal curvature are also considered. In order to facilitate the search, points of the second range image are organized in a table indexed by their curvature values. Then, considering a point M in the first view, whose curvatures are (k_1, k_2), the set of points in the second view, whose curvatures are close to (k_1, k_2), can be quickly found and evaluated as the corresponding point. For every potential matching, the Euclidean motion that aligns P_1=($M_1, \overrightarrow{e}_{11}, \overrightarrow{e}_{21}, \overrightarrow{n}_1$) with P_2=($M_2, \overrightarrow{e}_{12}, \overrightarrow{e}_{22}, \overrightarrow{n}_2$) is computed, where M_i is a 3D point in the i image, $\overrightarrow{e}_{1i}, \overrightarrow{e}_{2i}$ are the principal directions of the curvature at the M_i point, and $\overrightarrow{n_i}$ is the normal vector at that point. Thus, two rigid displacements, D and D', are computed, where D corresponds to align P_1 to P_2, while D' aligns P_1 to P_2'=($M_2, -\overrightarrow{e}_{12}, -\overrightarrow{e}_{22}, \overrightarrow{n}_2$). Finally, the transformation matrix (R,t) defining the Euclidean motion between both views can be easily computed as follows:

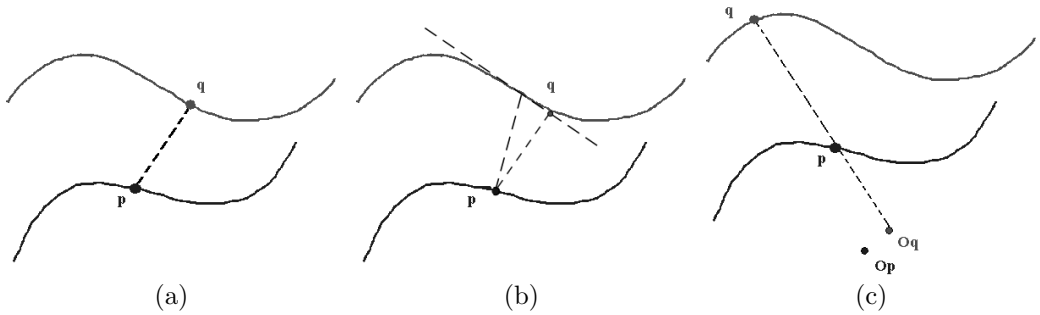
$$D = (BA^t, M_2 - BA^tM_1) \tag{2.16}$$

$$D' = (B'A^t, M_2 - B'A^tM_1) \tag{2.17}$$

where A is the 3x3 matrix whose columns are ($\overrightarrow{e}_{11}, \overrightarrow{e}_{21}, \overrightarrow{n}_1$), B is ($\overrightarrow{e}_{12}, \overrightarrow{e}_{22}, \overrightarrow{n}_2$) and B' is ($-\overrightarrow{e}_{12}, -\overrightarrow{e}_{22}, \overrightarrow{n}_2$). Then, every transformation matrix is evaluated. A ratio is computed by considering the number of points in the transformed surface (computed from the Euclidean motion) which have a corresponding point in the second surface at a distance smaller than a threshold related to the total number of points of the range image. If both D and D' do not reach the termination criteria, the algorithm is repeated using an alternative initial point. Otherwise, the transformation matrix computed is considered to be a good estimation.

Although the curvatures and principal directions are not always known, the author described a method to determine them based on the use of a parametrization of the surface similar to the polynomial fitting of Tarel (see Section 2.3.5).

The main problem of this method is that it is not robust. Only a good correspondence



(a) (b) (c)

Figure 2.7: Different kind of distances used in the Fine Registration: a) Point to Point; b) Point to Plane; c) Point to Projection

is search for, and when a possible solution is found, the algorithm stops. However, as the correspondence is not validated, it might be a false matching, obtaining a bad initial registration that may satisfies the rate of overlapping.

2.4 Fine Registration techniques

The term *fine registration* is used when an estimation of the motion is previously known and used as an initial guess to iterate and converge to a more accurate solution. In order to solve this problem, a distance function is minimized. Some authors used the distance between point correspondences, while others used the distance of a given point in the first view to a plane tangent to its closest point in the second view. There exist also the correspondences point-to-projection. However, results of this technique are very inaccurate and it is the reason it is not very used. A scheme of each one is presented in Fig 2.7.

In recent years, some methods have been presented : a) Iterative Closest Point; b) Chen's Method; c) Signed Distance Fields; and d) Genetic Algorithms, among others.

In the registration process, different methodology can be used independently of the technique chosen, which are hereafter briefly related:

Control Points Although some authors use all points in the registration step [Besl and McKay, 1992] [Chen and Medioni, 1991], others use Uniform subsampling [Turk and Levoy, 1996] [Masuda, 2001], Random Sampling [Masuda *et al.*, 1996] or Normal Sampling [Rusinkiewicz and Levoy, 2001]. In the common case, the better solution

is the Normal Sampling[2] for two reasons. As a sampling is applied, the time required is smaller. Furthermore, as points from significant parts of the surfaces are more important compared to parts of uniform surfaces, usually better results are obtained, specially in cases where the surfaces are not very shaped (see Fig. 2.8).

Points Weight Although in some situations all points have the same weight [Besl and McKay, 1992], in other situations weights are introduced depending on: distance between point-correspondences [Godin *et al.*, 1994], compatibility of normals [Godin *et al.*, 1994] and the uncertainty of the covariance matrix [Z.Zhang, 1992], among others.

Rejecting Pairs Some authors used all point correspondences to determine the motion because they usually work with image-to-model registration. However, when an image is not a subset of the following in the sequence, some correspondences are outliers and must be rejected.

Now, the surveyed fine registration techniques are discussed in the following sections.

2.4.1 Iterative Closest Point (ICP)

The ICP method was presented by Besl [Besl and McKay, 1992]. The goal of this method is to obtain an accurate solution by minimizing the distance between point correspondences, known as closest point. When an initial estimation is known, all the points are transformed to a reference system applying the Euclidean motion. Then, every point in the first image (p_i) is taken into consideration to search for its closest point in the second image (m_i), so that the distance between these correspondences is minimized, and the process is iterated until convergence (see Eq 3.2).

$$f = \frac{1}{N_p} \sum_{i=1}^{N_p} \| \overrightarrow{m_i} - R(\overrightarrow{q}_R) \overrightarrow{p}_i - \overrightarrow{t} \|^2 \tag{2.18}$$

[2]Normal sampling is based on sampling points by means of the direction of the vector normal to the surface at these points.

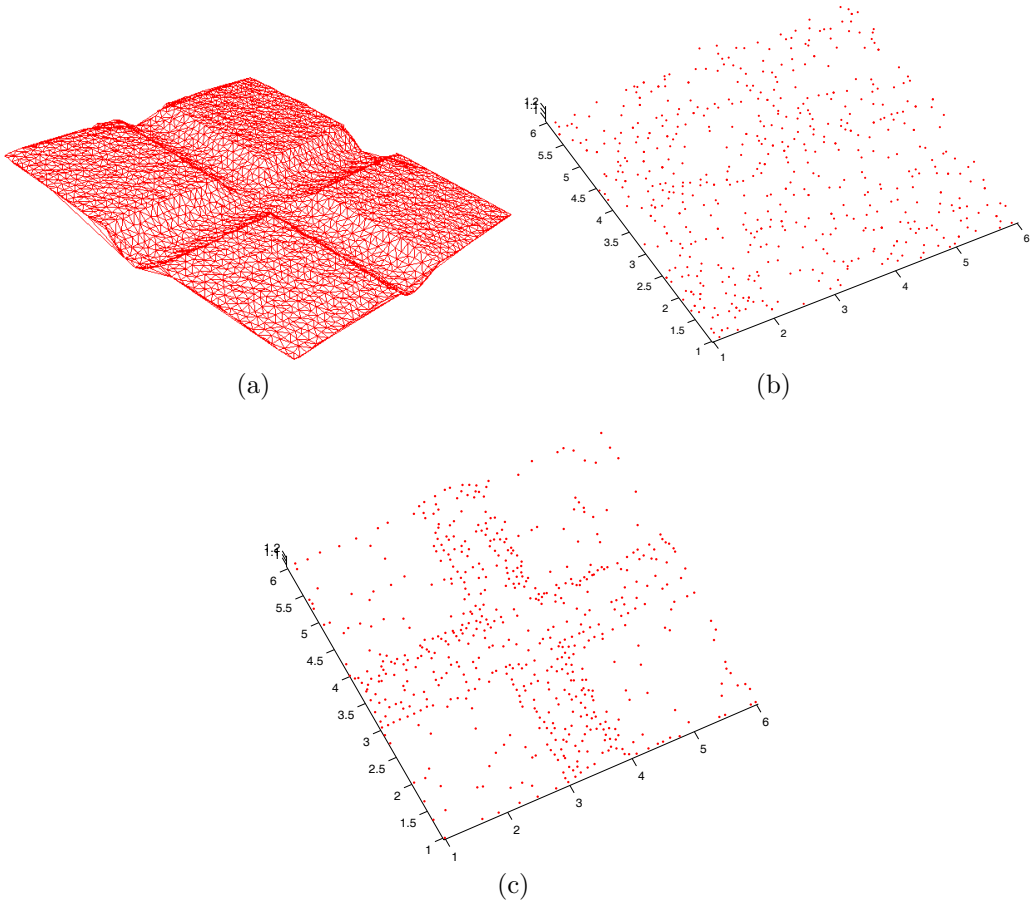


Figure 2.8: Effects of sampling: a) Original surface; b) Random sampling; c) Normal sampling.

In order to minimize the previous equation, a symmetric 4 x 4 matrix Q($\sum_{pm}$) is defined as follows,

$$Q(\Sigma_{pm}) = \begin{bmatrix} tr(\sum_{pm}) & \Delta^T \\ \Delta & \Sigma_{pm} + \Sigma_{pm}^T - tr(\Sigma_{pm}) I_3 \end{bmatrix} \tag{2.19}$$

where tr is the trace, $\Delta = [A_{23} A_{31} A_{12}]^T$ is computed from the anti-symmetric matrix $A_{ij} = (\sum_{px} - \sum_{px}^T)_{ij}$, Δ^T is the transpose of Δ, I_3 is the identity matrix and $\sum_{pm}$ is the cross-covariance matrix of the points P and M given by:

$$\Sigma_{px} = \frac{1}{N_p} \sum_{i=1}^{N_p} [\overrightarrow{p}_i \overrightarrow{m}_i] - \overrightarrow{\mu}_p \overrightarrow{\mu}_m \tag{2.20}$$

The unit eigenvector $\overrightarrow{q}_R = [q_0 \; q_1 \; q_2 \; q_3]^t$ corresponding to the maximum eigenvalue of the matrix Q is selected as the optimal rotation expressed in quaternions.

Once R is computed (see Eq. 2.21), the translation vector can be easily computed (see Eq. 2.22) and the motion determined by R and t.

$$R = \begin{bmatrix} q_0^2 + q_1^2 - q_2^2 - q_3^2 & 2(q_1 q_2 - q_0 q_3) & 2(q_1 q_3 + q_0 q_2) \\ 2(q_1 q_2 + q_0 q_3) & q_0^2 - q_1^2 + q_2^2 - q_3^2 & 2(q_2 q_3 - q_0 q_1) \\ 2(q_1 q_3 - q_0 q_2) & 2(q_2 q_3 + q_0 q_1) & q_0^2 - q_1^2 - q_2^2 + q_3^2 \end{bmatrix} \tag{2.21}$$

$$t = \overline{\mu}_m - R\overline{\mu}_p \tag{2.22}$$

The motion is applied to the first surface and the process is repeated until distances between corresponding points decrease below a threshold. ICP obtains good results even in the presence of Gaussian noise. However, the main drawback is that the method can not cope with non-overlapping regions because the outliers are never removed. Moreover, when starting from a rough estimation of the motion, the convergence is not guaranteed.

Some modifications of ICP have been presented in recent years. In 2001, Greenspan [Greenspan and Godin, 2001] applied the Nearest Neighbor Problem to facilitate the search of closest points. The first range image is considered as a reference set of points, which is preprocessed in order to find, for every point, the neighborhood of points in the second view located at a certain distance. The points of the neighborhood are sorted according to that distance. The use of this pretreatment leads to consider the closest point

of the previous iteration as an estimation of the correspondence in the current iteration. If this estimation satisfies the spherical constraint, the current closest point is considered to belong to the neighborhood of the estimated. A property of the spherical constraint holds that any point that is closer to $\overrightarrow{q}$ (a point in the second range image) than the current estimation $\overrightarrow{p_c}$ (the closest point obtained in the last iteration) must be located inside a sphere centered at $\overrightarrow{p_c}$ with a radius of $2\|\overrightarrow{q} - \overrightarrow{p_c}\|$. This pretreatment decreases the computing time drastically. A year later, Jost [Jost and Hugli, 2002] presented the Multi-resolution Scheme ICP algorithm, which is a modification of ICP for fast registration. The main idea of the algorithm is to solve the first few iterations using down sampled points and to progressively increase the resolution by increasing the number of points considered. The author divides the number of points by a factor in each resolution step. The number of iterations in each resolution step is not fixed, so that the algorithm goes to the next resolution when the distance between correspondences falls below a threshold.

In the same year, Sharp [Sharp *et al.*, 2002] proposed the *ICP using invariant features ICPIF*. In this case, points are matched using a weighted feature distance as follows,

$$d_\alpha(p, m) = d_e(p, m) + \alpha^2 d_f(p, m) \tag{2.23}$$

where d_e is the Euclidean distance, d_f is the distance in the feature space between each correspondence points and α controls the relative contribution of the features. Different invariant features were proposed: a) curvature; b) moment; and c) spherical harmonics. Experiments reported by the author pointed out that the spherical harmonics provided the best convergence rate, while the traditional ICP provided the worst one.

One year before, Godin [Godin *et al.*, 2001] presented a similar work, where color and curvature were used as feature information. As color is used as a matching constraint, symmetric objects can be well registered. However the author only presents results with simple objects without shadows, occlusions or luminance changes.

In addition, other authors proposed some improvements to increase the robustness of ICP. For instance, Trucco [Trucco *et al.*, 1999] implemented the RICP method making use of the Least Median of Squares approach. The method is based on executing the registration with just a sample of random points (m points), computing this operation a sufficient number of times with the aim of finding a registration without outliers. The Monte Carlo algorithm was used to estimate the number of executions. Once all the potential registra-

tions were computed, the one that minimizes the median of the residuals is chosen as the solution. Finally, the correspondences with a residual larger than 2.5σ were removed and the transformation between both views was computed using only the remaining points (inliers). Note that σ was estimated by using a robust standard deviation [Rousseeuw and Leroy, 1987].

Moreover, Zinsser [Zinsser and Schnidt, 2003] proposed a robust method based on outlier thresholding known as the Picky ICP algorithm. The main difference with respects to the previous methods is that at every iteration only the pairs of correspondences with the smallest distances are used in the motion computation. The threshold was fixed at a given multiple of the standard deviation.

Overall, ICP is the most common registration method used and the results provided by authors are very good. However, this method usually presents problems of convergence, lots of iterations are required, and in some cases the algorithm converges to a local minimum. Moreover, unless a robust implementation is used, the algorithm only can be used in surface-to-model registration.

2.4.2 Method of Chen

In 1991, Chen [Chen and Medioni, 1991] proposed an alternative to the ICP algorithm, which was based on minimizing the distance between points and planes. The minimization function was selected to be the distances between points in the first image with respect to tangent planes in the second. That is, considering a point in the first image, the intersection of the normal vector at this point with the second surface determines a second point in which the tangent plane is computed.

Despite other authors that considered this algorithm just an improvement of ICP, we have considered it a new technique for several reasons. First, the paper was presented at the same time of Besl's approach. Second, the search of correspondences is an important aspect of the algorithm, consequently it should be considered a different method.

A new algorithm to find these intersections between lines and range images was proposed, bearing in mind that the intersection of a given line with a cloud of points requires a high computational cost. A more accurate technique to find such a distance was proposed by Gagnon [Gagnon *et al.*, 1994].

Finally, once the distances between points and planes were obtained, the motion that

minimizes these distances was estimated by least squares. The process was repeated until convergence was attained.

In 2001, Rusinkiewick [Rusinkiewicz and Levoy, 2001] presented several variations of this algorithm to improvement the precision of the algorithm. The author proposed the *Normal Space Sampling*. The main idea is to select more points in the region where the normal is different from the other parts of the surface. Using this sampling technique better results are obtain in low shaped surfaces.

In 2003, Park [Park and Subbarao, 2003] proposed a fast variant based on recursively computed the intersection between the line and the 3D surface. This technique is based on a iterative point-to-projection, obtaining fast and accurately results.

Point-to-plane distance is normally more difficult to compute compare to point-to-point. When no normal information is given, the plane must be computed using neighborhood information, which requires a lot of time to detect this neighborhood, and not always with sufficient accuracy in the estimation of the normal vector. However, nowadays this estimation can be obtained directly from most part of range finders. This method is more robust to local minimum and, in general, results are better. Despite no robust method are applied, this method is less sensible in the presence of non-overlapping regions. The reason is that only the control points that their normal vector intersects the second view are considered in the matching. Moreover, point-to-point distance ensures a correspondence unless a distance threshold or other robust identifier is used. Chen's approach usually requires less iterations than the ICP approach.

2.4.3 Matching Signed Distance Fields

In 2001, Masuda [Masuda, 2001] [Masuda, 2002] presented a new registration algorithm, which was based on the Matching Signed Distance Fields. The method was a robust one so the outliers were removed, and all the views of a given object were registered at the same time, which means a *multi-view registration*.

First, all views (α) were transformed to a reference coordinate system using the initial estimation of the motion. A set of key points was generated on a 3D grid of buckets with a fixed size δ. Then, the closest point from every key point was searched for in order to establish the correspondences, limiting the distance between points to $\sqrt{3}\delta$.

The process was composed by the following steps:

1. Initialization: Compute the initial values of the motion parameters, T_0^α=$[R_0^\alpha\ t_0^\alpha]$
2. Determine the closest point p_i^α to every key point p.
3. Compute the new motion parameters T_{i+1}^α using the correspondences between p and p_i^α

Steps 2 and 3 are repeated until there is convergence and an integrated model of the object made by the p points is obtained.

At every iteration the closest point was computed as follows:

$$p_i^\alpha = (T_i^\alpha)^{-1}p - (n_p^{\alpha^T}(T_i^\alpha)^{-1}p - k_p^\alpha)n_p^\alpha \tag{2.24}$$

where n_p^α represents the normal vector of p and k_p^α represents the signed distance, which is computed as:

$$k_p^\alpha = n_p^{\alpha^T} c_p^\alpha - d_p \tag{2.25}$$

where c_p^α is the closest point to each key point for every range image α, and d_p is the average distance between the point p and each c_p^α.

$$d_p = \frac{1}{\sum_{1\leq\alpha\leq N_R} w_p^\alpha} \sum_{1\leq\alpha\leq N_R} w_p^\alpha d_p^\alpha \tag{2.26}$$

where d_p^α represents the distance between point p and the nearest key point and w_p^α is the weight of point p.

Finally, once the solution has converged, it is possible to compute the points (signed distance field) and their normal vectors, obtaining the final model as follows:

$$SDF_p = p + n_p d_p \tag{2.27}$$

$$n_p = \frac{\sum_{1\leq\alpha\leq N_R} w_p^\alpha R^\alpha n_p^\alpha}{\| \sum_{1\leq\alpha\leq N_R} w_p^\alpha R^\alpha n_p^\alpha} \tag{2.28}$$

The advantage of this method is that it registers all the views at the same time, which implies a more robust solution and avoids the error propagation problem present in *pair-wise registration* methods. On the other hand, this algorithm can not be used in real time as *localization and mapping* because not all views are already available when the motion must be computed.

Another advantage of this algorithm is that the final result is directly an integrated model, compared to Pair-wise registration techniques that require and off-line algorithm to eliminate the overlapping region and triangulate all the registered surfaces.

2.4.4 Genetic Algorithms

Chow [Chow *et al.*, 2004] presented a dynamic genetic algorithm to solve the registration problem. The goal of this method is to find a chromosome composed of the 6 parameters of the motion that aligns a pair of range images accurately. The chromosome is composed of the three components of the translation vector and the three angles of the rotation. In order to minimize the registration error, the median is chosen as the fitness function, as follows:

$$F(T) = Median(E_i) \tag{2.29}$$

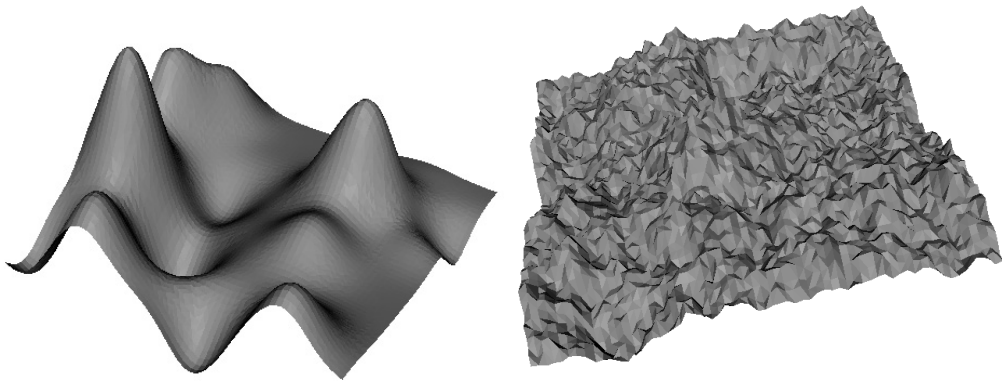
$$E_i = min|Tp_i - m_j| \tag{2.30}$$

where T is the transformation matrix composed by 6 motion parameters and p_i and m_i the points of both surfaces.

Therefore, only a sample of points of the first image were used to compute the error with the aim of decreasing the computing time. The cross-over operation consisted of combining genes made by two chromosomes to create a new chromosome. The author randomly selected the number of genes to be swapped. The cross-over operation works well when the chromosome is far from the final solution but it is useless for improving the solution in a situation close to convergence. Therefore, the mutation operation was defined as follows: a gene is randomly selected and a value randomly obtained between the limits $[-MV, +MV]$ is added. The limits are very wide at the beginning and become narrower at every step in order to guarantee the convergence in the final steps.

A similar method was proposed the same year by Silva [Silva *et al.*, 2003]. The main advantage of this work is that a more robust fitness function is used and the initial guess is not required. The author defined the Surface Interpenetration Measure (SIM) as a new robust measurement that quantifies visual registration errors. Another advantage of Chow's method is the multi-view registration approach. Finally, the hillclimbling strategy was used to speed up the convergence.

Overall, the use of genetic algorithms has the advantage of avoiding local minima which



(a) Wave surface (b) Fractal landscape

Figure 2.9: Test scenes used in the experiments

is a common problem in registration, especially when the initial motion is not provided or it is given with low precision. This algorithm also works well in the presence of noise and outliers given by non overlapping regions. The main problem of this algorithm is the time required to converge. Additionally, fitness function must be theoretically computed using real correspondences, however, as they are unknown, temporally correspondences are used using the estimated motion. However, as the motion change iteratively, correspondences must be searched several times (for example, every time a better result is found), resulting again in a lot of computing time.

2.5 Experimental Results

Although many authors evaluated the accuracy of their methods, very few provided a comparison between already published approaches [Rusinkiewicz and Levoy, 2001], [Dalley and Flynn, 2002], [Matabosch *et al.*, 2004*b*]. In order to provide such a useful comparison, some methods to evaluate the accuracy have been tested which allow us to decide the best method in every situation. The computing time has also been considered, which might be critical in some applications. The measures used to determine the accuracy are the following: *rotation error*, *translation error* and *RMS (root mean square)*.

In order to evaluate the accuracy in rotation, it is represented as a directional vector and a rotating angle around such a vector, which both can be easily extracted from any rotating matrix. Then, the estimated rotation is compared to a real one. So, the error is determined as the angle between both directional vectors (γ) and the discrepancy between

both angles of rotation (α - β). The translation error is defined as the distance between the origin of the coordinate system estimated by registration with respect to the real one, that is, the norm of the difference between both translation vectors. The RMS error is defined as the mean of the distances between point correspondences considering all views in the same reference frame.

In order to compare the main algorithms, synthetic data is used to evaluate the registration errors. First of all, synthetic test scenes are generated. Additionally more realistic synthetic data is used, and finally experiments with real data are realized.

The first test scenes used are *Wave* and *Fractal landscape* (see Fig. 2.9). These surfaces have been used by other authors [Zinsser and Schnidt, 2003] [Rusinkiewicz and Levoy, 2001] to test their algorithms. These surfaces are contained in a range of 1 in X and *Y-axis* and approximately 0.5 in *Z-axis*. Although these scenes certainly do not cover all sort of scanned objects, they are quite representative of many real scannings. Several experiments are performed to evaluate the precision in front of different error sources, as shape measurement error (noise, surface sampling, etc), and correspondence error (occlusion, outliers, etc.), among others. Although it is very difficult to distinguish these error sources, because they are usually related, these experiments are organized to show basically the effects of each source of error.

As final results can change depending on the implementation of the algorithm, details of this implementation are briefly commented. Ransac-Based DARCES approach is programmed using a subset of 5% of the points of the first view to search for correspondences using all the points in the second view. Moreover, as distances between primary, secondary and auxiliary points can not be exactly found, a variation of a 2% of such a distance is permitted. Furthermore, the distance between the 3 control points is limited to a 15% of the maximum distance to speed up the process. In the Spin Image algorithm, the number of correspondences searched is fixed to the 5% of the number of points, and 25% of points are used to compute the Spin Image. In the implementation of the Genetic Algorithm of Brunnström, only 5% points of one surface are searched in the second. In this situation, the second surface have all points, differing from the author that used only 50 points in every surface. However, it is still difficult to work with accuracy using such large number of points.

In fine Registration algorithms, a threshold of 10^{-5} is fixed to guarantee convergence. However, a maximum number of iterations is fixed to 20 in all the algorithms, except in

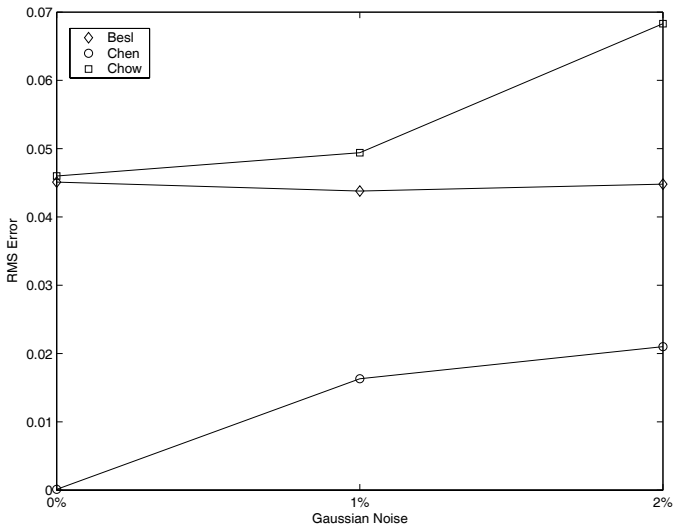


Figure 2.10: Effects of Gaussian Noise on the accuracy of Fine Registration Algorithms

the case of Chow which number of iterations is fixed to 600 due to the fact that the method is based on genetic algorithms.

In order to evaluate the effects of shape measurement error, two experiments have been done. First of all, gaussian noise is independently added into the test scenes. Experiments are performed using different values of standard deviation error; the first with an error of 1% of the range dimension in each angle direction, and in the second one with 2% of standard deviation.

In Coarse registration algorithm fractal surface and wave surface are subsampled to 512 points. The motion between both surfaces is a rotation of 45° in each angle and a displacement of 1 unity in each axis.

The results obtained shown that the final result is not very influenced by the noise, because this noise is negligible with respects to the errors in coarse Registration techniques.

In order to test the Fine Registration algorithms, a sequence of surfaces of 2000 points are registered. Every surface is related to the previous by a rotation of 5° around every axis and a displacement of 0.2 unities in the Z-axis . This initial motion is similar to the pre-alignment used by Rusinkiewicz [Rusinkiewicz and Levoy, 2001]. The same percentage of noise is independently added in each surface.

Results obtained in Fig. 2.10 shown that while the error obtained by Chen algorithm is directly related to the noise introduced, Besl and Chow have always a residual error not

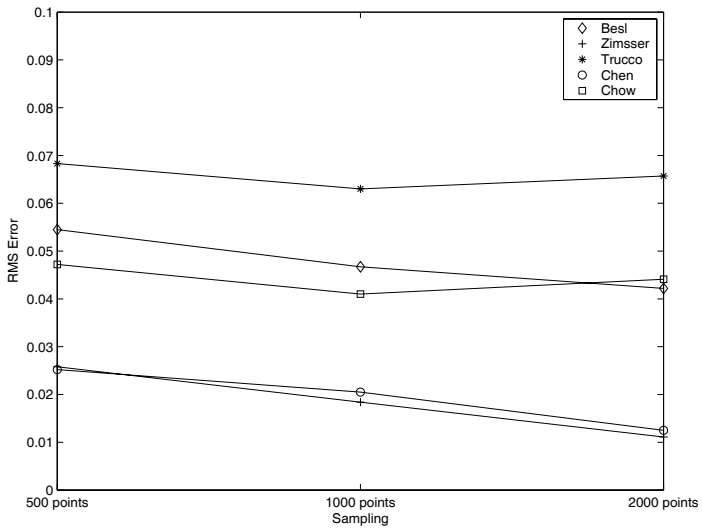


Figure 2.11: Evolution of the RMS error in Fine Registration algorithms in wave surface with respects to the surface sampling.

depending on the noise introduced. This error is usually consequence of the convergence to a local minima. As Chen approach solves better the problem of local minima, excellent results are obtained without noise.

Another case of misalignments is the surface sampling. Theoretically when surfaces are scanned with low resolution, worse results are obtained. As surfaces have less points, points correspondences are found with low precision. In order to test the effects, original surfaces are independently subsampled into different resolutions: a) (12.5% and 25% of the initial size (2048 points) in coarse registration; and b) 12.5% and 25% of the initial size (4096 points) in Fine Registration).

As is shown in Fig. 2.11, errors in Fine registration algorithms extremely depends on the sampling surface. While the variations in Trucco, Besl and Chow are not significative, in Chen and Zinsser algorithms, the errors using surfaces of 2000 points are approximately the half than using a subsampling of 500 points.

Furthermore, in the case of low sampling (500 points), Zinsser and Chen algorithms presented the best solutions. As the Chen approach use a point-to-plane distance, the accuracy is better compared to point-to-point because the plane interpolation decrements the effects of the subsampling. Although Zinsser algorithms presents similar results in RMS errors, the rotation errors of Zinsser's are considerably larger compared to Chen's (see Table 2.2). The fact is that not always RMS errors are representative of a correct

Table 2.2: Error in Fine registration methods using Synthetic data

Method	RMS error	Translation error	α-β(rad)	time(s)
Besl	0.052	0.029	-0.0297	5.47
Zinsser	0.006	0.28	-0.013	2.08
Trucco	0.053	0.056	-0.028	3.48
Chen	0.022	0.006	-0.001	2.96
Chow	0.043	0.033	-0.146	473.47

registration, especially in robust variant when correspondences related with large distances are removed, and only the correspondences with small distances remains.

On the other hand, in most part of results obtained with Coarse Registration methods, sampling does not effect considerably the final results, and low resolution surfaces can be used in this situation. In this experiment only Ransac-based DARCES is clearly effected by the sampling. This is because it is difficult to find 3 points in both surfaces that correspondences exactly to the same position. Besides, PCA does not use directly the points to computed the motion, and Spin Image and the Genetic algorithm use more points reducing the error in the motion estimation. However, experiments realized with more complex objects show that in this situation, results are more influenced by the sampling.

In both experiments realized, 100% of overlapping area is used. However, in most part of real applications both surfaces contains non-overlapping regions. In these situations, false correspondences are very common, and they affect considerably the final resolution. In this experiment, several percentages (5%, 10%, 20% and 50% of the surface) of non-overlapping region are introduced.

Results obtained shown that the presence of non-overlapping regions do not effect significantly Coarse Registration techniques. In general, coarse Registration techniques only use a few part of points to obtain the motion, and, if these points are correctly selected, the result is similar to the one obtained without outliers. On the other hand, Principal Component Analysis obtained similar eigenvalues from the surfaces with or without non-overlapping regions because they are almost planar.

In fine registration techniques, errors in Besl and Chow algorithms increase directly proportional to the percentage of non-overlapping region. This change is specially significant with 50% of outliers (see Fig. 2.12). On the other hand, Zinsser algorithm can cope with 5% and 10% of outliers, however, the accuracy decreases. Finally, Chen approach

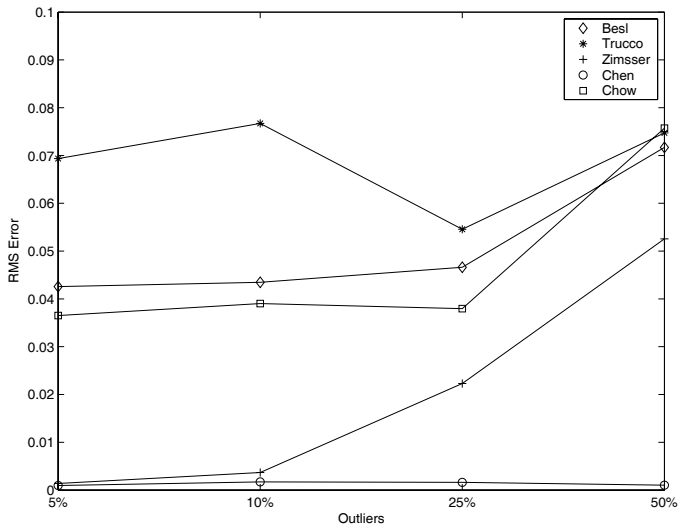


Figure 2.12: Effects of Outliers in the quality of the Fine Registrations Algorithms.

is robust to outliers. This is because point-correspondences are computed using the projection of points in a grid Z space, and points whose projection belongs to an empty cell are not considered. Most part of outliers are removed and motion is only computed using points belonging to both surfaces.

Other experiments are realized using more realistic objects. Tables 2.3 and 2.4 show the experimental results obtained by using synthetic clouds of 3D points of Fig. 2.13, taken from the database of RapidForm2004.

Finally, real scanned objects are used to take into account the effects of the shape measurement error. Results are shown in Tables 2.5 and 2.6. It can be shown in Fig. 2.5 that the executing time of Ransac-based DARCES algorithm is surprisingly smaller in the case of 500 points compared to the case of 250 points. This is because in the case of 250 points, the distance between three points is larger compare to the case of 500 and no good correspondences might be found without increasing the searching distance.

Although the time required is very important, all the methods have been programmed using Matlab 6.5 in a Pentium IV $2,6$GHz because we are just interested in the comparison among the methods and Matlab guarantees an easy implementation. Furthermore kd-tree or other velocity improvements are not used in these experiments to easy again the comparison.

The 3D clouds of points have been acquired using a Minolta Vivid 700 laser scanner,

Figure 2.13: 3D Model from RapidForm2004 used in the synthetic experiments

obtaining up to 8 views of a given object. The registration results obtained are shown in Fig. 2.14.

The real positions of the object are known and so the accuracy of the methods compared because the real objects have been scanned on a turnable table with a the precision of less than 1°.

Several objects have been registered and we have observed that the shape of the object affects the accuracy of the registration considerably and independently of the registration method used. Overall, when the views have poor details, planarity or symmetry, the registration does not converge (see Fig. 2.15). Moreover, a robust method is essential to register real range images. See, for instance, Fig. 2.14a compared to the rest of registrations (b, c and d) in which the registrations are more accurate because only the points belonging to the overlapping region have been used in the computation.

2.6 Conclusions

In this section, most representative Range Image Registration algorithms have been surveyed, programmed and tested with different surfaces: synthetic images, test scenes used by other authors and real images. Experiments performed shown the main characteristics of each method. Main advantages and drawbacks are summarized in Table 2.7.

Coarse registration techniques are used when an initial estimation of the Euclidean motion is unknown. The main interest in these techniques is the algorithm used to find the

Table 2.3: Error in Coarse registration methods using Synthetic data

Points	Method	Translation error (mm)	α-β (rad)	time (s)
200	Ransac-Based	50.613	0.98573	1.062
	PCA	21.229	2.9533	0.016
	SpinImage	149.84	2.8911	7.969
	GenAlg	81.359	0.0	4.281
400	Ransac-Based	21.394	0.30569	106.08
	PCA	18.777	3.0037	0.016
	SpinImage	81.19	2.3465	50.703
	GenAlg	33.537	1.6721	122.56
700	Ransac-Based	27.557	0.096427	32820.0
	PCA	16.566	0.21871	0.016
	SpinImage	64.254	0.90339	169.76
	GenAlg	18.017	2.9687	249.89

correspondences between clouds of points, which can be based on points, curves, surfaces and directional vectors. Besides, fine registration methods are based on converging to a solution from an initial estimation of the rigid motion. Depending on the method used, a quite accurate initial guess is required because some methods have convergence problems due to the presence of local minima.

The surveyed methods are explained in detail focusing the analysis on the main contribution of each technique and especially detailing the pros and cons between them. Experimental results using both synthetic and real data are reported providing a useful comparison of the methods, which is a worthwhile information that is not usually provided by authors.

Analyzing the results of coarse registration methods, we conclude that in the presence of low resolution views Chen's RANSAC-based DARCES method is the one that obtains better results, especially in synthetic models where the points are coincident and only related to a transformation matrix. Besides, in real images where points might not coincide, the results obtained are less accurate (see Fig. 2.16). Moreover, when the amount of points used in the registration increases, the computing time required to get the estimation is really a problem. In this case the best method becomes Spin Image, whose accuracy depends on the number of points used (see Fig. 2.17) while the computing time remains important. If computing time is critical, the Principal Component Analysis is the best fast method. The problem of this method is that it is based on aligning the three

Table 2.4: Experimental results using synthetic data data by Fine registration methods

Points	Method	Translation error (mm)	α-β (rad)	RMS error (mm)	time (s)
700	Besl	0.37	0.01	1.59	0.91
	Zinsser	0.25	0.003	2.64	3.28
	Jost	0.82	0.006	4.85	0.38
	Trucco	0.87	0.011	2.27	4.13
	Chen	0.0	0.0	1.49	25.03
1500	Besl	0.41	0.007	1.48	4.77
	Zinsser	0.093	0.002	1.75	9.28
	Jost	0.257	0.009	4.22	1.30
	Trucco	0.768	0.016	5.44	112.8
	Chen	0.0	0.0	1.68	145.06
6000	Besl	0.166	0.002	0.71	47.86
	Zinsser	0.169	0.002	1.86	406.05
	Jost	0.801	0.019	3.01	8.51
	Trucco	0.458	0.015	2.91	198.28
	Chen	0.0	0.0	1.36	1217.3

principal axes of the cloud of points, so that the given solution may suffer from symmetries of 90° or 180°. This problem can be solved modifying the method to find the solution that maximizes the number of overlapping points. However, in this case, the solution obtained is very bad when the overlapping region is not significant. Finally, a genetic algorithm is robust to noise and the results are quite good, but again a lot of time is required to reach a solution.

In addition, considering all the fine registration methods, the Chen method is the best one from our point of view. This method solves the local minima problem presented by ICP variants. Although the point-to-plane distance is theoretically difficult to compute, the iterative algorithm proposed by Chen is very fast and efficient. Furthermore, this method can work with non-overlapping regions, because points whose projection is far from the correspondences are not considered.

On the other hand, results provided by the method of Besl have the problem of local minima, presenting errors even without the presence of noise. Moreover, Besl's approach can not work with non-overlapping regions. In order to solve this problem, it is necessary to use a robust method like the ones proposed by Zinsser or Trucco. The main difference between the two of them is the way of computing the overlapping region. Although both

Table 2.5: Experimental results of two real range images in Coarse registration methods

Points	Method	Translation error(mm)	α-β (rad)	γ (rad)	time (s)
250	Ransac-Based	69.406	2.3562	1.3855	2.86
	PCA	67.464	1.9923	1.2985	0.016
	SpinImage	53.861	0.91301	1.9844	24.156
	GenAlg	72.094	2.1475	1.5708	44.766
500	Ransac-Based	127.03	2.3562	1.5708	2.453
	PCA	69.87	1.9703	1.3306	0.016
	SpinImage	40.147	0.54215	0.68824	101.59
	GenAlg	82.698	2.1988	1.5708	243.94
1000	Ransac-Based	40.53	0.17204	1.5236	1657.7
	PCA	69.893	1.9236	1.3262	0.203
	SpinImage	23.133	0.54339	1.5708	565.45
	GenAlg	38.059	1.7571	0.85619	1051.6

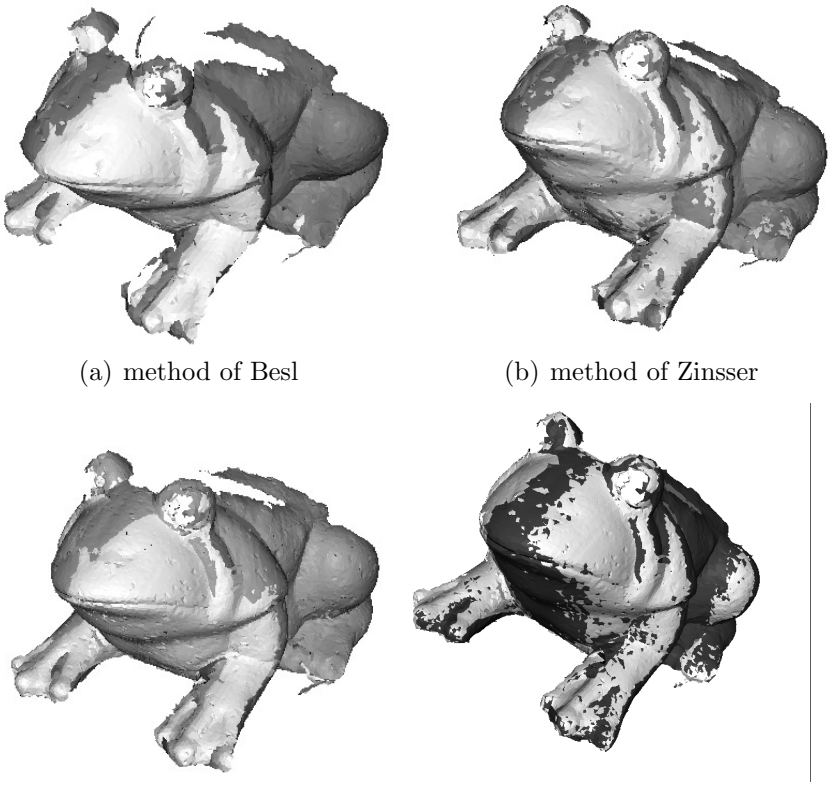
strategies are good, the results obtained indicate that Zinsser's method is more accurate than the one proposed by Trucco's. However, Trucco's method is faster in the presence of high resolution surfaces. The variant proposed by Jost obtains good results considering that the process has been speed up considerably.

Table 2.6: Experimental results of Fine registration methods with real data

Points	Method	Translation error (mm)	α-β (rad)	RMS error(mm)	time (s)
500	Besl	3.34	-0.09	4.18	1.41
	Zinsser	0.68	-0.06	2.35	1.95
	Jost	2.01	-0.07	4.54	0.36
	Trucco	2.87	-0.05	3.93	19.33
	Chen	1.3373	-0.008	2.1608	18.391
	Chow	0.27	0.0	4.60	154.05
1000	Besl	3.47	-0.09	3.78	5.22
	Zinsser	0.47	-0.02	1.42	7.97
	Jost	2.38	-0.08	3.70	1.38
	Trucco	2.71	-0.05	5.13	33.95
	Chen	0.29957	0.003	1.7305	67.954
	Chow	0.12	-0.01	4.51	281.61
5000	Besl	3.12	-0.08	3.20	57.36
	Zinsser	0.26	-0.02	0.85	281.58
	Jost	2.24	-0.09	2.83	9.14
	Trucco	2.57	-0.05	1.77	149.59
	Chen	1.2535	-0.015	1.2543	993.09
	Chow	0.06	0.0	3.57	1776.00

Table 2.7: Summary of pros and cons of registration techniques

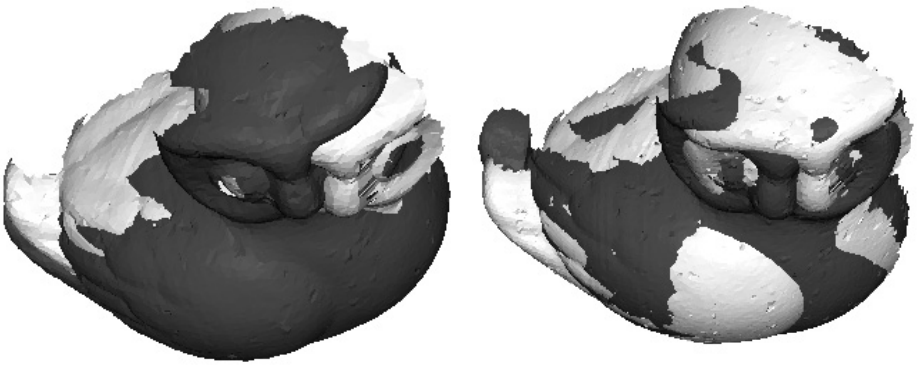
	Methods	Advantages	Drawbacks
Coarse	Ransac Based	Quite accurate	Large consuming time Similar sampling is required
	PCA	Very fast	Large overlapping area is required Not accurate
	Spin Image	Accurate	Normal vectors are required High resolution is required
	Genetic Algorithms	Robust	Huge consuming time
Fine	ICP	Only points are required	Converge to local minima
	Chen	Converge to the correct solution Few iterations	Complex to determine correspondences
	Masuda	Robust Global minimization	All views required Large computational cost
	Genetic Algorithms	Robust Global minimization	All views required Huge computational cost



(a) method of Besl (b) method of Zinsser

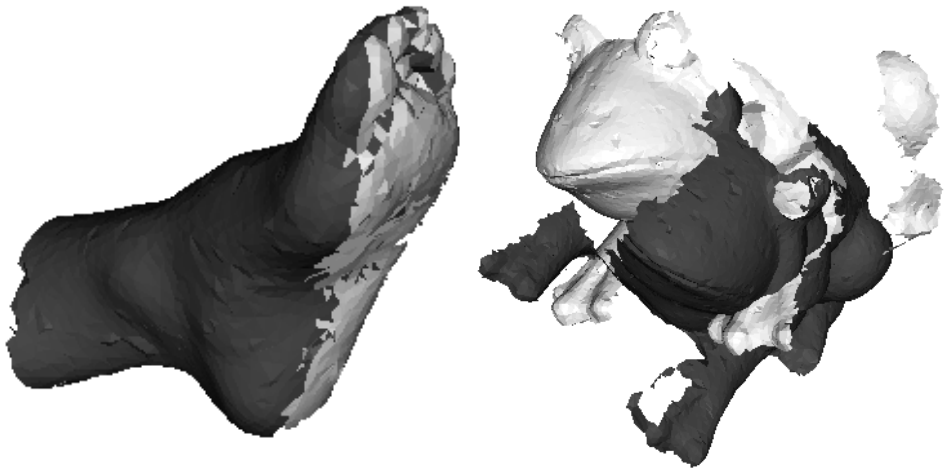
(c) method of Trucco (d) method of Chow

Figure 2.14: Pair-wise registration of two real data sets



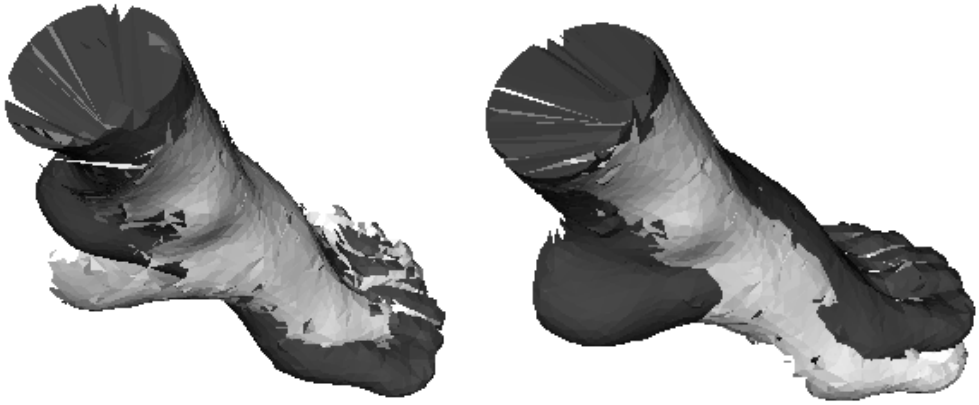
(a) method of Besl (b) method of Zinsser

Figure 2.15: Inaccurate pair-wise registration of two real range images in the presence of surfaces with few shape details: a) method of Besl; b) method of Zinsser.



(a) Synthetic data (b) Real data

Figure 2.16: Coarse Registration by RANSAC-Based



(a) Using 400 points (b) Using 1500 points

Figure 2.17: Influence of the number of points in the error with Spin Image

Chapter 3

A new multi-view approach based on cycles minimization

Pair-wise registration is a very important step in surface alignment. However, due to the propagation of the errors, this technique is not good enough when several acquisitions are done. In this chapter, a new approach is presented taking into account the advantages of multiview techniques and the use of cycles to reduce the number of views in the minimization.

3.1 Introduction

One-to-one alignment of views in a sequence causes a drift that is propagated throughout the sequence. Hence, some techniques have been proposed to reduce the propagating error benefiting from the existence of cycles and re-visited regions and considering the uncertainty in the alignment.

In order to minimize the propagating error, some authors improved their algorithms by adding a final step that aligns all the acquired views at the same time. This approach spreads one-to-one pair-wise registration errors throughout the sequence of views, being known as multi-view registration. Early approaches proposed the aggregation of subsequent views in a single metaview which is progressively enlarged each time another view

is registered [Chen and Medioni, 1991]. Here, the main constraint is the lack of flexibility to re-register views already merged in the metaview due to the greedy approach of the technique. In 1999, Pulli proposed an ICP relaxation method based on the previous metaview approach but considering all the potential alignments between views before proceeding with the multi-view registration. In addition, this method takes into account the information of all the overlapping areas and the already registered regions can be analyzed again for further transformations [Pulli, 1999]. Later on, N*ü*chter proposed a global relaxation method based on Pulli's approach with the main difference that no iterative pair-wise alignment is required. However the success of this method drastically depends on an accurate and previously known estimation of the pose [N*ü*chter *et al.*, 2004].

A different approach was proposed by Bergevin [Bergevin *et al.*, 1996], who presented a multi-view registration technique based on the graph theory: views are associated to nodes and transformations to edges. The authors consider all views as a whole and align all them simultaneously. The same idea was proposed later on by Silva [Silva *et al.*, 2006], Huber [Huber and Hebert, 2003] among others [Matabosch *et al.*, 2005*c*] [Krishnan *et al.*, 2005]. Besides, Masuda presented a multi-view registration algorithm based on the Matching Signed Distance Fields in which outliers are automatically removed obtaining a more robust method [Masuda, 2001]. Overall, multi-view techniques suffer two main drawbacks: a) the whole set of 3D views have to be previously acquired before the algorithm starts; b) an accurate estimation of the motion between views is needed as initial guesses to ensure convergence. Thus, multi-view techniques are not considered for on-line applications.

Few authors have faced the challenge of registering 3D views in a sequence while they are acquired avoiding or at least controlling error propagation. For instance, Sharp [Sharp *et al.*, 2004] proposed the registration of pairs of consecutive views until a cycle is found. Since only pair-wise registration is required, the method becomes very fast. Here, the interest is the way of distributing the motion (and hence the propagation error) among the different views. The author proposed to use weights directly related to the residue obtained in the pair-wise registration. Actually, this is not very accurate especially in the presence of misalignments between end views in the cycle as a matter of noise and object occlusions. In this case, the whole motion of such a cycle is also distributed to all the views increasing the error in the registration. Additionally, Lu works also with cycles, however the minimization is done when all views are acquired and the relations between them established [Lu and Milios, 1997].

In the last few years, a photogrammetric technique called Bundle Adjustment has increased popularity in the computer vision community and it is growing in interest in robotics. Bundle adjustment is the problem of refining a visual reconstruction to produce jointly optimal 3D structure and viewing parameters (camera pose and/or calibration) estimates [Triggs *et al.*, 2000]. Therefore, bundle adjustment techniques can be used in both robot/camera localization and 3D mapping in many fields such as camera calibration, robot navigation and scene reconstruction. Since bundle adjustment is a non-linear minimization problem, it is solved by means of iterative non-linear least squares or total squares methods such as Levenberg-Marquardt or M-estimator techniques [Fitzgibbon, 2001][Salvi *et al.*, 2007]. Although bundle adjustment is commonly classified as a multiview technique, some authors have used it in consecutive pairwise alignment as a technique to reduce error propagation [Pollefeys *et al.*, 2000].

In summary, we conclude that analytic methods based on the metaview approaches present good results when initial guesses are accurate and the surface to be registered does not have a large scale. Otherwise, the method suffers a large propagation error producing drift and misalignments and its greedy approach usually falls in local minima. The use of methods based on graphs has the advantage of minimizing the error in all the views simultaneously but they usually require a previous pairwise registration step, which accuracy can be determinant in the global minimization process. Besides, closing the loop strategies provide trustworthy constraints for error minimization but require a huge amount of memory and usually involve a high computational cost. Bundle adjustment techniques provide good results in the presence of outliers, but need a good enough initial guess and it is hardly used in large robot missions or large scale objects.

All these pros and cons of the existing methods have been considered to present a new surface registration technique which is presented and discussed in this section.

3.2 Review of Sharp's method

Due to the fact that our proposal can be considered a robust variant of the method of Sharp [Sharp *et al.*, 2004], this section briefly summarizes Sharp's method with the aim of illustrating the drawbacks and the points to improve.

Sharp's method is based on the minimization of a set of views conforming a cycle to

decrease the effects of the drift. Initially the author is focused on a single cycle, then, the algorithm is modified to be adapted to the presence of multiple cycles.

The main idea of cycle minimization is to constrain the global motion in a cycle to be null, i.e. the product of motion matrices in a cycle is constraint to the identity matrix. The discrepancy between the overall motion and the identity matrix is defined as the motion error. The author distributes this motion error through out all the views of the cycle, closing the cycle and obtaining quite good results. The author decouples rotation and translation in order to distribute the errors properly. Rotation matrix is transformed to axis-angle representation, so that the angle of rotation is weightily distributed. Lagrange multipliers are used to distribute the translation errors. Deferring from other multi-view proposals, point correspondences are not used in cycle minimization leading to a reduced computing time, though robustness is not guaranteed.

Besides, in Sharp's method, relationships among views conforming cycles are given, though usually cycle detection is a crucial step.

In next section, we propose a cycle minimization technique, which improves Sharp's one since we introduce a cycle detection module and we consider point correspondences in the minimization process to increase robustness. Finally, we also consider the overlapping among views and not only consecutive views in the minimization process, leading to more accurate results.

3.3 New cycle minimization strategy

This section describes the proposing method for continuously registering a sequence of 3D views while they are acquired. The method first aligns the consecutive views by means of point-to-plane pair-wise registration. When a cycle is detected, a multi-view technique is applied only in the views conforming to the cycle, leading to fast and accurate results and preserving the on-line registration for many and varied applications (see Fig. 3.1).

3.3.1 Pair-wise Registration

Pair-wise registration is divided into a first, coarse registration to estimate an initial alignment followed by a fine registration computed by means of minimization techniques

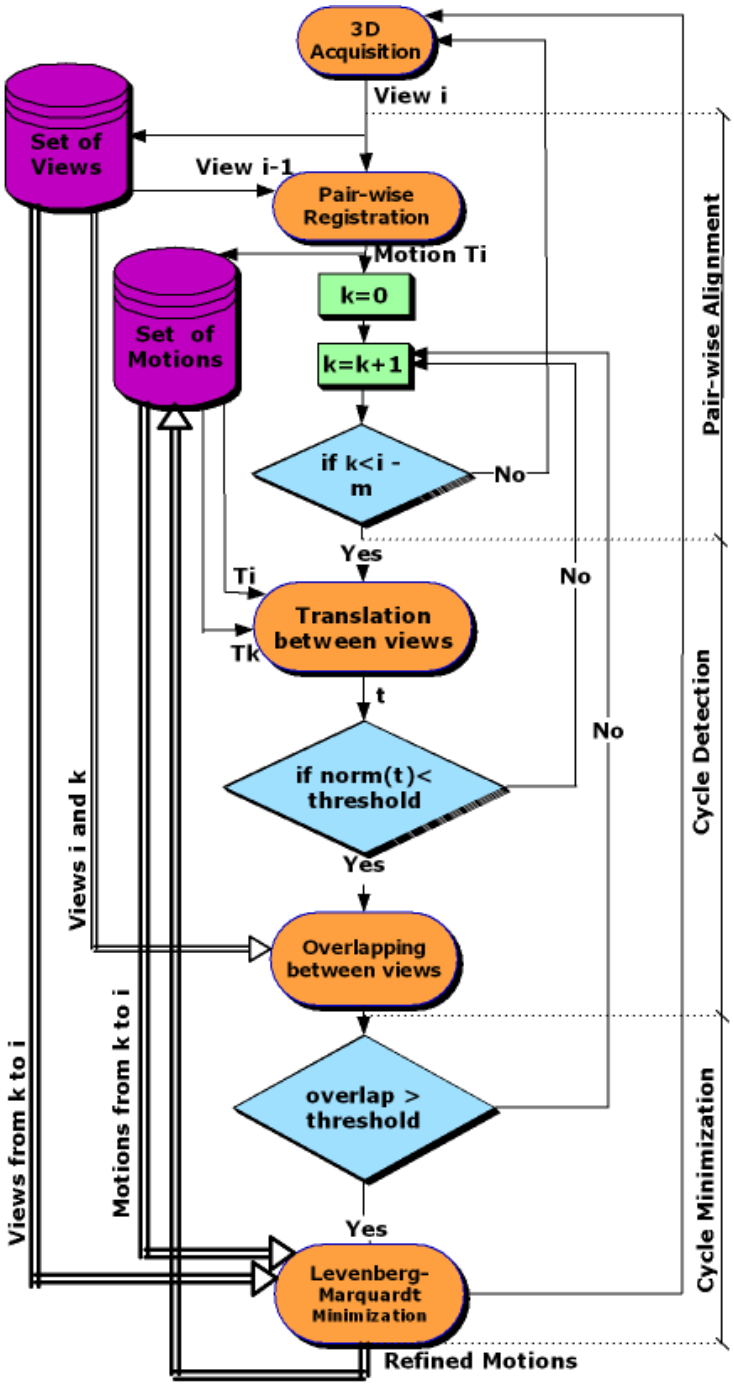


Figure 3.1: Flow diagram of the cycle minimization approach

as explained in the previous section. In our case, as views are acquired consecutively and a slight movement between views is assumed, we initialize fine registration considering

motionless views, avoiding the expensive computation required to compute initial guesses and preserving a high accuracy, as demonstrated in the following paragraphs and shown in the experimental results (see Chapter 5).

Point-to-plane has been chosen as the most suitable fine registration technique as discussed in the previous section. The technique we propose is based on the fast variant proposed by Park [Park and Subbarao, 2003] from the original point-to-plane registration proposed by Chen [Chen and Medioni, 1991], although some modifications have been implemented to increase accuracy.

First, we remove the non-overlapping area of the present view before this view is registered with the former. In theory, this area is unknown because the movement is also unknown. However, as the views are taken in a sequence with slight movements between them, we can assume that points located in the center of the view are good candidates for the matching. Besides, most of the points located in the boundary of the surface might be hardly matched. In consequence, the boundary area of the present view is not considered in the fine registration step. In fact, the bounding area coincides with the boundary in the image formed by projecting the present view to the XY plane of the camera (orthogonal to the focal axis), so the selection of points to remove becomes very easy. In the image plane, the bounding box is computed. A rectangle whose dimensions are 80% of the bounding box is centered to the image projection and all points out of this rectangle are not taken into account in the registration step.

Second, only a sample of the remaining points of the present view is preserved for the fine registration. There are several types of sampling: uniform sampling [Masuda, 2001] [Turk and Levoy, 1996], random sampling [Masuda *et al.*, 1996], and normal sampling [Rusinkiewicz and Levoy, 2001], among others. Although sampling is normally used to speed up the algorithm by selecting a reduced set of points, sampling can be also used to increase accuracy by selecting also the most appropriate points. Note that, in smooth surfaces with even shape registration becomes difficult. In this situation, only a small percentage of points give useful shape information. For instance, consider a flat surface with two perpendicular cuts. If all the points are considered in the registration, results are not accurate because of the low influence of points in cuts with respect to the rest of the points. However, if the registration is done with a high percentage of points on the uneven area, accuracy increases. More details about sampling are presented in section 2.4

The goal of our sampling is to select the most representative points to increase the

quality of the registration, so that a normal sampling is used. Hence, all points are first transformed to a 2D space defined by α and β as follows:

$$\begin{aligned} \alpha &= atan2(n_x, \sqrt{n_z^2 + n_y^2}) \\ \beta &= atan2(n_y, n_z) \end{aligned} \tag{3.1}$$

where α and β are the coordinates in the normal space, and n_x, n_y and n_z are the three components of the normal vector of each point. Then, every point is placed in a 2D grid. Finally only one point from every grid cell is randomly selected, so that a single point is chosen among all points with similar normal vectors. These selected points actually conform to the reduced set of points used to register the present surface.

As stated before, the fine registration technique we propose is based on the fast variant proposed by Park [Park and Subbarao, 2003] from the original point-to-plane registration proposed by Chen [Chen and Medioni, 1991]. Here we use a recursive method to compute the intersection between lines and surfaces which is actually the main difficulty of the method. Hence, initially the selected points of the previous view are projected orthographically onto the XY plane of the camera. A grid composed of $50x50$ square cells is scaled so that it contains the projection of all points. Second, a point p_0 of the current view is projected to such a grid, in whose cell we search for the closest point obtaining the point q_{p_0} in the previous surface. The projection of point q_{p_0} to the normal vector of p_0 defines a new point p_1, which is actually an approximation of the intersection. This approximation is refined recursively by projecting new points p_i until $norm(p_i - q_{p_i})$ is smaller than a threshold (see Fig. 3.2). Finally, the process is repeated for all the points conforming to the current view and a set of correspondences is obtained.

Once correspondences are established, minimization is applied to compute the motion between both surfaces (the previous and the current) as defined by Eq. 3.2.

$$f = \frac{1}{N_p} \sum_{i=1}^{N_p} \|m_i - Rp_i - t\|^2 \tag{3.2}$$

where N_p is the number of correspondences; m_i is the set of points selected in the former view that have a correspondence in the present view; p_i are the correspondences of m_i in the present view; and R and t are the rotation matrix and the translation vector that align both views.

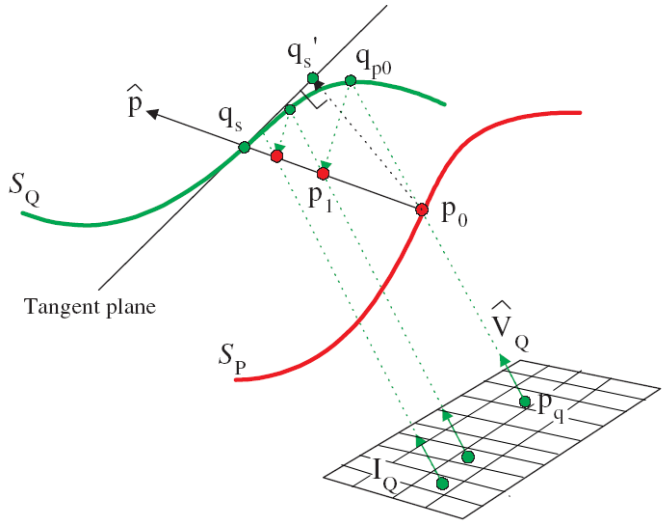


Figure 3.2: Strategy used to compute the intersection between the tangent plane and the surface S_q along the orthogonal vector $\hat{p}$. See Park [Park and Subbarao, 2003] for a extended review.

Eq. 3.2 is minimized by means of quaternions [Besl and McKay, 1992] so that R and t are refined iteratively. The algorithm stops when the mean of the square errors (distances between correspondences) is smaller than a given threshold.

Note that the views are registered consecutively, so that every registered view is referenced with respect to the first by means of the product of all the consecutive Euclidean motions defined by the sequence of views. Hence, registration inaccuracies are propagated through the sequence. In the following sections, we aim to minimize the propagation error by detecting cycles and minimizing the views conforming the cycle all together.

3.3.2 Cycle detection

Now the interest is to detect every time the scanner re-visits the same object surface obtaining cycles of views that are used to reduce the propagation error significantly.

Note that once any two views are registered, the Euclidean transformation between them is known and a link established. These links form paths through the views in which the motion of the scanner can be estimated from the product of the consecutive Euclidean transformations. Hence, the translation vector of such movement is considered, so that if this vector is close to null and the views are not neighbors, a potential cycle is considered.

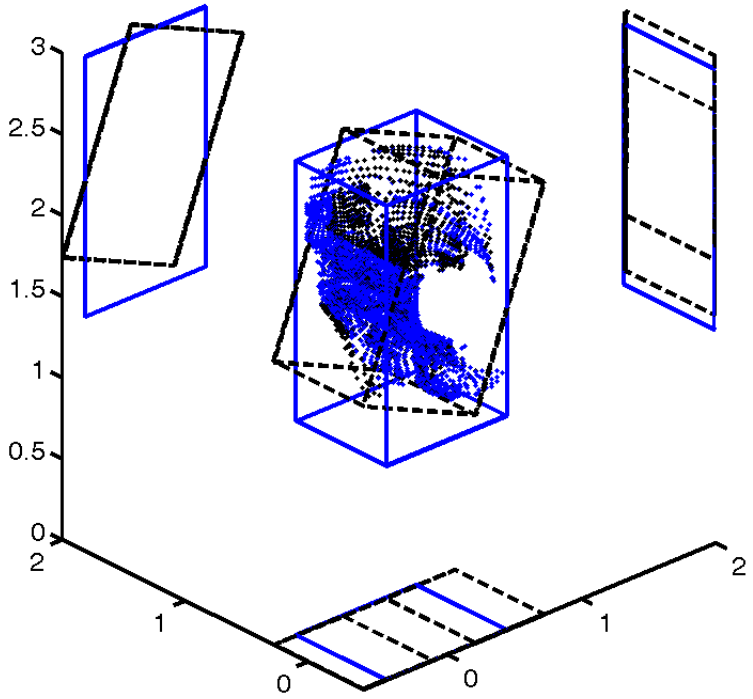


Figure 3.3: Example of the projection of the bounding boxes of two different views in the XY, XZ and YZ planes. The grey area represents the overlapping.

However, a sequence of views with an overall null translation does not always conform a cycle, especially when rotation is relevant. Hence, a cycle is considered if both end views also share a common surface that is a significant overlapping area.

The accurate computation of the percentage of overlapping would imply the fine registration between both end views and the computation of corresponding points. In order to avoid this expensive step, a fast technique is proposed based on the overlapping of bounding boxes, which is just an approximation of the convex hull of both surfaces, but accurate enough to detect cycles.

The bounding box of a given surface is defined as the minimum parallelepiped that contains all the points of the surface. The intersection of 3D bounding boxes is complex so that it is alleviated by projecting such boxes to the planes XY, XZ and YZ (see Fig. 3.3), defining two 2D bounding boxes in every plane and thus computing three overlapping areas. If the maximum of the three overlapping areas exceeds a given threshold of the total area and the distance between both bounding box centers is small enough, a cycle is considered.

Finally, we compute the motion matrix that closes the cycle, i.e. the matrix that aligns both end views. If this matrix is close to the identity, a cycle is considered. Rotation and translation errors are independently analyzed, so that rotation error is the discrepancy between such matrix and the identity, and translation error is constraint to be smaller than a threshold weighted considering the scale of the object and the number of views in the cycle.

The reason to choose the maximum overlapping value among the three planes instead of the product of overlapping values is in virtue of preserving the detection of potential cycles in the presence of almost flat surfaces. In this case, the bounding boxes in some of the three planes are usually not relevant.

3.3.3 Cycle minimization

Cycle minimization consists of a simultaneous minimization of all the correspondences between points of all the views that conform the cycle. In cycle minimization we assume that the overall motion in the cycle is null and hence the position of both end views coincides. This is actually impossible and that is the reason why a virtual view is added between both end views. This virtual view is nothing other than the first view of the cycle registered to the last one. We can assume that the overall motion in the cycle is null which means that the motion between both end views must be zero.

The significant points for every view are used to search for correspondences among all the other views in the cycle. A threshold in the relative motion between views is used to ensure a significant overlapping area between views and hence many point correspondences. Obviously, this decision leads to a quite fast method without losing robustness. Otherwise, the algorithm wasted a lot of time searching for correspondences where it was known they are either not available or not significant.

Finally, a Levenberg-Marquardt minimization is applied to determine a more accurate registration among views in the cycle. The minimizing parameters are the rotation matrices (represented as quaternion vectors) and translation vectors of the Euclidean transformations between consecutive views. The minimizing function is the sum of distances between point correspondences which is constrained to be zero, as shown in the following equation:

$$min\{\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\sum_{k=1}^{N_p}(P_i(k) - T_j^i \times P_j(k) + T_i^j \times P_i(k) - P_j(k))\} \tag{3.3}$$

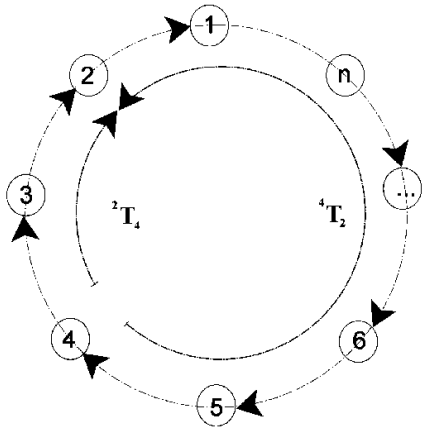


Figure 3.4: Difference between matrices ${}^iT_j$ and ${}^jT_i$

where $P_i(k)$ and $P_j(k)$ are the points that configure the k correspondence between views i and j; N_p is the number of points correspondences; N is the number of views; and T_i^j and T_j^i are the Euclidean motions (see Fig. 3.4) that transform points from i to j and from j to i, respectively, computed as follows,

$$T_j^i = \prod_{k=i+1}^{j} T_k^{k-1} \tag{3.4}$$

and

$$T_i^j = (\prod_{k=j}^{N-1} T_{k+1}^k) T_1^N (\prod_{k=2}^{i} T_k^{k-1}) \tag{3.5}$$

where $j > i$;

This minimization is done forcing the close-loop constraint to ensure that the product of all matrices of the cycles is the null motion. This constraint is expressed as follows:

$$\varepsilon_{cr} = \varepsilon_R + s_f \varepsilon_T \tag{3.6}$$

where ε_R is rotation constraint, ε_T is the translation constraint, and s_f is the scale factor that express the translation in the same range of the rotation parameters.

$$\varepsilon_R = sum(abs(R_{accum} - I_{3x3})) \tag{3.7}$$

where R_{accum} is the product of all partial rotation matrices, and I_{3x3} is the identity matrix.

The translation constraint is computed as the norm of the translation vector obtained as a multiplication of all partial motions. The rotation constraint is computed as:

$$\varepsilon_T = norm(t) \tag{3.8}$$

where t is the translation vector between initial and final view of the cycle.

The whole process leads to quite accurate results, but if they are not good enough, they can be refined and refined just selecting for new significant point correspondences at the end of every refinement.

3.4 Fast approach

Although our approach is fast with respect to the traditional multi-view approaches, if cycle contain lots of views, lot of time can be used to establish correspondences inside the cycle. In this section, a fast approach is presented based on the minimization of only the correspondences between neighbor views.

The main idea is the same as the generic method, however, the exhaustive computation of correspondences is avoided. During the pair-wise registration, correspondences are stored. When a cycle is found, motion parameters are minimized taking into the account only the consecutive views whose correspondences are stored. Modifying eq.3.3, a reduced one is obtained.

$$min\{\sum_{i=1}^{N-1}\sum_{k=1}^{N_p}(P_i(k) - T_{i+1}^{i} \times P_{i+1}(k) + T_i^{j+1} \times P_i(k) - P_{i+1}(k))\} \tag{3.9}$$

Despite only consecutive views are taking into account, the cost function takes into

account all the views on the cycle and the constraints of it, obtaining also good results decreasing the computational cost time. This function is also minimized under the close-loop constraint (see eq. 3.6).

3.5 Conclusions

The introduction overview at the beginning of this chapter shows the drawbacks of the pair-wise registration techniques. Registration errors are propagated for each register view, avoiding accurate reconstructions. To overcome this problem, multiview techniques are used. This group of techniques simultaneously minimizes all the views to get a global minimum instead of a set of local minima.

Although the accuracy of them, several drawbacks are presented. In general, multi-view techniques are constrained by the following drawbacks: a) all the views must be first acquired before the aligning algorithm starts leading to off-line applications; b) guesses to roughly align the views are needed to initialize the algorithm so that an expensive coarse registration technique is needed; and c) matching is searched among all the views without considering neighborhood which is inefficient and computing intensive, especially in large data sets. Besides, multi-view techniques are not suitable for registering views that form sequences and loops because of the error propagation problem.

There are several techniques to register a set of views, though most of them are based on the multi-view approach. This section presents a new multi-view registration technique which includes cycle minimization and it is updated in the measure that new views are acquired. Although the technique can be applied in short sequences of views, it is designed to deal with large data sets and with the presence of multiple cycles. First, a fast point-to-plane with normal space sampling and non-overlapping area removal is applied between consecutive views to obtain an accurate alignment. Second, in the measure that new views are acquired, the method searches for cycles considering neighborhood and overlapping percentage. Finally, once a cycle is detected it is minimized by means of a Levenberg-Marquardt approach, so that the system always ensures the most accurate global registration. Deferring from other authors, our approach presents a robust solution because the global error on the cycle is minimized.

Additionally, a fast approach is also presented. This approach avoids the search of

correspondences inside the loop, because only correspondences of consecutive views are considered. As these correspondences are established in the pair-wise registration computed a priori, Levenberg-Marquardt minimization can be directly computed when a cycle is found.

Chapter 4

A 3D Hand-held sensor for large surface reconstruction

In this chapter we present a prototype of a 3D hand-held sensor. The proposed sensor, based on the laser triangulation principle, is able to acquire surfaces in one single shot. This fact avoids misalignment due to object motions or sensor vibrations. Thanks to its dimensions and weight, this sensor can be used in some applications where most part of commercial scanners present some problems due to big sizes or enough degrees of freedom. Additionally, the sensor prototyped is a cheap solution compare to commercial ones.

4.1 Introduction

Most part of commercial scanners are considerably big, some of them are coupled to a big structure, so that the acquisition of objects is limited to a little objects than can be moved near the range finder. Examples of this situation are the range finder coupled to a translation mechanism (XC50 Cross Scanner). A similar problem is presented by systems based on the rotation tables. The size of the table and the weight of the object limit the number of applications. Furthermore, in some situations, it is not possible to move the scene to acquire due to the fragility of the object or the object is attached to a fixed structure.

The use of portable range finders increase the number of applications. Examples of these sensors are the Minolta Vivid 700 laser scanner, PS-Series Portable 3D Laser Scanner, etc. These sensors permit the acquisition *in-situ*. The scanner can be placed where the scene is and acquired in a few minutes. Although these scanners are portable, their weight is important and they need additional structures to support it. The scanner can be transported but it can not be manually moved to get the best orientation to acquire the surface. In case of big objects, occlusions can be a problem difficult to solve. The "*Digital Michelangelo Project*" realized by the Stanford University gets the 3D model of the statue of Michelangelo [Levoy *et al.*, 2000]. A mechanical structure of more than 7 meters was required to scan the object. It is obvious that the accuracy of the mechanical system was fundamental to get a good reconstruction. In addition, lot of time is required to install all the set up.

In recent years, some hand sensors have been developed, K-Scan(Metris), G-Scan RX2 (Romer), Leica T-Scan (Geosystems), FastScan Cobra among others. However, most of these scanners are coupled to a mechanical system to get the position of the sensor in each scan. Therefore, the mobility of the scanner is limited to the degrees of freedom of the mechanics. Otherwise, the pose information is computed by using magnetic tracker, which are sensible in some kind of environments.

In this chapter, a cheap prototype of a 3D hand-held sensor based on a laser emitter and a USB-camera is presented. Deferring from the others, only visual information is used to determine the position and orientation of the sensor. First of all, technique information is detailed. Secondly, the laser segmentation is described. As a set of laser slits is emitted, each detected peak must be labeled with the correct plane of light. This problem is known as stripe indexing and is detailed in section 4.4. Computation of depth information is detailed section 4.5. Finally, some accuracy results and reconstruction examples are presented in section 4.6.

4.2 Set up

The prototyped sensor is based on laser projection. Deferring from most part of commercial sensors, more than one plane is simultaneously projected. A set-up consisting of an on-the-shelf CCD camera, a 6 mm lens, a 635 nm LASIRIS laser emitter and an optical lens which spreads the laser beam into 19 planes has been arranged conforming

the imaging system. This system acquires images of 1200×900 pixels. Both camera and laser are located on a portable platform where their optical axis form an angle of 60^o and the distance between them is approximately 20cm. With this configuration, surface can be ideally acquired between 10 and 30 cm to the axis of the camera.

An illustration of the set-up is shown in Fig 4.1 The imaging system has been calibrated by using the complete quadrangle approach and a perspective calibrating technique published in a previous paper [Matabosch *et al.*, 2006*b*].

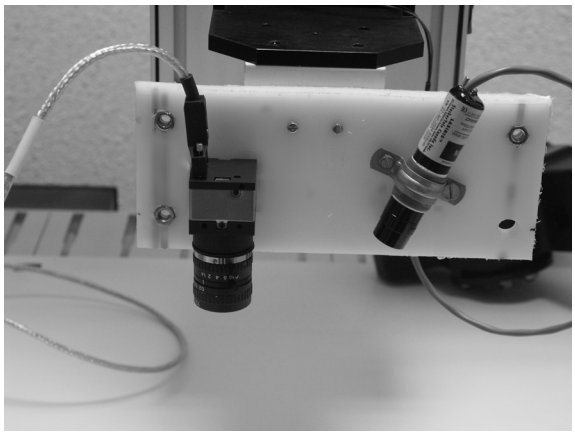


Figure 4.1: 3D hand-help prototype used in the experiments

The process of calibration consists of finding a relation between 3D points on the measuring surfaces with the projection of these points in the acquired image. This relation can be linearly approximated by the following equation:

$$\begin{bmatrix} sX \\ sY \\ sZ \\ s \end{bmatrix} = {}^{W}T_{L} \cdot \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} \tag{4.1}$$

where u and v are the pixel coordinates, X, Y and Z are the 3D points, s is a unknown scale factor and ${}^{W}T_{L}$ is the calibration matrix.

Once ${}^{W}T_{L}$ is known, 2D points in the image frame can be directly transformed to 3D points in the world reference frame. Obviously, the parameters t_{ij} of matrix ${}^{W}T_{L}$ should be estimated as precisely as possible in order to maximize the accuracy in the reconstruction. According to equation 4.1, the expressions for sX, sY, sZ and s are obtained and shown

in equation 4.2

$$\begin{aligned} sX &= t_{11} \cdot u + t_{12} \cdot v + t_{13} \\ sY &= t_{21} \cdot u + t_{22} \cdot v + t_{23} \\ sZ &= t_{31} \cdot u + t_{32} \cdot v + t_{33} \\ s &= t_{41} \cdot u + t_{42} \cdot v + t_{43} \end{aligned} \tag{4.2}$$

Arranging the terms and grouping, a homogeneous system of three equations with 12 unknowns (t_{11} to t_{43}) is obtained as shown in equation 4.3

$$\begin{aligned} t_{11} \cdot u + t_{12} \cdot v + t_{13} - t_{41} \cdot u \cdot X - t_{42} \cdot v \cdot X - t_{43} \cdot X = 0 \\ t_{21} \cdot u + t_{22} \cdot v + t_{23} - t_{41} \cdot u \cdot Y - t_{42} \cdot v \cdot Y - t_{43} \cdot Y = 0 \\ t_{31} \cdot u + t_{32} \cdot v + t_{33} - t_{41} \cdot u \cdot Z - t_{42} \cdot v \cdot Z - t_{43} \cdot Z = 0 \end{aligned} \tag{4.3}$$

If several correspondences between 3D points and 2D pixels are known, calibration parameters can be estimated.

In order to search for correspondences, the complete quadrangle is used [Forest, 2004]. The original method has been adapted to calibrate the set of 19 planes obtaining the 19 transformation matrices which describes the geometry of the sensor. For each laser plane, the following steps are processed:

- Detection of the points of the laser profile in the image plane,
- Find the correspondences between points in the image plane and 3D points in the calibrating plane,
- and Compute the T matrix using the correspondences given by the previous step.

The description of the laser profile is detailed in the following section 4.3.

4.2.1 Correspondences between points in the image and 3D points

The methodology is based on the *complete quadrangle* [Chen and Kak, 1987]. The principle of this method is the cross-ratio between the complete quadrangle and the acquired image

of this quadrangle (see fig. 4.2a)).

$$\frac{\overline{A'P'_A}}{\overline{A'G'}} = \frac{\overline{AP_A}}{\overline{AG}} \tag{4.4}$$

As A,B are known 3D points, and A', B' and P'_A can be found analyzing the acquired image, P_A can be determined by the cross-ratio principle. The same principle is applied with point P_B. If the quadrangle is moved along the Z-axis, a set of 2D-3D correspondences can be found for each Z position. Using this set of correspondences, eq. 4.1 can be solved determining the transformation matrix.

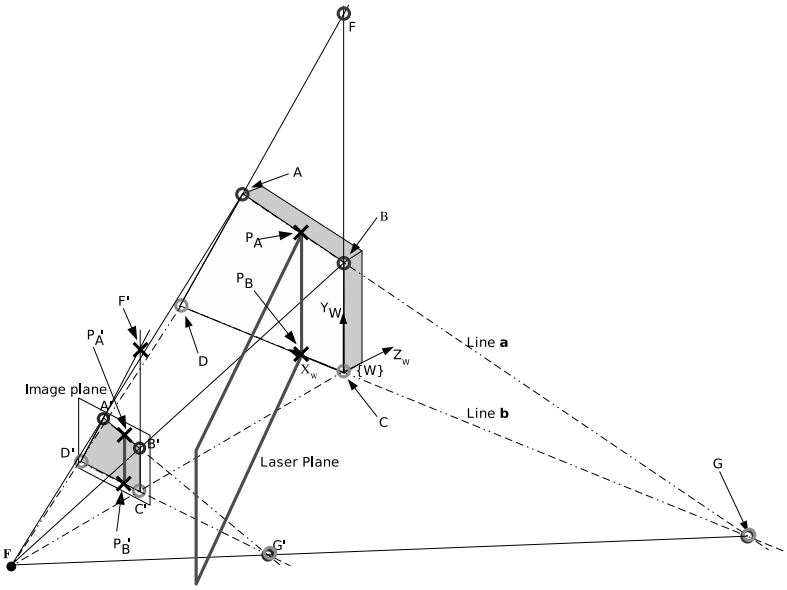
In general, only two points are used for every plane position. Note that calibration accuracy is related directly to the number of correspondences used. In order to improve the accuracy, a set of points along the laser stripe are selected. To do this, arbitrary points (P'_L) are selected in the quadrangle (see fig. 4.2b)). The pencil of lines that joint these points with point G' are created. The intersection of these lines with the laser stripe gives us the auxiliary points of the calibration. The process to determine the 3D correspondence points is the same as in the first situation. More details are presented in [Forest, 2004].

4.2.2 Compute T matrix using known correspondences

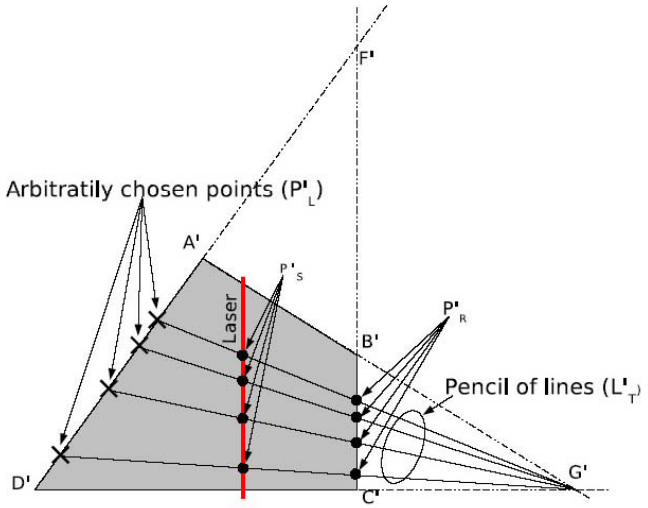
Now the transformation matrix can be obtained by minimizing eq. 4.5 which has been easily obtained arranging eq. 4.3, where t_{ij}'s are the parameters of the ${}^{W}T_L$ matrix, u_i and v_i are the pixel coordinates and X_i, Y_i and Z_i are the coordinates of the 3D position. The solution is obtained from the computation of the vector θ that minimizes equation $A \cdot \theta = 0$. A good estimation using Total Least Square technique is computed from the eigenvector corresponding to the smallest eigenvalue of matrix $A^T \cdot A$.

4.3 Laser segmentation

The laser segmentation consists in the extract the laser points from the image. As a laser filter is coupled to the camera, the observed scene is a black image with white stripes representing the observed laser profiles. The goal of this step is to accurately determine the position of all laser points observed by the camera. As the precision is fundamental



(a) Cross-ratio and the complete quadrangle used to determine 2D-3D correspondences



(b) Generation of other points to increment the quality in the calibration

Figure 4.2: Calibration process

$$\begin{bmatrix} \vdots & & \vdots & & \vdots & & \vdots & & \vdots \\ u_i & v_i & 1 & 0 & 0 & 0 & 0 & 0 & -u_i \cdot X_i & -v_i \cdot X_i & -X_i \\ 0 & 0 & 0 & u_i & v_i & 1 & 0 & 0 & -u_i \cdot Y_i & -v_i \cdot Y_i & -Y_i \\ 0 & 0 & 0 & 0 & 0 & 0 & u_i & v_i & 1 & -u_i \cdot Z_i & -v_i \cdot Z_i & -Z_i \\ \vdots & & \vdots & & \vdots & & \vdots & & \vdots \end{bmatrix} \cdot \begin{bmatrix} t_{11} \\ t_{12} \\ t_{13} \\ t_{21} \\ t_{22} \\ t_{23} \\ t_{31} \\ t_{32} \\ t_{33} \\ t_{41} \\ t_{42} \\ t_{43} \end{bmatrix} = \begin{bmatrix} \vdots \\ 0 \\ 0 \\ 0 \\ \vdots \end{bmatrix} \quad (4.5)$$

to get a good 3D reconstruction, subpixel accuracy is required.

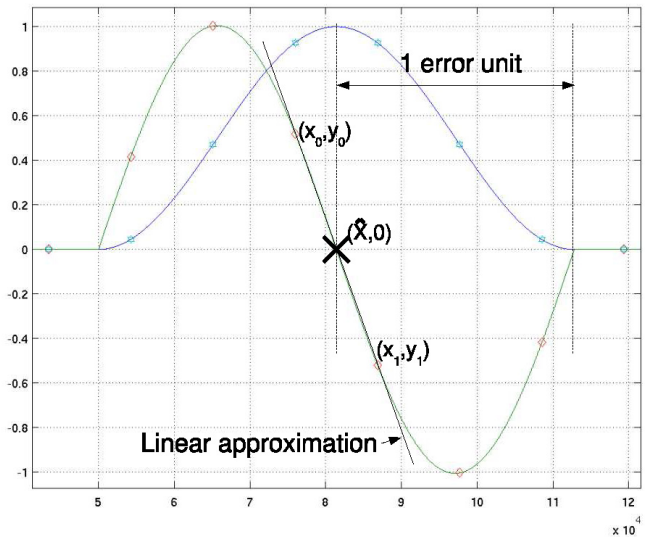


Figure 4.3: Use of zero crossing in the first derivative to determine laser peak position

The laser segmentation used is based on the laser peak detector [Forest *et al.*, 2004]. The principe of the method is to search the position of the maximum as the position where the first derivative is zero. This point can be interpolated between the smallest positive element and the greatest negative element (see Fig. 4.3) . However, this method is designed for one peak detection, in our application, 19 peaks per row might be detected. Hence, some modifications must be done to adapt the method to this problem.

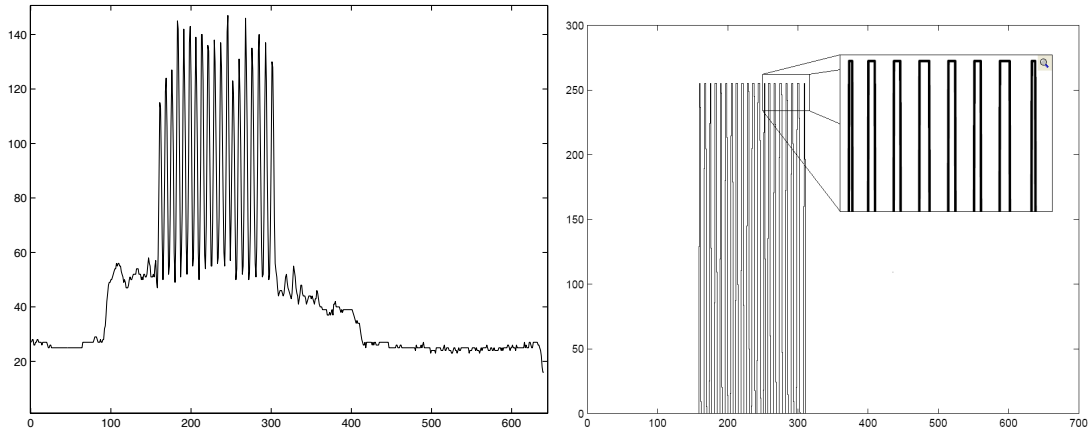


Figure 4.4: Laser Peak detection problem. a) Intensity profile of a row of the image; b) Binarized second order derivative

First of all, an approximate location of the candidate peaks is determined. Although maximum values can be taking as an initial approximation of the laser position, the second order derivative is computed because is less sensible to noise. The second derivative is binarized, 0 or 1 depending on a given threshold [Pages *et al.*, 2005]. Then, each center of all non-null regions is selected as a peak approximation (see Fig. 4.4).

A local first derivative is computed by applying the convolution with vector $[\ -1 \quad -1 \quad -1 \quad 0 \quad 1 \quad 1 \quad 1\ ]$. The zero crossing in the derivative gives us the sub-pixel position of the laser profile. This cross-zero is interpolated between the smallest positive value and the greatest negative value (see Fig. 4.3).

4.4 Stripe Indexing

Stripe Indexing problem is a new topic where only few researchers are working. Most part of applications that work with multi-stripe patterns are based on coded structured light, where labelling is determined by means of codification. Besides, in uncoded structured light there are no additional information to help us in the indexing problem as all stripes are identical. Therefore, this labelling is very complex, and some assumptions are needed. First of all, we assume that any laser plane can not appear twice in a row. This assumption is true under some circumstances as prove Robinson et al.[Robinson *et al.*, 2004]. Secondly, local smoothness assumption is needed in order to suppose that the order of the stripes can not be modified. Therefore, it is possible to find less stripes than laser planes projected,

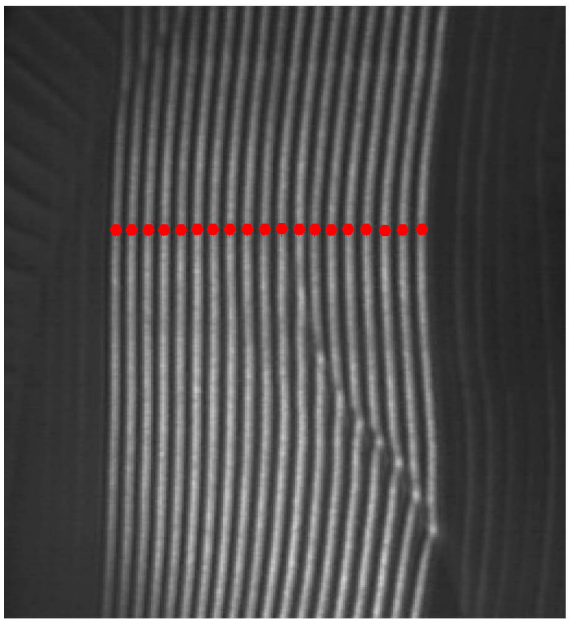


Figure 4.5: Example of a detected laser peaks on a row

but it is impossible to find them in a disordered distribution. The proposed algorithm is a variation of the one proposed by Robinson et al.[Robinson *et al.*, 2004], adapted in the case where we know the number of laser planes projected.

First of all, an auxiliary image is constructed where the value of a pixel is -1 if there is a peak in this position in the real image, otherwise the value is zero.

In order to remove noise from the image, a filter is applied to the auxiliary image as follows:

$$I(a,b) = \begin{cases} -1 & \text{if } \sum_{i=a-4}^{a+4} \sum_{j=b-4}^{b+4} I(i,j) < -1 \\ 0 & \text{otherwise} \end{cases} \tag{4.6}$$

where I represents the auxiliary image, and a and b the rows and columns of such image.

After this noise removal, all pixels labelled as -1 represents a projection of a laser plane in the image. Therefore, if a row with 19 labelled peaks is found, and considering the assumptions commented before, we can directly associate peaks i with label i. Note that when we talk about peaks, we are referring to all non-zero pixels in a row. On the other hand, if any row of the image is not composed of 19 peaks, the algorithm can not

be used, and this image can not be indexed. However, this is a problem not very common in our application.

When all rows with 19 stripes are identified (defined as completed rows), we need to select one of them to begin the algorithm. The selected row is the one that is surrounded by more completed rows. After this initial preprocessing the algorithm goes on as follows:

Table 4.1: Stripe Indexing Constraints

- Any i-th peak can be classified as a plane j, where $j < i$
- Any $i-th$ peak in a row of n peaks can be classified as a plane j, where $n - i > 19 - j$
- Peak i can only be classified belonging to a laser plane j if all peaks k(from 1 to i-1) are classified as a plane l, where $l < j$
- Peak i in a row of n peaks can only be classified belonging to a laser plane j if all peaks k(from i-1 to n) are classified as a plane l, where $l > j$

Table 4.2: Stripe Indexing Rules

- Peak i is classified as stripe i when peak j is classified as stripe j and $j > i$
- Peak i, in a row of n peaks, is classified as stripe $k = 19 + i - n$ if peak j is classified as stripe l, $n - j = 19 - l$ and $i > j$
- Peak i is classified as stripe $l = m + i - j$ if peaks j and k are classified as stripes m and n, respectively and $n - m = k - j$

If the actual row contains 19 peaks, for each one, the tracking algorithm is applied. The goal of this step is to label all peaks of the same profile. If a peak is detected in the precedent row closer than 1.5 pixels, this peak is labelled with the same label as the initial peak. The process stops when no more peaks are found, or when the constraints (Shown in Tab. 4.1) introduced are not satisfied. This constraints helps us to detect the change of laser plane, when the continuity in the stripe is visually correct. The process is repeated from top to down.

After all peaks in the actual row are analyzed, another row is selected. This row is randomly selected in all rows with non-labelled pixels. Using a random function to select

the point, we avoid to enter in a loop by selecting a row that have enough information to label any pixel of it.

Using this new actual row, the process goes exactly on if this row contains 19 peaks. In our cases, some rules are used in order to tend to index some peaks of the row. This rules are presented in Tab. 4.2. If any peak can be labelled, another view is taking into account, otherwise, the tracking algorithm is applied for all labelled pixels. An example of the algorithm is shown in Fig. 4.7.

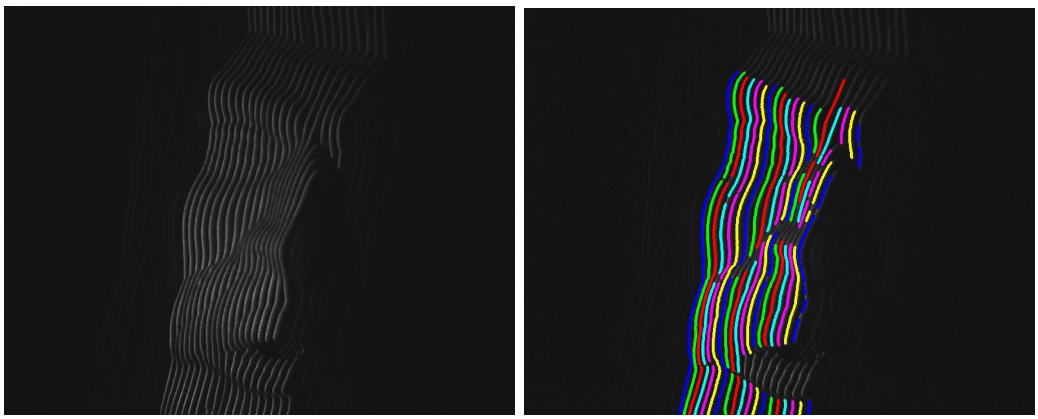


Figure 4.6: Stripe Indexing Problem. a) Original image; b) Labelled stripes onto the image

The algorithm stops when all non-zero pixels are labelled or the number of iterations reaches a fixed value. One labelled image is shown in Fig. 4.6. Although some stripes are not labelled, the most important is that labelled ones are correct.

4.5 3D computation

When laser profiles are labelled, depth information can be directly obtained by multiply each labelled point by the corresponding calibration matrix of its profile:

$$\begin{bmatrix} sX \\ sY \\ sZ \\ s \end{bmatrix} = {}^{W}T_{C}^{i} \cdot \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} \tag{4.7}$$

where X, Y and Z are the 3D coordinates of the reconstructed point, s is a scale factor,

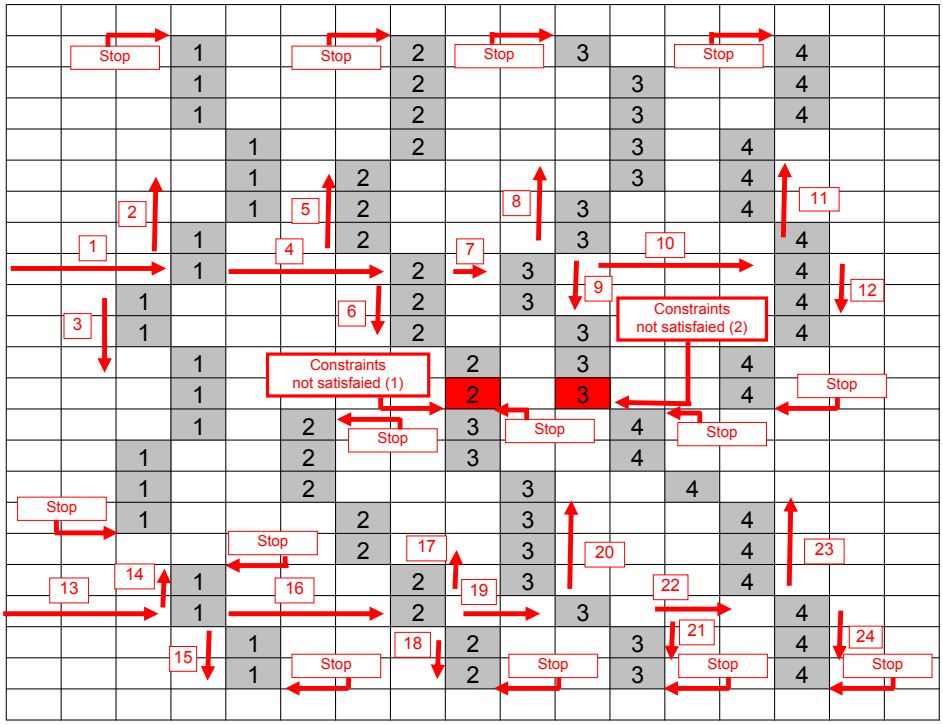


Figure 4.7: Example of stripe indexing algorithm

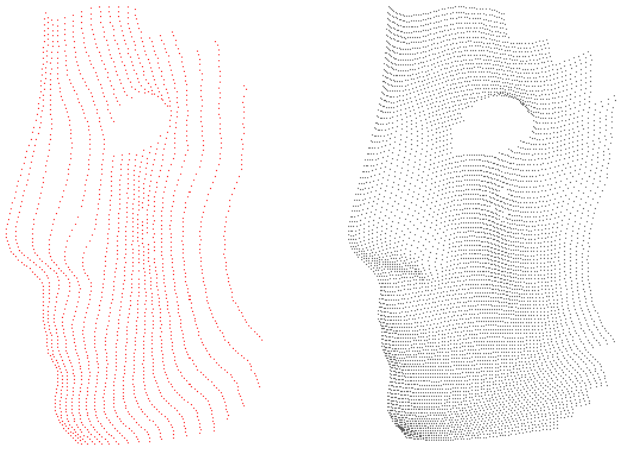
u and v are the pixel coordinates of the laser onto the image and ${}^{W}T_{C}^{i}$ is the calibration matrix. The index i corresponds to the label of the actual pixel.

For each laser plane, its calibration matrix is found by calibrating the relation between camera and plane [Forest, 2004]. Now, a cloud of points is obtained, however, as only 19 planes are projected, the cloud of points is not so much uniform and can not be considered as a continuous surface. Therefore, convergence problems can appear in the registration. To avoid this problem, and also to obtain a better visualization of the surface, points are interpolated between consecutive curves. In the general case, the interpolation in a 3D unorganized cloud of points is a complex problem. However, we simplify the problem by organizing the points, taking into account that rows are almost perpendicular to the stripes, so that the interpolation must be done perpendicularly to the stripe and a unique 3D curve is interpolated for each row.

Between consecutive points in a row, a spline curve is interpolated. In order to estimate correctly all the parameters of the curve, five consecutive points are used. if there are less than 5 consecutive points, interpolation is not performed.

$$\begin{aligned} x(t) &= a_x t^3 + b_x t^2 + c_x t + d_x \\ y(t) &= a_y t^3 + b_y t^2 + c_y t + d_y \\ z(t) &= a_z t^3 + b_z t^2 + c_z t + d_z \end{aligned} \tag{4.8}$$

where $x(t)$, $y(t)$ and $z(t)$ parameterizes a 3D curve between two consecutive 3D points of the same row, $a_x \cdots d_z$ are the terms of the 3D curves and t is the parameter of the function: 0 for the initial point and 1 for the end point. Given t values between 0 and 1, 3D points are obtained between the initial and final 3D points. Two examples of a surface with and without interpolation are given in Fig. 4.8 and Fig. 4.9.



(a) (b)

Figure 4.8: Effects of interpolation: a) initial cloud of points; b) Cloud of points obtained after splines interpolation

4.6 Quantitative and Qualitative evaluation

In this section, some experiments are done in order to evaluate the accuracy of our scanner. First of all, some images of the calibration plane are taken in different positions. As a translation table is used, the depth of each plane is known. When laser is segmented and indexed, the 3D reconstruction is done. Finally, the error is computed by comparing the z-coordinate of each point. The calibration complete quadrangle (see Fig. 4.10) has been located at several distances from the scanner in increments of 2mm. The closest plane is approximately located at 20cm from the sensor. Accuracy in the computation of every plane is shown in Fig. 4.11. The error usually increases proportional to the distance.

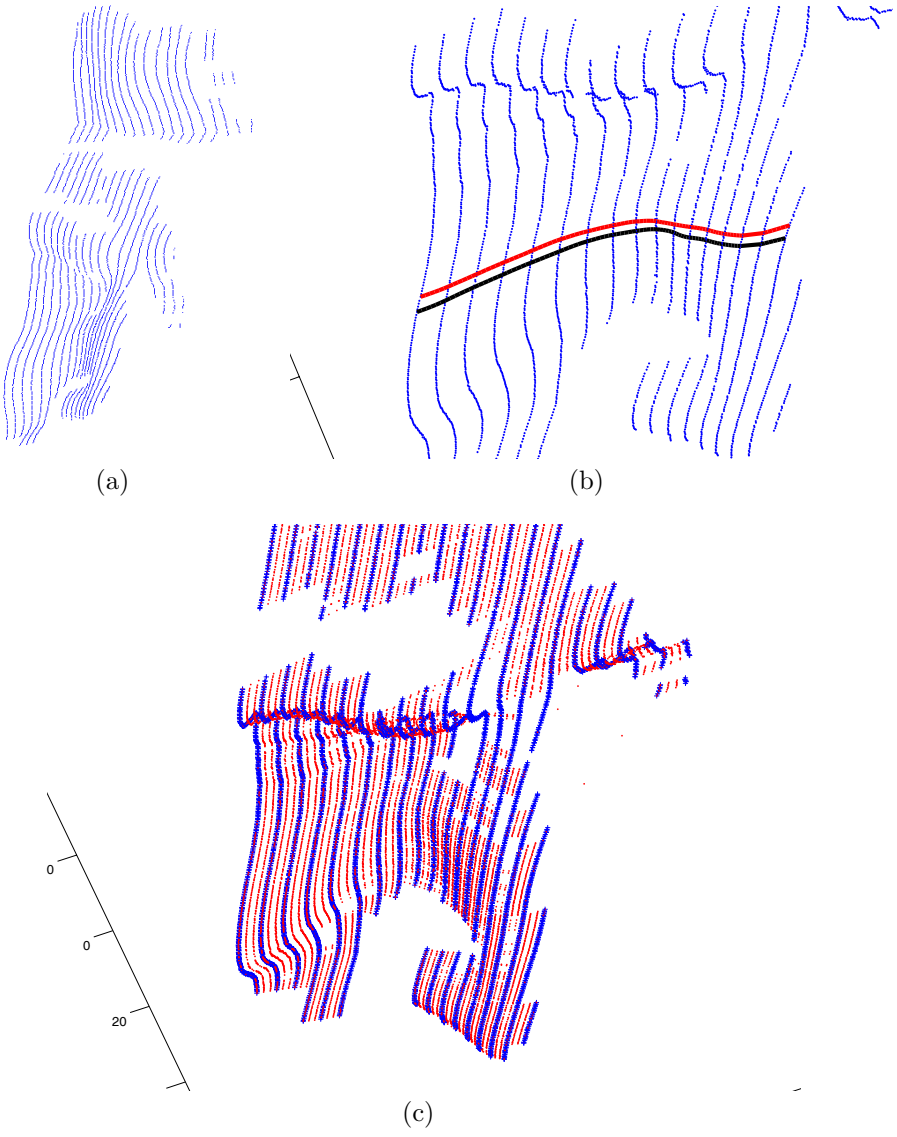


Figure 4.9: Acquisition examples: a) Acquired cloud of points from the 19 slits; b) Spline curve computation (in blue the acquired profiles, in red and black two samples of splines); c) Cloud of points obtained after spline sampling (in blue the original points, in red the new points computed)

However, when the object is near to the camera, the acquired image becomes blur due to the focal distance and, therefore, laser peak is inaccurately obtained, producing errors in the reconstruction. A reconstruction of a calibration plane is shown in Fig. 4.12.

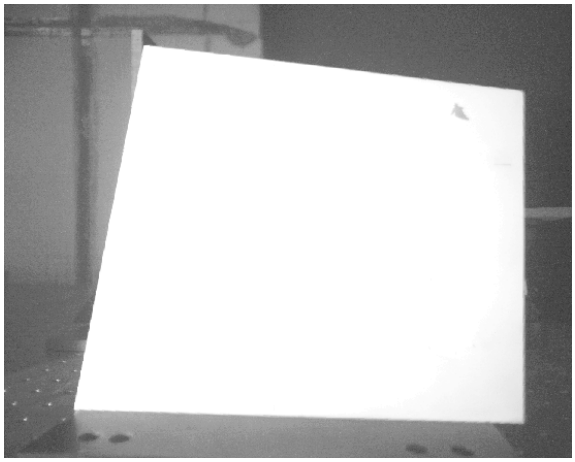


Figure 4.10: Calibration complete quadrangle

Finally, the indexing algorithm is tested by taking several images and computing the percentage of indexed points with respects to all segmented laser peaks. As can be shown

Table 4.3: Percentage of indexed points

View	Detected peaks	indexed peaks	%Indexing
1	13628	12746	93.53
2	14936	14714	98.51
3	15061	14893	98.88
4	15007	14556	96.99
5	15548	15142	97.39
6	13670	11138	81.48
7	16386	11758	71.76
8	12846	10288	80.09
9	11850	8495	71.69
10	12874	10744	83.46

in the Table 4.3, the percentage of indexed points is very high, but the most important is that 100% of indexed points are correctly indexed. Some of the images used in this experiment are shown in Fig. 4.13.

Finally, some examples of acquired objects are presented in Fig 4.14

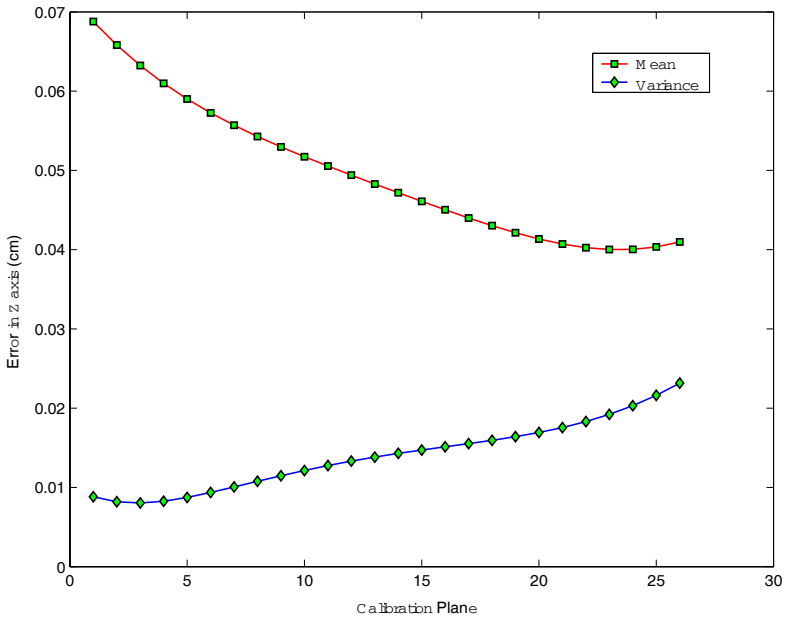
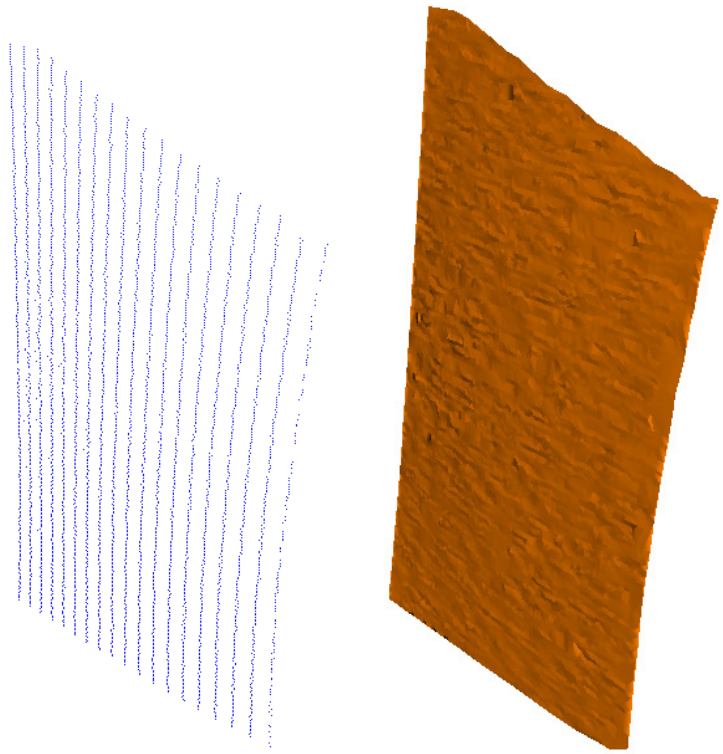
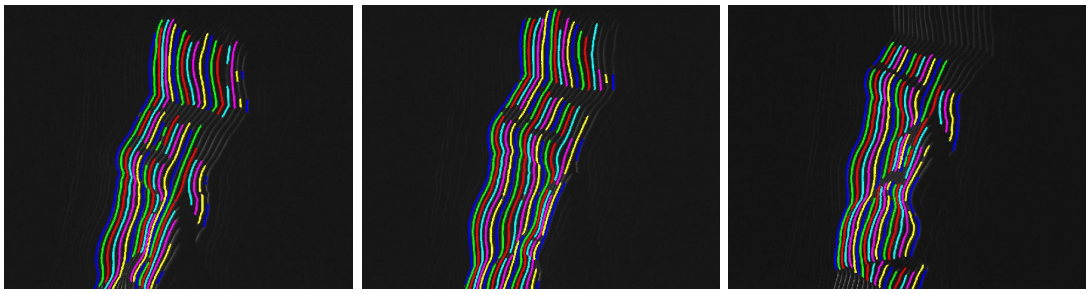


Figure 4.11: The accuracy obtained related to depth, i.e. the Z-axis



(a) (b)

Figure 4.12: Reconstruction of a plane: a) clouded points, b) 3D surface



(a) (b) (c)

Figure 4.13: Some of the images used to test the indexing algorithm

4.7 Conclusions

In this section a new 3D hand-held sensor is presented. This sensor is based on the multi-stripes projection. Therefore, one shot is enough to obtain a 3D acquisition. So that, acquisitions are not sensible to vibrations and moving scenes can be acquired without any control of the motion. Furthermore, due to the dimensions of the sensor, it can be freely moved in order to search for the best camera orientation avoiding occlusions and shadows.

As uncoded structured light is used, an indexing algorithm must be implement to determine the origin of each detected laser profile. A variant of Robinson [Robinson *et al.*, 2004] is presented adding some rules and constraints to help the indexing problem.

Additionally, a spline interpolation is performed to obtain a continuous acquisition of the scene. Although the interpolation lets us to obtain a better visualization of the surface, the main goal of the interpolation is to get a continuous surfaces to increase the accuracy in registration.

The 3D reconstruction of a plane lets us to determine the accuracy on the depth estimation. This accuracy is quite good compared to other one shot sensors.

Experimental results shows both the accuracy in the reconstruction and the percentage of indexed points. Although not all detected laser peaks are indexed, the most important is that false indexation is avoided. Note that if any profile is bad indexed, the error in 3D reconstruction might be difficult to detect and remove. On the other hand, non-indexed points do not decrease the accuracy of the reconstruction, they only decrease the area iof the reconstructed surface. However, this is not a problem when several images are acquired. Finally, some reconstructions are shown detailing the accuracy on the acquisition.

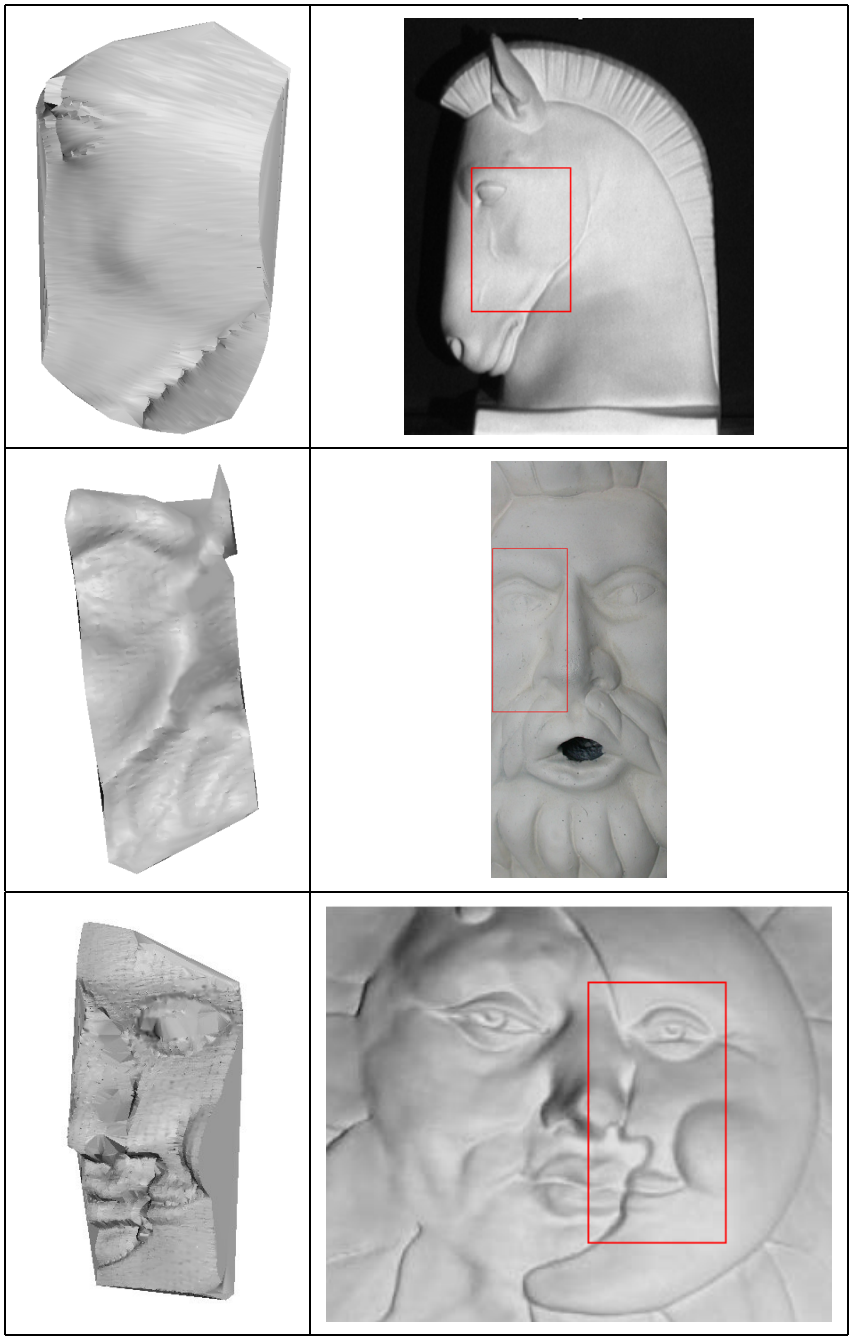


Figure 4.14: Set of acquisitions of the 3D hand-held sensor. Acquisitions of the left and real objects on the right. The box shows the acquired area

Chapter 5

Experimental Results

This chapter presents some experiments realized in order to validate our registration techniques. Experiments are divided in two groups (synthetic and real data). First of all, synthetic data is used. This experiment lets us to easy compare the estimated motion with the real one. Although gaussian noise is added to the models, synthetic data not always reflects all the problems of real data. The second part of this chapter uses real data to evaluate the accuracy of our approach. Both qualitative and quantitative results are performed.

5.1 Synthetic data

Although real experiment are preferred to better evaluate the accuracy of the algorithms, these can not be always done. In our case, estimated motion and real motion of the 3D sensor must be compared. The problem is the difficulty to determine the real motion as we explain in the next section. On the other hand, by using synthetic data, the motion is exactly known, so that, all registration error is because of inaccuracies of the algorithm. In order to use available models, Synthetic data is obtained from the 3D synthetic models courtesy of INRIA[1].

As a complete 3D model is get, a virtual scanning is applied in order to get a set of consecutive partial acquisitions of the object. This virtual scanning is performed in

[1]http://www-c.inria.fr/gamma/download/download.php

Figure 5.1: 3D model used in the experiments

Matlab to acquire only the points observed by the camera. In order to simulate a real problem, these acquisitions are represented in the sensor coordinate frame, and gaussian noise is added to the model. The trajectory of the sensor is based on a set of consecutive cycles shown in Fig. 5.3.

Both translation and rotation errors are represented in Table 5.3. Translation errors are obtained as the discrepancy between the real translation (given by XYZ-Table) and the estimated one (obtained by registration). Rotation errors can be analyzed by the discrepancy between the norm of both axis of rotation[2].

$$Error_{rotation} = \vec{n}_{est} e^{\theta_{est}} - \vec{n}_{real} e^{\theta_{real}} \tag{5.1}$$

where $\vec{n}_{est}$, $\vec{n}_{real}$ represent the estimated and real axis of rotation and θ_{est} and θ_{real} are the estimated and real angle of rotation, respectively.

Additionally, the MSE (Mean Squared Error) is computed. For each point of the

[2]Both Rotation matrices are represented in axis-angle representation

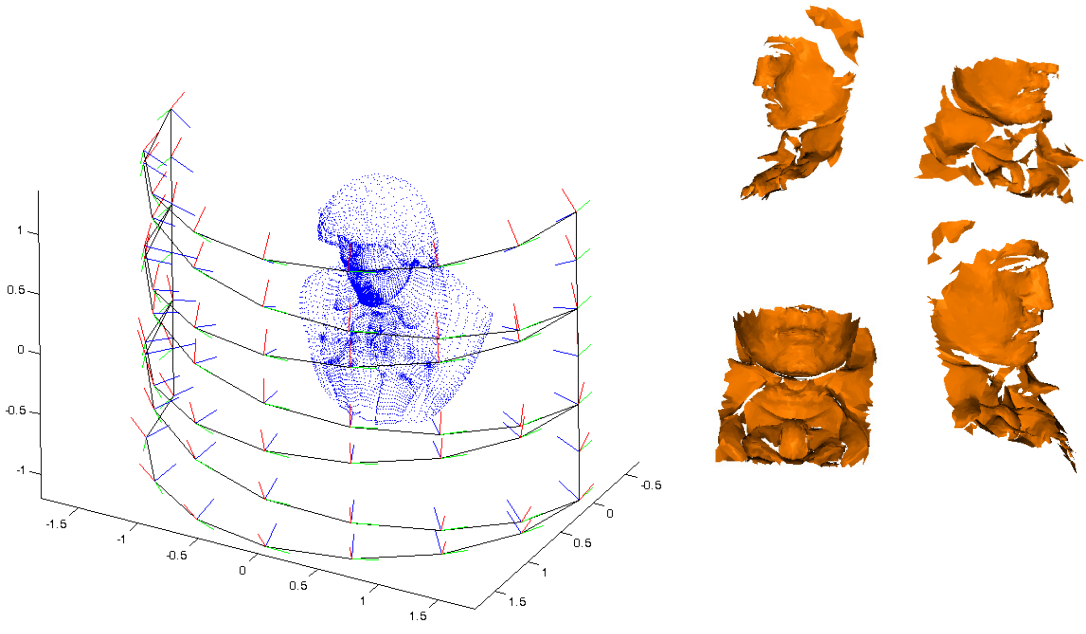


Figure 5.2: Left: Path described by the simulator to scan a synthetic object. Right: Some of the acquired synthetic images

registered acquisition, the nearest point in the set composed by the rest of acquisitions is found, determining a correspondence. The mean of all distances between correspondences give us the estimation of the discrepancy between registered views.

Experimental results are shown in Table 5.3. Every cell shows the mean and standard deviation of the mean of errors (rotation, translation and MSE) of all acquisitions. Additionally to our proposal and Sharp's one, another fast proposal is tested. This proposal is a modification of our approach. The main difference is in the cycle minimization where only neighbor views are considered. Therefore, no additional correspondences must be computed (only between initial and final view) decreasing considerably the computational time. This fast proposal consumes approximately the same time as Sharp's approach.

Although the errors without noise are similar, due to the robustness of our approach, differences when noise is big are significant. When the noise level is low, the result of pair-wise registration is quite good, without considerably propagation errors. So that, results of both techniques are really good. However, when noise is important, pair-wise registration is not good enough, and Sharp's approach badly distribute the motion error. On the other hand, our approach can deal with noise problem producing a good registration.

Table 5.1: Comparison of multi-view registration methods Both our original method and its fast variant are compared to the method of Sharp: errorR is the norm of the difference between both axes of rotation; errorT is the norm of the difference between both translation vectors (distance between the points of origin of both coordinate systems); MSE is the mean squared error. Every table cell indicates the mean (up) and standard deviation (down) of the error for a set of synthetic experiments varying the Gaussian noise (σ) and one experiment with real data.

	Our Method			Fast Variant			Sharp's Method		
Scene	$error_R$	$error_T$	MSE	$error_R$	$error_T$	MSE	$error_R$	$error_T$	MSE
σ=0	0.516	0.008	0.003	0.339	0.079	0.001	0.511	0.074	0.002
	1.120	0.004	0.003	0.867	0.191	0.006	1.006	0.034	0.002
σ=1.25%	0.675	0.154	0.004	1.177	0.459	0.006	2.225	4.403	0.026
	1.115	0.305	0.005	1.388	0.265	0.006	1.385	1.196	0.051
σ=2.5%	1.1286	0.4698	0.005	1.202	0.316	0.005	1.472	1.367	0.001
	1.1905	0.2149	0.005	1.410	0.217	0.005	1.202	0.704	0.002
σ=3.75%	0.246	0.056	0.003	1.552	0.875	0.007	2.601	3.485	0.026
	0.732	0.024	0.002	1.169	0.425	0.006	1.134	2.106	0.046
σ=5.0%	1.570	0.890	0.005	1.533	0.828	0.007	2.753	3.126	0.017
	1.284	0.682	0.005	1.144	0.497	0.006	1.212	2.225	0.020

Finally, an in order to evaluate the accuracy of the final surface, a ground truth test is done. This test consists in compare the final registration with the 3D synthetic model. Results of this test are shown in Tab. 5.2. In free noise test, all methods are similar. However, when noise is important, pair-wise alignment is not good enough to correctly determine the cycles. In this case, Sharp's approach is not robust enough to correct this error, and the given solution is far from the real one.

5.2 Real data

As not always synthetic data represents problems of real acquisitions, a 3D sensor is also used to test the different methods. However, in real applications is not easy to determine the motion of the sensor between different acquisitions. Mechanical systems can be used to obtain this information. Two different mechanisms have been used during the experiments. First of all, a 3 DoF translation table is used. Secondly, a 6 DoF robotic arm is also used.

Table 5.2: Ground Truth test of multi-view registration methods Both our original method and its fast variant are compared to the method of Sharp. This test consists in compare the 3D final registration with the 3D synthetic model used. Every table cell indicates the mean (up) and standard deviation (down) of the error for a set of synthetic experiments varying the Gaussian noise (σ) and one experiment with real data.

Noise	Our Method	Fast Variant	Sharp's Method
σ=0	0.0283	0.0076	0.0339
	0.0189	0.0175	0.0230
σ=1.25%	0.0705	0.0493	4.4129
	0.0564	0.0482	1.8869
σ=2.5%	0.0679	0.0403	0.7437
	0.0537	0.0335	0.5649
σ=3.75%	0.0143	0.0403	3.2935
	0.0104	0.0335	1.2683
σ=5.0%	0.0479	0.0946	2.1505
	0.0392	0.0659	0.9575

Both experiments are now detailed.

5.2.1 Translation table

This experiment consists in coupling the 3D sensor to a translation table, scanning an object following a known trajectory and compare this trajectory to the estimated one given by the registration algorithm.

A plaster object is used as a target, and 29 consecutive views are acquired by translating the sensor position. The motion between two consecutive acquisitions is 1 cm. The 3DoF translation table is shown in fig. 5.4.

As pose of the sensor can be determined by means of the translation table controller, estimated motion can be compared to the real one to get an approximation of the accuracy of each method.

Additionally to the motion errors, the MSE of both techniques is also computed. Figure 5.5 shows that our method is suitable to reduce the propagation error in the presence of cycles. Although Sharp's method obtains similar results at the end of the

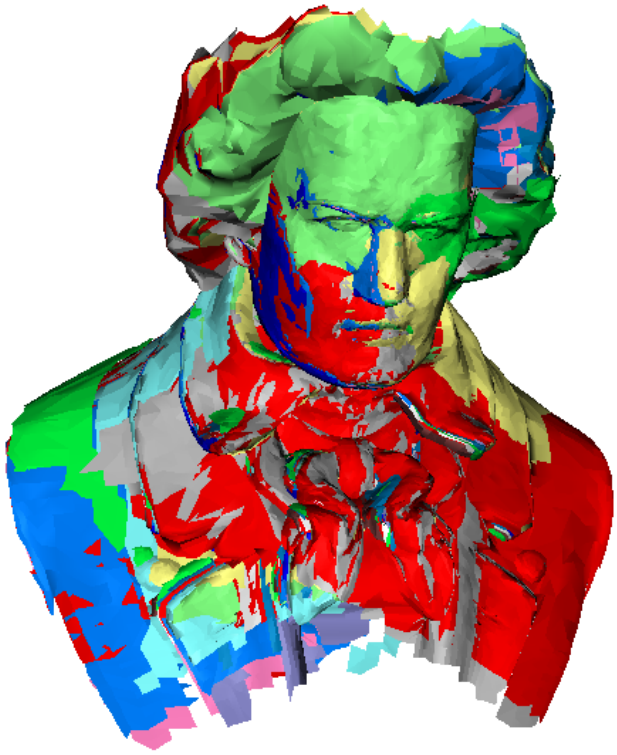


Figure 5.3: Registration of a set of partial acquisition of beethoven

Table 5.3: Comparison of multi-view registration methods in real objects

	Our Method			Fast Variant			Sharp's Method		
Scene	$error_R$	$error_T$	MSE	$error_R$	$error_T$	MSE	$error_R$	$error_T$	MSE
	1.2804	5.063	0.334	1.3485	5.4103	0.389	1.3863	4.640	0.432
	0.303	2.459	0.335	0.285	2.425	0.529	0.291	2.308	0.380

cycle (view 21), the error is worse distributed inside the cycle compared to our approach. Note that after view 21 the error increases in both methods till another cycle is detected.

5.2.2 Robotic arm

In order to evaluate the performance of the methods, it is also useful to observe the registration of a real object and analyze it from a qualitative point of view. In this experiment,

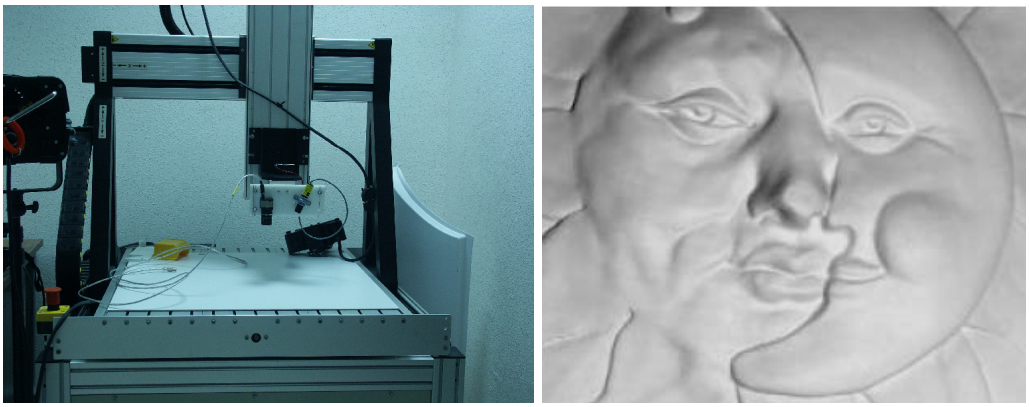


Figure 5.4: Set-up used in the experiments: a) Translation table and b) plaster object

the one-shot hand-held scanner is coupled to a FANUC industrial manipulator. The manipulator describes a trajectory so that a given object is scanned obtaining a sequence of views. As the kinematics of the manipulator is known, the views can be aligned without applying any registration and hence such raw alignment is provided for comparison.

Note that the kinematics of the manipulator provides the position of the robot hand H with respect to the coordinate frame of the robot base R (See Fig. 5.7). Besides, registration is referenced with respect to the frame S of the camera of the one-shot hand-held scanner. The rigid transformation between H and S is unknown and hence has to be first estimated.

There is a topic in robotics to estimate this transformation that is based on solving equation $AX = XB$, where X is the matrix we are looking for. So, X transforms points from the coordinate frame of the scanner S to the coordinate frame of the hand H, A is the motion of the hand between two different positions of the robot given by the robot control system, and B is the motion computed by means of triangulating the movement in the image of the one-shot hand-held scanner.

There are several papers addressing the computation of $AX = XB$ [Fassi and Legnani, 2005] [Shiu and Ahmad, 1989]. In our case, we have acquired 10 views of a calibrating pattern and the X matrix is estimated by using the algorithm of Shiu [Shiu and Ahmad, 1989]. First, the algorithm determines a set of A and B matrices from every view. Then, a system of equations with the form $AX - XB = 0$ is defined and solved. Theoretically X can be computed with only 3 views, though it is more accurate to solve the equation of an over-determined system by using singular value decomposition.

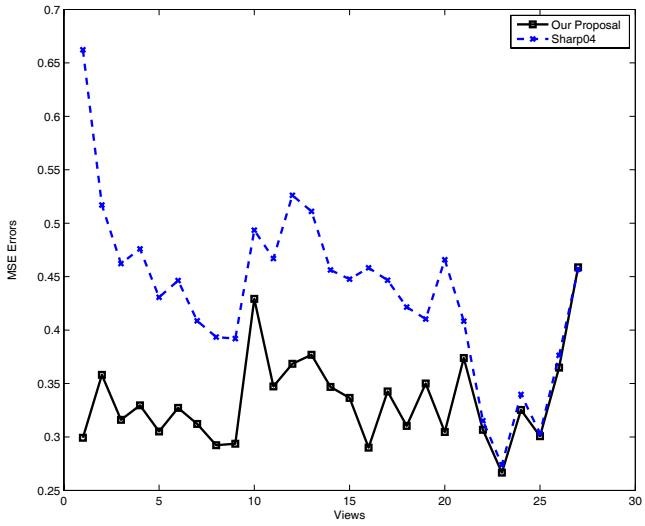


Figure 5.5: Evolution of the MSE registration errors. Scale of the measured object: 180mm (width) $\times$ 200 mm (height) $\times$ 56 mm (depth)

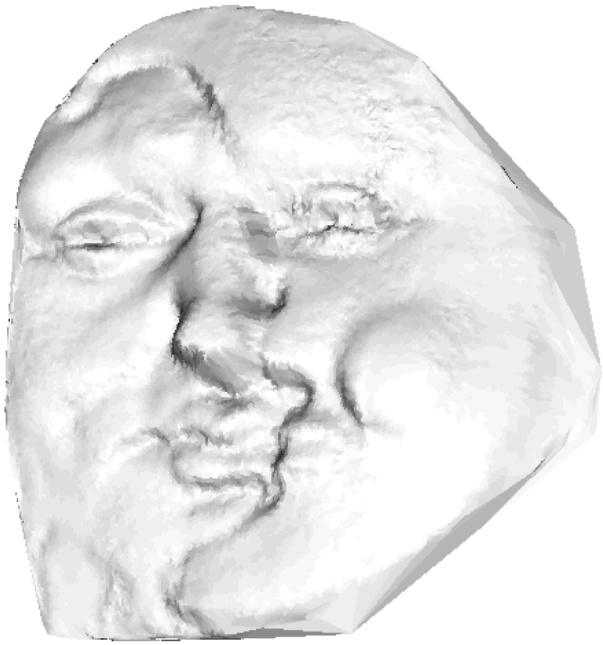


Figure 5.6: 3D reconstruction of the plaster object by means of the registration of 27 acquisitions

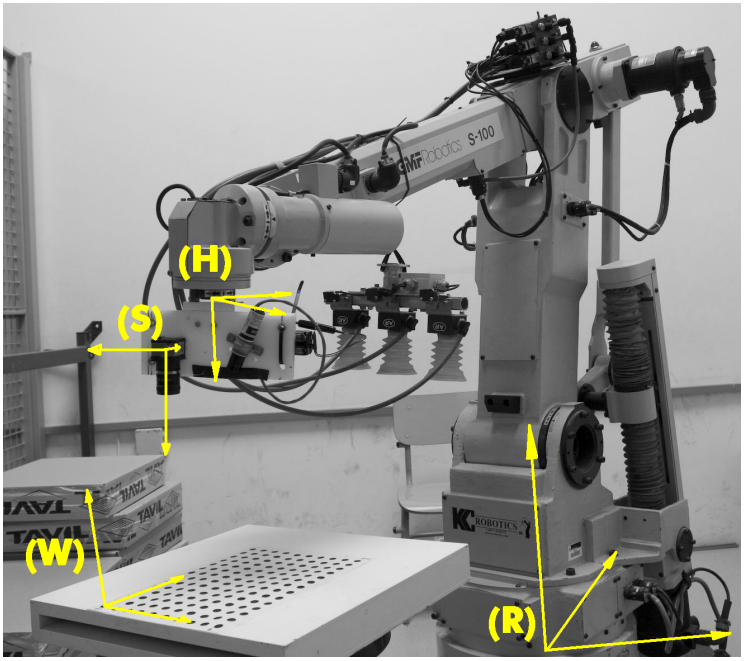


Figure 5.7: Industrial manipulator used in experiments. The four coordinate frames are represented: W (world), R (robot), H (Hand) and S(Scanner)

Once X is known, all views can be represented in the same reference using the following equation:

$$^{W}T_{S} = ^{W}T_{R} \times ^{R}T_{H} \times X \tag{5.2}$$

where $^{W}T_{S}$ is the Euclidean motion that transforms points in S to the world coordinate system W (used by the one-shot hand-held scanner to refer 3D points), $^{W}T_{R}$ is the Euclidean motion that relates the world coordinate system W to the robot base R, $^{R}T_{H}$ is the motion given by the kinematics of the robot arm, and X is the Euclidean transformation between the camera of the one-shot hand-held scanner and the robot hand. In the Fig 5.8, the result of the calibration is shown. Circles represent the real position of the calibration points, and asterisks represents the transformation of the camera acquired control points to the world coordinate frame. Errors on the estimation of X and in the measures of the position of the hand produces misalignments in the final position. These errors will produce misalignments in the mechanical alignment.

Now we can proceed with the experiment. The manipulator has been programmed so that an 8-shape trajectory is done over a ceramic object acquiring up to 41 images and

hence 41 3D partial views of the object. Note that the trajectory ensures cycles which will be used in the registration. First, all the views are referenced with respect to the same frame by means of the X matrix. Second, a volumetric integration algorithm is applied to get a continuous surface [Curless and Levoy, 1996]. Third, the sequence of views are aligned according to: a) the registration algorithm proposed in this article; b) the multiview algorithm proposed by Sharp [Sharp *et al.*, 2004]; and c) the kinematics of the robot. Finally, any surface smooth technique is applied to enhance the visualization. Qualitative results are shown in Fig. 5.9.

Registration really improves the alignment provided by the kinematics of the robot, which suffer not only from inaccuracies given by the mechanics but especially inaccuracies in the computation of X. Besides, the experiment also shows that our approach provides a surface with more details and less artefacts compared to the method proposed by Sharp.

5.3 Conclusions

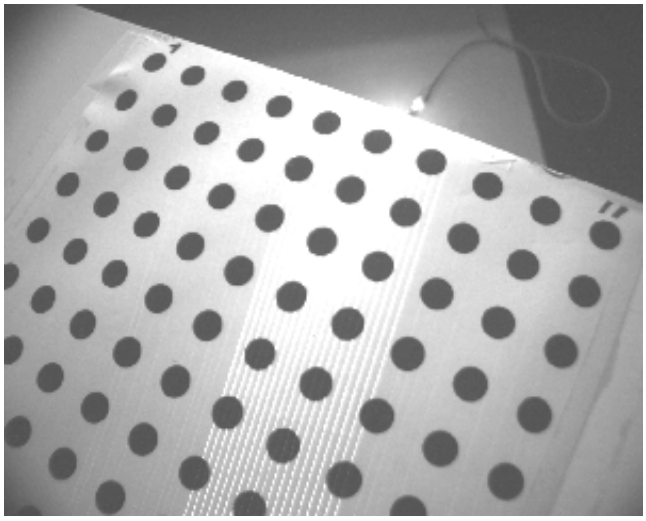
This chapter presents experimental results obtained using our multiview registration algorithm. Firstly, synthetic data is used to compare our method with an already published one. Synthetic data is preferred to real acquisitions due to the facility of obtaining the motion between different acquisitions.

In order to simulate real problems, Gaussian noise is added and acquisitions with non-overlapping area are considered. Errors are presented in terms of motion errors (rotation and translation) and MSE. Despite the results obtained with the tested algorithms are very similar when noise is small, our approach presents better results due to the robust minimization of the cycle.

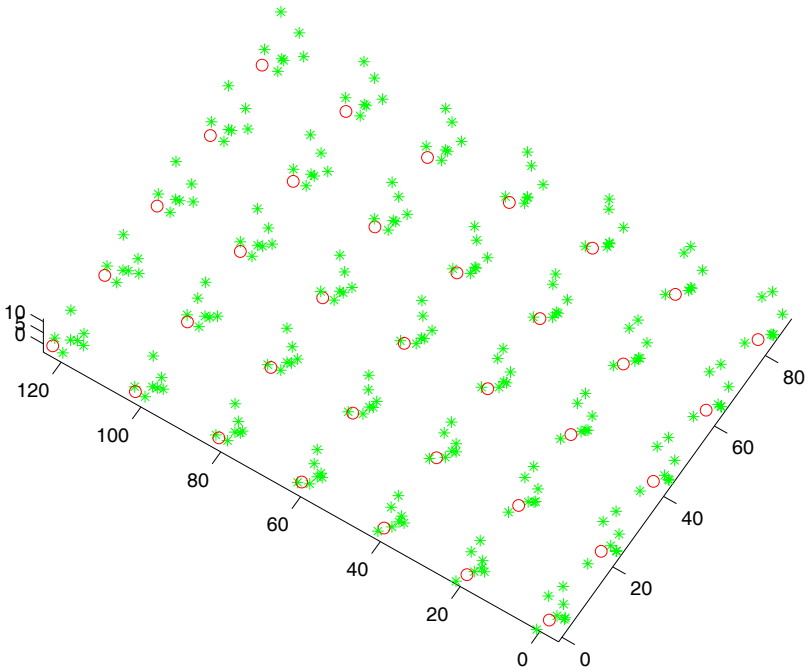
Additionally, real acquisitions are also registered to validate our approach in worse conditions with respects to synthetic ones. Mechanical systems are used to get information about the motion between acquisitions. However, due to inaccuracies on the set-up calibration, motion matrices can not always be accurately obtained. Some results demonstrate that registration algorithms are better than mechanical alignment. Despite this experiment, it is also true that depending on the mechanical system used, accurate alignments can be obtained.

Several quality results are shown to visualize the difference in the final reconstruction

depending on the method used.



(a)



(b)

Figure 5.8: Results of the robot arm calibration: a) calibration pattern; b) 3D alignment of the calibration points where every cross represents the center of the circles computed in each acquisition and the dot represents the theoretical position of them

(a) (b) (c) (d)

Figure 5.9: Results of the registration: a) Our proposal; b) Sharp's proposal; c) Mechanical alignment; c) Real object

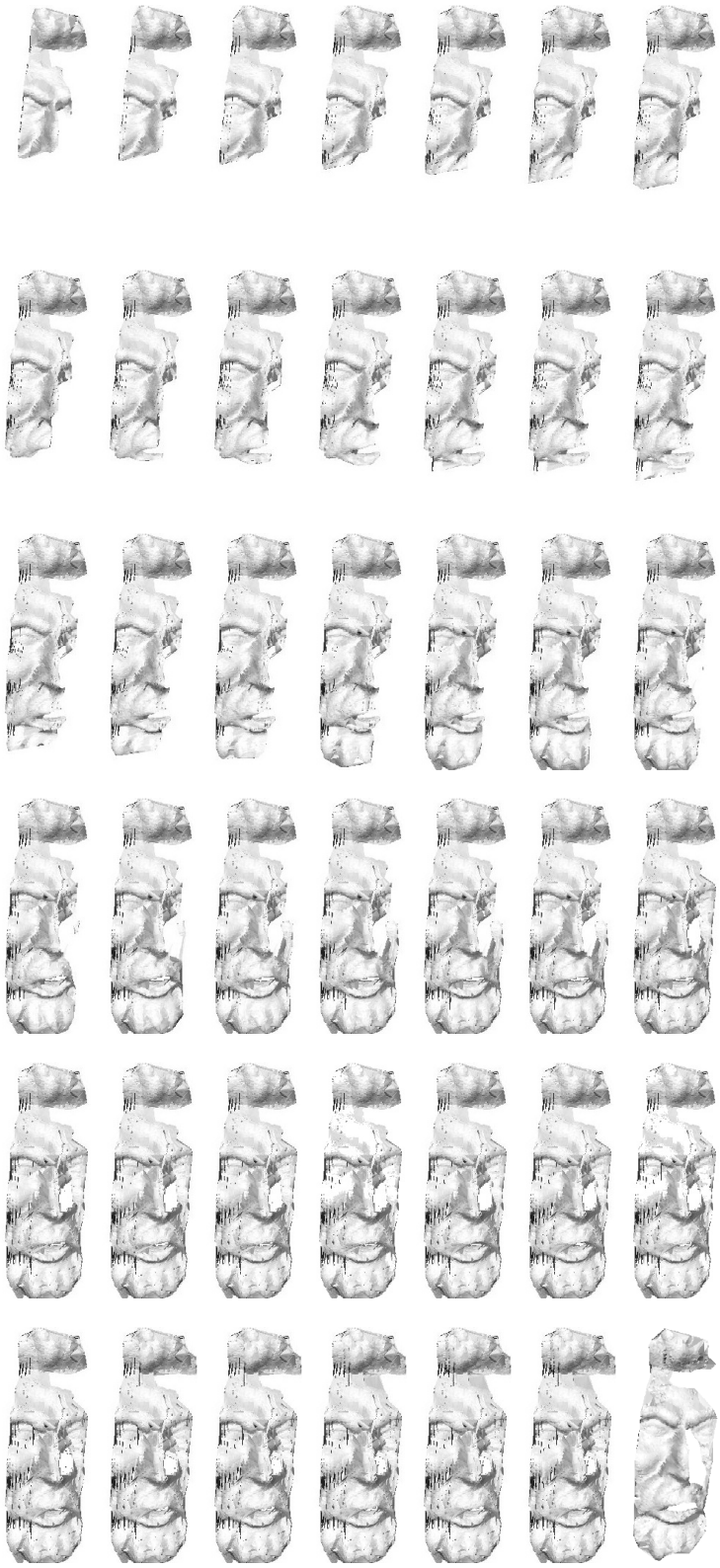


Figure 5.10: Results of the addition of every view in the sequence to the registered surface. Last view shows surface integration

Chapter 6

Conclusions and further work

This chapter presents the conclusions and some perspectives opened by this work. The scientific contributions of the thesis are first discussed. Afterwards, the list of publications related to this work is presented as well as the scientific collaborations involved during its preparation. Finally, further work and future perspectives are discussed.

6.1 Conclusions

This thesis is focused on 3D registration as a method to achieve a complete acquisition of a given object from a set of partial acquisitions. Registration is based on computing the Euclidean motion of every acquisition that minimizes the global alignment. Most of the techniques are focused on aligning two single views. However, in real applications, more views are required to achieve a complete acquisition. The solution offered in this thesis combines the consecutive registration of two views with cycle minimization techniques to deal with error propagation problems in on-line registration.

The thesis starts with a comprehensive look at the state-of-the-art of registration techniques, focusing especially on the alignment of two sets of data points. Registration is usually divided into two steps depending on the information available. If an estimation of the motion between views is available, the complexity decreases considerably and fine registration techniques can be applied directly. Otherwise, coarse registration techniques must be applied first to get an approximation of the motion. This part of the work

presents a classification of both approaches analyzing their pros and cons. One of the contributions of this thesis is the proposed classification and the comparative evaluation of surveyed techniques. Results are shown in the presence of both synthetic and real data and compared in terms of resolution of the data sets, percentage of outliers and Gaussian noise. The extensive classification and explanation of each group of techniques provides valuable guidelines for easily deciding which technique must be used depending on the application requirements.

The survey on coarse registration techniques has shown that computing time is one of the main problems. These techniques are based on the selection of correspondences through exhaustive searches in the 3D space, which deals with costly algorithms concerning computing time. Although principal component analysis presents the fastest solution, the accuracy is not usually good enough. Experiments demonstrated that the spin image approach presents the best ratio between time and accuracy.

Fine registration techniques are based on the minimization of the distances between point sets. The main difference between them is the distance to minimize: while some techniques use point-to-point distances, others use point-to-plane distances. The main drawback of point-to-point is the convergence to local minima.

Moreover, former versions of ICP[1] were not useful in the presence of non-overlapping areas and outliers and point-to-plane uses distance threshold to remove outliers. Finally, genetic algorithms also obtain accurate results, but with a high computational cost. In general, techniques based on point-to-plane distances are preferred to the others. However, this set of techniques are not good enough when more than two views must be aligned. In this situation, propagation of the registration errors produces misalignments in the final registration.

A second important contribution of this thesis is an overview of the existing multi-view registration algorithms that confront propagation errors. This set of techniques simultaneously registers all views, minimizing the global error. However, the simultaneous minimization of all the views has some drawbacks. Most of the multi-view techniques require a good initial alignment, which sometimes is given by mechanics, manual alignment or even pairwise coarse registration. Moreover, as all acquisitions are required in the registration, online applications cannot be considered. Additionally, as all views are used, a lot of computing time is required to achieve a complete registration.

[1]Iteration Closest Point

The main contribution of this thesis is the proposal of a new registration technique based on the combination of pair-wise and multi-view registration. This approach takes the best of both pair-wise and multi-view to obtain an accurate registration in online applications. It is based on the pair-wise registration of consecutive applications. As the motion between consecutive acquisitions is small, no initial estimations of such motions are required, and the computation of intensive coarse registration is avoided. In order to reduce the propagation errors, global minimization is applied each time the scanner revisits an already acquired view. Then, a cycle is defined between both views. A minimization is performed taking into account the views of the cycle. Additionally, constraints are added to increase the accuracy of the minimization. As only the views conforming the cycle are simultaneously minimized, computing time is not wasted searching for correspondences among views that do not have any overlapping area. This is especially useful in the presence of large data sets.

Additionally, a robust algorithm is proposed to detect cycles while the scanner is moving and acquiring new views. With the aim of avoiding the complexity of computing the overlapping area between two acquisitions in the 3D space, both views are projected on the three orthogonal planes and the analysis of the projections becomes a 2D problem. When a cycle is found, all data points conforming the cycle are simultaneously minimized, reducing the propagation error. Moreover, a variant of this approach is presented to speed up the process of computing point correspondences.

Experimental results are performed to validate our approach. Accuracy in the registration is compared to a similar technique. Both synthetic and real data are used. In general, Sharp's approach obtains good results in the end-views of the cycle, but errors are not always well distributed through the cycle. Results show that our proposal is more robust.

Finally, a new 3D hand-held scanner is proposed. Unlike other commercial scanners, only visual information is used. Most part of available scanners are coupled to mechanical systems to attain positions. This decreases the degrees of freedom, limiting the workspace of the scanner. The main difference is that our scanner can acquire a surface from a single shot of the camera. Most sensors only project a single laser plane, so only one profile is acquired per image frame. The proposed scanner is based on the projection of a multi-stripe pattern onto the measuring surface. The pattern is not coded, so we have had to deal with the stripe indexing problem. Additionally, spline interpolation is applied to

increase the resolution of the acquired surface leading to dense acquisitions and preserving a high degree of accuracy.

Summarizing, the combination of our 3D hand-held scanner with our registration algorithm lets us obtain a complete acquisition of the measuring surface from a set of partial acquisitions.

6.2 Contributions

This section briefly spots the list of contributions of this thesis, though they have been discussed and analysed in the previous sections. The contributions are:

- A new state-of-the-art on 3D registration including new classification criteria and providing comparative results.
- A new approach on multi-view registration based on cycle detection and minimization.
- Design and development of a new 3D hand-held sensor.
- A modification in a known stripe indexing technique to be adapted to un-coded multi-slit patterns.
- The normal space sampling has been described in detail to provide additional information to the reader.

6.3 Publications and scientific collaborations

The work developed in this thesis has produced a few journal publications and several contributions to international conferences, which are presented in the following paragraphs. Finally, the scientific collaborations carried out during the thesis are also detailed.

Publications

The following articles have been published or submitted in international journals:

- J. Salvi, C. Matabosch, D. Fofi and J. Forest, A review of Recent Range Image Registration methods with accuracy evaluation . **Image and Vision Computing**, *In Press* [Salvi *et al.*, 2007] (JCR[2] = 1.383)
  This article surveys the most representative registration techniques and includes experimental results.

- C. Matabosch, D. Fofi J. Salvi, and J. Batlle, Registration of surfaces minimizing error propagation for a new hand-held laser scanner. **Pattern Recognition** *(Submitted)* [Matabosch *et al.*, n.d.]. (JCR = 2.153)
  A new registration algorithm is proposed, which major contribution is the reduction of error propagation in the registration of large data sets.

In addition, while working on the thesis the following contributions to international conferences were made:

- C. Matabosch, E. Batlle, D. Fofi and J. Salvi, A variant of point-to-plane registration including cycle minimization . Photogrammetric Computer Vision, **PCV06**. Bonn, Germany, 20-22 September 2006 [Matabosch *et al.*, 2006*a*]
  This article presents the first variant of our multi-view registration approach based on cycles minimization.

- C. Matabosch, J. Salvi, D. Fofi and F. Meriaudeau, A Refined Range Image Registration Technique for Multistripe Laser Scanner. Proceedings of the SPIE -The International Society for Optical Engineering, Volume 6070, Machine Vision Applications in Industrial Inspection XIV, **MVA06**, San Jose, California, USA, 15-19 January 2006 [Matabosch *et al.*, 2006*b*]
  This article presents the developed hand-held sensor detailing calibration, segmentation and indexing steps.

- R. Garcia, R. Prados, T. Nicosevici, F. Garcia, C. Matabosch, Dense 3D Modelling of Underwater Structures Using an Optical Sensor, Eos Trans. AGU, 87(52), Fall Meet. Suppl., Abstract OS31B-1641, 2006. [Garcia *et al.*, 2006]
  3D registration techniques are applied on underwater images.

[2]Journal Citation Report (the larger the number, the greater the impact factor of the journal).©Thomson Corporation.

- C. Matabosch, D. Fofi, J. Salvi, Reconstruction et Recalage de Surfaces en Mouvement par Projection Laser Multi-lignes, 9èmes journées ORASIS'05, Puy-de-Dôme, France, 24-27 May, 2005 [Matabosch *et al.*, 2005*a*]
  First results of the 3D sensor were presented in this French conference, including scanner calibration and laser segmentation.

- C. Matabosch, D, Fofi and J. Salvi, Registration of moving surfaces by means of one-shoot laser projection. Lecture Notes in Computer Science, 3522 (I), pp. 145-152 [Matabosch *et al.*, 2005*b*]
  The thesis proposal was introduced at this conference. This is our first article that refers to the idea of a hand-held sensor to register a set of consecutive views.

- C. Matabosch, J. Salvi and D. Fofi, A new proposal to register Range Images, 7th International Conference on Quality Control by Artificial Vision, QCVA'05, Nagoya, Japan, 18-20 May 2005 [Matabosch *et al.*, 2005*c*]
  This paper presents a multiview registration algorithm based on the minimization of a cycle by means of the Dijkstra algorithm.

- C. Matabosch, J. Salvi, D. Fofi and F. Meriaudeau, Range image registration for industrial inspection. Proceeding of the SPIE -The International Society for Optical Engineering, Volume 5679, Machine Vision Applications in Industrial Inspection XIII MVAII'05, San Jose, California, USA, 16-20 January 2005 [Matabosch *et al.*, 2005*d*]
  Results of the range image registration techniques were detailed at this conference.

- C. Matabosch, J. Salvi, X. Pinsach and J. Pagès, A comparative survey on free-form surface registration. 1a Jornada de Recerca en Automàtica, Visió i Robòtica, AVR'04, pp 308-312, Barcelona, Spain, February 2004 [Matabosch *et al.*, 2004*a*]
  This paper presents the first proposed classification of the range image registration techniques.

- C. Matabosch, J. Salvi, X. Pinsach and R. García, Surface Registration from Range Image Fusion. IEEE International Conference on Robotics and Automation, **ICRA04**, pp 678-683, New Orleans, USA, April 26-May 1, 2004 [Matabosch *et al.*, 2004*b*]
  This paper presents the first conclusions about the state-of-the-art of range image registration

- C. Matabosch, J. Salvi and J. Forest, Stereo rig geometry determination by fundamental matrix decomposition. Workshop on European Scientific and Industrial Collaboration, WESIC'03, pp. 405-412, Miskolc, Hungary, 28-30 May 2003 [Matabosch *et al.*, 2003]
  This article is my introduction to 3D Imaging. Specifically, it is a study of Kruppa equations to reconstruct a scene from a set of three images of the scene acquired by an uncalibrated camera.

In addition to this publications, some contributions were made with other colleagues of the laboratory. These articles are the result of the first months in the VICOROB group, where I was introduced to computer vision. I was basically involved in setting-up some experiments related to other thesis.

- J. Pagès, J. Salvi and C. Matabosch, Implementation of a robust coded structured light technique for dynamic 3D measurements, IEEE International Conference on Image Processing, ICIP'03, pp. 1073-1076, Vol 1, Barcelona, Spain, September 2003 [Pagès *et al.*, 2003*a*]

- J. Pagès, J. Salvi, R. García and C. Matabosch, Overview of coded light projection techniques for automatic 3D profiling. IEEE International Conference on Robotics and Automation, ICRA'03, pp 133-138, Taipei, Taiwan, 14-19 September 2003 [Pagès *et al.*, 2003*b*]

- J. Pages, J. Salvi and C. Matabosch, Robust segmentation and decoding of a grid pattern for structured light, 1st Iberian Conference on Pattern Recognition and Image Analysis, IbPRIA'03, pp.689-696, Mallorca, Spain, June 2003 [Pagès *et al.*, 2003*c*]

Scientific Collaborations

The thesis was developed during the four-year period of 2003-2006. Due to the collaboration between the University of Girona and the University of Burgundy, the thesis has been partly done in both centers. The major part has been carried out in the VICOROB group of the University of Girona, while 8 months were spent in teh Le2i in Le Creusot (France), split into the following stays:

- 4-month stay in the Le2i group. Period: 06/03/04 - 30/06/04. Supervisor: Dr. David Fofi.

- 4-month stay in the Le2i group. Period: 01/02/05 - 31/05/05. Supervisor: Dr. David Fofi.

The work carried out during each stay let me to export knowledge from our group to the Le2i group. Specifically, the laser calibration principle based on the complete quadrangle was used by one of the Le2i students to calibrate his sensor. Additionally, other approaches concerning 3D computed vision began in that laboratory. For example, Thierry Molinier is developing a thesis titled *Approche Coopérative pour l'Acquisition et l'Observation de Formes Tridimensionnelles*, which is being supervised by Dr. Fofi and Dr. Salvi.

In september 2206, I started collaborating with AQSense S.L. basically on technology transfer. Since January 2007, I have a full-time job in AQSense and I'm involved in programming registration algorithms in FPGA devices for real-time visual inspection applications.

6.4 Future work

The topic of 3D hand-held sensors remains mainly unexplored. Therefore, the present thesis requires some further work and opens several perspectives for future research.

Fist of all, some further work based on the proposed approach should be done in order to improve its efficiency. The main improvement should be based on cycle strategy minimization. Our proposal adds constraint to force the motion in the studied cycle to be null. However, no historical information is used, so that, the minimization of one cycle can affect views from an already minimized cycle without any kind of constraint. Imagine, for example, that the sensor describes a trajectory similar to the number 8 beginning from the top of the number. First of all, the lower cycle is found, after the minimization, the scan goes on until the last view is acquired, and a new cycle is determined. This cycle contains all the views, so that all motion parameters can change. If a constraint is added, we can assure that the lower cycle will be closed; otherwise, this is not assured.

Another important improvement can be the study of the propagation of the uncertainty

like is done in SLAM techniques. This study can be use to reduce the propagation errors and additionally to better detect the cycles.

Another option is to include the study of non-rigid registration algorithms. In spite of large applications on rigid registration, the future of 3D is directly related with medical applications and biometrics. The main difficulty of these topics is that the objects are not rigid, increasing the difficulty of the alignment.

Additionally, referring to the pair-wise registration, sampling can be improved. Normal space sampling can be modified to take into account also the spatial distribution of selected points. When selected points are sampled one far from the others, registration is more accurate.

Other improvements are directly related to the 3D hand-held sensor. Nowadays, most commercial sensors give color/texture information, allowing more realistic reconstructions to be obtained. Nevertheless, color/texture information can be used to improve registration results. In order to obtain color/texture, a second camera can be added to the sensor so that the first camera (with a narrow bandpass filter) will capture the laser information and the second one captures the color of the scene. Another solution is to use a single camera to get all the information. However, in this case, a high power laser is required because, otherwise, the segmentation of the laser becomes difficult.

In addition to the texture, the study of the different integration techniques is required to generate a continuous and smooth surface without artifacts. In this thesis, only a few techniques were studied to obtain a visualization of the registration algorithms. An in-depth study can increase the accuracy concerning the visualization acquisitions.

Although the proposed laser-stripe indexing algorithm shows good results, the indexing becomes complex when surface contains some discontinuities and the percentage of indexed points decreases. This performance can be increased if color is used to code the stripes. The first perspective was attained using non heavy projectors and laser emitters. Nowadays, the dimensions, the weight and the cost of video-projectors are being progressively reduced. Therefore, it is likely that in a few years such devices will be suitable for being integrated in robot end-effectors or located on mobile robots. This will give great flexibility when designing patterns for robot positioning and 3D acquisition.

In addition to these remaining tasks, at this moment, the PhD student Elisabet Batlle is starting her thesis facing some of the future works here stated. Ms. Batlle is studying in a generic way the multi-view approaches to merge techniques from different topics like

SLAM[3] in robotics, bundle adjustment in photogrammetry or multi-view in computer vision.

[3]Simultaneous Localization And Mapping

Bibliography

Aloimonos, Y., I. Weiss and A. Bandopadhay (1987). Active vision. *International Journal of Computer Vision* **1**(4), 333–356.

Armangué, X. and J. Salvi (2003). Overall view regarding fundamental matrix estimation. *Image and Vision Computing* **21**(2), 205–220.

Armangué, X., H. Araújo and J. Salvi (2003). A review on egomotion by means of differential epipolar geomety applied to the movement of a mobile robot. *Pattern Recognition* **36**(12), 2927–2944.

Batlle, J., E. Mouaddib and J. Salvi (1998). A survey: Recent progress in coded structured light as a technique to solve the correspondence problem. *Pattern Recognition* **31**(7), 963–982.

Bergevin, R., M. Soucy, H. Gagnon and D. Laurendeau (1996). Towards a general multiview registration technique. *PAMI* **18**(5), 540–547.

Besl, P.J. and N.D. McKay (1992). A method for registration of 3-d shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **14**(2), 239–256.

Brunnström, K. and A. Stoddart (1996). Genetic algorithms for free-form surface matching. In: *International Conference of Pattern Recognition*. Vienna. pp. 689–693.

Carmichael, O., D. Huber and M. Hebert (1999). Large data sets and confusing scenes in 3-d surface matching and recognition. In: *Second International Conference on 3-D Digital Imaging and Modeling*. Ottawa, Ont. , Canada. pp. 258–367.

Chantler, M. (1994). The Effect of Variation in Illuminant Direction on Texture Classification.. PhD thesis. Dept. Computing and Electrical Engineering, Heriot-Watt University.

Chen, C. and A. Kak (1987). Modelling and calibration of a structured light scanner for 3d robot vision. In: *IEEE conference on robotics and automation*. pp. 807–815.

Chen, C and I. Stamos (2005). Semi-automatic range to range registration: a feature-based method. In: *International Conference on 3-D Digital Imaging and Modeling*. pp. 254 – 261.

Chen, C-S., Y-P. Hung and J-B. Cheng (1998). A fast automatic method for registration of partially overlapping range images. In: *International Conference on Computer Vision*. Bombay. pp. 242–248.

Chen, C-S., Y-P. Hung and J-B. Cheng (1999). Ransac-based darces: a new approach to fast automatic registration of partially overlapping range images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **21**(11), 1229–1234.

Chen, Y. and G. Medioni (1991). Object modeling by registration of multiple range images. In: *IEEE International Conference on Robotics and Automation*. pp. 2724 –2729.

Chow, C.K, H.T. Tsui and T. Lee (2004). Surface registration using a dynamic genetic allgorithm. *Pattern recognition* **37**(1), 105–117.

Chua, C.S. Jarvis R. (1997). Point signatures: A new representation for 3d object recognition. *International Journal of Computer Vision* **25**(1), 63–85.

Chung, D.H., Yun D. and S.U. Lee (1998). Registration of multiple-range views using the reverse-calibration technique. *Pattern Recognition* **31**(4), 457–464.

Curless, B. and M. Levoy (1996). A volumetric method for building complex models from range images. In: *SIGGraph-96*. pp. 303–312.

Dalley, G. and P. Flynn (2002). Pair-wise range image registration: A study in outlier classification. *Computer Vision and Image Understanding* **87**(1-3), 104–115.

Fassi, I. and G. Legnani (2005). Hand to sensor calibration: A geometrical interpretation of the matrix equation ax= xb. *Journal of Robotic Systems* **22**(9), 497–506.

Faugeras, O.D. (1993). *Three-Dimensional Computer Vision*. MIT Press.

Favaro, P. and S. Soatto (2002). Learning shape from defocus. In: *European Conference on Computer Vision, volume 2 of Lecture Notes in Computer Science*. Copenhagen, Denmark. pp. 735–745.

Feldmar, J. and N. Ayache (1994). Rigid, affine and locally affine registration of free-form surfaces. Technical report. Technical Report of INRIA, Sophia Antipolis.

Fitzgibbon, A. W. (2001). Robust registration of 2d and 3d point sets. *Proceedings of the British Machine Vision Conference*.

Forest, J. (2004). New methods for triangulation-based shaped acquisition using laser scanners. PhD thesis. University of Girona.

Forest, J., J. Salvi, E. Cabruja and C. Pous (2004). Laser stripe peak detector for 3d scanners. a fir filter approach. In: *International Conference on Pattern Recognition*. Vol. 3. Cambridge, United Kingdom. pp. 646 – 649.

Gagnon, H., M. Soucy, R. Bergevin and D. Laurendeau (1994). Registration of multiple range views for automatic 3-d model building. In: *Computer Vision and Pattern Recognition*. pp. 581–586.

Garcia, R., R. Prados, T. Nicosevici, F. Garcia and C. Matabosch (2006). 3d modelling of underwater structures using an optical sensor. In: *Eos Trans. AGU, Fall Meet.*. Vol. 87(52). pp. 546–549.

Gibbins, D. (1994). Estimating Illumination Conditions for Shape from Shading.. PhD thesis. The Flinders University of South Australia Bedford Park, Adelaide, Australia.

Godin, G., D. Laurendeau and R. Bergevin (2001). A method for the registration of attributed range images. In: *3DIM01 Third International Conference on 3D Digital Imaging and Modeling*. Québec, Canada. pp. 179–186.

Godin, G., M Rioux and R. Baribeau (1994). Three-dimensional registration using range and intensity information. In: *Proceedings of SPIE Videometric III*. Vol. 2350. pp. 279–290.

Greenspan, M. and G. Godin (2001). A nearest neighbor method for efficient icp. In: *Third International Conference on 3-D Digital Imaging and Modeling*. Quebec City, Canada. pp. 161–168.

Grossmann, E.G. (2002). Maximum Likelihood 3D Reconstruction From One or More Uncalibrated Views Under Geometric Constraints.. PhD thesis. Instituto Superior Técnico, Portugal.

Hilton, A. and J. Illingworth (2000). Geometric fusion for a hand-held 3d sensor. *Machine Vision and Applications* **12**(1), 44–51.

Huber, D and M. Hebert (1999). A new approach to 3-d terrain mapping. In: *International Conference on Intelligent Robotics and Systems*. pp. 1121–1127.

Huber, D. and M. Hebert (2003). Fully automatic registration of multiple 3d data sets. *Image and Vision Computing* **21**(7), 637–650.

Johnson, A. (1997). Spin-images: A Representation for 3-D Surface Matching.. PhD thesis. Carnegie Mellon University, USA.

Johnson, A.E. and M. Hebert (1999). Using spin images for efficient object recognition in cluttered 3d scenes. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **21**(5), 433–449.

Jost, T. and H. Hugli (2002). A multi-resolution scheme icp algorithm for fast shape registration. In: *First International Symposium on 3D Data Processing Visualization and Transmission*. pp. 540– 543.

Kim, S., C. Jho and H. Hong (2003). Automatic registration of 3d data sets from unknown viewpoints. In: *Worshop on Frontiers of Computer Vision*. pp. 155–159.

Krishnan, S., P.Y. Lee, J.B. Moore and S. Venkatasubramanian (2005). Global registration of multiple 3d point sets via optimization-on-a-manifold.. In: *Symposium on Geometry Processing*. pp. 187–196.

Kruppa, E. (1913). Zur ermittlung eines objektes aus zwei perpektiven mit innerer orientierung. *Sitz.-Ber. Akad. Wiss., Math. Naturw.* **IIa**(122), 1939–1948.

Levoy, M., J. Ginsberm, J. Shade, D. Fulk, K. Pulli, B. Curless, S. Rusinkiewicz, D. Koller, L. Pereira, M. Ginzton, S. Anderson and J. Davis (2000). The digital michelangelo project: 3d scanning of large statues. In: *Proceedings of the 27th annual conference on Computer graphics and interactive techniques*. New Orleans, Louisiana, USA. pp. 131–144.

Longuet-Higgins, H.C. (1981). A computer algorithm for reconstructing a scene from two projections. **293**, 133–135.

Lu, F. and E.E. Milios (1997). Globally consistent range scan alignment for environment mapping. *Autonomuos Robots* **4**(4), 333–349.

Masuda, T. (2001). Generation of geometric model by registration and integration of multiple range images. In: *Third International Conference on 3-D Digital Imaging and Modeling*. pp. 254–261.

Masuda, T. (2002). Object shape modelling from multiple range images by matching signed distance fields. In: *First International Symposium on 3D Data Processing Visualization and Transmission*. pp. 439–448.

Masuda, T., K. Sakaue and N. Yoyoka (1996). Registration and intergration of multiple range images for 3-d model construction. In: *ICPR '96: Proceedings of the 1996 International Conference on Pattern Recognition*. Vol. 1. Washington, DC, USA. pp. 879 – 883.

Matabosch, C., D. Fofi and J. Salvi (2005*a*). Reconstruction et recalage de surfaces en mouvement par projection laser multi-lignes. In: *9èmes journées ORASIS'05*. Puy-de-Dôme, France.

Matabosch, C., D. Fofi and J. Salvi (2005*b*). Registration of moving surfaces by means of one-shot laser projection. In: *2nd Iberian Conference on Pattern Recognition and Image Analysis*. Estoril, Portugal. pp. 145–152.

Matabosch, C., D. Fofi, J. Salvi and J. Batlle (n.d.). Registration of surfaces minimizing error propagation for a new hand-held laser scanner. *Pattern Recognition*.

Matabosch, C., E. Batlle, D. Fofi and J. Salvi (2006*a*). A variant of point-to-plane registration including cycle minimization. In: *Photogrammetric Computer Vision*.

Matabosch, C., J. Salvi and D. Fofi (2005*c*). A new proposal to register range images. In: *7th International Conference on Quality Control by Artificial Vision*. Nagoya, Japan.

Matabosch, C., J. Salvi and J. Forest (2003). Stereo rig geometry determination by fundamental matrix decomposition. In: *Workshop on European Scientific and Industrial Collaboration*. pp. 405–412.

Matabosch, C., J. Salvi, D. Fofi and F. Meriaudeau (2006*b*). A refined range image registration technique for multi-stripe laser scanner. In: *Machine Vision Applications in Industrial Inspection XIV*. Vol. 6070. SPIE. pp. 60700R–1–8.

Matabosch, C., J. Salvi, D. Fofi and F. Meriaudeau F. (2005*d*). Range image registration for industrial inspection. In: *Machine Vision Applications in Industrial Inspection*. San Jose, California, USA.

Matabosch, C., J. Salvi, X. Pinsach and J. Pagès (2004*a*). A comparative survey on free-form surface regist. In: *AVR04 1a Jornada de Recerca en Automàtica, Visió i Robòtica*.

Matabosch, C., J. Salvi, X. Pinsach and R. Garcia (2004*b*). Surface registration from range image fusion. In: *IEEE International Conference on Robotics and Automation*. New Orleans. pp. 678 – 683.

Nüchter, A., H. Surmann, K. Lingemann, J. Hertzberg and S. Thrun (2004). 6d slam with an application in autonomous mine mapping. In: *IEEE International Conference on Robotics and Automation*. Vol. 2. ICRA '04. pp. 1998 – 2003.

Pagès, J., J. Salvi and C. Matabosch (2003*a*). Implementation of a robust coded structured light technique for dynamic 3d measurements. In: *International Conference on Image Processing*. Vol. 1. Barcelona, Spain. pp. 1073–1076.

Pages, J., J. Salvi, C. Collewet and J. Forest (2005). Optimised de bruijn patterns for one-shot shape acquisition. *Image and Vision Computing* **23**(8), 827–849.

Pagès, J., J. Salvi, R. García and C. Matabosch (2003*b*). Overview of coded light projection techniques for automatic 3d profiling. In: *IEEE International Conference on Robotics and Automation*. Taipei, Taiwan. pp. 133–138.

Pagès, J., J. Salvi, R. García and C. Matabosch (2003*c*). Robust segmentation and decoding of a grid pattern for structured light. In: *1st Iberian Conference on Pattern Recognition and Image Analysis*. Mallorca, Spain. pp. 689–696.

Park, Soon-Yong and Murali Subbarao (2003). A fast point-to-tangent plane technique for multi-view registration.. In: *3DIM, 4th International Conference on 3D Digital Imaging and Modeling*. pp. 276–284.

Peng, X., Z. Zhang and H. J. Tiziani (2002). 3-d imaging and modeling - part ii: integration and duplication. *Optik - International Journal for Light and Electron Optics* **113**(10), 453–458.

Pollefeys, M., M.R. Koch, M. Vergauwen and Van Gool L. (2000). Automated reconstruction of 3d scenes from sequences of images. *Photogrammetry and Remote Sensing* **55**, 251–267.

Pulli, K. (1999). Multiview registration for large data sets. In: *3-D Digital Imaging and Modeling*. Ottawa, Ont., Canada. pp. 160–168.

Robinson, A., L. Alboul and M. Rodrigues (2004). Methods for indexing stripes in uncoded structured light scanning systems. *Journal of Winter School of Computer Graphics*.

Rousseeuw, P.J. and A.M. Leroy (1987). *Robust Regression and Outlier Detection*. Wiley. New York.

Rusinkiewicz, S. and M. Levoy (2001). Efficient variant of the icp algorithm. In: *3rd International Conference on 3-D Digital Imaging and Modeling*. pp. 145–152.

Salvi, J., C. Matabosch, D. Fofi and J. Forest (2007). A review of recent range image registration methods with accuracy evaluation. *Image and Vision Computing* **25**(5), 578–596.

Salvi, J., J. Pagès and J. Batlle (2004). Pattern codification strategies in structured light systems. *Pattern Recognition* **37**(4), 827–849.

Sharp, G.C., S.W. Lee and D.K. Wehe (2002). Icp registration using invariant features. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **24**(1), 90–102.

Sharp, G.C., S.W. Lee and D.K. Wehe (2004). Multiview registration of 3d scenes by minimizing error between coordinate frames. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **26**(8), 1037–1050.

Shiu, Y.C. and S. Ahmad (1989). Calibration of wrist-mounted robotic sensors by solving homogeneous transform equations of the form $ax = xb$. *IEEE Transactions on Robotics and Automation* **5**(1), 16–29.

Silva, L., O.R.P. Bellon and K.L. Boyer (2003). Enhanced, robust genetic algorithms for multiview range image registration. In: *3DIM03. Fourth International Conference on 3-D Digital Imaging and Modeling*. pp. 268–275.

Silva, L., O.R.P. Bellon and K.L. Boyer (2006). Multiview range image registration using the surface interpenetration measure. *Image and Vision Computing* **25**(1), 114–125.

Solomon, F. and K. Ikeuchi (1996). Extracting the shape and roughness of specular lobe objects using four light photometric stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **18**, 449–454.

Soucy, M. and D. Laurendeau (1995). A general surface approach to the integration of a set of range views. *PAMI* **17**(4), 344–358.

Stamos, I and M. Leordeanu (2003). Automated feature-based range registration of urban scenes of large scale. In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. Vol. 2. pp. 555 –561.

Stoev, S.L. and W. Strasser (2002). A case study on automatic camera placement and motion for visualizing historical data. In: *IEEE Visualization*. Boston, Massachusetts. pp. 545– 548.

Tarel, J.P., H. Civi and D.B. Cooper (1998). Pose estimation of free-form 3d objects without point matching using algebraic surface models. In: *Proceedings of IEEE Workshop on Model-Based 3D*. pp. 13–21.

Triggs, B., P. McLauchlan and A. Hartley, R. Fitzgibbon (2000). Bundle adjustment - a modern synthesis. In: *International Workshop on Vision Algorithms*. London, UK. pp. 298 – 372.

Trucco, E., A. Fusiello and V. Roberto (1999). Robust motion and correspondences of noisy 3-d point sets with missing data. *Pattern Recognition Letters* **20**(9), 889–898.

Turk, G. and M. Levoy (1996). Zippered polygon meshes from range images. In: *SIGGRAPH '94: Proceedings of the 21st annual conference on Computer graphics and interactive techniques*. Orlando, Florida. pp. 311–318.

Wong, L., C. Dumont and M. Abidi (1999). Next best view system in a 3-d object modeling task. In: *International Symposium on Computational Intelligence in Robotics and Automation*. pp. 306–311.

Woodham, R. (1978). Refectance map techniques for analyzing surface defects in metal castings. Technical report. Rapport technique AI-TR-457,MIT A.I. Laboratory.

Wyngaerd, Joris Vanden (2002). Automatic crude patch registration: Toward automatic 3d model building. *Computer Vision and Image Understanding* **87**, 8–26.

Zinsser, T. and H. Schnidt, J. Niermann (2003). A refined icp algorithm for robust 3-d correspondences estimation. In: *International Conference on Image Processing*. pp. 695–698.

Z.Zhang (1992). Iterative point matching for registration of free-form curves. Technical report. Technical Report of INRIA, Sophia Antipolis.