Universitat de Girona

CATADIOPTRIC STEREO BASED ON STRUCTURED LIGHT PROJECTION

Radu ORGHIDAN

ISBN: 84-690-1427-7
Dipòsit legal: GI-1230-2006

Department of Electronics, Computer Science and Automatic Control

PhD Thesis

Catadioptric Stereo based on Structured Light Projection

Thesis presented by Radu Orghidan,
to obtain the degree of:
European PhD in Computer Engineering.

Supervisors:
Dr. Joaquim Salvi
Dr. El Mustapha Mouaddib

Girona, December 2005

Acknowledgments

Probably the most important part of a Ph.D. dissertation is the the acknowledgments section. This is the only section that everyone who picks up your thesis will read. No matter if they will casually find your dissertation in the library or will browse it by curiosity, he or she will flip through the first few pages, looking for a juicy acknowledgments section. This is the place where to put either things that you really want to share with all the persons that will open this book or be very mysterious so that everyone wonders what you're hiding. I prefer to take advantage of this page and clearly name the persons and the extraordinary sequence of happenings that made this thesis possible.

First of all, I want to thank Dr. Joaquim Salvi, my thesis director, that introduced me to computer vision and initiated me in the art of writing articles. Without his support, both scientific and financial, this work wouldn't have been possible.

Special thanks to Dr. El Mustapha Mouaddib for offering me the opportunity to work in his laboratory and for sharing with me his experience. I had the chance to benefit of three stages in Amiens that turned into veritable milestones in the progress of my research.

Many thanks to Dr. Yasushi Yagi who kindly received me in his laboratory and patiently listened to my omnivision problems. I am grateful for his valuable advices and hints that undoubtedly pushed forward my research.

I can not forget the important help that Dr. José Gaspar generously offered me when I was struggling with the calibration.

I would like to express my gratitude to Dr. Joan Batlle that introduced me in the University of Girona and guided me during my first months through the bureaucratic labyrinth. I feel indebted with him for trusting in me in difficult moments.

There are more people that I have the pleasure to name in this section. I will name them in chronological order, as they appeared in my story. Actually, it all started when Joaquim Salvi proposed me to study the effects of projecting a laser light into a shinny surface and simultaneously observe the scene reflected by that surface. We quickly called it omnidirectional vision with embedded structured light and I was quickly called the guy with the mirrors. I guess I have to thank for this

fancy title to Josep Forest that always managed to make me smile even when the things were not at all as shinny as my curved mirrors.

During my first months in the research group I shared the office (but, unfortunately, not the passion for soccer) with Xevi Cufi and with Lluis Pacheco. I used to spend lots of time in the office of the department secretaries Mar Ortega and Sonia Soler, and later, with Marta Geli and Anna Renart testing their endurance with my questions in English and even worse with my poor catalan. Other favorite victims of mine were Josep Tomàs, Toni Verdú, Lluís Magí, Íngrid Azorín and Marc Rodríguez that I constantly pissed off with my technical demands. By that time I met Joan Martí, Jordi Freixenet, Jordi Pages, the Xevis (Armangue, Muñoz and Llado), Rafa Garcia and Pere Ridao. With all of them I spent instructive moments at the coffee machine and my casual catalan improved noticeably. During the next years I met Carles Matabosch, Arnau Oliver, David Raba, Anna Bosch, Silvana and Armagan. I especially thank Carles Matabosch for being there every time I needed his help no matter if I was asking for his opinion or for his handcraft ability.

I remember with pleasure all the people I met during my stages in Amiens. My visits to the city of Jules Verne wouldn't have bee the same without the company of Cedric, Josselin, Arnaud, Khaled, Cyril, Pascal, Asli, Gerard, the Cristians, Adrian, Daniela, Alex and Dennis. I can not forget our short trips around Amiens and the dinners that we had together. I am especially grateful to Pierre who kindly helped me with the weird configurations that I imagined for my experiments.

I feel gratitude for my family from Romania and Spain. Thanks to all of you for the advices and the unconditional support in all my decisions. It's great to receive love from an area larger than 3000 km!

Finally, I extend my deepest gratitude and love for my wife Cristina who stayed next to me during all these years from the very beginning of the thesis. Thank you Cristina for your endless patience and for the sacrifices you made for me.

Catadioptric Stereo based on Structured Light Projection

Abstract

Computer vision is the study of methods which allow computers to "understand" images in a similar way the humans do. Vision perception is enhanced when a large field of view is available. This thesis is focused on the visual perception of depth by means of omnidirectional cameras. The 3D sensing is obtained in computer vision by means of stereo configurations with the drawback of feature matching between images. The solution offered in this dissertation uses structured light projection for solving the matching problem.

The objectives of the thesis lead in a first step to a survey on omnidirectional vision systems. The survey reveals the main trends for obtaining omnidirectional images and brings valuable guidelines for the choice of a camera depending on the application requirements. Then, the sensor design is addressed and we present the particular stereo configuration of the proposed sensor formed by two components: the omnidirectional camera and the laser projector. An accurate model is obtained by a careful study of both components of the sensor. Since the model parameters are difficult to measure directly, a set of calibration methods have been developed and validated by experimental results.

Several higher level algorithms were carried out using the sensor parameters obtained by calibration. These algorithms were finally used in two real applications. The first application was focused in modelling the scene geometry for an accurate environment reconstruction while the second application was designed for the interaction with a human operator charged with surveillance tasks.

The results obtained are encouraging and prove that the sensor can be used in depth perception applications such as scene modelling, pipe inspections, robot navigation, etc.

Catadioptric Stereo based on Structured Light Projection

Resum

La visió per computadora és l'estudi de mètodes que permeten als ordinadors comprendre les imatges d'una manera semblant als humans. La percepció per visió és millorada quan es pot gaudir d'un camp de visió ampli. Aquesta tesi es concentra en la percepció visual de la profunditat amb l'ajuda de càmeres omnidireccionals. La percepció 3D s'obté generalment en la visió per computadora utilitzant configuracions estèreo amb el desavantatge del cost computacional elevat a l'hora de buscar els elements visuals comuns entre les imatges. La solució que ofereix aquesta tesi és l'ús de la llum estructurada per resoldre el problema de relacionar les correspondències.

Els objectius de la tesi han determinat en primer lloc la realització d'un estudi sobre els sistemes de visió omnidireccional. L'estudi descobreix les principals tendències per a la obtenció d'imatges omnidireccionals i dóna directrius per poder escollir una càmera en funció dels requisits de l'aplicació desitjada. El següent pas és el disseny del sensor. S'han avaluat vàries configuracions estèreo utilitzant una càmera omnidireccional i un projector làser i, finalment, se n'ha escollit la millor. El model matemàtic del sensor s'obté estudiant ambdós components. Els paràmetres del model són difícils de mesurar directament i, en conseqüència, s'ha desenvolupat una sèrie de mètodes de calibració que s'han validat mitjançant experiments de laboratori. La interacció amb el sensor es realitza amb l'ajuda d'algoritmes d'alt nivell que es basen en els paràmetres del model. Aquests algoritmes s'han utilitzat en dues aplicacions reals. La primera proporciona el model 3D de l'escena enriquit per la textura dels objectes. La segona aplicació ha estat dissenyada per utilitzar la càmera omnidireccional per a la interacció amb un operador encarregat de supervisar una àrea de seguretat.

Els resultats obtinguts són prometedors i demostren que el sensor pot ésser utilitzat en aplicacions per a la percepció de la profunditat com serien el modelatge de l'escena, la inspecció de canonades, navegació de robots, etc.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The capacity to capture large fields of view is often encountered in the vision systems of the animals. This feature motivated researchers to study and to replicate the systems from nature. This chapter is aimed to offer an overview on vision from a sensorial point of view and on panoramic vision sensors from the perspective of robotics. The reader will find the following sequence of arguments: among the senses, vision is usually the best choice; among the vision sensors, the ones that replicate nature models are the most versatile; among the vision sensors inspired by nature the omnidirectional ones are the most suitable for 3D reconstruction and navigation tasks. The reader will also find a classification of the most common omnidirectional sensors together with a description of their characteristics. In addition, the structured light technique and its applications are introduced. This chapter continues with a presentation of the context of the thesis and the motivation that was at the base of this research work. Finally, the objectives of the research are listed and the chapter ends with an outline of the thesis.

1.1 Panoramic vision

Vision is certainly the most important of our five senses. It is a complicated process that requires numerous components of the human eye and brain to work together. An essential component of the vision system, either human, animal or artificial, is the vision sensor that gathers data from the environment. Because of our sensors, the eyes, human vision has a very limited field of view (FOV) compared to other

visual organs found in nature. The massive computational power offered by the cerebral cortex enables humans to smoothly perform navigation and recognition tasks despite the lack of visual information. During their life, other animals also need to find their way in unpredictable environments, to recognize food or to avoid predators among others vital tasks. Nevertheless, insects brain, for example, has 10^5 to 10^6 neurons compared with the 10^{11} of a human brain so they dispose of substantially less processing power. The visual information is processed by insects using a neural network which is a biological onboard computer that weights less than one milligram and has the size of a pin-head [Franceschini *et al.*, 1991]. It is plausible that the memory and the computational capacity of such a tiny computer are lower than the ones provided by the powerful computers of our days. However, humans are not yet capable to build similar view systems. Therefore, it is logic to think that the performances of these perfect flying systems are improved by the special construction of their eyes.

The majority of arthropods benefit from a wide field of view given by their compound eyes, as shown in Figure 1.1. A compound eye is a visual organ that consists of between 12 and 1,000 *ommatidia* which are tiny sensors that distinguish between brightness and darkness, and sometimes can detect color. The image perceived by the arthropod is a combination of inputs from the numerous ommatidia, which are oriented to point in slightly different directions in order to cover a large field of view. In contrast to other eye types, there is no central lens or retina, resulting in poor image resolution; however, it can detect fast movement and, in some cases, the polarization of light.

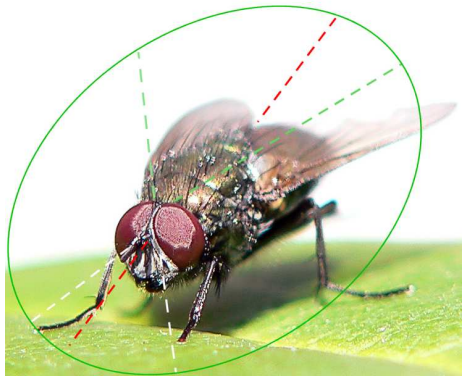
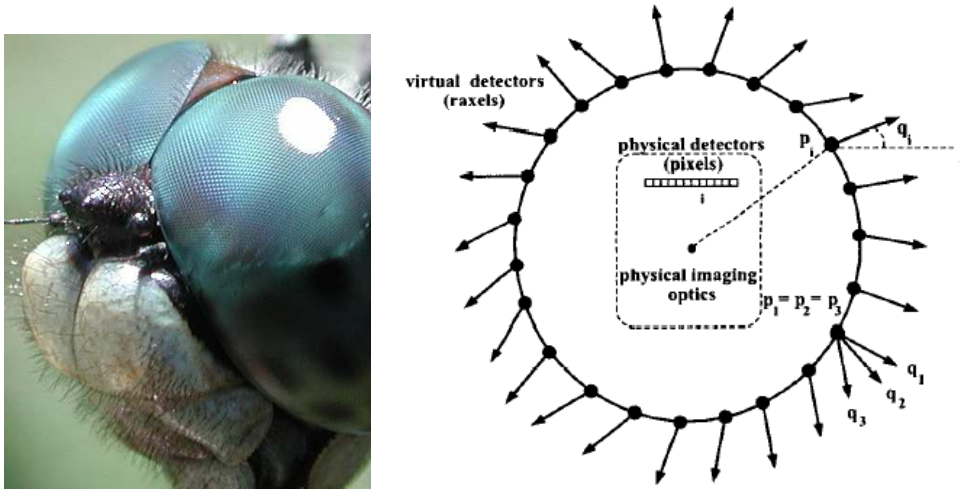


Figure 1.1: The azimuthal filed of view of a fly is almost 360 degrees. *Courtesy of Adrian Dusa http://adriandusa.smugmug.com/*

Three main classes of compound animal eyes with panoramic vision ability have been identified: diurnal and nocturnal insects and some species of crustaceans [Benosman and Kang, 2001]. While the eyes of the diurnal insects are using photoreceptive cells that cover a maximum vision area, the eyes of the nocturnal insects focus all the incident rays into one focal point which enhances the night vision. In the compound eyes of diurnal arthropods, each ommatidium is separated from its neighbours by pigmented, or iris, cells under all conditions of illumination. As a result, the rhabdom[1] of each ommatidium receives light only through its own corneal lens; light from the lenses of other ommatidia is blocked by the pigment. In the compound eyes of nocturnal arthropods, the rhabdoms are deep within the eye, far from the cornea and crystalline cone which creates a gap between the cornea and the rhabdom. The result is an image at half the radius of the eye, i.e. at the tips of the rhabdoms. Therefore, putting together many such ommatidia sensors in a controlled manner leads to an extremely flexible and functional imaging system. The compound eye of a dragon fly, shown in Figure 1.2, is a complex vision system that might have inspired the sensor with similar structure and behavior proposed by Grossberg and Nayar [Grossberg and Nayar, 2001].

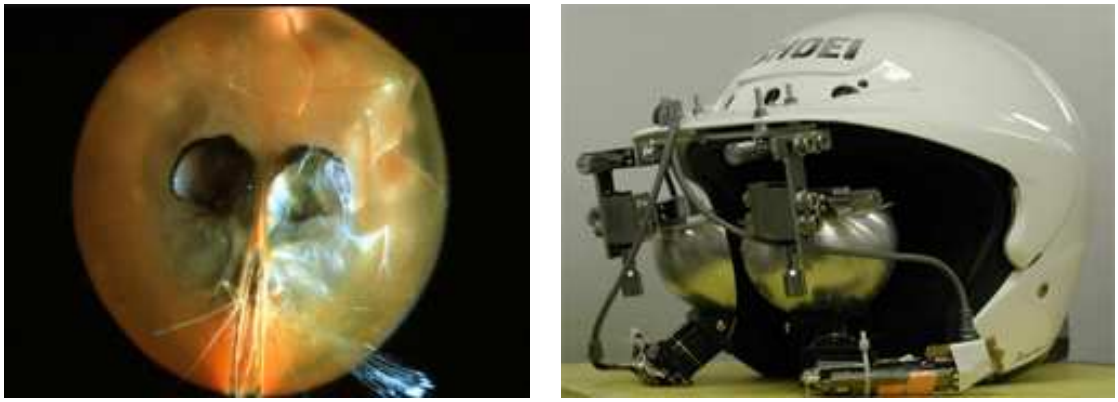


(a) (b)

Figure 1.2: Compound insect eye vs an abstract imaging system. (a) Compound eye of a dragonfly. *Courtesy of David L. Green* (b) An imaging system modelled as a set of light sensors on a sphere. *Proposed by Grossberg M.D. and Nayar S.K*

[1] A transparent rod in the center of each ommatidium in the compound eye of an arthropod.

The crustaceans eyes are composed by combinations of mirrors and refracting surfaces that focus the light at different points on the retina. The ostracod Gigantocypris shown in Figure 1.3 (a) is a crustacean with large reflecting eyes having a structure similar to optics used for telescopes. Its eyes are adapted to the very dim light conditions found in the oceans depths. The parabolic mirror that composes its eyes works in a similar way to the parabolic mirrors used by the Head Mounted Display built by Nagahara et al. [Nagahara *et al.*, 2003], presented in Figure 1.3 (b).

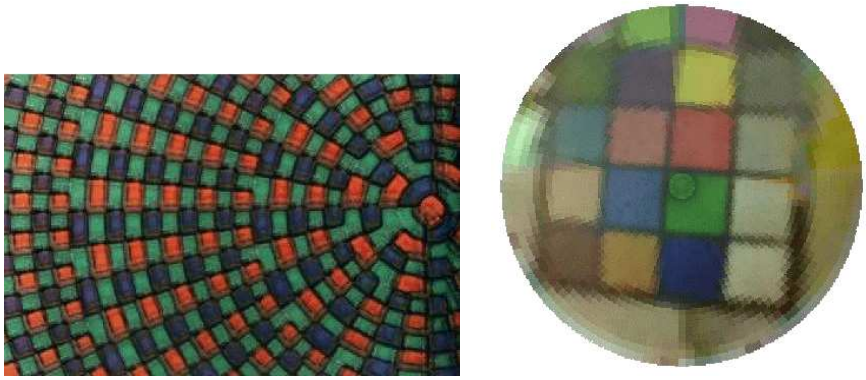


(a) (b)

Figure 1.3: Gigantocypris versus Head Mounted Display. (a) The ostracod Gigantocypris is a crustacean with large reflecting eyes that work as catadioptric devices. (b) Head Mounted Display for immersive vision.

The design of new vision sensors was inspired not only by animal eyes but also by the human eye. The eye's sharpest vision occurs when light is focused on the small part of the retina called the fovea centralis. This region has a higher density of photosensitive cells than anywhere else in the retina. The space-variant visual sensors [Sandini *et al.*, 2000] use the analogy with the human eye: the CCD of such a sensor is composed by photocells distributed as a function of the eccentricity. The density of the photocells is higher in the center which means that the sensor resolution is high in the central region and decreases monotonically towards its bounds as shown in Figure 1.4. A retina-like configuration makes a compromise between resolution and field of view, reduces the volume of information transmitted and processed by the central unit and thus requires less computational power.

As we saw in the above examples, people realized that the models of current cameras are too simple for many of the tasks required nowadays. Actually, the off-



(a) (b)

Figure 1.4: Space-Variant Visual Sensor with Color Acquisition (SVAVISCA). (a) Image obtained with the SVAVISCA sensor. (b) Photocells distribution.

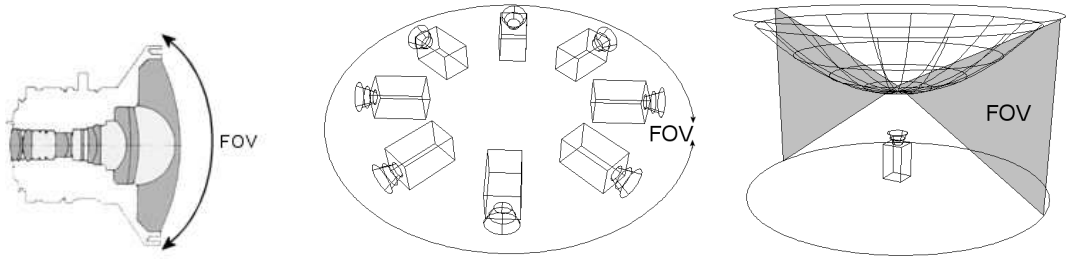
the-shelf photographic cameras are variations of the first pinhole cameras designed and built over one hundred years ago. The traditional cameras are designed to capture images to be printed in paper and not for pattern recognition, robot navigation or surveillance tasks. Therefore, an important number of algorithms and techniques have been developed in order to overcome the problems risen by perspective cameras. The study of the autonomous systems found in nature suggested that by adapting the design of the vision sensors to the desired tasks can spare the effort of processing the images obtained by means of traditional cameras.

1.2 Omnidirectional vision sensors

We call omnidirectional sensor or omnidirectional camera a vision sensor that has a full (360 degrees) azimuthal FOV. In literature these sensors are also known as panoramic because they actually do not cover the whole sphere of vision. A common classification of existing omnidirectional sensors is related to the technology applied in their fabrication, see Figure 1.5. Thus, for enhancing the field of view of conventional cameras researchers used:

- Special lenses.
- Multiple image acquisition systems by means of rotating cameras or using structures of many cameras with complementary fields of view.

- Catadioptrics which are combinations of cameras and mirrors (planar, hyperbolic, parabolic, spherical, dual convex and conic).



(a) (b) (c)

Figure 1.5: Omnidirectional sensors classification with respect to the technology applied in their fabrication. (a) Special lenses. (b) Multiple image acquisition systems. (c) Catadioptrics.

If the scene is observed from a single point in space the sensor possesses a unique center of projection i.e a Single View Point (SVP). The SVP is one of the most used classification criteria of cameras with large field of view [Yagi, 1999; Svoboda and Pajdla, 2000].

Every type of omnidirectional sensors is analyzed in the following sections.

1.2.1 Special lenses

Fish-eye lenses are imaging systems with a very short focal length which gives a hemispherical field of view, see Figure 1.6 (a). The obtained images benefit of a good resolution on the center but have poor resolution on the marginal region, as shown in Figure 1.6 (b). In human vision, a central zone of the retina, called the fovea centralis, provides high quality vision while the peripheral region deals with less detailed images. Therefore the vision acquired by a fish-eye lens is somehow similar to human vision from the point of view of resolution distribution. In addition, fish eye lenses introduce radial distortion which is difficult to remove. Another important drawback of the fish-eye lenses is their lack of a SVP having instead a locus of viewpoints, known as "diacaustic" [Grossberg and Nayar, 2001].

(a) (b)

Figure 1.6: Fisheye lenses. (a) Nikon Fisheye Converter FC-E8 mounted on a Coolpix 990. (b) Image captured using fisheye optics.

Despite the above mentioned drawbacks, the large vision field offered by the fish-eye lenses made this type of dioptric elements an attractive choice for several researchers. In 1993, Basu et al. [Basu and Licardie, 1993] published a study providing two mathematical solutions for compensating the fish-eye distortion. The fish eye transformation is defined as a variable resolution mapping function that generates an image with high resolution in the central area and nonlinearly decreasing resolution towards the periphery. It was also emphasized that the variable resolution of the images can be turned into a useful feature as it reduces the quantity of information to be processed keeping in the same time a large field of view. Moreover, combining variable resolution with active vision can simplify some vision problems.

Precise measurements are obtained by means of an accurate mapping function that links world points to their projections on the image plane. Shah and Aggarwal [Shah and Aggarwal, 1994*b*] designed a set of experiments for cameras with fish-eye lenses and carried out a calibration method that determines the mapping function and the parameters of the model, namely the distortion coefficients, the optical center, the effective focal length and the one pixel width on the image plane. Relying on the results of their calibration algorithm, Shah and Aggarwal [Shah and Aggarwal, 1994*a*] performed depth estimation using two fish-eye lenses in an indoor structured environment. After removing the distortion they used a line detector based on vanishing points to extract the main features of the image. Then, by means of a stereo technique, the one to one correspondences between the two images were detected. Finally, the range detection was performed by means of triangulation.

In 1997, Xiong and Turkowski [Xiong and Turkowski, 1997] designed a self calibration algorithm for fish-eye lens cameras. A remarkable feature of Xiong's algorithm is that the camera calibrates itself while recording images which makes that the quality of the information extracted from the image improve iteratively. The projection function of the scene 3D points onto the image plane was approximated by the so called "equidistance" model which has the advantage that allows an arbitrarily large field of view. Using four overlapped pictures the 34 parameters of the model are found by minimizing the projection function.

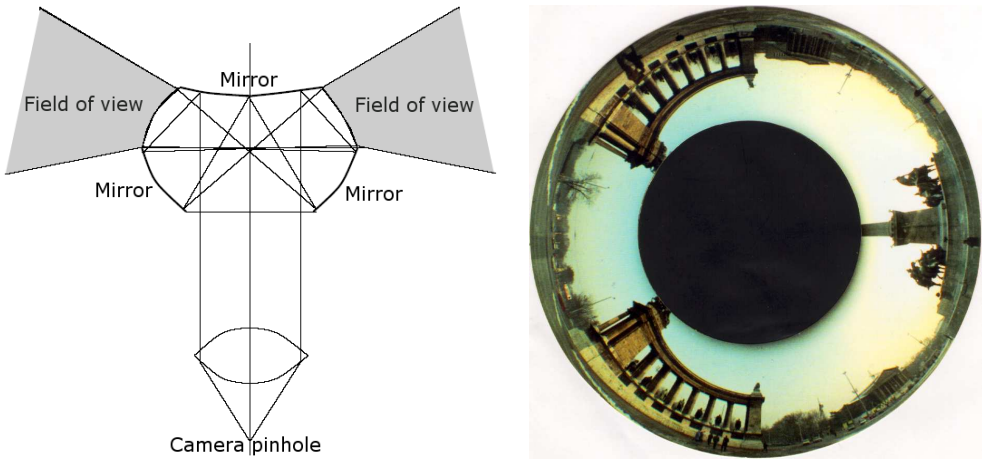
A binocular fish-eye lens active vision system was developed later on by Kurita et al. [Kurita *et al.*, 2000]. They realized a real-time gaze control system that had the capability to detect and follow objects moving within the field of view. Due to the non-linear resolution perception of the world the fish eye lenses are suitable for this kind of applications since movements in the scene can be detected with lower computation costs than using traditional cameras, i.e. with uniform resolution. Furthermore, the fish-eye lens cameras benefit of higher resolution in the center so, if the camera is set to image the moving object in the central region, it can be perceived clearly.

Another study about the correction of fish-eye lens distortion was issued by Bräuer-Burchardt and Voss [Brauer-Burchardt and Voss, 2001]. They compensate the image distortion using a single image taken with a non calibrated camera. The method calculates the radial distortion coefficient by inspecting the arcs of circles in the image plane corresponding to straight lines from the scene.

In 2003, Bakstein and Pajdla [Bakstein and Pajdla, 2003] presented a full omnidirectional mosaicing with a fish-eye lens camera. Having a field of view larger than 180^{o}, the whole mosaic is obtained by rotating the camera. This paper introduces the model of a camera with spherical retina and defines its projection function. Besides, the calibration algorithm and a set of experimental results are also provided.

A particular configuration, proposed by Greguss [Greguss, 1984], forms a cylindrical projection of the scene using the panoramic annular lens (PAL) with mirror surfaces shown in Figure 1.7 (a). The image provided by this optic is presented in Figure 1.7 (b).

Going over the main points discussed above we see that wide angle lens cameras can provide images suitable for a number of applications. However, the modelling of these cameras is a complicate process because of specific shortcomings such as



(a) (b)

Figure 1.7: Panoramic Annular Lens, by Greguss. (a) The PAL configuration. (b) Image of Budapest captured using the PAL optics.

the radial distortion, the low non-uniform resolution and the lack of a unique view point which makes impossible unwrapping of the images.

1.2.2 Multiple image acquisition systems

High resolution panoramic views can be obtained using a mosaic formed by multiple images provided by a rotating camera. For instance, Peer and Solina [Peer and Solina, 2002] realized range measurements of the scene with a single camera attached to a rotating arm, as shown in Figure 1.8. A full omnidirectional image is obtained by putting the snaps together by means of a time consuming procedure which makes this method worthless in realtime applications.

The omnidirectional image construction is accelerated by means of cameras with a very narrow field of view. The camera spins at constant angular velocity while taking narrow, vertical scan lines from different images and joins them to form a panoramic view. A binocular omnidirectional vision sensor built with two line ccd cameras was proposed by Benosman et al. [Benosman *et al.*, 1996; Benosman and Devars, 1998]. The line scan technique alleviates matching computation and the proposed sensor can be used in real time applications. The horizontal resolution is

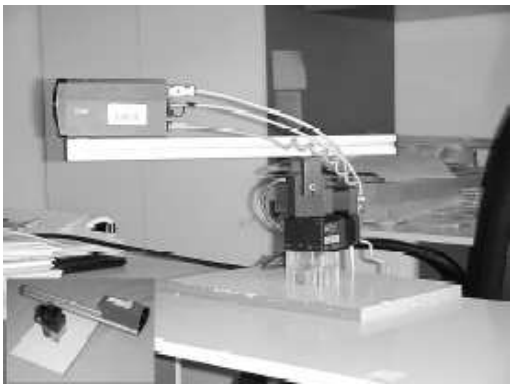


Figure 1.8: 3D depth measurement sensor using a camera mounted on a rotating robot arm. *Courtesy of Peer P. from the University of Ljubljana*

established by the angular speed of the rotation and not by the resolution of the camera. However, in some applications moving parts are not wanted or impossible to set up.

The compound eyes of insects have been a source of inspiration for a new type of omnidirectional systems. In 2000, Fermuller et al. [Fermuller *et al.*, 2000] developed an omnidirectional vision system compound from several cameras looking in different directions. Since each camera can be seen as an eye with a limited field of view, the structures resulting from putting together many cameras were called "eyes from eyes", see Figure 1.9.

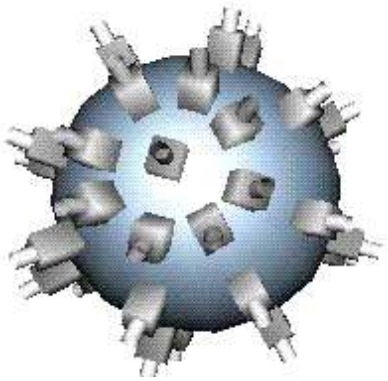
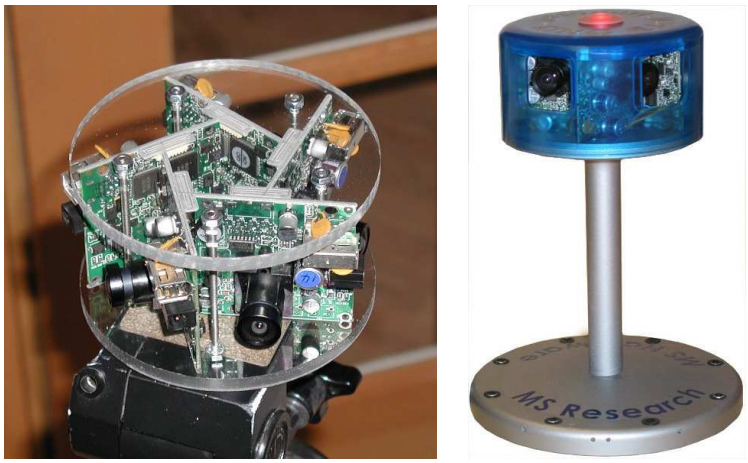


Figure 1.9: Eyes from eyes are structures of cameras looking in different directions built in order to obtain an omnidirectional vision field. *Courtesy of Fermuller, University of Maryland.*

Neuman et al. [Neumann *et al.*, 2002] studied the relationship between the cam-

era design and the quality of the visual information over the time. The complete information obtained by a vision sensor, formed by a camera or a structure of cameras, is described by the plenoptic function. By examining the time variation of the plenoptic function, Neuman et al. classified the vision sensors in relation with the stability and complexity of the computations necessary to estimate structure and motion. They obtained a hierarchy that starts with the pinhole camera while at the top level is placed the full field of view polydioptric camera.

A full omnidirectional field of view is not always necessary. Applications such as video-conferencing need only a limited vertical view angle. R. Cutler et al. [Cutler *et al.*, 2002] designed a model of panoramic camera which covers a full panoramic field of view by arranging the cameras in the same horizontal plane as shown in Figure 1.10 (a). Another model of this topology is presented in Figure 1.10 (b) where the system also contains an embedded microphone array. This kind of arrangement offers only the required part of the omnidirectional field of view reducing the number of cameras and, consequently, the manufacturing costs.



(a) (b)

Figure 1.10: (a) A ring of cameras with complementary fields of view. (b) A version of the ring of cameras that has an integrated microphone array in the base of the camera and a privacy shutter. *Courtesy of Ross Cutler (http://research.microsoft.com/ rcutler/)*

Still cameras with complementary fields of view overcome some problems of the moving cameras but the sensor becomes more difficult to calibrate since it involves several cameras and the corresponding image acquisition cards.

Summarizing, the advantage of using rotating cameras or configurations of sev-

eral cameras covering the scene offer very good image quality at a high resolution. However, this useful feature is payed-off by a set of drawbacks. In the case of rotating a camera, the part of the scene behind the image plane is always hidden so this type of cameras can not be used in real time applications. The omnidirectional view obtained from several partial images are difficult to setup, expensive and can contain discontinuities.

1.2.3 Catadioptrics

Another way to obtain omnidirectional images is to point a camera looking into a specially designed mirror. The combination of mirrors (catoptrics) and lenses (dioptrics) is known as catadioptrics.

In history, the use of catadioptric systems has been testified in the ancient Greece where the properties of the curved mirrors of focusing the light have been exploited for burning the enemy crafts [Geyer and Daniilidis, 2000*b*]. Catadioptrics were also used in the XVIIth century by Isaac Newton who developed a reflective telescope (the Newtonian Reflector) that used mirrors instead of lenses in order to correct the chromatic aberration (rainbow halo). Today there are three main types of commercial telescopes: the refractors, the reflectors and the Schmidt-Cassegrains. The last type, also known as catadioptrics, is a cross between the refractors and reflectors and was invented in the 1960s. The Schmidt-Cassegrains model uses a secondary mirror that bounces light through a hole in the primary mirror to an eyepiece as illustrated in Figure 1.11.

Observing the surrounding scene by means of reflecting surfaces is not a new idea and a classical example is the drawing of the Dutch graphic artist Escher (1898-1972), shown in Figure 1.12. However, what really triggered off the research on catadioptric systems was the use of modern cameras that provided a good quality image of the reflected scene.

An early use of a catadioptrics for a real application was proposed by Rees in 1970 [Rees, 1970]. He invented a panoramic television camera based on a convex, hyperbolic-shaped mirror. Twenty years later, researchers focused again their attention on the possibilities offered by the catadioptric systems, mostly in the field of robotics vision. In 1990, the Japanese team from Mitsubishi Electric Corporation

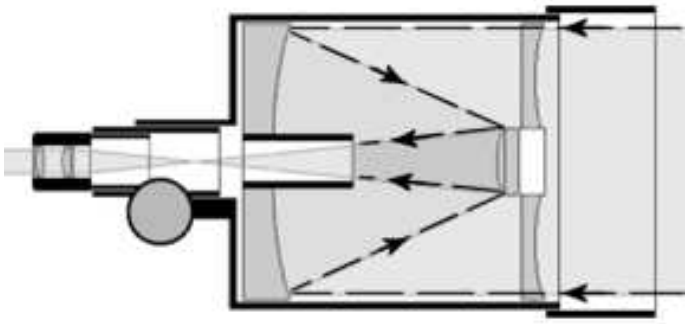


Figure 1.11: A Schmidt-Cassegrains telescope uses a combination of reflective and refractive optic elements to obtain a sharp image over a wide field.

Figure 1.12: Drawing by the Dutch artist M.C. Escher (1898-1972). The artist uses a spherical mirror to image the surrounding scene.

lead by Y. Yagi [Yagi and Kawato, 1990] studied the panoramic scenes generated using a conic mirror-based sensor. The sensor, named COPIS [2], was used for generating the environmental map of an indoor scene from a mobile robot. The conic mirror shape was also used, in 1995, by the researchers from the University of Picardie Jules Verne, lead by Mouaddib. Their robot was provided with an omnidirectional sensor, baptized SYCLOP [3], that captured 360° images at each frame and was used for navigation and localization in the 3D space [Pegard and Mouaddib, 1996].

Starting with middle nineties the attention given to omnidirectional vision and its knowledge base has grown continuously as the numbers of researchers involved in the study of omnidirectional cameras has increased. New mathematical models for catadioptric projection and, consequently, better performing catadioptric sensors have appeared.

Summarizing, the large interest generated by catadioptrics is due to their specific advantages when compared to other omnidirectional systems, especially the price and the compactness. For a detailed discussion about the theory and practical realizations of catadioptrics the reader is referred to the chapter 2 of this thesis.

1.3 Structured light

Structured light is a technique widely used to measure 3D surfaces by the projection of a sequence of light patterns. The main advantage of structured light is that the projected features are easily distinguished by the camera. The measured surface is emphasized by the apparent deformations of the imaged pattern from the projected one [Serrat *et al.*, 2001]. In order to avoid ambiguities while matching the projected features in the image the patterns can be codded. A surveys containing the most common techniques in coded light methods was written by Salvi et al. [Salvi *et al.*, 2004]. Structured light is used in dense range sensing, industrial inspection, object recognition, 3D map building, reverse engineering, fast prototyping etc.

The light patterns are projected by means of slide projectors, digital projectors or laser light emitters which forms different shape of patterns. Slides are difficult to use with calibrated systems because of the deformations caused by the functioning temperature of the projector lamp. On the other hand, digital projectors have

[2] COnical Projection Image Sensor
[3] Conical SYstem for LOcalization and Perception

focusing problems which limits the depth range of the measuring area. An example of the reconstruction of a horse head statue, performed by means of digital projection of coded light patterns is presented in Figure 1.13.

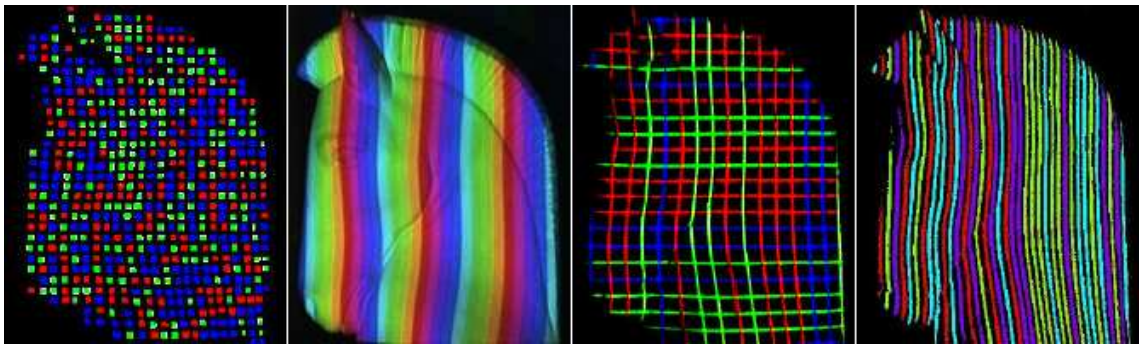


Figure 1.13: Reconstruction of a 3D shape using structured light. *Courtesy of Jordi Pages, University of Girona.*

Finally, monochromatic laser light offers attractive features since it permits the use of optical filters for a straight-forward image segmentation. Laser projectors are thin and can be placed inside compact devices. Uniform intensity laser projectors are especially useful for structured light applications, including industrial inspection, alignment and machine vision. Structured light lasers used in inspection minimize process variation by drawing attention to parts that do not conform to specifications. They can pick out vegetables with blemishes from food processor lines or can ensure that the right colored capsule goes in the correct bottle in drug packaging lines. Another laser application is alignment. In computer assembly, a laser system can help an operator determine if a computer chip is perfectly positioned on a circuit board. Machine vision is a combination of structured lighting, a detector, and a computer to precisely gather and analyze data. For example, it is used on robots as 3-D guiding systems to place or insert a part on a car, such as a windshield wiper or a door.

Visual range measurement is a task usually performed by stereo imaging systems. Depth is calculated by triangulation using two or more images of the same scene. The previously calibrated cameras are used for crossing the rays coming from the same point in the observed scene. However, the search of correspondences between images is generally a difficult task, even when taking into account epipolar constraints. A solution to this problem is offered by using the information from a structured light pattern projected in the scene. Usually this is achieved by a combination of a pattern

projector and a camera.

1.4 Context of this thesis

The research of this thesis started and has been carried out mostly within the 3D perception laboratory, part of the VICOROB[4] research group of the University of Girona (UdG). The VICOROB group is part of the Institute of Informatics and Applications (IIiA) of the UdG. The Institute investigation covers fields such as computer graphics, computer vision, robotics, distributed systems, communications, artificial intelligence and advanced control. The VICOROB group is composed of a total of 26 members between fulltime researchers and PhD students which are involved in research on underwater robotics and vision, mobile robotics, 3D perception and image analysis. A great deal of work has been done in areas such as object, pattern, color and shape recognition and tracking of mobile objects with immediate application in the control quality of industrial products, medical imaging and security-related systems. The robots developed by the group are endowed with computer vision systems with the aim to improve their autonomy and mobility in industrial and research applications. For instance, the Underwater Robotic Intelligent System (URIS) has been used in the inspection of a hydroelectric dam among other missions. The robot URIS is presented in Figure 1.14.

Figure 1.14: Underwater Robotic Intelligent System (URIS) built by the VICOROB group.

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

The 3D perception laboratory is involved in the study of the geometrical relations which allow the perception of the three-dimensional world by means of computer vision. The laboratory acquired a solid background in perspective camera modelling and calibration, epipolar geometry and fundamental matrix estimation. Depth perception is performed using perspective stereo techniques or structured light systems based on coded patterns. The systems realized in the 3D perception laboratory have been applied in object reconstruction, active visual servoing, camera localization and positioning and supply-chain visual inspection. The GRILL robot is a Pioneer 2 mobile robot equipped with a pair of cameras forming a stereo vision head, see Figure 1.15. It is used in robot localization and mapping tasks.

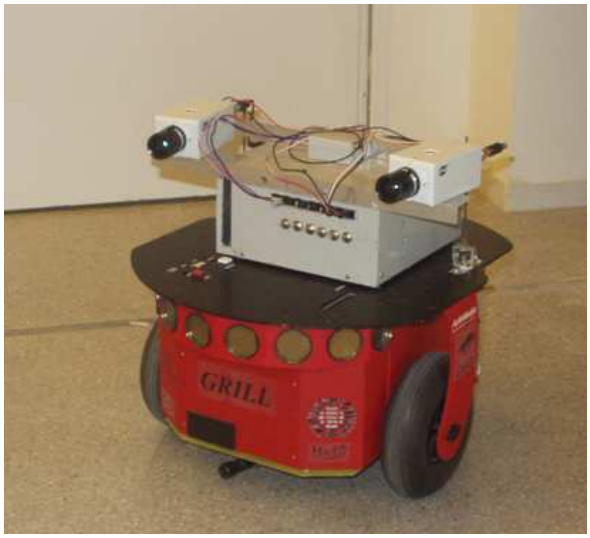


Figure 1.15: The GRILL robot was built at the 3D percetion laboratory.

In this context, the wide field of view offered by the omnidirectional cameras showed up as a possible solution to specific problems introduced by the use of traditional cameras. Moreover, the usefulness of structured light techniques and the challenge of their application to omnidirectional vision motivated us to start a research in the area of panoramic vision sensors.

The study of omnidirectional vision was the origin of a new collaboration with the laboratory of the Centre of Robotics, Electrotechnics and Automation (CREA) part of the University of Picardie Jules Verne (UPJV). The CREA is a a first order research centre with a wide experience in the field of omnidirectional vision. It is worth to say that the researchers of CREA leaded by Dr. El Mustapha Mouaddib are one of the pioneers in the use of omnidirectional vision sensors for robot perception.

The SYCLOP robot was built more than ten years ago and was used for localization of mobile robots, see Figure 1.16.

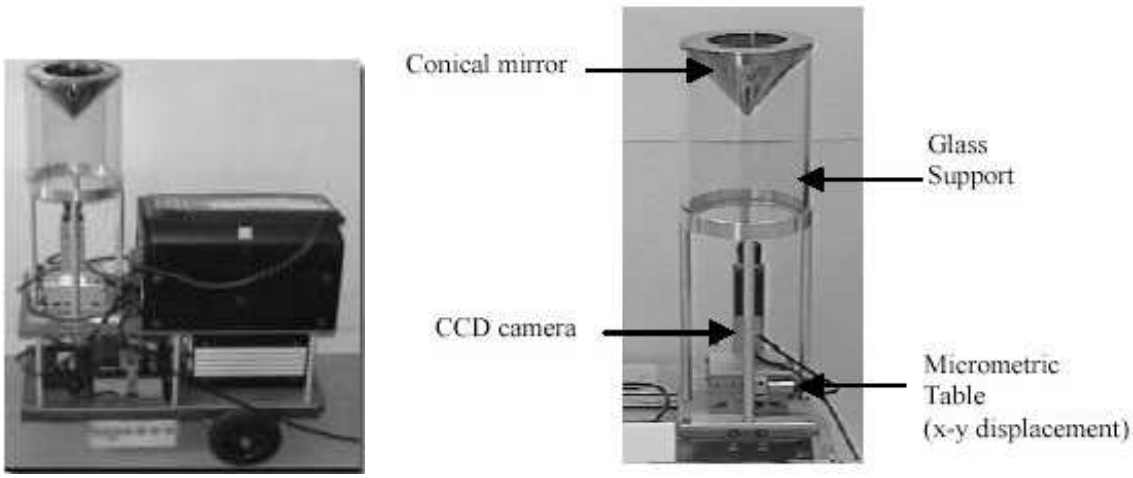


Figure 1.16: The SYCLOP robot was built at the CREA laboratory. (a) Full view of the mobile robot. (b) The conic mirror based omnidirectional camera used by the robot.

The first contact with CREA has been realized by Dr. Salvi who carried out his PhD thesis coordinated by Dr. J. Batlle and by Dr. E.M. Mouaddib. Therefore, the collaboration with Dr. Mouaddib appeared as a natural continuation of the relations established between the two research centres and as an excellent opportunity for the VICOROB group to take advantage of the knowledge base of the CREA laboratory in omnidirectional vision. The first step was a three months research stage in 2002 where the fundamentals of omnidirectional vision have been studied. Two subsequent stages of four and one month during 2004 and 2005 respectively completed the fruitful collaboration between the two laboratories.

1.5 Motivation of the research

The use of robots is an attractive option in places where human intervention is too expensive or hazardous. Robots have to explore the environment using a combination of their onboard sensors and eventually process the obtained data and transform it in useful information for further decisions or for human interpretation. Therefore, it is critical to provide the robot with a model of the real scene or with the ability to

build such a model by itself. The construction of a visual, non intrusive environment model motivated our research.

The omnidirectional vision enhances the field of view of traditional cameras by using special optics and combinations of lenses and mirrors. Besides these obvious advantages offered by a large field of view, in robot navigation the necessity of employing omnidirectional sensors also stems from a well known problem in computer vision: the motion estimation algorithms can sometimes mistake a small pure translation of the camera for a small rotation, and the possibility of error increases if the field of view is narrow or the depth variations in the scene are small. An omnidirectional sensor can reduce this error since it receives more information for the same movement of the camera than the one obtained by a reduced field of view sensor.

However, the omnidirectional vision sensors have a major drawback which is their lower resolution compared to perspective cameras. This shortcoming affects the correspondence search since it becomes more difficult to find details in the images.

Structured light based techniques are a particular case of stereo vision used to solve the correspondence problem in which one of the cameras is replaced by a pattern projector [Salvi *et al.*, 2004; Pagès and Salvi, 2004]. The use of this technique is similar to placing visible landmarks in the scene so that image points can be identified and matched faster and more accurately. The extensive experience accumulated by our group in this field during the last years motivated us to rely on the use of structured light.

Therefore, the simultaneous use of both omnidirectional vision and structured light can overcome the shortcomings of each technique gathering their own advantages. This powerful combination was the idea that triggered our investigation.

1.6 Objectives

The main objective of this research work is to obtain an omnidirectional sensor that provides 3D information using structured light. The proposed sensor must combine the qualities of omnidirectional vision and of the structured light projection for solving the matching problem.

A compulsory step for obtaining the sensor is the study of the wide range of omnidirectional vision sensors. A survey in this direction is fundamental in order

to compare the features of the different omnidirectional cameras and eventually to chose the more appropriate model. Therefore, we consider this survey as the first goal of the thesis which provides a way to analyze the advantages and the drawbacks of the existing techniques.

The second goal should be the survey of structured light patterns but has been overpassed since a recent survey was recently published by the group in the context of another PhD thesis.

The third goal is the construction of the real prototype and comes as the next logical step to take once we establish which are the most appropriate components and configuration for the sensor.

Since we need to use the resulting prototype for further measurements and calculations it has to be calibrated. Therefore, the fourth goal is the study of the calibration of omnidirectional vision coupled with structured light emitters in order to make a contribution in the field of mapping in mobile robot navigation.

Finally, we would like to use the sensor in real applications which permits us to validate the prototype and its model. This new goal leads to a set of software tools that have to be built for processing the data stream provided by the sensor.

1.7 Thesis outline

This thesis deals with the omnidirectional cameras with 3D perception capabilities and proposes a new sensor that alleviates the specific problems of this topic. The dissertation is structured in five chapters. The first two chapters present notions about omnidirectional vision sensors while the new sensor is described in detail Chapter 3. Even though the subjects treated in Chapter 4 and Chapter 5 are strongly interconnected a separation between them was established since Chapter 3 deals with the theory involved in image processing while Chapter 5 is related to the practical issues in experimental results.

Chapter 1 presents an overview on omnidirectional vision. Several important qualities of large FOV cameras are discussed and a general classification is introduced with examples of the most common omnidirectional sensors. Then, an overview of the structured light technique is presented. This chapter also introduces the notion of single view point. The single view point is a very important property that converts the omnidirectional cameras into wide field of view pinhole cameras.

A state of the art of the catadioptric omnidirectional sensors is presented in **Chapter 2**, in which the sensors are classified by their optical properties and by the technology applied in their fabrication. The chapter starts with a discussion about the unified projection model. Then, the chapter presents a survey regarding the omnidirectional cameras which are compared in terms of resolution, SVP property, isotropy and their capacity of providing 3D measurements. The epipolar geometry for catadioptric cameras is also presented here.

The proposed omnidirectional stereo sensor is introduced in **Chapter 3**. The calibration of catadioptric sensors in general and of the proposed sensor in particular is addressed in this chapter. The sensor is calibrated in two steps and the calibration procedure of each component is followed by a set of experimental results.

The software tools that process the data provided by the proposed sensor are detailed in **Chapter 4**. Several algorithms that perform functions such as image segmentation, laser detection, the 3D scene reconstruction, line tracking or the image unwrapping among other are presented here.

Chapter 5 gives a technical view of the sensor and shows practical applications of the algorithms described previously. The sensor specifications and the accuracy estimation open this chapter. The chapter continues with the presentation of two applications in which the sensor is used for 3D scene modelling and for surveillance.

Finally, the conclusions of this thesis and the contributions of the entire research work presented in the dissertation are discussed in **Chapter 6**. The publications resulted from the research are also listed in here. Moreover, the financial support offered by different institutions is mentioned in this chapter. We also discuss new research paths derived from the results of this work.

Chapter 2

Catadioptric omnidirectional cameras

Catadioptric cameras act like analog computers performing transformations from 3D space to the 2D image plane through combinations of mirrors and lenses. The mirrors used in catadioptric cameras must cover the full azimuthal FOV and thus are symmetric revolution shapes, usually with conic profile. This chapter starts with an overview of the conic sections and continues with a survey of the existing catadioptric cameras. The cameras are first classified with respect to the SVP property and then according to the mirror shapes used in their fabrication. We focus on the omnidirectional cameras with depth perception capabilities that are highlighted among the other catadioptric configurations. Finally, we present the epipolar geometry for catadioptric cameras.

2.1 Introduction

As stated earlier, catadioptrics are combinations of mirrors and lenses arranged in order to obtain a wider field of view than the one obtained by conventional cameras. In catadioptric systems, the image suffers a transformation due to the reflection in the mirror. This alteration of the original image depends on the mirror shape, therefore a special care was given to the study of the mirror optical properties.

There are several ways to approach the design of a catadioptric sensor. One method is to start with a given camera and find out the mirror shape that best fits its

constraints. Another technique is to start from a given set of required performances such as field of view, resolution, defocus blur, image transformation constraints etc., then search for the optimal catadioptric sensor. In both cases a compulsory step is to study the properties of the reflecting surfaces.

Most of the mirrors considered in the next sections are *surfaces of revolution*, i.e. 3D shapes generated by rotating a two-dimensional curve about an axis. The resulting surface therefore always has azimuthal symmetry. Moreover, the rotated curves are *conic sections*, that is, curves generated by the intersections of a plane with one or two nappes of a cone as shown in Figure 2.1. For instance, a plane perpendicular to the axis of the cone produces a circle while the curve produced by a plane intersecting both nappes is a hyperbola. Rotating these curves about their axis of symmetry a *sphere* and a *hyperboloid* are obtained, respectively. For a plane which is not perpendicular to the axis and which intersects only a single nappe, the curve produced is either an ellipse or a parabola and the corresponding surfaces of revolution are called *ellipsoid* and *paraboloid*.

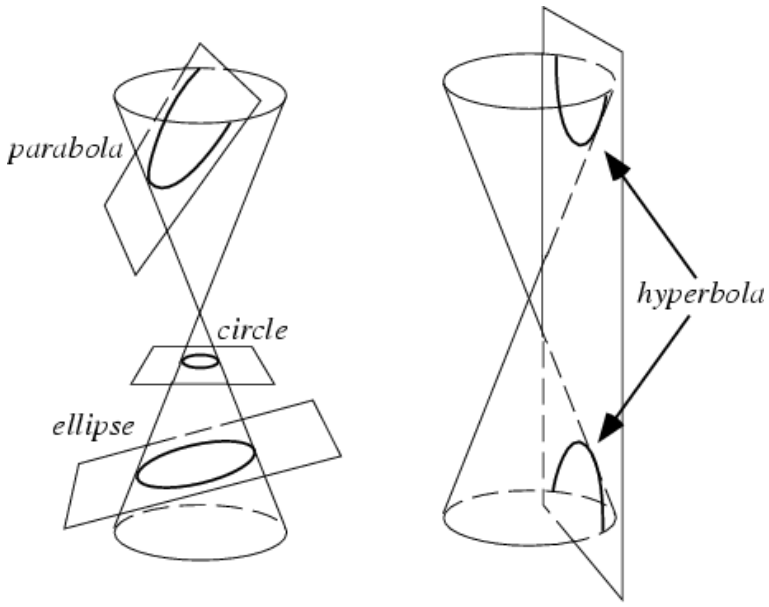


Figure 2.1: Examples of conic sections: the parabola and the ellipse, formed by the intersection of a single nappe of the cone with a plane which is not perpendicular to the axis; the circle, produced by the intersection of the cone with a plane perpendicular to its axis; and the hyperbola, formed by a plane intersecting both nappes of a cone. *Courtesy of http://mathworld.wolfram.com/*

Two usual terms when working with conical sections are the *focus* (whose plural

is *foci*) and the *directrix*.

The focus is a point related to the construction and properties of conic sections. For example, hyperbolas and non-circular ellipses have two distinct foci while the parabola has only one focus.

The directrix is the line which, together with the focus, defines a conic section as the locus of points whose distance from the focus is proportional to the distance from the directrix. Note that if the above ratio is $r = 1$, the conic is a parabola, if $r < 1$, it is an ellipse, and if $r > 1$, it is a hyperbola. Hyperbolas and non-circular ellipses have two associated directrices, each directrix being perpendicular to the line joining the two foci. The parabola has only one directrix.

In this chapter we will analyze the mathematical models and the physical realizations for the existing omnidirectional catadioptric vision sensors. The chapter is organized as follows: Section 2.2 addresses the of Single View Point (SVP) property of cameras also known as projection center or central point. The catadioptric sensors are classified according to different criteria and the classification is concentrated in the table contained in section 2.3. The class of SVP complying sensors is presented in section 2.4 and the non-SVP sensors are discussed in section 2.5. The catadioptric configurations that can measure depth are presented in section 2.6. An overview of the epipolar geometry of panoramic cameras is given in section 2.7. The chapter ends with the concluding notes presented in section 2.8.

2.2 The unified projection model

A true omnidirectional sensor is an imaging device capable to simultaneously watch the world in all directions. If the scene is observed from a single point in space the sensor possesses a unique center of projection i.e a Single View Point (SVP). Continuous scene monitoring can also be done from various locations or from a locus of view points. Consequently, the omnidirectional sensors can be classified upon their capacity to have or not a SVP.

When the images captured by a vision sensor are intended to be used for further calculations there must be defined a coherent mapping function. This function enables the reprojection of each 2D image point along the line that contains the imaged 3D object point. In the case of non SVP omnidirectional sensors the obtained images do not respect the perspective projection laws which leads to complicated mapping

functions. Because of this feature, the images produced by this kind of sensors are difficult to process and the traditional computer vision techniques turn out to be if not useless at least very hard to use. However, there are applications that suit the use of non-SVP sensors due to specific requirements such as resolution distribution, sensor size or price [Fiala and Basu, 2002*a*; Hicks and Bajcsy, 2000]. Recently, a new model of non-SVP sensors was proposed by Grossberg and Nayar [Grossberg and Nayar, 2001]. They designed a general imaging model composed by virtual sensing elements called ray pixels or raxels. This model does not have a SVP but a locus of points where the incoming rays converge, called caustic.

The SVP omnidirectional sensors are designed so that the captured image is formed only by rays passing through an unique point. In 1996 Bruckstein [Bruckstein and Richardson, 2000] compared several mirror shapes and studied thoughtfully the parabolic and hyperbolic profiles, the most frequent mirrors in SVP sensors. Later, Baker and Nayar [Baker and Nayar, 1998] explored the catadioptric sensors that produce SVP and have mathematically derived the class of single-lens single-mirror catadioptric configurations that poses this property. Using geometrical relations they found the equation of the single view point constraint. The solutions of this expression represent the entire class of mirrors that respect the constraint. While some of these solutions are impractical, others define the following set of feasible mirrors: planar, conical, spherical, ellipsoidal, hyperboloid and parabolic. Baker defined the single view point as the only point through which the catadioptric sensor measure the intensity of light. This is an extremely useful property that enables distortion-free reconstruction of panoramic images in a familiar form for the human users. Furthermore, a properly unwrapped omnidirectional image can be analyzed just as the images of traditional cameras.

The ideal omnidirectional sensor is a system having the capability to see simultaneously 360^o in all directions from a single point as if the user's eye is placed in the center of a sphere and receives the image through all the surface of the sphere. Actually, in mathematics, this is called "projection on a sphere". Geyer and Daniilidis [Geyer and Daniilidis, 2000*b*] introduced an unifying theory for all central catadioptric systems that covers perspective, elliptic, parabolic and hyperbolic projections. In the case of the catadioptric sensors that poses a SVP the image transformation can be modelled as a isomorphic projection from a symmetry axis of a sphere to a plane perpendicular to that axis. For instance, it is known from geometry that if this point is the north pole and the plane is tangent to the south

pole, a stereographic projection is produced. In fact, this is the case of the parabolic mirror shapes. It was proved [Geyer and Daniilidis, 2001] that in the elliptic and hyperbolic cases the corresponding point lies between the north pole and the center of the sphere while for the planar mirrors the projection point is located at the center of the sphere. A modified version of the unifying model for central catadioptric imaging was also proposed by Barreto [Barreto, 2003].

Besides the obvious advantage of unifying the geometrical model for the SVP catadioptric sensors, the mapping on the sphere opens new perspectives to calibration since it reduces the number of independent unknown parameters. As an application of this method, Geyer et al. [Geyer and Daniilidis, 1999] realized a calibration algorithm with an input of (at least) two sets of parallel lines. Recently, Barreto et al. [Barreto and Araujo, 2005] studied the projective geometry of lines in SVP catadioptric sensors and performed the calibration of a paracatadioptric sensor using the image of three lines. Such a calibrations are impossible to perform for classical perspective imaging systems. Hereafter, we present a state of the art of the catadioptric sensors.

2.3 Classification of catadioptrics

A lists of catadioptric configurations along with several common characteristics is presented in Table 2.1 . The sensors are first divided in three main categories according to the technology applied in their fabrication as explained in section 1.2. Each technology can be implemented in different manners or using different components. For instance, the catadioptric sensors can be built by combinations of single or multiple mirrors with different shapes placed in front of one or more cameras. The most common models of omnidirectional cameras were surveyed and classified. The following features are evaluated for each model: if it can measure depth from an unique image (I_1), if it was used for measuring depth using several images obtained either by moving the sensor or by using several cameras (I_n), its resolution, the SVP property and the capacity of producing a panoramic view at once (isotropy). Moreover, for each model a set of bibliographic references are provided in Table 2.2.

Many of the omnidirectional cameras included reviewed in Table 2.1 are presented in the following sections in which we'll highlight the models that have been used for 3D perception which is our subject of interest.

Table 2.1: Classification and comparison of omnidirectional cameras

Item	**Type**	**Construction**	**Depth**		**Resolution**	**SVP**	**Isotropic**
			I_1 [a]	I_n [b]			
	Special lenses						
1		Monocular fish eye	no	yes	low	no	no
2		Binocular fish eye	no	yes	low	no	no
3		Annular	no	yes	low	no	yes
	Multiple image acquisition						
4		Rotating cameras	no	yes	high	yes	no
5		Linear stereo	no	yes	high	no	no
6		Multiple cameras	no	yes	high	yes	yes
	Catadioptrics						
7	Single Camera	Conic mirror	no	no	medium	no	yes
8	with	SVP conic mirror	no	yes	medium	yes	yes
9	Single Mirror	Hyperbolic mirror	no	yes	low	yes	yes
10		Parabolic mirror	no	yes	low	yes	yes
11		Spherical mirror	no	no	low	no	yes
12		Special mirror [c]	no	no	low	no	yes
13	Single Camera	Planar mirrors	yes	yes	low	yes	no
14	with	Parabolic mirrors	yes	yes	low	yes	yes
15	Multiple Mirrors	Ellipsoidal mirrors	yes	no	low	yes	yes
16		Spherical mirrors	yes	yes	low	no	yes
17	Multiple Cameras	Pyramidal mirrors	yes	yes	high	no	yes
18	with	Conic mirrors	yes	yes	medium	no	yes
19	Multiple Mirrors	Hyperbolic mirrors	yes	yes	low	no	yes
20		Hybrid sensors [d]	yes	yes	high	no	no
21		Parabolic mirrors	yes	yes	low	no	yes

[a] Indicates if the sensor can perceive depth from a single image.
[b] Shows if the sensor was used for depth perception using several images.
[c] Specially designed mirrors that produce required projections.
[d] Various sensors combining omnidirectional and binocular vision.

Table 2.2: Bibliography related to the classification table

Item	**Construction**	**References**
1	Monocular fish eye	[Basu and Licardie, 1993; Shah and Aggarwal, 1994*b*; Xiong and Turkowski, 1997; Brauer-Burchardt and Voss, 2001; Bakstein and Pajdla, 2003]
2	Binocular fish eye	[Shah and Aggarwal, 1994*a*; Kurita *et al.*, 2000]
3	Annular	[Greguss, 1984]
4	Rotating cameras	[Peer and Solina, 2002]
5	Linear stereo	[Benosman *et al.*, 1996; Benosman and Devars, 1998]
6	Multiple cameras	[Fermuller *et al.*, 2000; Neumann *et al.*, 2002; Cutler *et al.*, 2002]
7	Conic mirror	[Yagi and Kawato, 1990; Pegard and Mouaddib, 1996; Southwell *et al.*, 1996*b*; Cauchois *et al.*, 1999; Mouaddib and Marhic, 2000; Drocourt *et al.*, 2002]
8	SVP conic mirror	[Lin and Bajcsy, 2001*b*]
9	Hyperbolic mirror	[Yamazawa *et al.*, 1993; Yamazawa *et al.*, 1995; Shih-Chieh *et al.*, 1998; Onoe *et al.*, 1998]
10	Parabolic mirror	[Nayar, 1997; Gluckman and Nayar, 1998; Sturm, 2000]
11	Spherical mirror	[Zhang *et al.*, 1991; Hong *et al.*, 1991; Winters *et al.*, 2000; Gaspar *et al.*, 2000]
12	Special mirror	[Gaspar *et al.*, 2002; Hicks and Bajcsy, 2000; Hicks and Perline, 2002; Hicks and Perline, 2001]
13	Planar mirrors	[Baker and Nayar, 1998; Nene and Nayar, 1998; Gluckman and Nayar, 1999; Gluckman and Nayar, 2002]
14	Parabolic mirrors	[Nayar and Peri, 1999; Mouaddib *et al.*, 2005]
15	Ellipsoidal mirrors	[Nene and Nayar, 1998]
16	Spherical mirrors	[Southwell *et al.*, 1996*a*; Fiala and Basu, 2002*b*; Fiala and Basu, 2002*a*]
17	Pyramidal mirrors	[Nalwa, 1996; Kawanishi *et al.*, 1998; Hua and Ahuja, 2001; Kar-Han *et al.*, 2004]
18	Conic mirrors	[Lin and Bajcsy, 2003; Spacek, 2005]
19	Hyperbolic mirrors	[Ollis *et al.*, 1999]
20	Hybrid sensors	[Yachida, 1998; Yagi and Yachida, 2002]
21	Parabolic mirrors	[Gluckman *et al.*, 1998]

2.4 Single View Point catadioptrics

At the modelling step it is desirable to use a SVP sensor essentially because it allows the creation of correct perspective images. The transformed image is easier to understand and can be processed by traditional computer vision techniques. The interest generated by SVP catadioptrics is reflected by the large number of publications dedicated to this type of vision sensors. The SVP configurations together with the corresponding mirror shapes are discussed in the next sections.

2.4.1 Planar mirrors

Even if the planar mirrors are found as a feasible solution by Baker and Nayar [Baker and Nayar, 1998], they lack of practical value since the vision field is not enhanced by the use of a single perspective camera and a planar mirror. However, a wider field of view can be obtained by carefully placing several planar mirrors in front of a perspective camera.

In 1996, Nalwa [Nalwa, 1996] built a pyramidal structure of mirrors that, placed in front of the cameras, captures a SVP omnidirectional field of view at high resolution, as shown in Figure 2.2. Currently this technology is used by the "Full View" company.

A variant of Nalwa's omnidirectional camera was built by Kawanishi et al. [Kawanishi *et al.*, 1998] in 1998. The Japanese team obtained high resolution stereo omnidirectional images by putting back to back two hexagonal pyramidal structures of planar mirrors and twelve cameras, one for each mirror. The sensor had a SVP as the cameras were carefully calibrated and set in place. Omnidirectional video sequences can be produced when the cameras are synchronized. Stereo views are also obtained in the overlapped FOV of the two back to back sensors. A similar approach to Nalwa's panoramic vision via a pyramidal structure of planar mirrors was given by Hua and Ahuja [Hua and Ahuja, 2001]. They proposed a double mirror pyramid that consists of two truncated mirror-pyramid and two horizontal layers of cameras. The result is a wider vertical field of view.

In 2004, Kar Han et al. [Kar-Han *et al.*, 2004] addressed the issues of analyzing, designing and building mirror-pyramid based cameras. The authors proposed a two-view mirror pyramid configuration using four conventional cameras, two for each view-point. The prototype has the capability to provide several view points

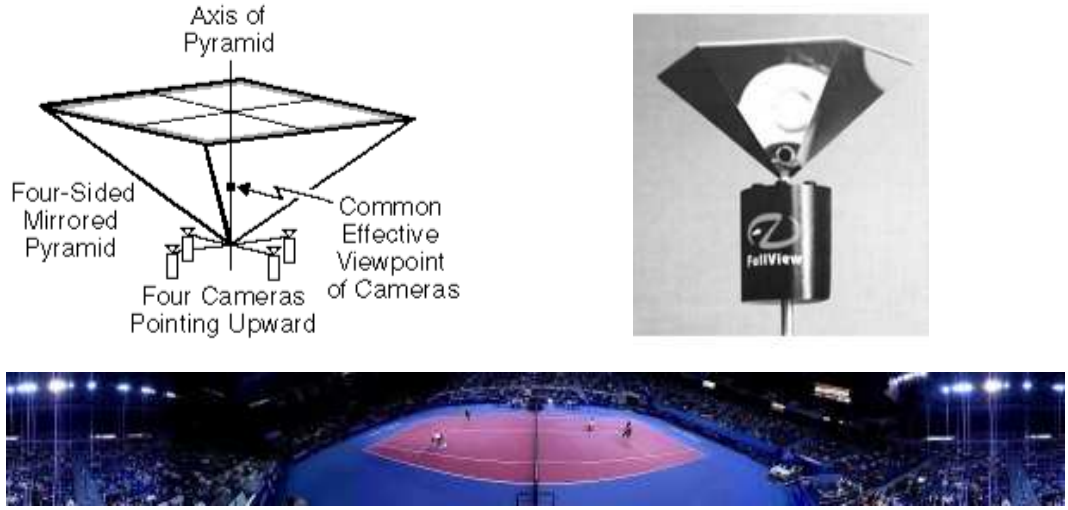


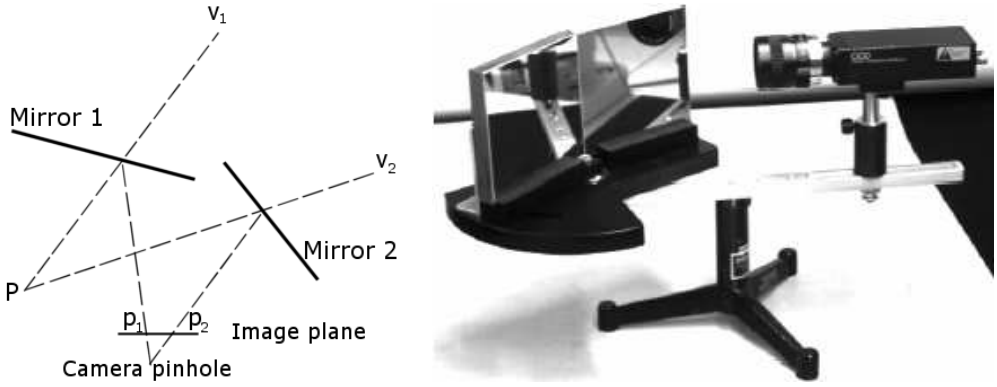
Figure 2.2: Nalwa's original design was composed by four mirrors arranged in a pyramidal configuration and the same number of cameras looking at each mirror. The prototype using six cameras is shown in the upper left corner and a panoramic image taken with Nalwa's camera is displayed at the bottom. *Courtesy of FullView® company (http://www.fullview.com/)*

using a single mirror pyramid. Hence, depth information can be retrieved at a given time instant by matching the features of the stereo images.

A 360º horizontal field of view is also achieved using a single camera looking into a convex mirror. In 1998, Nene and Nayar designed a stereo catadioptric system using all the class of SVP mirrors [Nene and Nayar, 1998]. A stereo system using two planar mirrors and a single camera was also proposed and analyzed. This approach has several important advantages compared with traditional stereo techniques due to the fact that there is only one camera involved and, consequently, only one set of parameters to calibrate. Moreover, the data acquisition is done faster than using two or more cameras.

In 1999, Gluckman and Nayar [Gluckman and Nayar, 1999] studied stereo catadioptric image formation using a single camera and planar mirrors (Figure 2.3). The stereo matching is completed using the epipolar geometry. Finally, they implemented a real-time stereo sensor and proved that it can provide good quality depth reconstruction. One year later, Gluckman and Nayar [Gluckman and Nayar, 2002] carried out a research about image rectification by means of catadioptric stereo sensors, including the ones using planar mirrors. A pair of stereo images is rectified if

the epipolar lines are aligned with the scan-lines of the images. Rectification needs to be done before stereo matching as it speeds up the search for correspondences. The geometry of the sensor must respect several constraints in order to ensure the creation of rectified images.



(a) (b)

Figure 2.3: (a) Stereo image formation with a single camera and two planar mirrors. A scene point P, reflected by mirrors $Mirror_1$ and $Mirror_2$ is imaged as if seen from two different viewpoints v_1 and v_2. (b) The prototype of the catadioptric stereo sensor with a single camera. *Courtesy of S.K. Nayar.*

2.4.2 Ellipsoidal mirrors

The ellipse is the curve on which are located all the coplanar points the sum of whose distances from two fixed and distinct points (the foci) is a given positive constant. The rotation of such a curve around the axis of symmetry is called ellipsoid.

The ellipsoidal mirrors can be used for building an imaging sensor with an unique point of view if the pinhole and the view point are placed at the two foci of the ellipsoid. This kind of mirrors are of a little practical use since the reflective part and the camera must be on the inner part of the surface providing an ineffective field of view. In 1998, a stereo system using ellipsoidal mirrors was proposed by Nene and Nayar [Nene and Nayar, 1998] (see Figure 2.4).

2.4.3 Hyperbolic mirrors

A hyperbola is a conic section defined as the locus of all points in the plane, the difference of whose distances from two fixed points (called the foci) is a given posi-

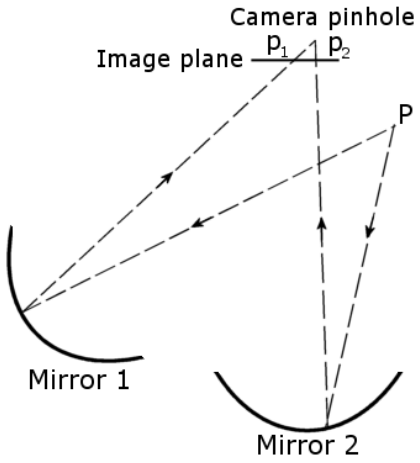


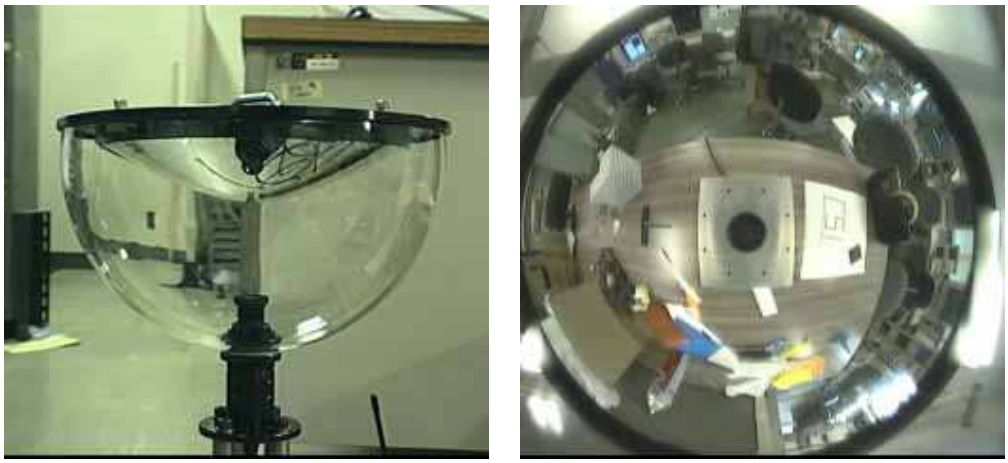
Figure 2.4: The projection of a scene point P into the two ellipsoidal mirrors $Mirror_1$ and $Mirror_2$ is imaged through the camera pinhole which is also the common focal point of the two conic sections.

tive constant. In the case of the hyperbolic mirrors, the SVP constraint is satisfied when the pinhole and the mirror view point are placed at the two foci of the hyperboloid. Therefore, the mirrors with this shape produce a large field of view but the components must be carefully set in place.

One of the first to use catadioptric systems with hyperbolic mirror was Rees who patented, in 1970, a prototype of a panoramic television viewing system.

In 1993, Yamazawa et al. [Yamazawa *et al.*, 1993] realized the prototype of HyperOmni, an omnidirectional vision sensor with a hyperbolical mirror, presented in Figure 2.5 (a). An image captured by this sensor is shown in Figure 2.5 (b).

The sensor was used for robot localization and, two years later, for obstacle detection [Yamazawa *et al.*, 1995]. In 1996, Yamazawa et al. [Yamazawa *et al.*, 1995] performed a mobile robot motion estimation in real-time by means of the omnidirectional optical flow. They exploited the fact that the radial component of the optical flow generated by the hyperbolic mirror has a symmetric behavior. Later, in 1998, Shih-Chieh et al. [Shih-Chieh *et al.*, 1998] used the HyperOmni and an ultrasonic sensor in a map building application. Another field where omnidirectional sensors can be used is the video surveillance. Onoe et al. [Onoe *et al.*, 1998], built in 1998, a monitoring system based on the omnidirectional video camera HyperOmni. The same sensor was used, one year later, by Yagi et al. to perform vision based navigation in man made environments. Such scenes typically contain straight lines



(a) (b)

Figure 2.5: The Omnidirectional sensor HyperOmni (a) The sensor is composed by a hyperbolical mirror in front of a perspective camera. (b) A scene imaged with HyperOmni. *Courtesy of Yagi Y. and the Robot Team of the University of Osaka*

that are projected as arcs of circle on the reflecting surface. An active contour model (ACM) was used for extracting smooth, continuous contours. In 2000, Yagi carried out an experiment for egomotion estimation and environmental map generation in a dynamic scene using the hyperbolic omnidirectional sensor. The method can detect a moving object and find occlusion and mismatching by evaluating estimation error of each object location.

The low resolution of most catadioptric sensors is caused by the mirror curvature that changes the number of reflecting points on the azimuth. Nagahara et al. [Nagahara *et al.*, 2001] proposed a method to improve resolution by means of a scene texture generation. Their approach creates a mosaic from a sequence of omnidirectional image captures and uses a texture weight table to rebuild the scene.

Ollis [Ollis *et al.*, 1999] simulated various configurations using two hyperbolic mirrors and one or two cameras. Depth measurements were obtained by triangulation. The correspondence problem was solved by window matching between both images by means of a correlation algorithm considering the curvature of the mirrors.

In 2002, the team from the Osaka University published the results obtained with two "multiple sensing robots" using the hyperbolic catadioptric sensor and two Pan-Tilt-Zoom cameras [Yagi and Yachida, 2002]. The movement in the scene was detected in real time from the omnidirectional video captured by the HyperOmniVi-

sion sensor. The resolution of the captured images was too low and did not permit the analyze of details in the scene. This limitation was overcome by the use of a stereo vision system built with the two high resolution cameras. The result was a hybrid sensor able to monitor the environment and, simultaneously, to acquire high resolution images. As a practical application, they developed a method for human silhouette and face tracking.

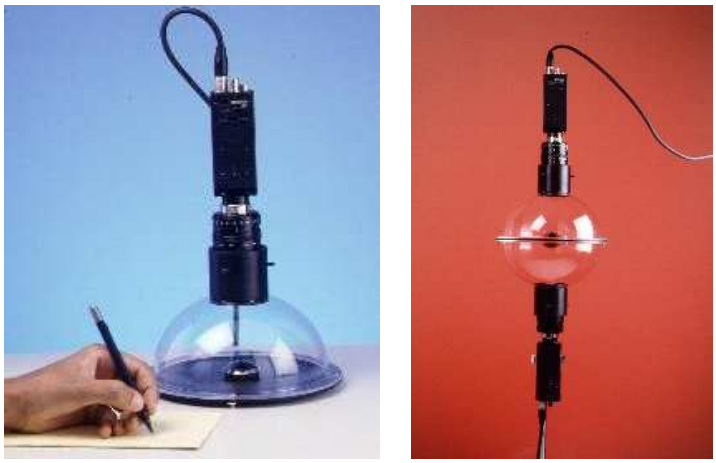
2.4.4 Parabolic mirrors

A parabola is the set of all points in the plane located at an equal distance from the directrix L, and the focus F. The focal parameter is the distance between the directrix and focus and is therefore given by $f = 2a$, where a is the distance from the vertex to the directrix or focus. The surface of revolution obtained by rotating a parabola about its axis of symmetry is called a paraboloid. However, the mirrors with this surface are commonly called parabolic mirrors.

The parabolic shape is a solution of the SVP constraint in a limiting case which correspond to orthographic projection. The parabolic mirror works in the same way as the parabolic antenna: the incoming rays pass through the focal point and are reflected parallel to the rotating axis of the parabola. Therefore, a parabolic mirror should be used in conjunction with an orthographic camera. A perspective camera can also be used if placed very far from the mirror so that the reflected rays can be approximated as parallel. Obviously, this solution would provide unacceptable low resolution and has no practical value.

One of the first to use parabolic mirrors was S.K. Nayar. In 1997 he presented the prototype of such a camera having a hemispherical field of view [Nayar, 1997], see Figure 2.6.

It was shown that, when using orthographic projection, the mappings between the image plane, the mirror and the scene are independent from the translation of the mirror with respect to the camera. This is an useful feature at the calibration stage since the above mentioned translation is not anymore a parameter to consider. Moreover, perspective image reconstruction is simplified. In 1998, Gluckman and Nayar [Gluckman and Nayar, 1998] used the parabolic-mirror based sensor to perform egomotion estimation . They adapted an algorithm for computing egomotion designed for planar perspective cameras and applied it to omnidirectional vision by mapping the optical flow field to a sphere. Later that year, Gluckman



(a) (b)

Figure 2.6: Omnidirectional cameras realized by Nayar in 1997. (a) Hemispherical field of view. (b) Full sphere field of view. *Courtesy of S.K.Nayar.*

et al. [Gluckman *et al.*, 1998] developed a panoramic stereo vision sensor. They aligned on the vertical axis two parabolic mirrors and the corresponding cameras. They calibrated the whole system by means of an iterative method and realized depth computation using epipolar geometry.

The parabolic mirrors have been used by Nayar and Peri [Nayar and Peri, 1999] for building a SVP folded catadioptric camera. They highlighted the advantages of using folded optics namely size reduction and the removal of undesirable optical effects due to the curvatures of large mirrors. Moreover, they proved that any catadioptric system that uses several mirrors with conics revolution shapes are geometrically equivalent with a system that uses a single conic. A dictionary of conics was presented and eventually a folded panoramic camera using two parabolic mirrors was realized.

In 2000, Strum [Sturm, 2000] set up a method for 3D reconstruction from a single panoramic image acquired with a parabolic-mirror omnidirectional sensor. The method gets as input information a set of constraints on the 3D structure, provided by the user. These constraints refer to coplanarity, parallelism and perpendicularity of objects in the image. The author described a simple calibration method by fitting a circle to the border of the scene image reflected by the mirror. The reconstruction algorithm uses the SVP property of this catadioptric configuration which permits the reprojection of the 2D image points back to the scene along a calculable line.

This feature together with the initial information enables the 3D reconstruction of points and planes. The presented method provides a reliable way of obtaining 3D models and can be easily adapted to other mirror shapes with known optical properties, such as the hyperbola.

When using parabolic mirrors with orthographic cameras the only constraint upon the camera position is that it must be aligned parallel to the optical axis of the mirror. The omnidirectional vision sensors with hyperbolic mirror produce SVP compatible images only when used in combination with a standard perspective camera whose pinhole is located at a precise point in space as described in the previous section. This features makes difficult to design the hyperbolic omnidirectional sensors. Moreover, the parabolic mirror system gives better results in the acquisition of non-blurred images and it can eliminate internal reflection by using a cylindrical or spherical glass around the mirror. This effect is due to the use of the telecentric lens, exploited to obtain the orthographic projection. However, the orthographic cameras are larger and more expensive than the conventional perspective ones.

2.5 Non - Single View Point catadioptrics

In section 2.4 we presented the advantages of SVP systems. Note that there are only a few shapes that enable SVP projections and their spreading out is certainly due to a simplification of the calculations at image processing stage as in the case of the pinhole model for cameras. While SVP and pinhole are valid approximations for certain applications there are other situations where these models are not sufficient. Another drawback of the SVP model is that it is based on a set of initial assumptions that are not always physically feasible and are difficult to check. For instance, in the case of a hyperbolic mirror with a perspective camera, the incoming rays do not converge anymore in an unique point if the focal point of the camera is not placed at the other foci of the hyperbola or if the camera and the mirror are misaligned. Moreover, there are other reflecting surfaces that do not possess a SVP but have attractive qualities such as a better resolution or a wider field of view than the shapes described in section 2.4. Relaxing the SVP constraint means modelling a catadioptric sensor so that the virtual center of projection is not singular but can be found on a 3D surface.

2.5.1 Conical mirrors

According to Baker et al. [Baker and Nayar, 1998], the conical mirrors have the single view point at the apex of the cone. This means that, in order to accomplish the SVP constraint, the camera's pinhole must be placed at the tip of the cone. In 1990, Yagi [Yagi and Kawato, 1990] built a catadioptric sensor, called COPIS (COnic Projection Image Sensor) and performed visual navigation in a real world environment using azimuth information extracted from the radial lines of the image. Later, in 1996, Mouaddib and Pegard [Pegard and Mouaddib, 1996] used a conical mirror in conjunction with a perspective pinhole modelled camera, as shown in Figure 1.16 to calculate the location of the robot. They used vertical lines in the world, which appear as radial lines in the image. If the positions of these lines are known then the lines can be seen as landmarks and can be used to estimate the robot position. They also found out that the mirror view point was placed on a caustics and not at a single location. The calibration of the system was published in [Cauchois *et al.*, 1999]. A method to obtain 3D using a single catadioptric camera was also presented by Drocourt et al. [Drocourt *et al.*, 2002] that used the SYCLOP sensor to perform simultaneous localization and map building. The data provided by the catadioptric camera was complemented by the dead-reckoning information obtained from the rigid translation of the sensor during the navigation.

In the special situations when a 3D scene model is available, the panoramic image can be mapped onto the model to obtain a virtual 3D environment. Researchers from the University of Alberta [Southwell *et al.*, 1996*b*] used a similar method for pipe inspection using a catadioptric camera with a conical mirror. The omnidirectional image provided by the catadioptric camera is mapped onto a 3D surface defined by a previously available model of the real surface or onto a complex developable surface formed by basic building blocks.

It is known that most omnidirectional cameras suffer from a rather low and non-uniform resolution which prevents the use of these cameras in imaging applications that require details analysis. In 1998, Yachida [Yachida, 1998] built a real-time high resolution omnidirectional sensor by combining the COPIS sensor with a binocular system formed by two conventional CCD cameras. This configuration, known as MISS (Multiple Visual Sensing System), compensates the weak resolution without losing the benefit brought by the large field of view.

Even though the conic mirror was traditionally discarded from the set of mirrors

that comply with the SVP constraint, placing the conical mirror configurations in the category of Non-SVP catadioptric sensors became arguable. Recently, Lin and Bajcsy [Lin and Bajcsy, 2001*b*] proved that the conical mirror can be used in a SVP catadioptric sensor if the camera model is not the pinhole but an extensive one that takes into account more complex optical phenomena. Lin and Bajcsy[Lin and Bajcsy, 2003] described an omnidirectional sensor that provides depth information by using two cameras, a beam splitter and a conical mirror. A similar sensor has been proposed by Spacek [Spacek, 2005] that used an omnidirectional stereo setup built with two coaxial conical mirrors.

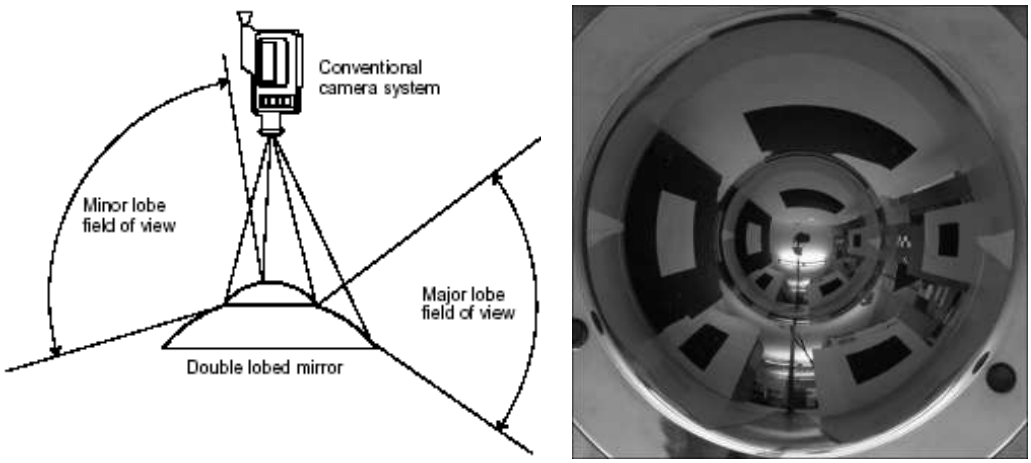
2.5.2 Spherical mirrors

The spherical mirrors respect the SVP constraint only in the case when the camera pinhole is placed at the center of the sphere. So, this kind of mirrors are useless in systems that are required to have an unique view point since the observer merely sees itself. Nevertheless, sometimes the spherical shaped mirrors represent an attractive choice due to features such as the low fabrication cost, specific characteristics over the azimuth (like density of image resolution as a function of elevation), their size etc.

In 1991, Hong et al. [Hong *et al.*, 1991] used a robot provided with a spherical mirror omnidirectional vision sensor to perform homing in an indoor environment. Homing is a navigation strategy based on a set of 3D locations. Their positions are known to the robot and it can find its way to any of these targets. Hong realized image-based homing for close targets, within the vision range. The same year, Zhang et al. [Zhang *et al.*, 1991] exploited a similar omnidirectional sensor for robot navigation. They used a method for matching previously known landmarks by analyzing qualitative geometric features for circular waveforms.

Southwell et al. [Southwell *et al.*, 1996*a*] used a non-SVP catadioptric sensor provided with a two lobbed spherical mirror to obtain two images with separated view points that allow panoramic 3D perception. Later, Fiala and Basu [Fiala and Basu, 2002*b*] used a similar sensor to obtain a stereo panoramic view in which horizontal and vertical lines were detected using the Panoramic Hough Transform, an extension of the Hough Transform adapted for catadioptric images. Depth is retrieved by matching the lines imaged by the two spherical mirrors through the radial symmetry. They used the mirror with two spherical lobes shown in Figure 2.7.

In 2000, Winters et al. [Winters *et al.*, 2000] performed robot navigation with a catadioptric sensor using a spherical reflecting shape. The robot was moved in an artificially created environment containing visible landmarks.



(a) (b)

Figure 2.7: Stereo spherical mirror-based omnidirectional sensor. (a) Two lobbed spherical mirror with a perspective camera. (b) Scene image reflected into the bi-lobed mirror. *Courtesy of M. Fiala, University of Alberta.*

2.5.3 Special designed mirror surfaces

Starting from the observation that the hyperbolic mirrors produce images suitable for perspective cameras while the parabolic mirrors are best imaged by orthographic lenses one can think that it is possible to design a mirror shape that produces a required projection. Moreover, a specially designed mirrors could correct some of the problems of traditional catadioptric sensors such as the low and non-uniform resolution on the azimuth or the imaging of non-interest areas like the central region that, generally, reflects the camera.

This type of "exotic" mirrors are extensively studied by A. Hicks. In 2000, Hicks and Bajcsy [Hicks and Bajcsy, 2000] published a study on catadioptric sensors capable to provide a panoramic field of view while still keeping some perspective properties. A catadioptric system with telecentric lenses was considered and the mirror shape was obtained by solving a differential equation with the distortion parameter constrained to be linear. The resulting mirror, shown in Figure 2.8(a), approximates perspective projection. An example of image produced by this mirror

is shown in Figure 2.8(b). The fact that the images respect the perspective projection laws enables faster image processing.

(a) (b)

Figure 2.8: Special mirror designed for perspective projection. (a) The special mirror designed by A. Hicks. (b) Imaged obtained with the mirror of Hicks suspended above a chessboard. *Courtesy of A. Hicks, Drexel University, Philadelphia.*

The transformation suffered by a scene imaged with a catadioptric sensor depends on the shape of the mirror and on the type of projection introduced by the camera. It is known that there are two main types of cameras: perspective and orthographic. Besides, the mirrors can have an infinity of shapes and, consequently, can produce an infinity of projections. A method for obtaining the analytical formula for the shape of a reflective surface with a prescribed projection of the world has been explored by Hicks and Perline in [Hicks and Perline, 2001] in 2001. They developed a general scheme for catadioptric sensor design with the aim of obtaining the equation of a mirror surface that, together with a specified camera, produce a desired projection. If there can not be found a precise formula for a mirror, the authors approximate it by means of a numerical minimization method, such as least-squares.

Another problem of classical catadioptric sensors is their non-uniform resolution. Hicks and Perline [Hicks and Perline, 2002] found a true equi-areal catadioptric camera or, in other words, an omnidirectional sensor whose mapping from the sphere to a plane is area preserving. The mirrors have special shapes that uniformly allocate pixels to the imaged regions of the scene. The cataoptric surfaces have been explored for both orthographic and perspective cameras.

Gaspar et al. [Gaspar *et al.*, 2000] presented the methodology for designing mirrors with linear projection properties. The shapes obtained produce images with

Figure 2.9: The special mirror designed by J. Gaspar. The mirror generates both vertical and horizontal constant resolution. *Courtesy of J. Gaspar, Instituto de Sistemas e Robótica, Lisboa.*

vertical, horizontal and angular resolution. A prototype of mirror that generates both vertical and horizontal constant resolution images was manufactured, see Figure 2.9. The outer part of the mirror was used for mapping linearly the distances from vertical planes while the inner part was employed for obtaining a linear mapping of the radial horizontal lines.

2.5.4 Viewpoints locus in non-SVP catadioptric cameras

A non-SVP catadioptric sensor has, by definition, more than one view point. The positions of these points in space are given by the shape of the reflecting surface. The locus of viewpoints of non SVP catadioptrics is called caustic [Swaminathan *et al.*, 2001]. Caustics of dioptric systems (ex. fish-eye lenses) are known as "diacaustics" and for catadioptric systems are named "catacaustics". In fact, the caustics are formed by the points of the space where the rays reflected by the mirror intersect. This can be observed by looking at a metal ring with a reflecting surface. As this shape does not possess a SVP, the projection centers form a caustic surface. Intersecting this surface with a plane results in a bright pattern, visible in Figure. 2.10.

It is obvious that each pixel from the image has a correspondence on the caustic surface since it is the reflection of a ray that crossed that surface. Therefore, a catadioptric system can be completely described by its caustic. The reciprocal of

Figure 2.10: The bright shape is the cross-section of the caustics with the plane of the table and is formed by the the close bunching together of light rays near the caustic's surface.

this property is that a catadioptric sensor can be designed according to a required caustic. For example, the SVP catadioptrics are systems that constrain the caustic surface to a single point.

Swaminathan et al. [Swaminathan *et al.*, 2001] provided a method to compute the caustics for catadioptric systems with a conical shaped mirror and a perspective or orthographic camera. The results have been verified by comparison with catadioptric systems with known behavior. It was shown that, as expected, the caustic surface of a panoramic camera with parabolic mirror and telecentric lens is a point.

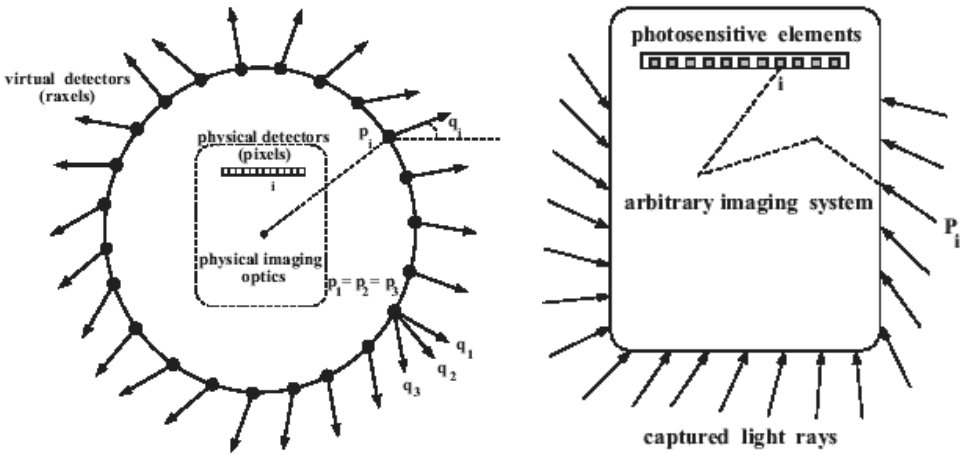
A particularity of the caustic surfaces are their singularities, called cusps. These correspond to points on the mirror where the reflected light rays observed by the camera coincide with the surface normal. The world point corresponding to such a feature of the caustic is the best approximation of the locus of viewpoints, or, in other words, is the best location that approximates the caustic by a point.

Using caustics Swaminathan et al. inspected the field of view, resolution and behavior of panoramic cameras. They also present a method of calibration that estimates the caustic surface and camera parameters from two images of an unknown scene.

2.5.5 Abstract imaging sensor

Relaxing even more the SVP constraint, Grossberg and Nayar [Grossberg and Nayar, 2001] studied a general imagining model. They presented a model for a sensor based on new virtual vision elements called "raxels". The name *raxel* comes from the fusion of two words: ray and pixel. The raxels are meant to be an abstract mathematical

equivalent of the physical pixels. A raxel is uniquely defined by the couple *position* and *orientation* and is placed so that it measures only the light intensity of one ray. Therefore, a raxel acts somehow as the view point of the catadioptric systems. Since every light ray has its own direction this determines the orientation of the raxels. The light intensity does not change within short distances, while the only medium that it crosses is the air, so a raxel can be placed anywhere along a ray. Every raxel has its own projection center, position and direction so it is clear that this model does not have a SVP. Raxels can be seen as perspective imaging systems with a very narrow field of view. Figure 2.11 (a) shows the raxel-model of a general imaging system. The arbitrary projection of the scene points on the image plane is illustrated in Figure 2.11(b).



(a) (b)

Figure 2.11: A vision system modelled by raxels. (a) Each raxel has its own position and direction. Note that multiple raxels can be located at the same point but with different orientations. (b) Incoming light is received by the raxels and is directed to the photosensitive elements. It is assumed that there is a correspondence between each ray entering the system and a light detecting element of the retinal plane. *Courtesy of S.K. Nayar, Columbia University.*

The authors also proposed a method for calibration.

The imagining model is formed by a set of raxels located onto a surface called "ray surface". It can be convenient to chose a spherical surface surrounding the imaging system. However, while possessing many advantages, the spherical surface is not an ideal one. An attractive option for the choice of a ray surface are the caustics.

The caustic is described as a surface on which the direction of a raxel is a function of its absolute position and the incoming rays are tangent to this surface. A method to compute the caustics of a particular imaging system modelled using raxels is to derive the mapping from the world to the image and compute it numerically. Another method is to obtain a discrete mapping and build a fuzzy-type correspondence function. In [Grossberg and Nayar, 2001] the authors propose a method for computing caustics. They also study the field of view of the raxel-model.

2.6 Panoramic catadioptric sensors for 3D perception

Stereoscopic vision makes that two separate images taken from the two eyes are combined into one image in the brain. Stereopsis gives us the depth perception ability, i.e. to visually perceive depth and three dimensional space, to visually judge relative distances between objects, a perceptual skill that aids accurate movement in three-dimensional space. Stereopsis is also a common technique used for passive scene range computation. Since we are interested in the omnidirectional configurations designed for scene depth measurement the topic of omnistereo systems made for immersive virtual scenes generation (that create the depth perception only in the viewer's mind) is not discussed here.

The main advantage of stereo vision is the possibility of range measurement. This can be done by observing the same point from two or more different locations. The matching of points into the images is time-consuming since this operation requires complex computations. A common method to speed-up the points matching is the use of epipolar geometry which is a technique that narrows the searching process of a point into the image of the whole scene to a line of that image. However, for catadioptric stereo the images suffer a set of transformations due to the relative orientation of the embedded mirrors. Undoing these changes involves even more calculations and degrades stereo data due to resampling of the images. A solution to this problem was given by Gluckman et al [Gluckman and Nayar, 2002] and consists in rectifying the mirrors. A pair of stereo images is rectified if the epipolar lines are aligned with the scan-lines of the images. Catadioptric systems can be built so that, by their internal geometry, the produced images are rectified.

Stereo techniques are based on the assumption that the feature correspondences

between images are accurately found. In the case of catadioptric sensors the matching is deteriorated because the resolution of the omnidirectional cameras is usually very poor compared to conventional cameras due to the different field of view. This shortcoming affects the matching since it becomes more difficult to find details in the images even when taking into account the epipolar constraints [Svoboda *et al.*, 1998]. Stereo catadioptric sensors are special structures of catadioptrics used to obtain depth.

Summarizing, omnidirectional stereo can be realized with different mirror profiles which determine the field of view. By using only one camera for obtaining stereo the sensors result cheaper and easier to calibrate. Moreover, the images are easier to deploy if the mirror geometry is known and the sensor is made such that it complies with the SVP constraint. The main drawbacks of omnidirectional stereo sensors are the low image resolution and the problems that rise at the matching stage.

2.7 Epipolar geometry for SVP omnidirectional cameras

The epipolar geometry for two perspective cameras, defined by the pinhole model, establishes a relation between a point in the image plane of one camera and a line, called epipolar line, in the image plane of the other camera. The relation between such a point and its corresponding epipolar line is given by the fundamental matrix. Since the searching of a point is performed on a line and not in the whole image, epipolar geometry permits faster correspondence calculations for calibrated cameras. A extensive surveys on fundamental matrix estimation with accuracy analysis were realized by Armangué and Salvi [Armangue and Salvi, 2003].

The epipolar geometry can be applied to panoramic vision sensors, too. Svoboda et al. [Svoboda *et al.*, 1998] proved that panoramic cameras that respect the SVP constraint allow the same epipolar geometry as the perspective cameras. In omnidirectional sensors the epipolar lines became epipolar curves as each epipolar plane intersects the mirror. In Figure 2.12 it is represented the epipolar plane that intersects both mirrors.

In [Svoboda *et al.*, 1997], the equation of an epipolar curve is calculated. The catadioptric sensor is composed by a hyperbolic mirror in conjunction with a perspective camera. The camera model is the pinhole.

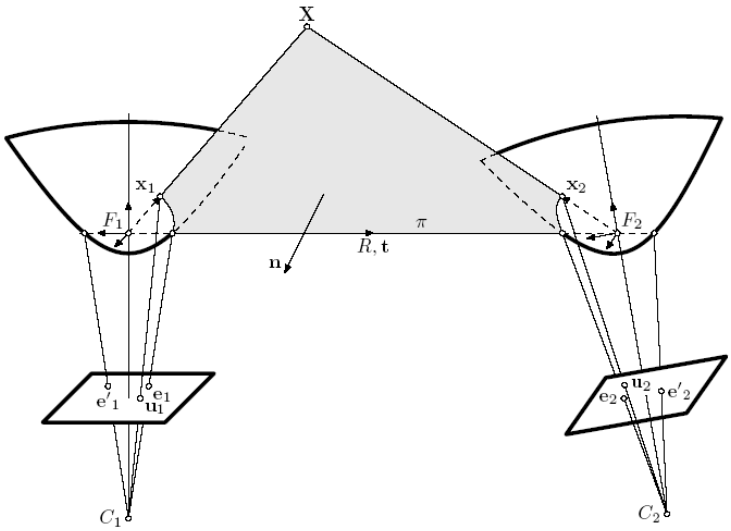


Figure 2.12: Epipolar geometry for panoramic cameras. *Courtesy of T. Svoboda, Czech Technical University.*

Gluckman et al. [Gluckman *et al.*, 1998] realized depth computation using a panoramic stereo vision sensor and applying epipolar geometry.

In stereo vision it is possible that not all the points satisfying the epipolar constraints belong to a real structure. According to Hartley's cheirality theory, as cited by Werner and Pajdla [Werner and Pajdla, 2001], particular conditions, in terms of epipolar lines and epipoles, must be added in order to ensure a robust correspondence between images. In [Werner and Pajdla, 2001], the authors detected that the panoramic sensors need supplementary constraints in order to satisfy this correspondence. Therefore, they extended Hartley's theory for wide field of view images and expressed the necessary constraints in terms of image points, epipoles, fundamental matrix and inter-image homography.

2.8 Conclusions

This chapter presented a classification of the existing catadioptric cameras. Moreover, particular examples are provided for the most common configurations. The omnidirectional field of view of these sensors is obtained by combining mirrors and conventional cameras. The mirror geometry, its position relative to the camera and the camera model divide the catadioptric configurations in two categories: SVP and

non-SVP. The SVP property transforms a catadioptric sensor into a very wide field of view pin-hole camera. The images of a SVP sensor can be unwrapped and thus directly interpreted by humans because of their similarity with the familiar pictures given by the traditional cameras. Besides, the non-SVP catadioptric configurations offer other advantages that compensate the effort required for image interpretation. Non-SVP sensors are modelled in a more flexible manner and, consequently, easier to manufacture. Mirror profiles can also be calculated in order to correct some of the optical errors produced by light reflection. Moreover, special designed mirror shapes open the application field of the catadioptrics by providing new, attractive, projections such as the projections that keep the uniform resolution or the ones that approximate the perspective model.

Another approach to omnidirectional vision through catadioptric projection is the combination of several mirrors. These configurations offer simultaneous views of the scene and therefore 3D reconstruction can be performed. We highlighted these configurations and the methods to obtain depth by means of stereo catadioptric sensors.

Summarizing, the use of one camera for the real-time capture of images with a 360º azimuth reduces the fabrication costs of the omnidirectional sensors. Their relatively low price, makes the catadioptric sensors competitive on the market of computer vision solutions which is a determinant factor for the growth of this branch of artificial sensing. As a consequence of their potential, the catadioptric cameras have been thoroughly investigated in the past ten years and nowadays they benefit of an important mathematical background.

Chapter 3

Design, modelling and calibration of the proposed sensor

The sensor presented in this thesis combines the advantages of both omnidirectional vision and structured light in order to obtain panoramic depth information. The resulting sensor is formed by a single catadioptric camera and an omnidirectional light projector. This chapter presents the sensor's design and the calibration of the laboratory prototype. The chapter starts with a study of the arrangement of the two components (the omnidirectional camera and the laser projector) and the optimal configuration is chosen among several alternatives. Subsequently, the geometry and the mathematical model of the sensor are presented. The chapter continues with a thorough study of the sensor calibration supported by experimental results. The chapter ends with conclusions about the sensor and the calibration method.

3.1 Introduction

In the previous chapters we presented the advantages that omnidirectional cameras bring to computer vision applications and it was shown that the wide field of view of these cameras alleviate some of the typical problems in pattern recognition. Besides, depth perception capacity is often needed in navigation and reconstruction tasks and can not be performed if the scene is observed from an unique position. The solution offered by traditional stereo vision is based on the feature matching between two images. The scene is observed from different positions and the common features

between the views are extracted. Depth can be obtained by triangulation for each feature found in two (or more) images. However, for indoor scenes the two images are really necessary only when the sensor is used to create a visual stereo immersion in the scene. If the main goal of the vision sensor is to measure depth the use of two images is inefficient because of the difficulty to search the features in both images. The structured light technique offers a more robust solution by replacing one of the cameras with a projector. Therefore, the projected pattern is immediately identified in the images provided by the camera. The glowing points of the pattern in the camera image are back-projected along their respective incoming 3D rays and the third dimension is found by crossing these rays with the projected pattern.

Traditionally, structured light systems use perspective cameras and the projected pattern is directed so that it covers most of the image captured by the camera. When using an omnidirectional camera we must take advantage of the whole FOV so a perspective projector is not appropriate. The projector must provide a wide FOV in a similar way as the omnidirectional camera so we propose the use of a catadioptric configuration for its construction. The resulting vision sensor will be composed of two components with an omnidirectional FOV: the camera and the laser projector. The first goal is to find the arrangement that ensures a maximum overlapping of the fields of view of the components as detailed in Section 3.2.

The image captured by the omnidirectional camera will be used later on for detecting the light pattern projected in the scene and for extracting the texture of the surrounding objects. Both tasks need to use the camera model whose parameters are obtained by calibration. A study of different calibration methods for omnidirectional cameras and catadioptric stereo configurations is presented in Section 3.3 where the advantages offered by each method are discussed. Among the catadioptric cameras, the SVP are highlighted because they enable correct image unwrapping. We present here two models for the omnidirectional camera: the reflection model and the sphere of equivalence model. Each model addresses a particular situation and the most convenient one can be selected depending on the configuration to be calibrated. An interesting issue at the camera calibration stage is related to the choice of the calibration points. The theoretical model deals with ideal 3D and 2D points but in reality these points are difficult to find with good accuracy. The choice of the calibration points in the image is difficult due to the low and non-uniform resolution of the camera. The image of the points depends on the calibration pattern and on the distance between the camera and the calibration planes. Therefore, Section 3.3

also provides the exploration of several situations which involve different calibration patterns placed at different distances from the camera. Finally, the calibration of the whole sensor is done in two steps: first the calibration of the camera is performed and then the laser projector is calibrated.

This chapter is structured as follows: Section 3.2 presents the sensor design and reviews several configurations of the sensor in order to select the most appropriate one. Several existing omnidirectional camera calibration methods are presented in Section 3.3. The proposed sensor is calibrated as described in Section 3.4 which also provides a set of experimental results obtained for different calibrations of the camera and the laser projector. Finally, the chapter ends with a section containing the conclusions.

3.2 Sensor design

As explained in the introductory section, the main goal of the proposed sensor is to retrieve 3D information from an omnidirectional image to overcome the matching problem present in stereo vision. Therefore, an omnidirectional camera must be coupled with a light projector. Due to their optical properties, catadioptric sensors having an SVP were chosen to build the omnidirectional camera of the proposed sensor. Commonly, the central projection constraint is fulfilled by the use of a parabolic mirror in conjunction with an orthographic camera or a hyperbolic mirror with a perspective camera. Both configurations have been evaluated [Orghidan *et al.*, 2003] and the observations obtained from this study helped us to choose the desired configuration [Orghidan *et al.*, 2005*b*].

A hyperbolic mirror forms a SVP image when is pointed by a perspective camera placed at the second focus of the hyperbola. This behavior is explained by the property of the hyperbola that a ray originating at a given focus (where is placed the camera) reflects in the opposite hyperbolic sheet (i.e. the mirror) so that the outgoing path lies along the line from the mirror focus through the point of intersection of the ray with the mirror surface. The main problem of this configuration is the difficulty to place the focal points of the mirror and camera at the right locations. Probably this is the reason why there are few catadioptric cameras using hyperbolic mirrors available on the market and only at high prices.

A parabolic mirror in conjunction with an orthographic camera form a cata-

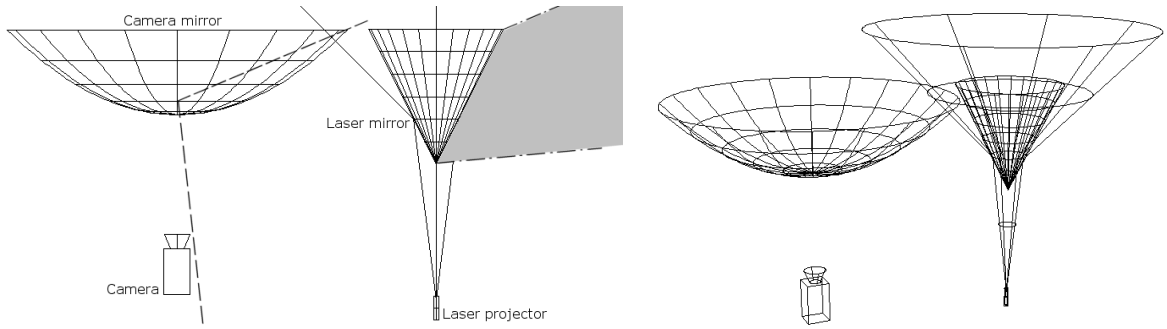


Figure 3.1: Stereo catadioptric configurations. Horizontal arrangement.

dioptric sensor that preserves the SVP independently of the translation between the mirror and the camera. This type of configuration was finally preferred for the imaging part of our sensor due to its simplicity in focusing the image. Moreover, there are commercial paracatadioptric cameras available at low cost.

The omnidirectional light projector is formed by a laser that emits a circle and is pointed at a mirror so that the projected light covers the entire field of view of the catadioptric camera. The mirror used to project the structured light in the scene must introduce a minimum blur in the light projection and does not require a SVP. The proposed projector can be seen as a reversed non SVP omni-camera where the light flows in the opposite direction. One of the most common mirrors used in non SVP catadioptrics are the conical ones. So, the projector can take advantage of the attributes revealed by previous studies of catadioptric cameras based on the conical mirror shape. Lin and Bajcsy [Lin and Bajcsy, 2001*a*] pointed out that the conical mirror can be used to built true SVP configurations with the advantage that it preserves image point brightness better than other mirrors since it does not distort the image in longitudinal directions. Yagi [Yagi, 1999] highlighted that the conical mirror on a vertical section behaves like a planar mirror and consequently provides a much better resolution than any other omni-mirror shape. Baker and Nayar [Baker and Nayar, 1998] proved that the curved mirrors (such as parabolic, hyperbolic, etc.) increase defocus blur because of their bend. Consequently, the cone turns out to be the ideal mirror shape to be used to built the structured light projector.

The third dimension of the scene is perceived by crossing the light rays emitted by the laser with the ones observed by the camera or, in other words, performing a triangulation. The laser should project a pattern that covers the whole scene and should be easily identifiable. The chosen pattern was the circle. The light

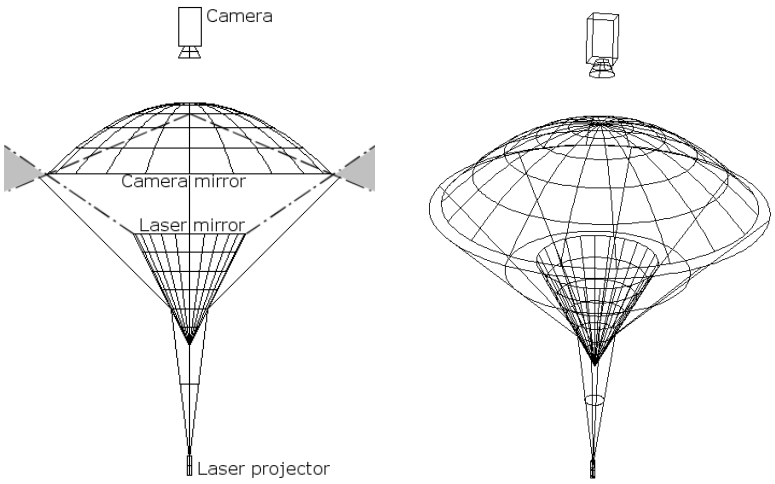


Figure 3.2: Study of stereo catadioptric configurations. Back to back arrangement.

of the projected laser pattern is reflected by the laser mirror and spread onto the scene. The camera mirror reflects the whole scene into the camera and all the spots illuminated by the laser can be immediately identified. With the models for both components of the sensor, a precise triangulation can be carried out.

In the following we present several possible configurations that are compared and the most desirable one is chosen for our sensor. The mirrors are assumed to be symmetrical about the z axis thus the bi-dimensional graphical representation is used for a clear understanding. We call the two mirrors *laser mirror* and *camera mirror*, as they work in conjunction with the laser projector and with the camera, respectively. The dark regions mark the common FOV of the two mirrors. The bounds of the FOV are represented by the dotted lines.

The first configuration places the two mirrors along a baseline, as presented in Figure 3.1. The light pattern is projected onto one of the mirrors while the other mirror grabs the scene. Such a setting makes invaluable a part of the image because of the scene occlusion caused by the laser projector into the field of view of the camera. Besides, part of the structured light pattern will be hidden or reflected by the camera mirror. So, even if this is a possible solution from the point of view of the stereo vision, it seems it is a bad option.The above problem can be solved by placing the two mirrors back to back along the rotation axis as shown in Figure 3.2. Such a system removes the mutual occlusion problem but the laser has to be projected on

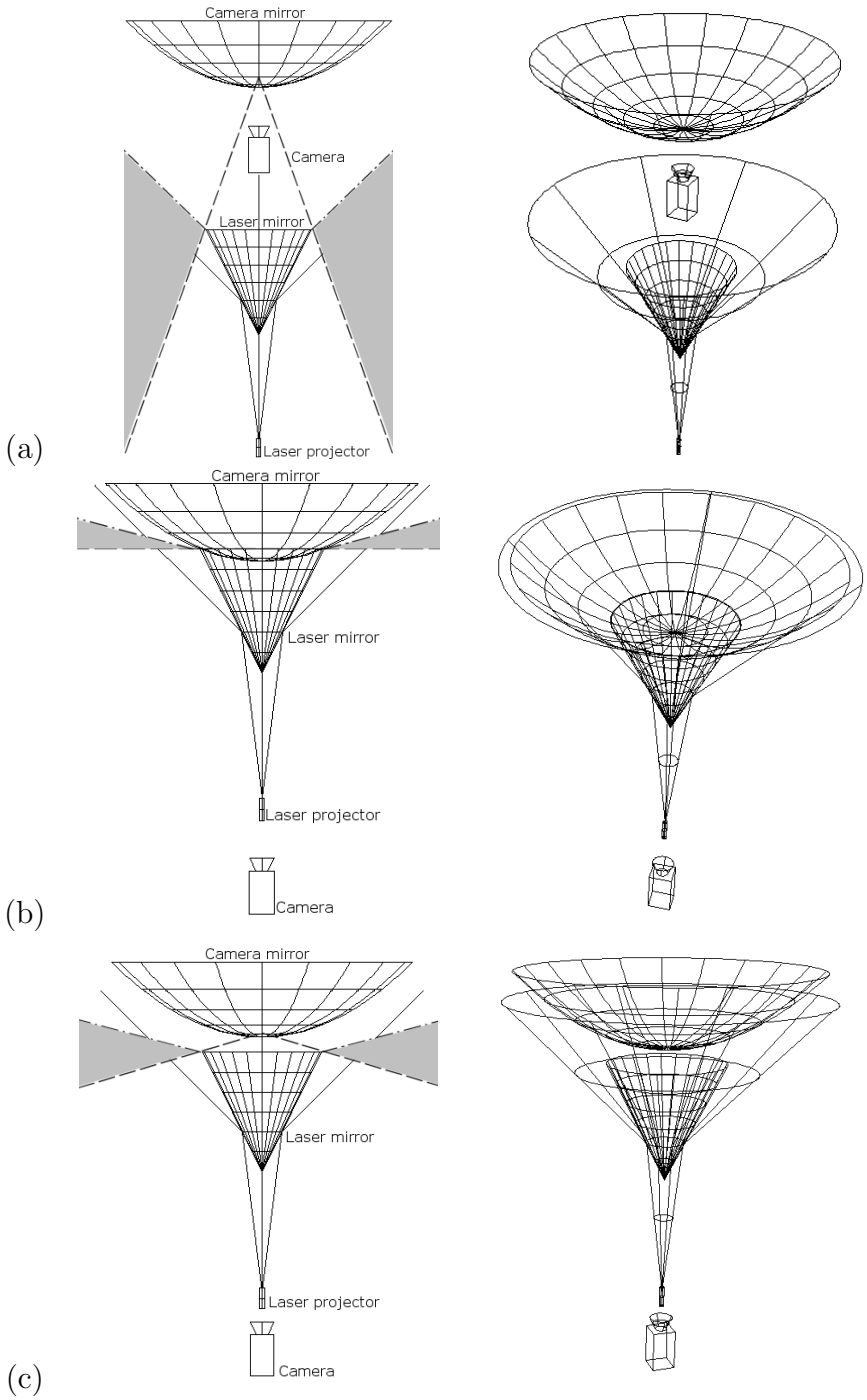
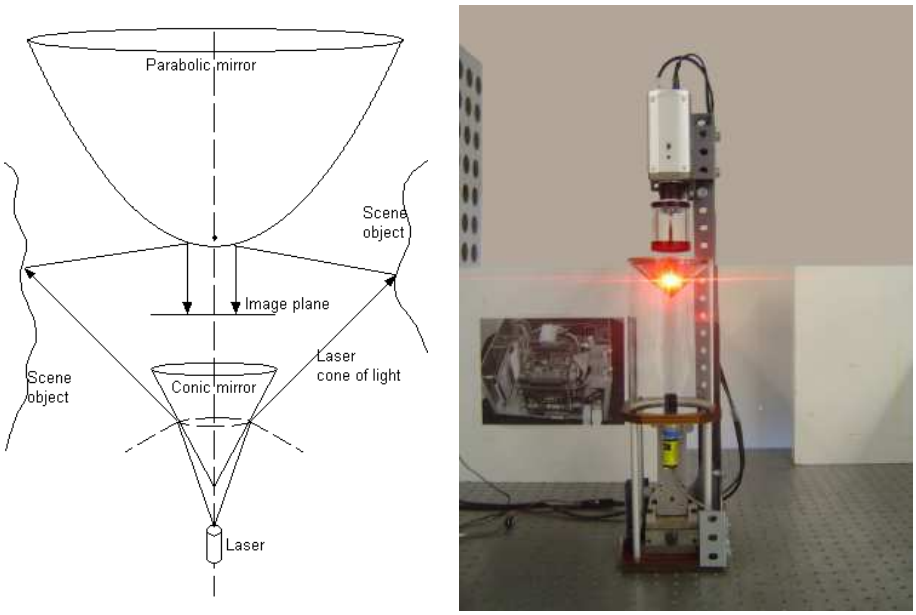


Figure 3.3: Stereo catadioptric configurations. Vertical arrangements.

the peripheral size of the mirror so that the light spots can fall in the field of view of the camera mirror. Therefore, a very small amount of stereo information will be provided by this a setting. Placing the two mirrors by a single translation along the vertical axis, as in Figure 3.3 (a), the laser is projected freely on the mirror and a large stereo FOV is obtained. A more compact sensor can be obtained by placing the laser mirror inside the "death cone" formed by the self reflection of the camera mirror, as in Figure 3.3 (b). This configuration has the inconvenient that the laser mirror occludes part of the vision field of the camera mirror. In order to avoid this problem and to enhance the baseline, the two mirrors can be separated; this configuration is presented in Figure 3.3 (c).

Taking into account the previous observations, we chose the configuration presented in Figure 3.3 (a) that provides a large stereo FOV and it is additionally easier to manipulate. The proposed catadioptric sensor [Orghidan *et al.*, 2005*b*; Orghidan *et al.*, 2005*a*; Orghidan *et al.*, 2005*c*] is presented in Figure 3.4.



(a) (b)

Figure 3.4: The proposed sensor. (a) Catadioptric omnidirectional camera with an embedded structured light projector. (b) Laboratory prototype.

3.3 Calibration of omnidirectional sensors

Any computer vision sensor can be modelled by a set of parameters through which its behavior can be predicted in any future situation. The computation of these parameters (called calibration) is usually done by means of an error minimization algorithm. An excellent survey of conventional camera calibration methods has been published by Salvi *et al.* [Salvi *et al.*, 2002]. The catadioptric cameras are a combination of traditional cameras and mirrors, therefore, the model must contain the camera parameters and also the ones modelling the mapping from the imaged scene to the mirror. Even though it seems that adding more parameters complicates the model and consequently the calibration, it results that catadioptric sensors have some nice properties that offer new perspectives for calibration. In this chapter we will discuss different approaches to catadioptric camera calibration and analyze the advantages offered by each of these methods.

3.3.1 Calibration methods for catadioptric cameras

The recovery of intrinsic and extrinsic parameters of omnidirectional catadioptric sensors is an important topic especially when the resulting image has to be used for further calculations. For instance, in the case of SVP catadioptrics, the distorted images can be unfolded since the pixels can be mapped to rays through the SVP if the calibration parameters are known.

Geyer and Daniilidis [Geyer and Daniilidis, 1999] proposed a method of calibration of a catadioptric system with parabolic mirror. This technique uses the principle of the vanishing points, well known in computer vision, which states that the perspective projection of a cube generates three vanishing points that determine the image center as being the center of the triangle formed by the vanishing points. Similarly, Geyer and Daniilidis are able to find the image center and then retrieve the information about the intrinsic parameters of the system. Actually, the image center encodes the information about the focal length of the parabola. The method requires the detection of two sets of parallel lines in the scene and assumes that the optical axis of the parabolic mirror is perpendicular to the imaging plane. When the pixel aspect ratio is equal to one, the images of the projections of these lines on the mirror are circles while, in the case of an aspect ratio different from one, the lines are mapped as ellipses with parallel axes and identical eccentricities. There

exist an unique transformation which maps such a family of ellipses to a family of circles. This transformation determines the image skew and aspect ratio.

The vanishing points of the parallel lines are transformed into intersections of the corresponding circular arcs. In the same article [Geyer and Daniilidis, 1999], the authors enounced and provided the proofs for three propositions that form the basis of their calibration mechanism:

The first proposition states that a line is imaged as a circular arc if it is not parallel to the optical axis of the sensor otherwise it is projected as a line. The next proposition shows that the projections of two parallel lines, imaged as two circles, intersect depending only on the common direction and not on the position of the lines. The image center lies on the line connecting the two intersection points. This proposition generates two corollary ideas: a) the centers of the circles representing the projections of a set of parallel lines are collinear; b) the image center is the intersection of the lines formed by the cross points of two sets of circular arcs corresponding to the projection of parallel lines with different directions. This means that the image center can be determined using only two sets of parallel lines. The third proposition affirms that the vanishing points of lines with coplanar directions lie on a circle. Finally, the fourth proposition proves that the focal length can be determined from the radius of the circle formed in the conditions described in the previous proposition.

Later, Geyer and Daniilidis [Geyer and Daniilidis, 2002] developed an algorithm for calibrating a paracatadioptric camera using a single view of at least three lines without any metric information. They consider a new property of the parabolic projection which states that a sphere whose equator is a line image contains two points that encode information about the focal distance of the parabola. Furthermore, three such spheres, corresponding to three line images, cross in a point of the optical axis located above the image plane at a distance equal to two times the focal distance. Since the intersection of three spheres in space in the presence of noise might be empty the authors use a point that minimizes the distance to the spheres. A necessary condition for calibrating using three lines is that the lines do not all intersect in a single point. The method retrieves the mirror center, the focal length, the aspect ratio and the skew and provides explicit formulas for all these parameters with exception of the skew estimation. When it can not be found a solution for the skew, the aspect ratio can not be expressed neither.

In 2001, Strelow et al. [Strelow *et al.*, 2001] studied a method of calibration for

non-SVP or misaligned omnidirectional sensors. The algorithm assumes that the camera is previously calibrated and, consequently, the intrinsic parameters are identified. The image of the mirror reflecting a set of points with known coordinates is analyzed and the error between the observed positions and the predicted reprojections is minimized using the Levenberg-Marquard algorithm. This calibration method can be used for any combination of a mirror and a camera. It also serves for checking if the camera and the mirror are correctly aligned.

Kang [Kang, 2000] realized the self-calibration of a paracatadioptric sensor using two methods. The first recovers the image center and the mirror parameters using the projection of the mirror boundary that must be visible on the image. The second method is based on moving the sensor and tracking points across a sequence of captured images. Kang finds the correspondences between images using Svoboda's [Svoboda *et al.*, 1998] epipolar geometry (developed for hyperbolic mirrors and perspective cameras) adapted to the parabolic reflective surfaces with orthogonal projection. One advantage of Kang's method is that it does not require finding particular structures in the images, such as straight or parallel lines, as in [Geyer and Daniilidis, 2002]. Moreover, this approach does not need to use any calibration pattern since it takes advantage of the trackable features from the images.

D. Aliga [Aliaga, 2001] calibrates a paracatadioptric sensor using two sets of coordinates corresponding to the 3D calibration points and their 2D projections on the image plane. By means of an optimization loop, he fits them to the mathematical model of the catadioptric system and estimates the internal and external parameters. The model takes into account the radial distortion and the mirror misalignment which permits the relaxation of the SVP constraint. The mathematical line of Aliga's calibration is similar to our methodology. Another noticeable feature is the detection of bright spots in the scene created in [Aliaga, 2001] by means of light bulbs.

A model for an arbitrary imagining system based on new virtual vision elements called "raxels" was studied by Grossberg and Nayar [Grossberg and Nayar, 2001]. In this case, calibration is equivalent to find the parameters for a "black box" imagining system. The general concept of the imagining system does not allow perspective projections therefore conventional calibration methods can not be used. Given that the only available information are the sets of 3D world points and their images, finding a relation between them is a compulsory step for calibration. The matching is done by means of a calibration pattern which enables an unique image point to

scene ray correspondence. Once this relation is established the parameters of the system are retrieved solving the three dimensional Jacobian equation of the general model.

3.3.2 Calibration of stereo catadioptric sensors

In [Gluckman and Nayar, 1999], Gluckman and Nayar studied the case of the stereo catadioptrics based on planar mirrors in an arbitrary configuration. The geometric properties of the catadioptric sensors are analyzed and the authors developed a method of retrieving the focal length from a single catadioptric image. Moreover, they show that the epipolar geometry and the relative orientation are constrained for the planar motion. Following the same research line, Gluckman and Nayar studied the rectification of catadioptric stereo sensors. In [Gluckman and Nayar, 2002], is presented a study to determine the number of mirrors to be used and the constraints that these mirrors must satisfy in order to obtain rectified stereo images with a single camera. From the calibration point of view the article does not bring a new algorithm for finding the parameters of the sensor but a set of design considerations and methods to avoid errors due to camera or mirrors misplacement.

In 2002, Fiala and Basu [Fiala and Basu, 2002*a*] performed feature extraction and calibration for stereo reconstruction using non-SVP optics in a panoramic stereo-vision sensor based on spherical mirrors. The parameters to be found were: the dioptric camera focal length, the distances from the camera focal point and each of the lobe centers and the mirror radii. The camera focal length was determined by focusing. The relation between a point of the scene and an image point of known radius was also calculated. Using the two previous information the calibration was performed by measuring the radius of scene points at known locations. Only radial symmetry was considered since vertical errors are dramatically amplified by the shape of the mirror. This algorithm was applied separately for each mirror lobe.

The Czech researcher T. Svoboda published several works dealing with the analysis of epipolar geometry for panoramic cameras. In 2002, Doubek and Svoboda [Doubek and Svoboda, 2002], performed 3D reconstruction of the scene from a moving platform provided with two panoramic cameras. The correspondence problem was solved by means of the epipolar geometry together with a heuristic method for selecting the most representative points for 3D reconstruction. The most important contribution of this paper is the study of the reliable reconstruction space.

It was proved that the reconstruction of a complex space is more efficient using 3 view sources even though, in theory, two of them are enough. They also observed that reliable 3D reconstruction is possible even if the cameras parameters are only approximated. Therefore, this is not a calibration method but a proof that stereo omnidirectional vision can, in some conditions, overcome the inconvenient of not possessing accurate parameters of the model.

Stereo catadioptric sensors benefit of the important feature of having only one camera, therefore, only one set of parameters to calibrate. This fact removes the inter-cameras uncertainty problem and improves the quality of stereo data at no additional cost. However, the fact that the sensor contains cataoptric elements complicates the calibration process. In literature there are not many works dealing with calibration issues for stereo catadioptrics but there exist methods of correctly aligning the different parts of the sensor and performance check methods. The extensive work about single-camera single-mirror catadioptric calibration provides helpful information for parameter estimation.

3.4 Calibration of the proposed sensor

The proposed sensor is composed by a paracatadioptric camera and a laser projector. Range detection is performed by triangulation if both the camera and the projector are calibrated. In the following, we describe the modelling process and parameter estimation (calibration).

3.4.1 Omnidirectional camera calibration

A camera model can be built and calibrated if the mirror surface is considered as a known revolution shape and is modelled explicitly. For instance, the reflecting surface can be a paraboloid placed in front of an orthographic camera. Another way of approaching camera calibration is assuming that the pair camera-mirror possesses a SVP. Thus, the omnidirectional camera can be modelled as a projection onto a sphere followed by a projection to a plane, as stated by Geyer and Daniilidis in [Geyer and Daniilidis, 2000*b*; Geyer and Daniilidis, 2000*a*]. Both models were tested and the comparative results were also reported in [Orghidan *et al.*, 2005*b*].

Calibration using the reflection model

The reflection model addresses the image formation based on the reflection of light into a mirror surface and assumes that the mirror is perfectly specular i.e. the angle of incidence is equal to the angle of reflection, see Hicks and Bajcsy [Hicks and Bajcsy, 2000] and Gaspar [Gaspar, 2002]. Therefore, for a given 3D point in space it is possible to find its correct reflection point on the mirror so that the point is imagined by the camera. Then, the projection of the mirror point onto the camera image plane is straight forward [Salvi *et al.*, 2002]. Note that, by using this method, the SVP property becomes a consequence and not a constraint which means that the model can be applied for any kind of mirror shapes no matter if they possess a SVP or not.

The coordinates of the world points are measured with respect to the world coordinate system $\{W\}$. On the other hand, the camera model must have an independent coordinate system related to the camera $\{C\}$. The transformation from $\{W\}$ to $\{C\}$ is done considering the rotation RPY (Roll-Pitch-Yaw) angles (Equation (3.1)) and the translations on each axis (Equation (3.3)). The transformation matrix is calculated from the rotation (Equation (3.2)) and the translation (Equation (3.3)) matrices as shown in Equation (3.5).

$$
\begin{aligned}
{}^{C}R_{XW} &= \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos(\alpha) & \sin(\alpha) & 0 \\ 0 & -\sin(\alpha) & \cos(\alpha) & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \\
{}^{C}R_{YW} &= \begin{pmatrix} \cos(\beta) & 0 & -\sin(\beta) & 0 \\ 0 & 1 & 0 & 0 \\ \sin(\beta) & 0 & \cos(\beta) & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \\
{}^{C}R_{ZW} &= \begin{pmatrix} \cos(\gamma) & \sin(\gamma) & 0 & 0 \\ -\sin(\gamma) & \cos(\gamma) & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}
\end{aligned}
\tag{3.1}
$$

$$
{}^{C}R_{W} = {}^{C}R_{XW} \cdot {}^{C}R_{YW} \cdot {}^{C}R_{ZW} \tag{3.2}
$$

$$ {}^{C}T_{W} = \begin{pmatrix} 0 & 0 & 0 & t_X \\ 0 & 0 & 0 & t_Y \\ 0 & 0 & 0 & t_Z \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{3.3} $$

$$ {}^{C}RT_{W} = ({}^{C}R_{W};\ \ {}^{C}T_{W}) \tag{3.4} $$

Finally, a point of the world ${}^{W}P_{w}$ represented with respect to the world coordinate system is transformed in the point ${}^{C}P_{w}$ camera coordinate system by

$$ {}^{C}P_{w} = {}^{C}RT_{W} \cdot \ {}^{W}P_{w} \tag{3.5} $$

In the following, the reflection model is described for a catadioptric configuration formed by a convex mirror surface with a perspective camera. Then two particular cases are addressed: a parabolic mirror with orthographic camera and to a hyperbolic mirror with perspective camera. All the calculations are done in the camera coordinate system $\{C\}$ therefore the notation of a point P is considered equivalent to ${}^{C}P$.

Reflection using a convex mirror with a perspective camera

The rotational symmetry of the mirrors studied in this thesis permits, without loss of generality, to reduce the study of the optical behavior of such a surface from three-dimensional Cartesian coordinates system (X, Y, Z) to a two dimensional one $\{r, z\}$, where $r = \sqrt{X^2 + Y^2}$.

Consider a function $F(r)$ that represents the curvature of a convex mirror as shown in Equation(3.6). Figure 3.5 shows the reflection of a beam of light in the mirror surface at a given point $P_h(r, z)$. The incidence angle α is expressed with respect to the tangent at the reflection point.

$$ z = F(r) \tag{3.6} $$

The slope of the tangent is also the local slope of the mirror profile $F'(r)$ as

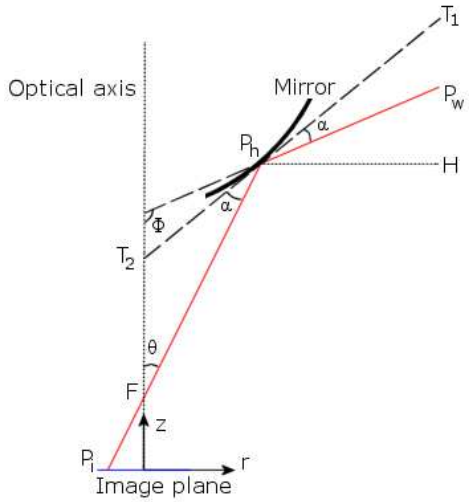


Figure 3.5: Image formation with a convex mirror and a perspective camera.

expressed in Equation(3.7).

$$\widehat{T_1 P_h H} = \arctan(F'(r)) \tag{3.7}$$

We can also write,

$$\widehat{T_1 P_h H} = \frac{\pi}{2} - (\alpha + \theta) \tag{3.8}$$

From the previous equations it follows that

$$2 \cdot \widehat{T_1 P_h H} = 2 \cdot \arctan(F'(r)) = \pi - 2(\alpha + \theta) \tag{3.9}$$

Therefore,

$$2 \cdot \arctan(F'(r)) + 2 \cdot \theta = \pi - 2 \cdot \alpha \tag{3.10}$$

From the geometrical configuration it also results that,

$$2 \cdot \alpha + \phi + \theta = \pi \tag{3.11}$$

or,

$$\phi + \theta = \pi - 2 \cdot \alpha \tag{3.12}$$

Therefore, from Equation(3.10) and (3.12) we obtain the projection function of a

known mirror shape, as expressed by Equation(3.13).

$$\phi = 2 \cdot \arctan(F'(r)) + \theta \tag{3.13}$$

The relation between a world point $P(r_w, z_w)$ and its reflection in the mirror $P_h(r, F(r))$ is highlighted when the angular variables are expressed in cartesian coordinates as shown in Equation(3.14) and (3.15).

$$\theta = \arctan(\frac{r}{F}) \tag{3.14}$$

$$\phi = \arctan(-\frac{r_w - r}{z_w - F}) \tag{3.15}$$

Then, replacing the cartesian coordinates in the projection function we obtain an equation that relates the world point, its image and the mirror function, see Equation(3.16).

$$\frac{\frac{r}{F} + 2 \cdot \frac{F'}{1-F'^2}}{1 - 2 \cdot \frac{r \cdot F'}{F(1-F'^2)}} = -\frac{r_w - r}{z_w - F} \tag{3.16}$$

If the world points P_w and their images are known then Equation(3.16) can be used for calculating the mirror shape. Hence, if the world points are set according to a given criteria we can determine a mirror shape that satisfies that criteria. This equation has been used by Hicks and Bajcsy [Hicks and Bajcsy, 2001] for the design of mirrors that remove the distortion from the image of a given plane, see Figure 2.8. These mirrors act like computational sensors that alleviate the processing effort providing undistorted images without software transformations. Similarly, Gaspar et al. [Gaspar *et al.*, 2000; Gaspar, 2002] obtained a mirror that provides vertical and horizontal constant resolution images, see Figure 2.9.

Equation(3.16) presents an effective criteria for computing the parameters of a given catadioptric system. However, exact solutions might not exist so a different method must be used for finding approximate solutions. In the following we present a variation of this model that ensures the convergence of the calibration and that was used for modelling the catadioptric camera.

Reflection model of a parabolic mirror with orthographic camera

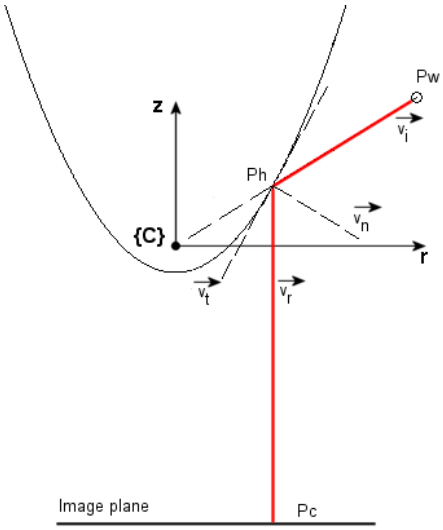


Figure 3.6: Image formation with a paracatadioptric camera. The device is composed by a parabolic mirror in front of a camera provided with telecentric lenses.

The parabolic shape is modelled in the $\{r, z\}$ coordinate system by Equation (3.17), where a is the parameter of the parabola.

Consider P_w a point on the scene, P_c the point in the image plane and the mirror surface described by a known function as shown in Figure 3.6.

$$z = \frac{r^2}{4 \cdot a} \tag{3.17}$$

Let P_h be the point on the mirror where the point P_w is reflected i.e. P_h belongs to the tangent $\overrightarrow{v_t}$ at the mirror surface whose normal $\overrightarrow{v_n}$ is the bisecting line of the angle $\widehat{P_w P_h P_c}$. Also said, the incident vector $\overrightarrow{v_i} = \overrightarrow{P_h P_w}$ is the symmetric of the reflected vector $\overrightarrow{v_r} = \overrightarrow{P_c P_h}$ with respect to $\overrightarrow{v_n}$. This relation is described by Equation (3.18).

$$\overrightarrow{v_i} = 2 \cdot (\overrightarrow{v_r} \cdot \overrightarrow{v_n}) - \overrightarrow{v_r} \tag{3.18}$$

Consequently, for finding the image of a point P_w with known coordinates the problem reduces to finding the point P_h on the mirror surface so that the vectors $\overrightarrow{v_i}$ and $\overrightarrow{v_r}$ satisfy Equation (3.18). The point P_h is calculated in the following.

Consider the origin of the camera coordinate system $\{C\}$ as $O_C(x_0, y_0, z_0)$. The

world point $P_w(x, y, z)$ can be written in two dimensions as $P_w(r, z)$, see Equation (3.19)

$$\begin{cases} r = \sqrt{(x - x_0)^2 + (y - y_0)^2} \\ z = z - z_0 \end{cases} \tag{3.19}$$

Let us consider P_{hi} a hypothetical mirror reflection point which can be expressed as shown in Equation (3.20).

$$P_{hi} = [r_i \quad \frac{r_i^2}{4 \cdot a}] \tag{3.20}$$

The tangent vector at the mirror surface at point P_{hi} is expressed in Equation (3.21).

$$\begin{array}{l} \overrightarrow{v_t} = \frac{dP_{hi}}{dr_i} \\ \overrightarrow{v_t} = [1 \quad \frac{r_i^2}{2 \cdot a}]^T \end{array} \tag{3.21}$$

We can now define a world point P_{wi} that represents the source of the ray reflected at P_{hi}. Therefore, $\overrightarrow{P_{hi}P_{wi}}$ is the symmetric of the vector $\overrightarrow{P_{ci}P_{hi}}$ where P_{ci} is the orthographic projection of P_{hi} on the image plane. The vector $\overrightarrow{O_C P_{wi}}$ is expressed by Equation (3.22).

$$\overrightarrow{O_C P_{wi}} = [r_i \quad z] + \frac{2 \cdot (P_{hi} - [r_i \quad z_c] \cdot v_t) \cdot v_t^T}{v_t^T \cdot v_t} \tag{3.22}$$

The distance from the point P_w to the vector $\overrightarrow{P_{hi}P_{wi}}$ is the error due to the choice of the mirror point P_{hi} as the reflection of the real world point P_w. Therefore, the error $\varepsilon = dist(P_w, \overrightarrow{P_{hi}P_{wi}})$ and is calculated as shown in Equation (3.23), where $\overrightarrow{P_{hi}P_{wi\perp}}$ is the normal to the vector $\overrightarrow{P_{hi}P_{wi}}$. Minimizing this error the correct reflection point P_h is found.

$$\varepsilon = (P_w - P_{hi}) \cdot \overrightarrow{P_{hi}P_{wi\perp}} \tag{3.23}$$

The reflection of a set of world points in the parabolic mirror is shown in figure Figure 3.7 (a). Even though the SVP was not used for modelling the reflection in the mirror, the rays that form the image intersect in an unique point if an orthographic camera is used. On the other hand, the caustic formed by observing a parabolic

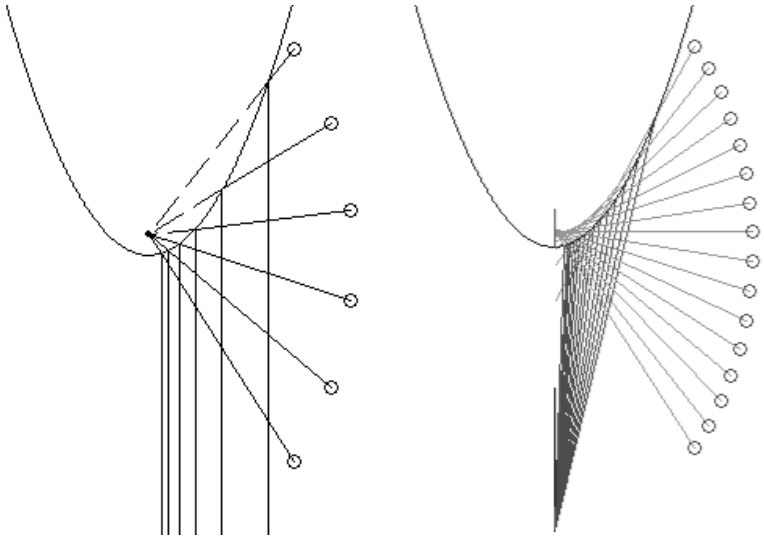


Figure 3.7: Reflection in a parabolic mirror. The mirror is pointed by (a) an orthographic camera. (b) a perspective camera.

mirror with a perspective camera is shown in Figure 3.7 (b).

Once the point P_h is correctly determined, its orthogonal projection on the image plane is immediate, as shown in Equation (3.24), where z_c is a constant value representing the translation between the mirror and the image plane.

$$P_c = [r_h \quad z_c]^T \tag{3.24}$$

Taking into account the intrinsic camera parameters, the pixel coordinates of the image point are calculated in Equation (3.25), where θ is the azimuth angle of the reflected world point P_w.

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{pmatrix} \alpha_u & 0 & u_0 & 0 \\ 0 & \alpha_v & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \cdot \begin{bmatrix} r_h \cdot \cos(\theta) \\ r_h \cdot \sin(\theta) \\ z_c \\ 1 \end{bmatrix} \tag{3.25}$$

The point obtained is represented in the computer image coordinate system $\{I\}$ located in the upper-left corner of the image plane.

Considering the world points with known coordinates, placed onto a calibration

pattern, their projection onto the image plane can be calculated using an initial guess for the parameters of the model. The difference between the positions of the calculated image points and the positions of the real image points is the calibration error of the model. Minimizing the above error by means of an iterative algorithm, such as Levenberg-Marquardt, the omnidirectional camera is calibrated. The parameters of the model are: a, the parameter of the parabola; α_u, α_v, u_0, v_0, the intrinsic camera parameters; α, β, γ, and t_X, t_X, t_Z, the six extrinsic parameters that model respectively the orientations and the translations between the world coordinate system and the mirror coordinate system. The three angles which define the rotation on each axis are expressed in radians while the translations are measured in millimeters.

Reflection model of a hyperbolic mirror with perspective camera

As stated before, the same method can be used for any mirror with known curvature. In the case of a hyperbolic surface (see Figure 3.8) a point P_{hi} on the mirror is defined as shown in Equation (3.26) where a and b are the hyperboloid parameters. The hyperbolic shape is sometimes described using the interfocus distance c and the excentricity k which are related with the parameters a and b as shown in Equation (3.27).

$$P_{hi} = [r_i \quad \frac{c}{2} - a \cdot \sqrt{1 + \frac{r_i}{b^2}}] \tag{3.26}$$

$$\begin{cases} a = \frac{c}{2} \cdot \sqrt{\frac{k-2}{k}} \\ b = \frac{c}{2} \cdot \sqrt{\frac{2}{k}} \end{cases} \tag{3.27}$$

The tangent at the point P_{hi} is

$$\begin{cases} \overrightarrow{v_t} = \frac{dP_{hi}}{dr_i} \\ \overrightarrow{v_t} = [1 \quad \frac{-a \cdot r_i}{b^2 \cdot \sqrt{1+\frac{r_i^2}{b^2}}}]^T \end{cases} \tag{3.28}$$

The vector $\overrightarrow{P_{wi}}$ is calculated taking into account the interfocus distance c:

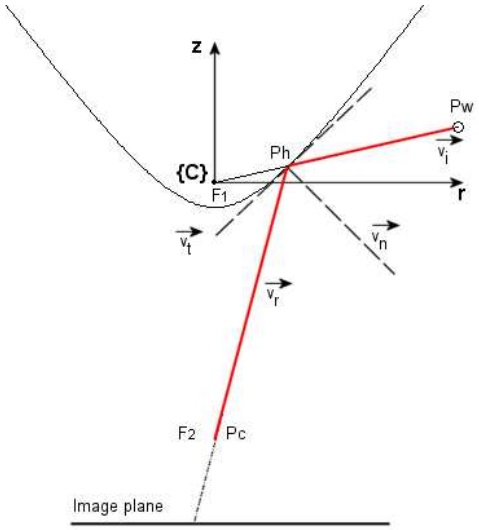


Figure 3.8: Image formation with a hypercatadioptric camera. The device is composed by a hyperbolic mirror and a perspective camera. The points $F1$ and $F2$ are the two hyperbolic foci.

$$\overrightarrow{P_{wi}} = [0 \quad c] + \frac{2 \cdot (P_{hi} - [0 \quad c] \cdot v_t) \cdot v_t^T}{v_t^T \cdot v_t} \tag{3.29}$$

Minimizing the error distance from the point P_w to the vector $\overrightarrow{P_{hi}P_{wi}}$ as shown in Equation (3.23) the correct reflection point P_h is found. The reflection of a several world points in the hyperbolic mirror is shown in figure Figure 3.9. The incoming rays have been artificially enlarged in order to emphasize the SVP.

After performing the perspective projection of P_h onto the image plane and considering the intrinsic camera parameters, the pixel coordinates of the image points are calculated. As in the previous case, the error to be minimized for calibrating the whole model is the distance between the positions of the calculated image points and the positions of the real image points. The parameters of the model are: a, b, the parameters of the hyperbola; α_u, α_v, u_0, v_0, the intrinsic camera parameters; α, β, γ, and t_X, t_X, t_Z, the six extrinsic parameters.

Summarizing, the previously described calibration method provides a simple way to model a catadioptric camera. However, it is constrained by the assumption that mirror curvature is perfectly determined by a known function.

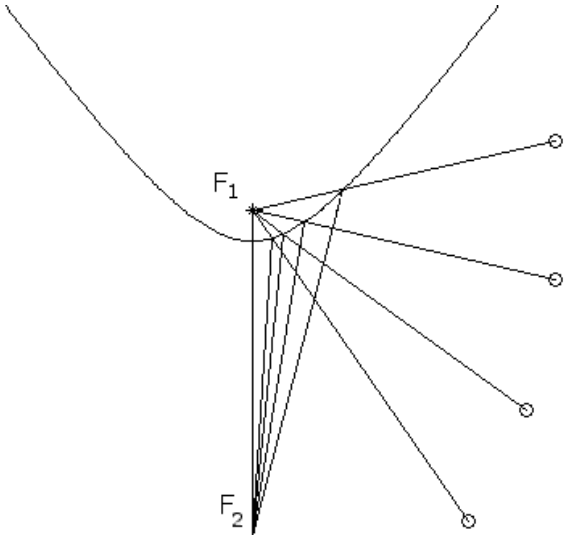


Figure 3.9: Reflection in a hyperbolic mirror. The mirror is pointed by a perspective camera. The points $F1$ and $F2$ are the two hyperbolic foci. The camera focal point is placed at $F2$.

Calibration using the sphere of equivalence model

The catadioptric camera calibration can be achieved using the projective equivalence between any central projection and the projection onto the sphere followed by the projection to a plane from a point [Geyer and Daniilidis, 2000*a*]. Remark that this model does not assume that the mirror shape is known but considers that the omnisystem possesses a SVP. Considering $P_s = [x_s, y_s, z_s]$ the intersection of the light ray emitted by the point P_w with the sphere of radius R, then the perspective projection of P_s on the image plane from a point $C = [0, \xi]$ produces a point $P_i = [x, y]$ as shown in Figure 3.10.

The relation between P_w and P_s is expressed in equation (3.30).

$$\begin{cases} x_s = \lambda \cdot x_w \\ y_s = \lambda \cdot y_w \\ z_s = \lambda \cdot z_w \end{cases} \tag{3.30}$$

Since the points belong to the sphere, equation (3.31) also holds.

$$x_s^2 + y_s^2 + z_s^2 = R^2 \tag{3.31}$$

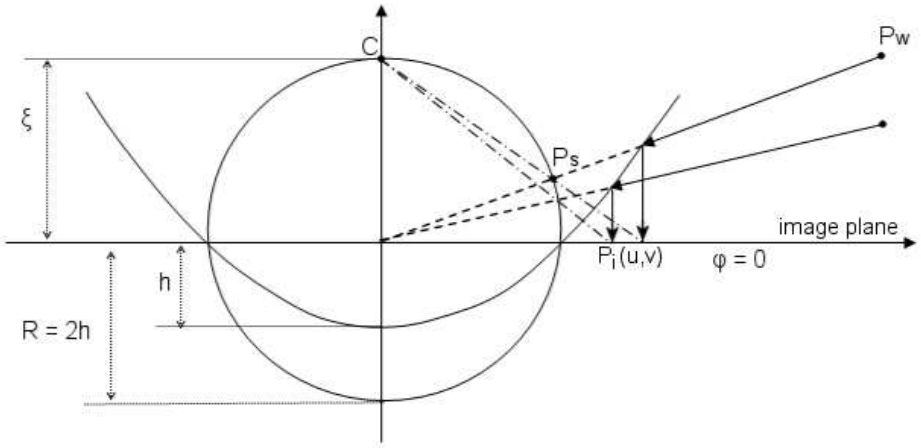


Figure 3.10: Image formation using the projective equivalence of a SVP catadioptric projection with the projection on the sphere.

The perspective projection of P_s on the image plane from a point $C = [0, \xi]$ produces a point $P_i = [x, y]$ as expressed in equation (3.32)

$$\begin{cases} \frac{x_s}{\xi - z_s} = \frac{x}{\xi + \varphi} \\ \frac{y_s}{\xi - z_s} = \frac{y}{\xi + \varphi} \end{cases} \tag{3.32}$$

Adding the intrinsic camera parameters (α_u, α_v, u_0, v_0) the pixel coordinates of the image points are shown in equation (3.33)

$$\begin{cases} u = \frac{\alpha_u(\xi+\varphi)x_w}{\xi\sqrt{x_w^2+y_w^2+z_w^2}-z_w} + u_0 \\ v = \frac{\alpha_v(\xi+\varphi)y_w}{\xi\sqrt{x_w^2+y_w^2+z_w^2}-z_w} + v_0 \end{cases} \tag{3.33}$$

The calibration is performed using a set of known 3D points, P_w, distributed on the four walls of a cube placed around the sensor. The parameters of the model are ξ, which depends on the eccentricity; φ which is a function of both the eccentricity and the scale; α_u, α_v, u_0, v_0, the intrinsic camera parameters; α, β, γ, and t_x, t_y, t_z, the parameters that model respectively the orientation and the translation between the world coordinate system placed in the upper corner of the first calibration plane and the camera coordinate system. The difference between the positions of the calculated image points and the positions of the real image points is the calibration error of the model. Minimizing the above error by means of a nonlinear iterative

algorithm such as Levenberg-Marquardt the model of the omnidirectional camera is calibrated.

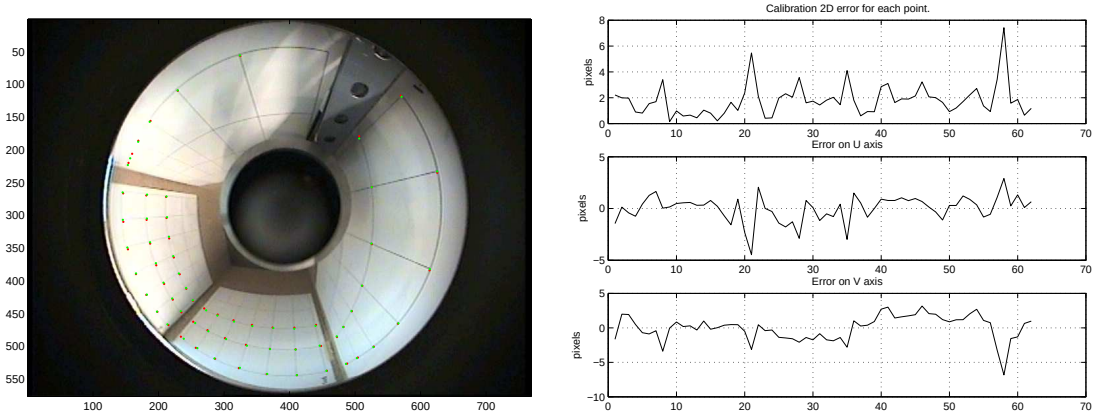
3.4.2 Experimental results

Section 3.4.1 describes two methods for calibrating an omnidirectional camera. Both methods use pairs of 3D and 2D points for fitting the corresponding models. The omni camera used for our prototype has a SVP but contains two reflecting surfaces so the calibration using the sphere of equivalence (shape independent) was preferred. This section offers an in depth view of the calibration of our catadioptric camera.

An important issue in calibration is the repeatability of the process. Note that the calibration must be performed every time when one or more intrinsic or extrinsic model parameters are changed. Usually these changes are due to physical factors such as changes of the relative position between the camera and the mirror or the variation of the focal length. So, a fast and reliable detection of the calibration points is very important.

An easy way to obtain the calibration points is to place 3D points with known coordinates around the catadioptric camera. This can be easily achieved by placing the camera inside a box having a grid-like calibration pattern fixed on its inner walls. However, the proposed sensor must observe the laser stripe which shows up in the omnidirectional image at unpredictable distances from the image center. Therefore, a more appropriate camera calibration technique should use points distributed on planes placed at different distances from the sensor. Both configurations have been tested and are presented in this section.

The choice of the calibration pattern is also a critical issue for obtaining good reference points and eventually a good calibration. In the following we will approach the detection of the calibration points in the omnidirectional image as part of the calibration process. Therefore, we tested the calibration using successively a grid, a dotted and a checkered pattern and the calibration performances with the three patterns are detailed in this section.



(a) (b)

Figure 3.11: The grid pattern calibration plane fixed on the inner walls of a box placed around the omnidirectional camera. (a) The reference points are shown with green while the calibrated points are plotted with red. (b) The error for all the 62 points used for calibration.

Calibration using a grid pattern

The camera calibration was performed using a box which displays four grids on its inner side. The 62 calibration points, manually selected by the user at the intersections of the lines in the grid, are shown in Figure 3.11 (a) superposed to the omni image captured by the camera. The overall 2D accuracy error is $\mu_{2D} = 1.79$ pixels and the sample standard deviation $\sigma_{2D} = 2.17$ pixels.

Calibration using a dotted pattern

The 2D calibration points are the centers of the dots in the calibration pattern, see Figure 3.12. The pattern was built so that the distance between two adjacent points on the same plane is 6cm. Four images used for calibration are presented in Figure 3.13.

The main advantage of using dots is the possibility to automatize the detection of their centers by means of a region growing algorithm. Since the dots that belong to the same row are placed collinearly in space, their centers lie along a curve in the omnidirectional image. By performing an ellipse fitting [Fitzgibbon *et al.*, 1999] it is possible to find an ellipse that intersects all the dots of a given row. Then, the region growing algorithm is applied for each dot placed on the elliptical trajectory and the

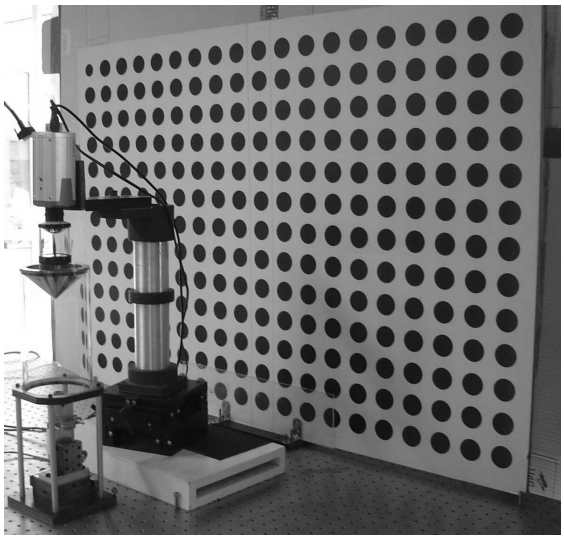


Figure 3.12: Experimental setup with the sensor placed in front of the dotted calibration plane.

gravity centers are extracted. The dots that belong to the same column of the pattern are also collinear and are projected along radial lines in the omnidirectional image. Therefore, for extracting all the calibration points it is sufficient to select a minimum number of dots to perform ellipse fitting on the the upper and lower rows in the pattern. This algorithm can be successfully used as a fast way to find the calibration points. For instance, for a calibration plane composed by 19 columns of 5 points, as the one shown in Fig. 3.14 (a), we can determine all the 95 image points by selecting only 10 points: five for the first row and five for the last one. A point on the calibration pattern (the center of a dot) can be selected by simply clicking the corresponding dot anywhere on its surface and by means of the region growing algorithm we calculate the coordinated of the gravity center of the dot with subpixel accuracy. Note that if the user provides the 10 required points the algorithm will work independently of the numbers of calibration points of the pattern since neither the number of rows nor the number of columns are limited.

The calibrated intrinsic and extrinsic parameters of the camera-model with the calibration pattern placed at different distances are listed in the Table 3.1 and Table 3.2, respectively. The values of the column *Dist.* represent the radius of the circle on the plane XY, with center in the origin of the cameras' reference system and tangent to the calibration plane, measured in cm. The overall 2D accuracy error is $\mu_{2D} = 1.64$ pixels and the sample standard deviation $\sigma_{2D} = 1.87$ pixels.

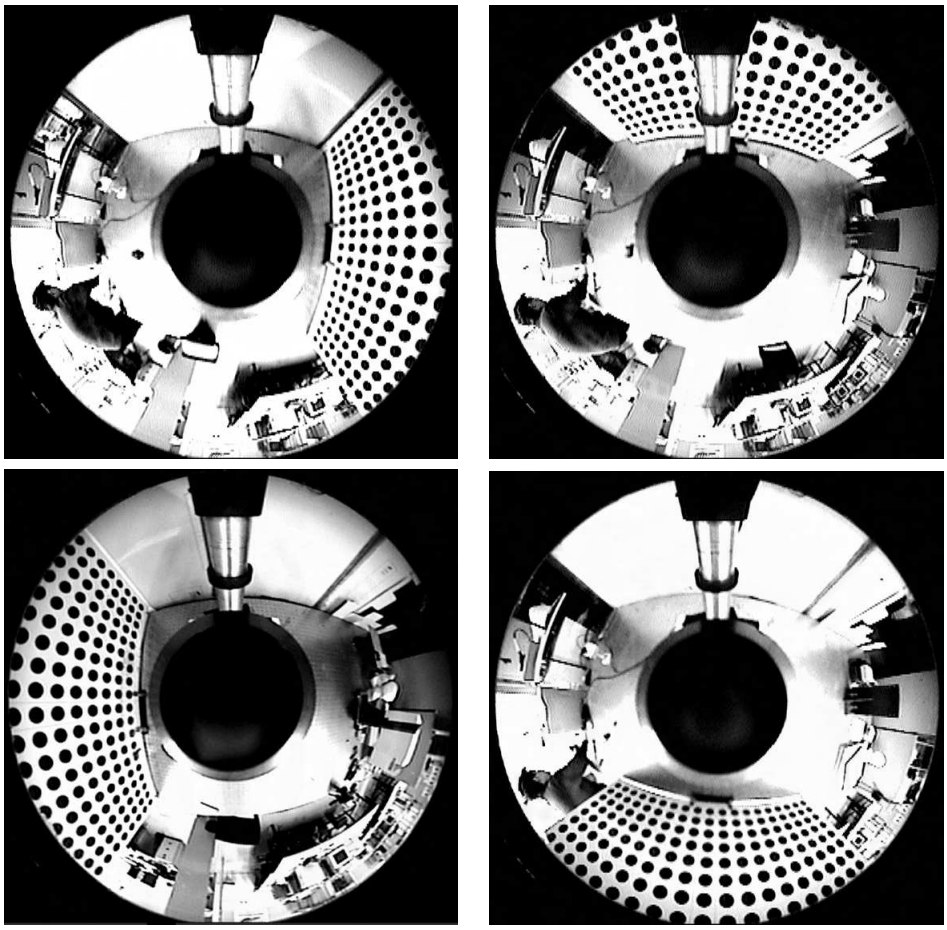
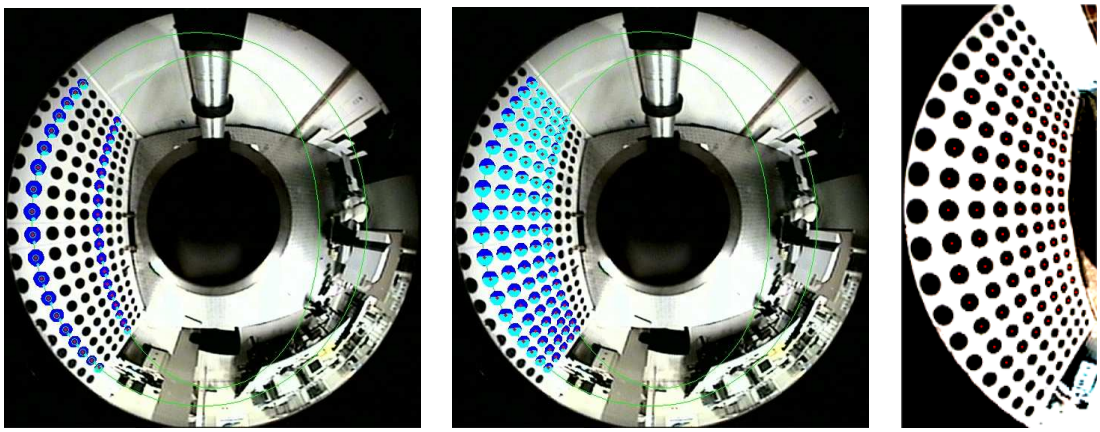


Figure 3.13: The dotted calibration plane placed around the omnidirectional camera.

Table 3.1: The intrinsic parameters of the omnidirectional camera calibration using a dotted pattern. The measurement units are: (ξ, φ), the mirror parameters in mm; (α_u, α_v, u_0, v_0), the intrinsic parameters in pixels.

Dist.	ξ	φ	α_u	α_v	u_0	v_0
40.64	1.07	-6.95	-47.66	49.34	389.72	288.03
43.18	1.07	-6.96	-47.71	49.30	388.43	288.45
45.72	1.07	-6.95	-47.73	49.26	388.31	287.80
48.26	1.07	-6.97	-47.75	49.27	388.03	289.17
50.80	1.07	-6.96	-47.76	49.24	387.80	288.69



(a) (b) (c)

Figure 3.14: Semi-automatic detection of calibration points. By clicking a dot anywhere on its surface the region growing algorithm finds its gravity center with subpixel accuracy. (a) Points selection and ellipse fitting on the the upper and lower rows in the pattern. (b) Automatic detection of all the calibration points. The centers of the dots are marked with red. (c) Detail of the calibration pattern with the centers of the dots highlighted.

Table 3.2: The extrinsic parameters of the omnidirectional camera calibration using a dotted pattern. The measurement units are: (α, β, γ) the rotation angles in radians and (t_x, t_y, t_z) the translations in mm.

Dist.	α	β	γ	t_x	t_y	t_z
40.64	-3.08	-3.11	-3.14	25.65	10.32	-412.20
43.18	-3.08	-3.10	-3.14	31.36	11.85	-413.12
45.72	-3.08	-3.10	-3.14	32.59	9.85	-411.79
48.26	-3.08	-3.10	-3.14	34.10	15.10	-413.20
50.80	-3.08	-3.10	-3.14	35.89	13.07	-413.32

Calibration using a checkered pattern

The 2D calibration points are the corners of the squares in a checkered pattern. The distance between two adjacent points on the same plane is 5cm. Four samples of the images used for calibration are presented in Fig. 3.15.

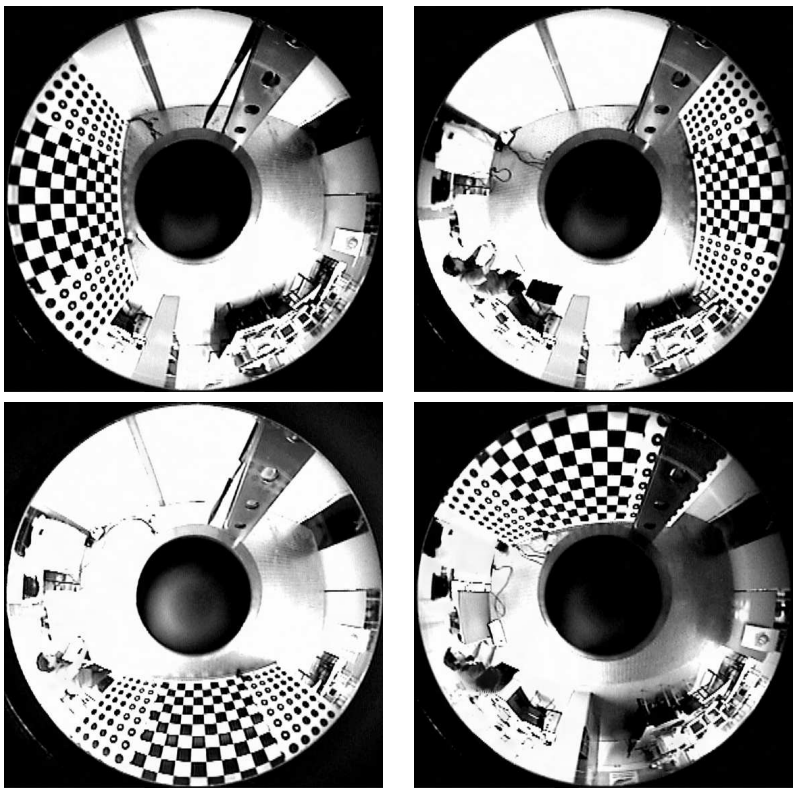
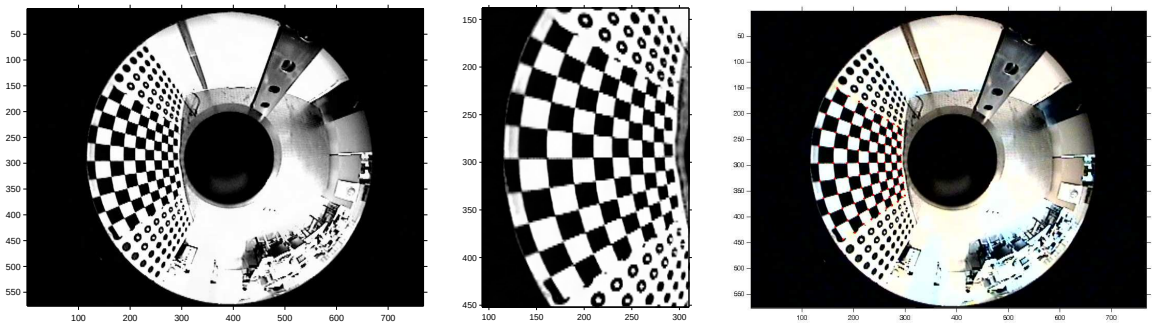


Figure 3.15: The calibration planes placed around the omnidirectional camera.

The image points are obtained using a Matlab programmed tool that enables the user to select manually the desired point. The calibration image is shown in Figure 3.16 while three samples of windows for user input are presented in Figure 3.17. Even though, this method is not automatic, as in the case of the dotted pattern, it proved to be more reliable and fast enough for our purpose.

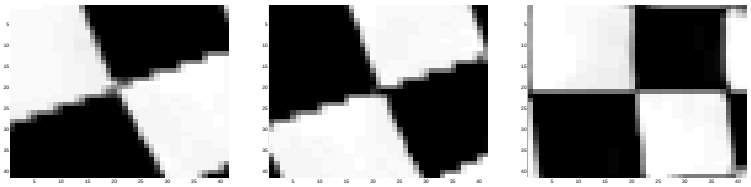
The calibrated intrinsic and extrinsic parameters of the camera-model with the calibration pattern placed at different distances are listed in Table 3.3 and Table 3.4 respectively.

The 2D image error and the 3D scene reconstruction error have been computed for a sample of 1320 calibration points located on planes placed around the sensor in



(a) (b) (c)

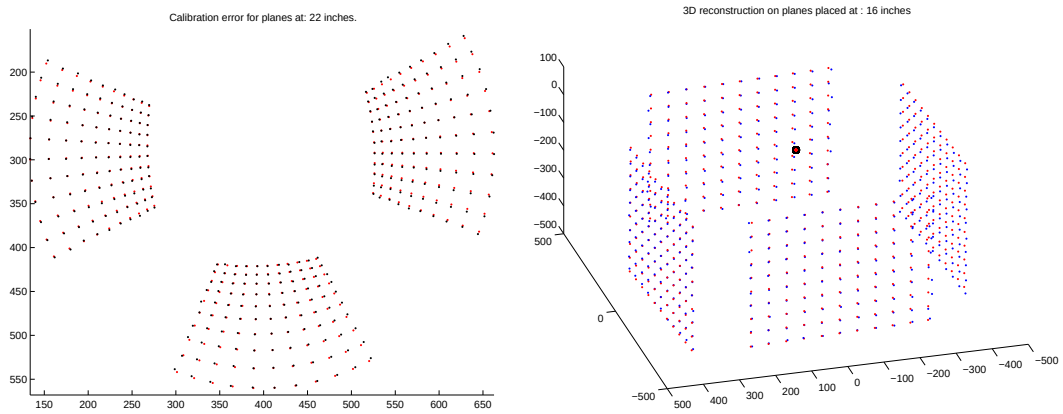
Figure 3.16: Calibration image. (a) A calibration plane placed in the sight of the omnidirectional camera. (b) Detail of the calibration plane. (c) The image of the same plane with the calibration points highlighted.



(a) (b) (c)

Figure 3.17: User input windows for the detection of the calibration points.

a range going from 40 cm up to 80 cm. The average 2D calibration error, computed as the distance in pixels from the imaged point and the modelled image of the same point, is $\mu_{2D} = 1.22$ pixel and the sample standard deviation $\sigma_{2D} = 1.48$ pixel. The 2D error can be observed by the superposition of the real image points and the points calculated using the calibration model, as shown in Figure 3.18 (a). The 3D calibration error was calculated as the distance between the backprojection of the image points on the calibration planes and the 3D coordinates of the corresponding real points from the calibration pattern. The average 3D calibration error is $\mu_{3D} = 3.09$ mm and the sample standard deviation $\sigma_{3D} = 2.75$ mm. The 3D error can be visually estimated from the backprojection of the image points onto the calibration planed presented in Figure 3.18 (b).



(a) (b)

Figure 3.18: Graphical representations for qualitative inspection of the calibration. (a) Image formed by the real points detected by the user (blue) and the calculated points using the calibrated model (red). (b) 3D reconstruction of the calibration points (red) using the calibrated model. The real 2D points are plotted with blue.

After analyzing the calibration results provided using the two calibration patterns we note that the calibration using a checkered pattern proved to be more accurate than the one performed using a dotted pattern. The error introduced by the use of the last mentioned pattern is mainly due to the deformation of the dots after the reflection in the mirror, deformation that invalidates the assumption that the gravity center of the dot coincides with its real center. Thus, the calibration of a catadioptric camera using a dotted pattern is suitable for finding accurate initial values for the calibration algorithm. A checkered pattern should be used for a more precise points detection method and thus a better calibration.

Table 3.3: The intrinsic parameters of the omnidirectional camera calibration using a checkered pattern. The measurement units are: *Dist.*, the radius of the circle on the plane XY, with center in the origin of the cameras' reference system and tangent to the calibration plane, measured in cm. The value 'all' stands for the calibration parameters obtained using the points from the planes at all experimental distanced.; (ξ, φ), the mirror parameters in mm; (α_u, α_v, u_0, v_0), the intrinsic parameters in pixels.

Dist.	ξ	φ	α_u	α_v	u_0	v_0
40.64	1.07	-6.17	-55.09	56.96	392.21	293.31
48.26	1.05	-6.12	-55.20	56.67	393.19	294.69
55.88	1.06	-6.14	-55.27	56.63	393.79	295.07
63.50	1.06	-6.04	-54.95	57.59	392.86	291.45
all	1.06	-6.12	-55.10	56.84	394.23	292.20

Table 3.4: The extrinsic parameters of the omnidirectional camera calibration using a checkered pattern. The measurement units are: *Dist.*, the radius of the circle on the plane XY, with center in the origin of the cameras' reference system and tangent to the calibration plane, measured in cm. The value 'all' stands for the calibration parameters obtained using the points from the planes at all experimental distanced.; (α, β, γ) the rotation angles in radians and (t_x, t_y, t_z) the translations in mm.

Dist.	α	β	γ	t_x	t_y	t_z
40.64	-3.08	-3.12	-3.12	17.14	24.52	-447.08
48.26	-3.08	-3.12	-3.12	13.44	28.94	-449.70
55.88	-3.08	-3.12	-3.12	12.39	31.22	-452.27
63.50	-3.08	-3.12	-3.14	13.48	15.44	-433.72
all	-3.08	-3.12	-3.13	10.22	20.94	-444.29

3.4.3 Omnidirectional laser projector calibration

The omnidirectional light projector is formed by a laser that emits a circle and is pointed to a conical mirror so that the projected light covers the entire field of view of the catadioptric camera. The bright spots on the scene are observed by the calibrated omnidirectional camera which possesses a unique center of projection. This property allows calculating the direction of the light source for each image point. Since the locations of the calibration planes are known, the 3D coordinates of the laser-stripe lying on those planes can be determined. A set of such points can be used for calibrating the laser-conic mirror pair. The 3D projected shape is represented using a parameterized model that was fitted to the experimental data.

A 3D surface can be mathematically described in the implicit form as stated in Equation (3.34), where a is the vector of parameters and (x, y, z) are the coordinates of points in a 3D space.

$$G(a, x, y, z) \cong z - F(a, x, y) = 0 \tag{3.34}$$

A general 3D curve can be obtained by intersecting two implicit surfaces. The choice of the two surfaces is not trivial and an unappropriate selection can lead to unnecessary large set of model parameters, situation known as overparametrization. The correct constraints have to be found and applied in order to remove the redundant degrees of freedom.

Ideally, when the laser is perfectly aligned with the conic mirror, the 3D shape formed by the reflected laser pattern can be imagined as a circular cone, called "laser cone of light". Unfortunately, the precision of obtaining the coordinates of the bright spots is bounded by the catadioptric camera calibration accuracy and by its resolution. Moreover, a perfect alignment of the laser and the conic mirror is difficult to guarantee so a more general shape than the circular cone should be considered. Since the perspective projection of a circle placed on a plane Π onto a plane that is not parallel to Π is an ellipse, it can be deduced that a suitable shape for modelling the laser-cone is a revolution surface whose intersection with the plane perpendicular to the omnidirectional camera optical axis is an ellipse. This shape, the elliptic cone, was used in [Orghidan *et al.*, 2005*b*] and proves to be more accurate than the circular cone. Still, for a large amount of noise, the elliptical cone can not be uniquely determined.

Therefore, the general quadratic surface was chosen for modelling the laser projection. Lets assume, without loss of generality, that the world reference system is placed such that the calibration planes are perpendicular to the X and Y axis. The intersections of the quadratic with the calibration planes are arcs described by a subinterval of the parameter domain: the arcs contained in the planes perpendicular to the X and Y axis provide information on the parameters of the quadratic with $x = ct$ and $y = ct$, respectively. Writing the quadratic as in Equation 3.35, its intersection with the planes X and Y are shown in Equations 3.36 and 3.37, respectively. The parameters of the arcs for each plane are obtained by fitting the corresponding points into the subsequent equations. Taking into account that the $3x3$ matrix is symmetric, most of the parameters of the quadratic surface can be retrieved from Equations 3.36 and 3.37. The parameters a_{12} and a_{21} correspond to the ellipse obtained by crossing the quadratic with the plane XY.

$$\begin{bmatrix} x & y & z \end{bmatrix} \cdot \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \cdot \begin{bmatrix} x \\ y \\ z \end{bmatrix} + \begin{bmatrix} x & y & z \end{bmatrix} \cdot \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} + f = 0 \quad (3.35)$$

$$\begin{bmatrix} y & z \end{bmatrix} \cdot \begin{bmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{bmatrix} \cdot \begin{bmatrix} y \\ z \end{bmatrix} + \begin{bmatrix} y & z \end{bmatrix} \cdot \begin{bmatrix} P_x \\ Q_x \end{bmatrix} + R_x = 0 \quad (3.36)$$

$$\begin{bmatrix} x & z \end{bmatrix} \cdot \begin{bmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{bmatrix} \cdot \begin{bmatrix} x \\ z \end{bmatrix} + \begin{bmatrix} x & z \end{bmatrix} \cdot \begin{bmatrix} P_y \\ Q_y \end{bmatrix} + R_y = 0 \quad (3.37)$$

With the added constraints, the number of parameters to be simultaneously minimized decreases which leads to a more accurate computation.

In a general case the light hitting a conic mirror (Figure 3.19 (a)) produces a conical shape, see Figure 3.19 (b). If there is a special configuration between the vertical angle of the mirror plane and the incident light angle (Figure 3.19 (c)), the projected pattern becomes a plane as illustrated in Figure 3.19 (d).

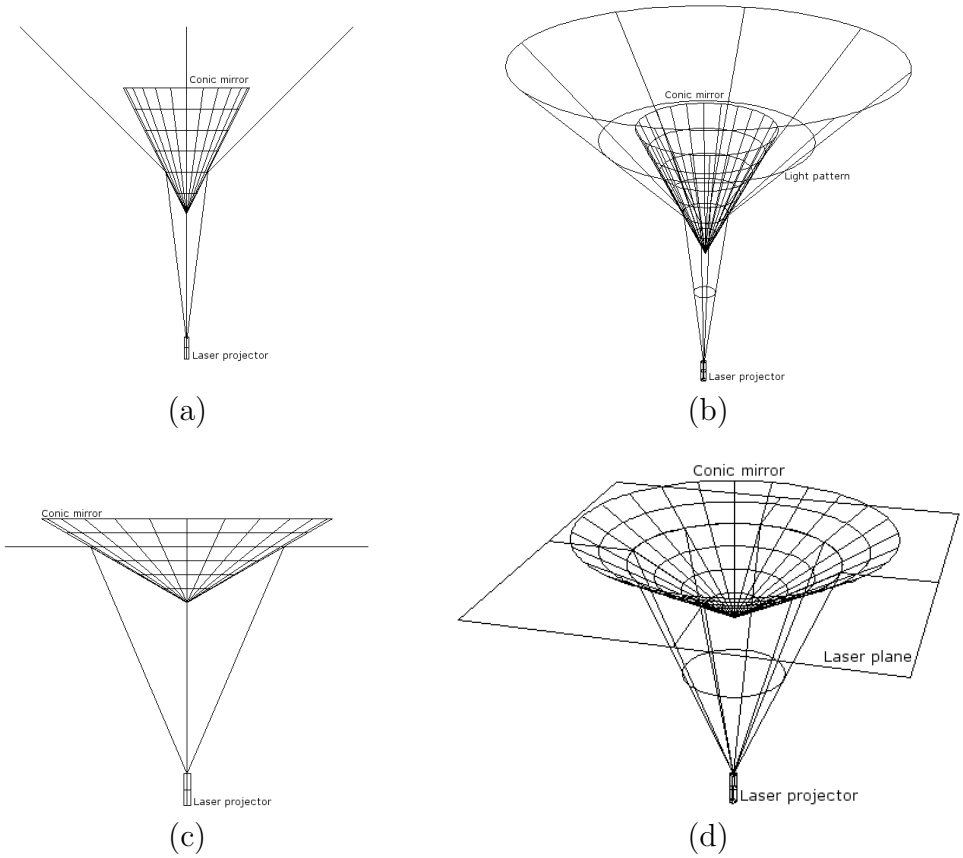
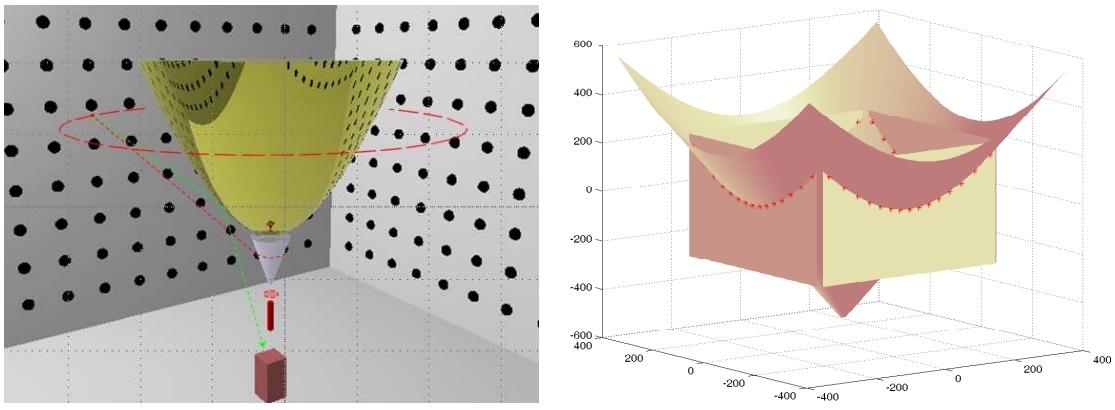


Figure 3.19: Omnidirectional projection of a structured light pattern. (a) Projection of a conical shaped pattern. (b) 3D view of the conical projection. (c) Projection of a planar pattern. (d) 3D view of the planar projection.

3.4.4 Experimental results

The bright spots on the scene are observed by the calibrated omnidirectional camera which possesses a unique center of projection. This property allows calculating the direction of the light source for each image point as shown in the simulation presented in Figure 3.20 (a). The robustness of the calibration method against noise was tested under simulation. We generated a set of 61 points representing the intersection of a cone with the walls of a calibration box, see Figure 3.20 (b). Then, the three coordinates of the points were perturbed with gaussian noise having the variance of σ.

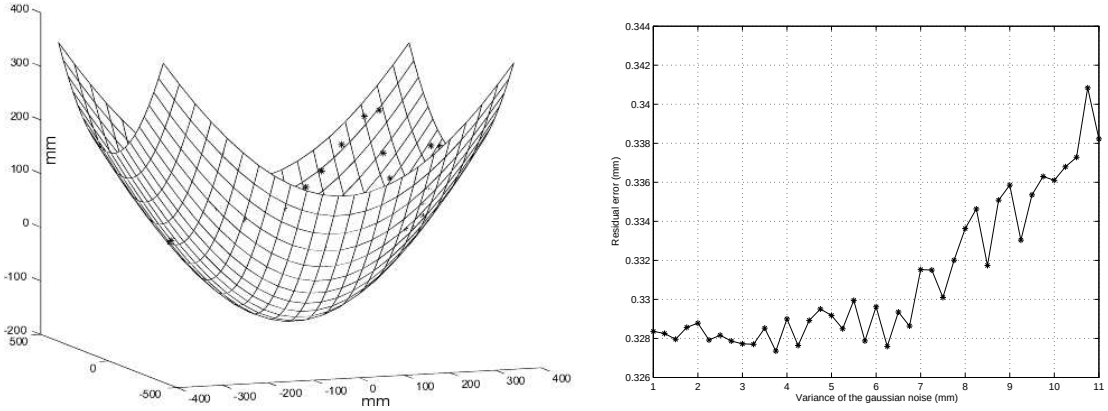


(a) (b)

Figure 3.20: Functionality simulation of the structured light projector. (a) A simulation of the laser projection and triangulation performed for depth perception. (b) Generation of points for the laser projector calibration. The points, marked with stars, are placed at the intersection between the conical shape and the walls.

The general quadratic shape, representing the projection of the circular laser pattern into the conic mirror, was fitted to the 3D points and the resulting surface is shown in Figure 3.21 (a). The error is considered to be the residual value obtained by plugging the 3D points into the quadratic equation. Consequently, for the points laying on the surface the residual value equals to zero and it increases for points placed away from the surface. The calibration was validated one hundred times for each value of σ in the interval [1..11] with a step of 0.25. The average of the absolute values of the residual errors was calculated at each step. The evolution of the error as a function of noise is presented in Figure 3.21 (b).

The conic mirror used for building the laboratory prototype has a height $h =$



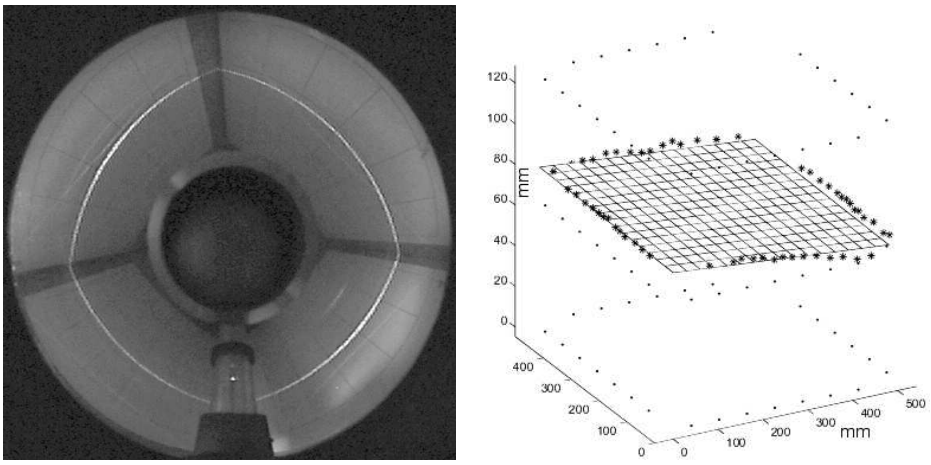
(a) (b)

Figure 3.21: Calibration of the general quadratic shape. (a) Calibrated quadratic fitted to the 3D points. In the illustrated case the points were affected by a gaussian noise with $\sigma = 5$. (b) The evolution of the error as a function of noise.

4.4 cm and the cone aperture angle is $\beta = 52$ degrees. The laser projects a circular cone with a fan angle $\alpha = 11.4$ degrees. Since the relation between the two angles is $\beta \approx 0.5(\alpha + \pi/2)$ the laser is reflected along a very flat surface which can be approximated to a plane: $ax + by + cz + d = 0$, see Figure 3.22 (b).

The 3D positions of the laser points are calculated using the reversed camera model for the back-projection of the image points on the known calibration planes. For a more accurate calibration, the laser plane was fitted to a cloud of points obtained by the back-projection of the laser image points on several calibration planes placed at different distances and orientations with respect to the sensor. The intersections between the laser plane and the planes of the calibration box are straight lines which are imaged as elliptical arcs by the omnidirectional camera. Hence, the calibration accuracy is increased by performing an ellipse fitting for each curve in the image, as shown in Figure 3.23. However, this approximation can only be done for those segments of the stripe that represent the projection of the laser pattern onto a plane on the scene.

The laser parameters obtained by calibration are: $a = -0.0212, b = 0.0019, c = 1$ and $d = 91.7291$.



(a) (b)

Figure 3.22: Calibration of the laser plane. (a) Omnidirectional view of the laser pattern. (b) Plane fitted to a set of points of the laser stripe. The dotted rectangles are the points on the planes around the camera.

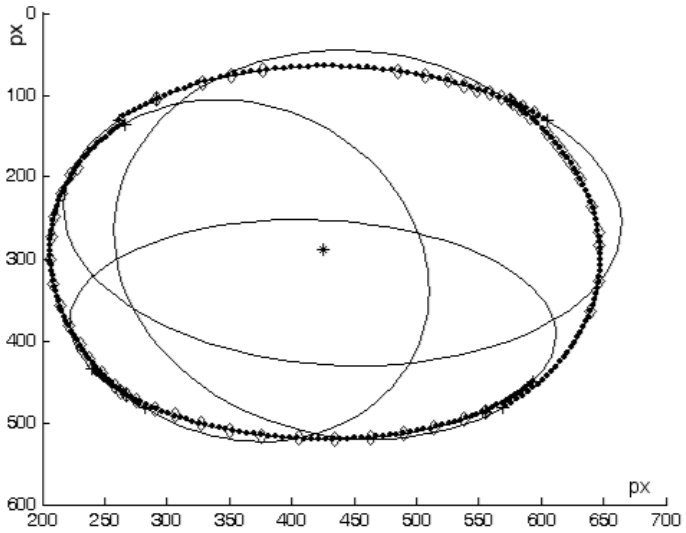


Figure 3.23: Ellipse fitting to laser stripes projected on the four calibration walls.

3.5 Conclusions

We presented the modelling and calibration of a 3D vision sensor which explores depth within the complete azimuthal FOV. The main contribution of the proposed sensor is the use of an omnidirectional light projector for a fast and accurate matching. The traditional stereo systems use either a search window or the epipolar geometry to detect the correspondences in the images. The use of the catadioptric cameras difficult the search of correspondences because of their variable resolution. This problem is alleviated by the structured light pattern which is immediately detected in the scene.

This chapter also presents the design of the sensor. The camera and the laser projector have been placed in several configurations that were analyzed and the one that produced the maximum FOV overlap i.e. the largest stereo region was chosen. Another advantage of the sensor comes from its construction since it has no moving elements. We obtained therefore a robust omnidirectional sensor capable to scan simultaneously all the horizon around it.

The calibration of catadioptric sensors is based on the advantages offered by their peculiar behavior at the image formation step. The reflection into the mirror transforms the scene image depending on the mirror shape. Since the mirrors generally possess a known geometry, the relation between the parameters of the equation modelling the specific shapes and the image formed by the scene points into the mirror can be calculated. Firstly, we presented a direct solution for computing the parameters of a catadioptric system. However, the proposed equation might not have exact solutions so a different method was chosen for modelling the catadioptric camera. The proposed solution is based on the reflection law and estimates for each world point the point on the mirror surface that approximates better the correct reflection point. Another calibration method presented in this chapter uses the equivalent sphere for modelling the catadioptric system. Considering a set of points on the scene with known locations or simply a calibration pattern, the parameters can be extracted by the minimization of the resulting systems of equations. The main part of this chapter deals with the calibration of the components of the sensor. Two calibration models have been analyzed for the camera and their residual errors were compared. The laser projector was finally represented by two models (the general quadratic and the plane) but other models were also evaluated (the perfect cone and the elliptic cone).

An important effort was dedicated to the choice of the calibration pattern which is critical for obtaining good reference points. Since the proposed sensor is meant to find depth in scenes with an unknown structure its model must perform well for the reconstruction of objects placed at unpredictable distances. Consequently, the calibration was performed using patterns placed at different distances from the sensor. The points of the pattern must be clearly visible in the image no matter where we place the calibration pattern. However, the clarity of a point image depends both on the camera resolution and on the texture of the scene. Since the camera resolution is fixed we evaluated three different calibration patterns. First, the calibration points were placed at the intersections of the lines of a grid. Then, the centers of a set of black coplanar dots were taken into account. Finally, the reference points were considered to be the corners of the black squares of a checkered pattern. The camera was calibrated using each pattern. The grid pattern fails at large distances because the image of the lines become too weak. Besides, the dotted pattern introduces unacceptable errors at short distances because the circles are deformed by the reflection in the mirror. Therefore, the checkered pattern was chosen because it permitted the selection of the points in the image with good accuracy independently of the distance between the calibration planes and the camera.

Finally, the sensor was calibrated using the optimal models for its components. The camera was modelled using the sphere of equivalence model and the checkered pattern. The projection of the laser was modelled by a plane.

The combination of omnidirectional vision with structured light is a powerful approach that leads to an overall configuration characterized by a better accuracy, improved robustness and simplicity of construction.

Chapter 4

Data acquisition and processing

The data obtained from the sensor is processed in order to accurately retrieve useful information. This chapter explains in detail the techniques involved in image segmentation, laser detection, the 3D scene reconstruction or the image unwrapping among other software tools for transforming the rough data into valuable information for robot and human use.

4.1 Introduction

The sensor described in the previous chapter provides omnidirectional images containing both the projection of the laser and the texture of the observed scene. These images can be regarded as rough data that needs to be interpreted. Our goal is to manage and analyze this data using proper processing software.

Due to their architecture traditional cameras work in a very similar manner to our eyes and form a perspective projection of the world. This is why the interpretation of perspective images is done almost effortlessly by humans. Besides, images acquired with a catadioptric camera appear distorted. For instance, the parallel margins of a table appear curved and seem to be separated by a variable distance while the vertical edges of a wall become radial lines in the image. Therefore, the "correction" of these images is equivalent to their transformation into images under-

standable by human subjects. We propose here two transformations: the *perspective* and the *bird-eye view* described in section 4.2. Even though both transformations are formed through a SVP and lead to perspective images we call bird-eye view transformed image an image that presents an upper view of the scene as opposite to the perspective transformed image which is an image that offers a directed view similar to a conventional camera.

The perspective transformation is a way to represent the curved image from the surface of the mirror on a flat surface such as the computer display. A 3D curved shape can provide the most accurate image representation for the scene reflected onto the mirror but it is difficult to use. Besides, flat projections enable the representation of some or all of the mirror's surface, at a wide variety of scales, on a flat, printable surface. Moreover, since the parameters of the SVP model are known, the image unwrapping also removes the distortion introduced by the curvature of the mirror. The bird eye view is an image unwrapping technique that offers the observer an upper view of the scene. The main advantage of this method is that the perspective effects of the ground floor are removed so the ground features can be tracked without effort by a human operator. Our sensor adds a new dimension to the scene observed by a human operator since it can produce light spots that serve as landmarks with known positions. Therefore, a precise location of the ground features can be performed visually.

Another source of information comes from the study of the deformation of the laser stripe in the omnidirectional image. The projected pattern intersects the objects from the scene and produces a visible contour around the sensor. For example, we can notice intuitively that the projection of the laser onto a plane produces a stripe with a different shape than if the laser was projected onto a cylinder. Thus, the deformation of the laser stripe encodes information about the shapes of the scene objects. Moreover, considering laser projections onto surfaces with the same shape located at different distances from the sensor we obtain also different stripes. Thus, the laser stripe also encodes information about the distances in the scene. Therefore, the first goal is to detect the laser stripe with a maximum accuracy. The laser stripe detection is described in section 4.3.

The laser stripe provides a set of image points that belong to objects in the scene. Since the camera model is known we can back-project the points of the laser stripe along the respective incoming rays modelled as 3D lines passing through the SVP. Besides, the laser points belong to the laser plane. The equations of the 3D rays

Table 4.1: List of data processing tools

Data processing tools	Description
Perspective transformation	The transformation of the curved image from the mirror surface to a flat surface.
Bird-eye view transformation	Orthographic projection of the image from the mirror surface on the ground plane.
Laser stripe detection	Method to filter the noise that affects the laser stripe.
Range scan computation	Computation of the 3D range scan using the back-projected laser peaks.
Mapping from range scan registration	The process of aligning several range scans in the same coordinate system.
Line tracking and segment detection	Line extraction from the cloud of points obtained by the range scans registration. The lines are then divided in segments that represent the projections of the vertical planes onto the laser plane.
Best view evaluation	Evaluation of the sensor position from which the scene plane corresponding to a given segment is observed under the most favorable angle.

and of the laser plane are known so the position of the 3D points enlightened by the laser can be determined, as shown in section 4.4. Since the transformation between the coordinate systems of the camera and the world are known we can measure the distances to the scene objects hit by the laser. We obtain a rage scan i.e. a set of distances measured around the sensor with a desired angular step.

Different range scans can be obtained by moving the robot in the scene. It is natural to think that we can obtain a richer knowledge about the surroundings combining the information obtained by the sensor from different positions. A single range measurement provides a primitive map of the scene but several range scans can be used for a more complex mapping, as described in section 4.5.

In structured environments the scene map obtained from different range scans contains mostly parameterized shapes which can be extracted, as shown in section 4.6. Therefore a parametric model of the map can be obtained. Once the surfaces in the environment are identified we can use the texture information extracted from the camera images for each object.

The methods to detect the laser stripe in the omnidirectional image, to compute

the range scan, to register several range scans, to detect the lines in the scene and then to split them in segments, to calculate the best view of a segment, the 2D and 3D map building and the texture mapping are described in detail in the following part of this chapter. A quick view of the data processing tools is offered in Table 4.1.

4.2 Panoramic and bird's eye view generation

The images acquired by the catadioptric sensors are distorted, basically due to the geometry of the mirror. The mirror curvature makes different world areas to be mapped with different image resolutions. Moreover, straight lines in the scene are generally projected as curves in the image. The only exception are the 3D lines that belong to vertical planes containing the omnidirectional camera's optical axis and are projected as straight (radial) lines. Therefore, the images provided by catadioptric sensors are difficult to understand by human operators. Nevertheless, since the image formation model is known by calibration these images can be "corrected". We present here two methods to remove the distortion from a catadioptric omnidirectional image: the perspective view and the bird's eye view.

In the following we will call the plane on which the perspective image is projected as destination plane. Figure 4.1 (a) shows how the region to be unwrapped and the distance R determine the width W of the destination plane. The distance R is a constant value chosen depending on the model parameters and the desired scale. The vertical FOV, shown in Figure 4.1 (b), is bounded by the useful part of the mirror. The inner circular part that reflects the camera was removed. The 3D plane is scanned and the texture of each point $P_w(x_w, y_w, z_w)$ is retrieved from its projection $P_i(u, v)$ on the image plane using the projective model defined by Equation (3.33) and presented again here for the convenience of the reader. The model parameters are known since the camera is calibrated.

$$\begin{cases} u = \frac{\alpha_u(\xi+\varphi)x_w}{\xi\sqrt{x_w^2+y_w^2+z_w^2}-z_w} + u_0 \\ v = \frac{\alpha_v(\xi+\varphi)y_w}{\xi\sqrt{x_w^2+y_w^2+z_w^2}-z_w} + v_0 \end{cases} \tag{4.1}$$

The mirror is rotationally symmetric thus the geometrical computations can be

reduced to a plane determined by the optical axis, a given 3D point from the destination plane and its corresponding image point. Consequently, the search can be modelled by a non-injective and surjective function which maps the vertical coordinate of the points from the destination plane to the radial coordinate of the image points. Since the parameters of the camera are known this correspondence can be established off-line and stored into a lookup table. This technique considerably accelerates the unwrapping process and permits fast generation of perspective images.

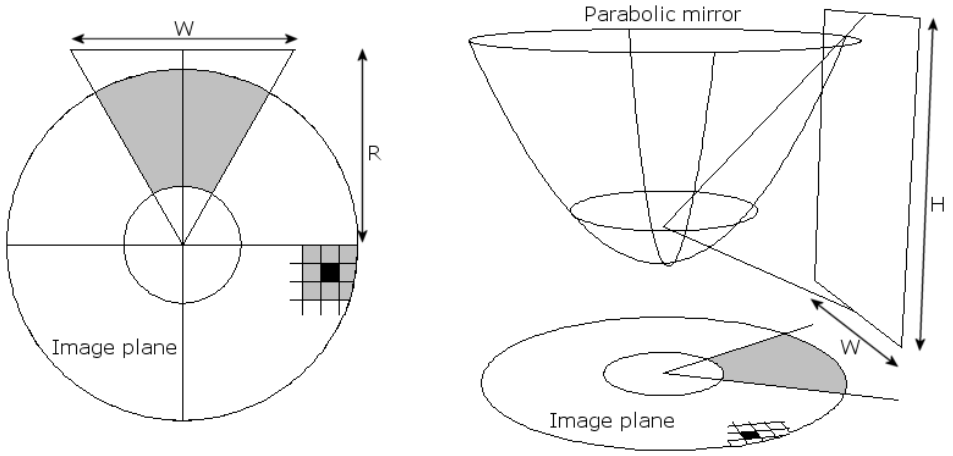
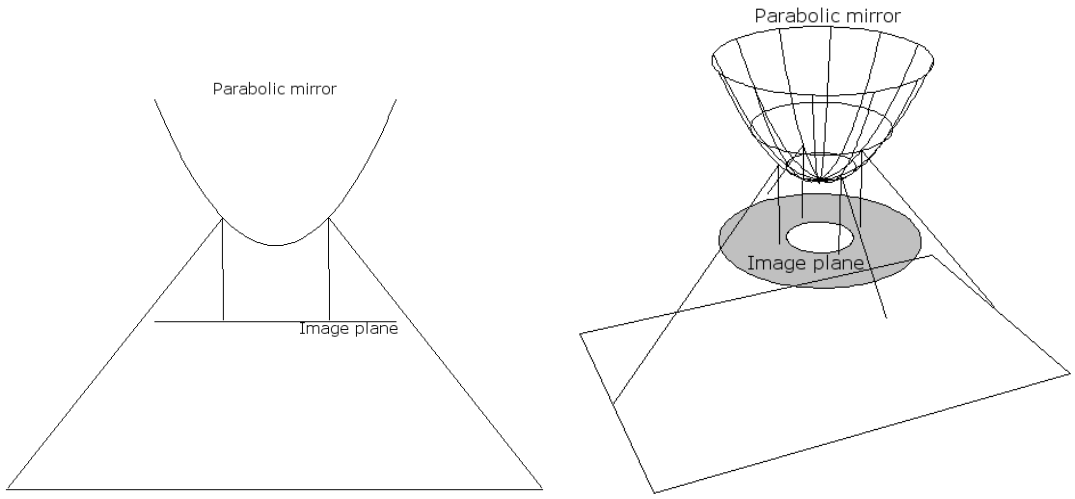


Figure 4.1: Perspective transformation of the omnidirectional image. (a) Upper view of the image and destination planes. The azimuthal FOV is highlighted. The shaded region corresponds to the patch to be unwrapped. (b) Lateral view of the omnidirectional camera. The vertical FOV is given by the useful mirror region.

The image unwrapping must also cope with the non-uniform image resolution. The fact that several points from the destination plane have the same correspondence point in the image plane can create artificially colored images. Therefore, the color of each point of the unwrapped image is calculated by interpolation from the colors of the neighbor pixels of the source image point. Figure 4.1 (a) shown this situation: the black square stands for a point in the image and its neighbors are used for obtaining the final color value.

The bird eye view is similar with the above transformation with the main difference that this technique projects orthographically the image on the ground plane, as shown in Figure 4.2.



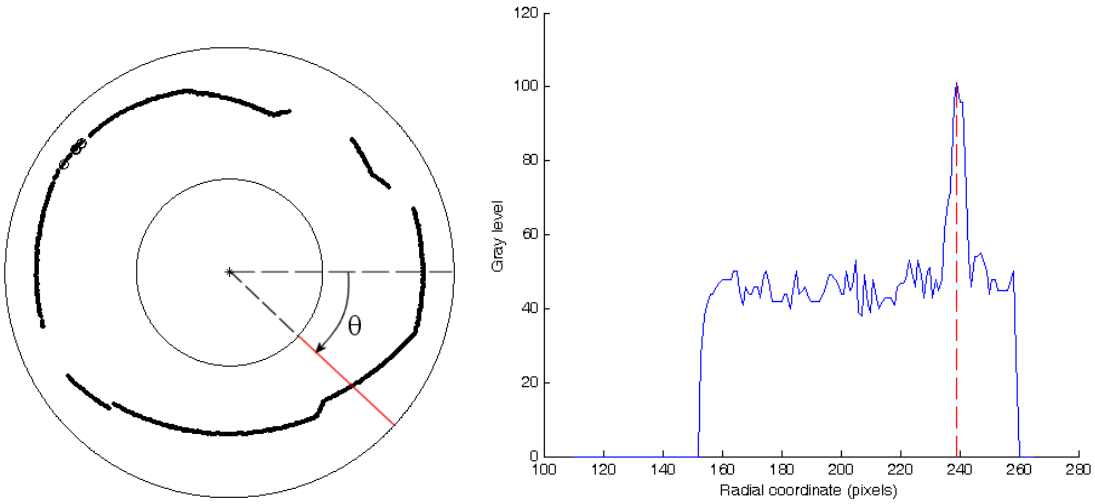
(a) (b)

Figure 4.2: Bird eye view transform. Points on the horizontal plane are related to their projections in the image.

4.3 Automatic laser stripe detection with sub-pixel accuracy

Once the sensor has been calibrated, the image acquisition process is initiated. In each shot the laser stripe in the image is automatically detected. In the omnidirectional image, the laser stripe is affected by the superposition of a certain amount of undesired lighting peaks. This fact is due to noise produced by four main sources: electrical noise, quantization noise, speckle and image blur. The first two noise sources are associated with the image sensor while the speckle is related to the reduced wavelength of light compared to the surface roughness and the monochromatism of the laser light. The image blur is inherent in the catadioptric cameras due to the mirror curvature. An accurate reconstruction needs precise laser detection therefore we used the peak detection method described by Forest et al. [Forest *et al.*, 2004] in order to automatically extract the peak of the laser with sub pixel accuracy. In the original paper, Forest used perspective images containing vertical laser lines. In our case, the omnidirectional image provided by the sensor is scanned radially in order to find the light peaks along the circular laser profile projected on the scene.

As explained, the laser stripe in the image can be considered as a signal affected



(a) (b)

Figure 4.3: Radial scan in an omnidirectional image. (a) A radial line at an angle θ is extracted from the image. (b) The gray level profile along the radial line and the laser peak detected.

by noise with a roughly Gaussian profile. In this case, the maximum light intensity of the signal corresponds to the zero crossing point of the first derivative of the signal intensity. However, because of the noise, the shape of the real signal looks quite jerky rather than like a perfectly smooth Gaussian curve, see Figure 4.3 (b). A previous filtering is therefore necessary.

Consider a radial line in the omnidirectional image as the source signal $x(n)$ for extracting the laser peak, see Figure 4.3 (a). The gray level profile along the selected radial line is presented in Figure 4.3 (b).

First, the signal is filtered by applying a finite impulse response filter $h(n)$ i.e. performing the convolution of the signal $x(n)$ with the coefficients of the filter, see Equation (4.2).

$$y(n) = \sum_{k=0}^{M} x(k) \cdot h(n-k) = x(n) * h(n) \tag{4.2}$$

Based on the Fourier analysis of the signal, the cut-off frequency and the transition bandwidth can be established and the filter coefficients are calculated accordingly. Then, the first derivative of the filtered signal is computed as the convolution of $x(n)$ and the first derivative of $h(n)$, as show in Equation (4.3). With the filtered

first derivative, the method for obtaining the zero crossing point is straight forward. As shown in Figure 4.4, the laser peak corresponds to the zero crossing of the line that joins the last positive and the first negative gray level values of the signal.

$$\begin{aligned}
\frac{dy}{dn} &= \frac{d[x(n)*h(n)]}{dn} \\
\frac{dy}{dn} &= \frac{d\sum_{k=0}^{M} x(k)\cdot h(n-k)}{dn} \\
\frac{dy}{dn} &= \sum_{k=0}^{M} \frac{d[x(k)\cdot h(n-k)]}{dn} \\
\frac{dy}{dn} &= \sum_{k=0}^{M} \frac{dx(k)}{dn}\cdot h(n-k) + \sum_{k=0}^{M} x(k)\cdot \frac{dh(n-k)}{dn} \\
\frac{dy}{dn} &= \sum_{k=0}^{M} x(k)\cdot \frac{dh(n-k)}{dn} \\
\frac{dy}{dn} &= x(n)\cdot \frac{dh(n)}{dn}
\end{aligned} \tag{4.3}$$

Equation (4.4) computes the zero crossing point.

$$\widehat{X} = x_0 - \frac{y_0\cdot(x_1-x_0)}{y_1-y_0} \tag{4.4}$$

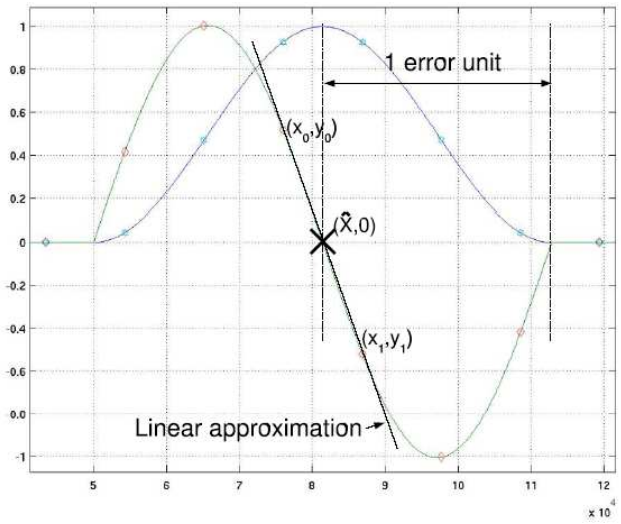


Figure 4.4: Laser peak and the first derivative. *Courtesy of Dr. J. Forest.*

(a) (b) (c)

Figure 4.5: Peak detection of a laser stripe in the omnidirectional image. (a) The image of the surrounding scene captured by the omnidirectional camera. (b) The laser profile corresponding to a real scene. In the the original image the bright laser stripe contrasts with the dark background. For printing convenience the image is presented in negative. (c) Detecting the laser profile. The laser peaks are marked with dots and the outliers are marked with small circles. The image centre is marked by a star. The two concentric circles are the bounds of the stereo field of view and therefore the limits of the automatic search of the laser in the image.

The entire process is summarized by Figure 4.5 which presents the image of the real scene, the laser stripe and finally the laser profile detected in the image.

4.4 Range scan computation

Since the image formation model is known the laser peaks detected in the image can be back-projected. We use here the camera model based on the equivalent sphere so the image points must be firstly projected on the unit sphere.

Given a laser point $P_i(u_i, v_i)$ in the image with the coordinates expressed in pixels. The point is represented with respect to the center of the image, see Figure 4.6. The same point will be expressed in millimeters in the image plane with respect to the camera coordinate system. The resulting point is called $P_{ci}(x_i, y_i)$ and is obtained using the intrinsic parameters of the camera model as shown in Equation (4.5).

$$\begin{aligned} x_i &= \frac{u_i - u_0}{\alpha_u} \\ y_i &= \frac{v_i - v_0}{\alpha_v} \end{aligned} \tag{4.5}$$

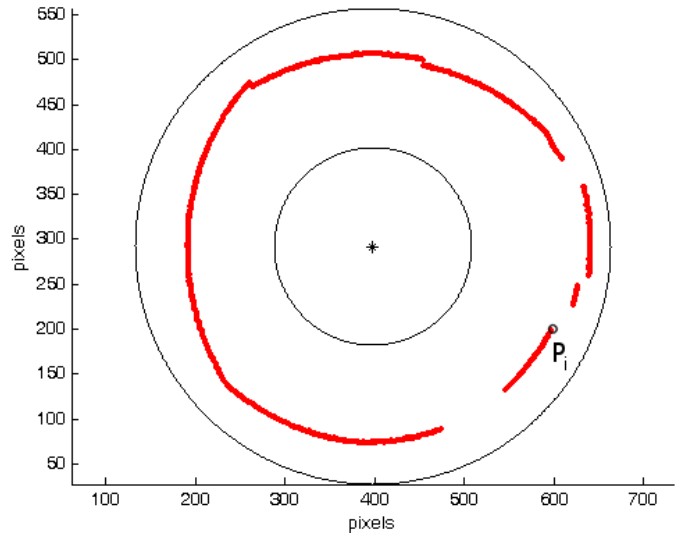


Figure 4.6: Laser stripe used for range scan computation. The point P_i is an arbitrary point that belongs to the laser stripe.

Then, the projection on the unit sphere of the point P_{ci} is called P_s and is obtained by applying Equation (4.6).

$$\begin{aligned} z_s &= \frac{(-k_1-\sqrt{k_1^2-4k_2k_0})}{2k_2} \\ x_s &= \frac{\xi-z_s}{\xi+\varphi}\cdot x_c \\ y_s &= \frac{\xi-z_s}{\xi+\varphi}\cdot y_c \end{aligned} \tag{4.6}$$

where

$$\begin{aligned} k_0 &= \xi^2\sqrt{x_c^2+y_c^2}-(\xi+\varphi)^2 \\ k_1 &= -2\xi\sqrt{x_c^2+y_c^2} \\ k_2 &= \sqrt{x_c^2+y_c^2}+(\xi+\varphi)^2 \end{aligned} \tag{4.7}$$

The projection of all the points of the laser stripe on the sphere is shown in Figure 4.7 where the point P_s is highlighted.

Finally, the point P_s is projected from the origin of the reference system $(O_c(x_0, y_0, z_0))$ onto the laser plane defined by Equation (4.8). The coordinates of the resulting point $P_c(x_c, y_c, z_c)$ are given by Equation (4.9).

$$a\cdot x + b\cdot y + c\cdot z + d = 0 \tag{4.8}$$

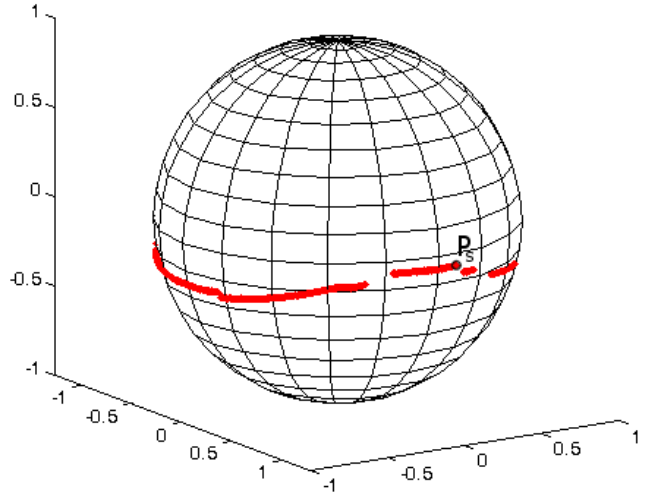


Figure 4.7: Laser stripe projection on the sphere for range scan computation. The point P_s is the projection on the sphere of an arbitrary point from the laser stripe. It is represented with respect to the camera coordinate system.

$$\begin{bmatrix} x_c \\ y_c \\ z_c \end{bmatrix} = \begin{bmatrix} x_0 \\ y_0 \\ z_0 \end{bmatrix} + \frac{ax_0 + by_0 + cz_0 + d}{a(x_0 - x_c) + b(y_0 - y_c) + c(z_0 - z_c)} \cdot \begin{bmatrix} x_c - x_0 \\ y_c - y_0 \\ z_c - z_0 \end{bmatrix} \tag{4.9}$$

The projection of all points from the sphere onto the laser plane leads to a 3D contour called range scan, as shown in Figure 4.8. Note that the range scan is available only in the regions where the laser appears in the image, in other words in the regions where the laser light met an obstacle.

Summarizing, the 3D range scan is computed by triangulation using the camera and the laser projector parametric models.

4.5 Mapping by means of registration

Figure 4.9 shows four range scans calculated at different positions in a scene. It can be seen that each range scan contains different details and even the surfaces that are seen in more than one range scan are detected with different degrees of accuracy.

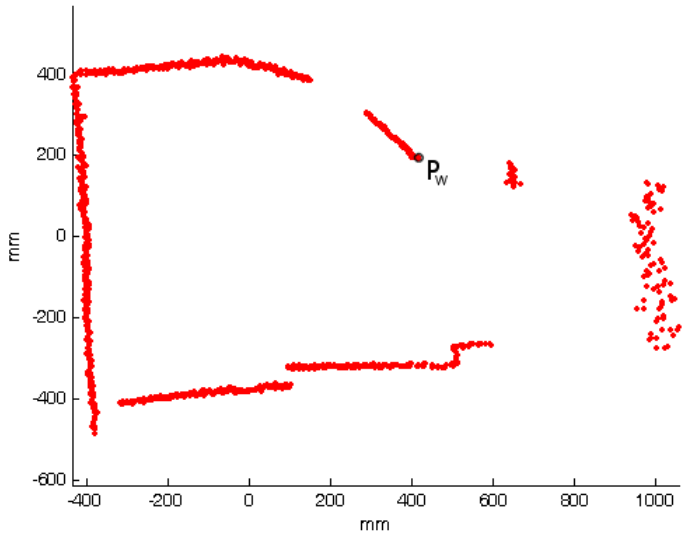


Figure 4.8: The range scan was obtained by the projection of all the points from the sphere onto the laser plane.

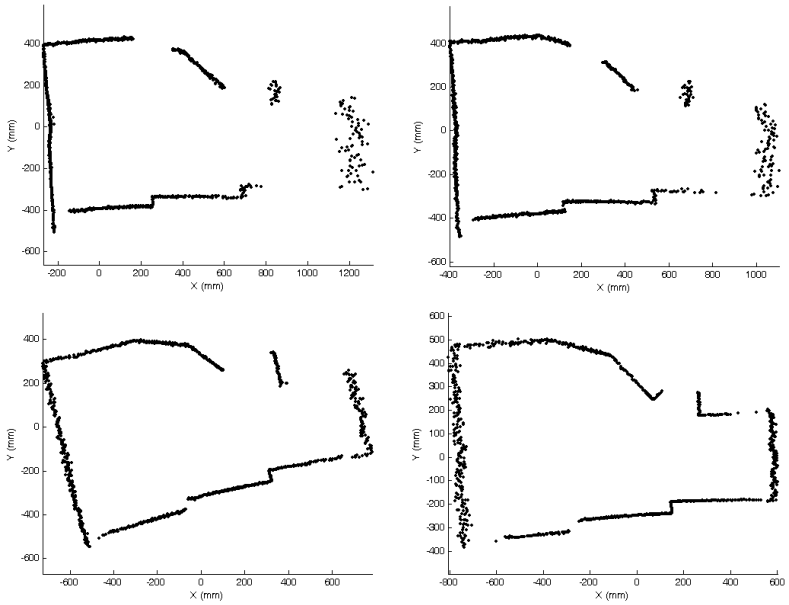


Figure 4.9: Four range scans calculated at different positions in a scene.

The single shot scene model already provides a rich representation but suffers from two inherent drawbacks: there are occluded parts of the scene that are obviously not modelled and the lower resolution of objects in the distance. These shortcomings can be alleviated if several points of view of the scene are used, i.e. by reconstructing a larger scene using different range scans with overlapping regions. In addition, using a large FOV is very useful for finding correspondences between parts of different images because the points remain longer in the sight of the sensor. The process of aligning several lines or surfaces in the same coordinate system is called registration. The most common method for performing registration is the Iterative Closest Point (ICP) algorithm, introduced by Besl and McKey [Besl and McKay, 1992]. This algorithm finds the motion between two clouds of points by minimizing the distance between them. Consider that the coordinates of the N_p pairs of points that belong to the two clouds are stored in the m_i and p_i vectors. The distance between the pairs of points in the two clouds can be expressed by the function f, described by Equation (4.10). The rotation R and the translation t, known as the motion parameters, between the two clouds of points are found by minimizing this function.

$$f = \frac{1}{N_p} \sum_{i=1}^{N_p} \| \overrightarrow{m_i} - R \cdot p_i - \overrightarrow{t} \|^2 \tag{4.10}$$

Although good results are obtained with free-form shapes the major shortcoming of this algorithm is that it can only be used in surface-to-model registrations. However, in real applications, surfaces usually contain non-overlapping regions. Several robust alternatives that use thresholds to detect false correspondences have been proposed by Trucco et al. [Trucco *et al.*, 1999] and Zinsser et al. [Zinsser *et al.*, 2003] in order to alleviate this problem. A different version of the ICP algorithm was also proposed by Chen and Medioni [Chen and Medioni, 1991]. The main difference between the two algorithms is the method to find the correspondences: Chen used the point-to-plane distance while Besl used the point-to-point distance. Generally, the point-to-plane distance solves the problem of local minima better since it overcomes the difficulty of finding correspondence between the points of the two clouds. These techniques, known as Fine Registration techniques, converge fast enough but require a good initial solution. When the initial solution is not available, the Coarse Registration techniques can be used to estimate it. We refer the reader to a in-depth comparison on surface registration recently published in [Matabosch *et al.*, 2005].

In our case, the motion between two consecutive positions is small and the wide field of view offers large overlapping areas which helps the ICP algorithm converge even when the initial solution is the identity matrix. Therefore, the ICP algorithm proposed by Chen was chosen.

A preliminary step for registration is transforming the coordinates of the 3D points of each range scan from the sensor to the laser plane reference system. Therefore, the third coordinate can be ignored and the registration is performed in the 2D space of the laser plane. Consider the laser plane described by the Equation (4.11) and a range scan point in the sensor reference coordinate system ${}^{C}P({}^{C}X_c,\ {}^{C}Y_c,\ {}^{C}Z_c)$. The transformation of the point ${}^{C}P$ into the point ${}^{L1}P({}^{L1}X_{l1},\ {}^{L1}y_{l1},\ {}^{L1}z_{l1})$ expressed in the reference frame $\{L1\}$ of the first laser plane is carried out by using Equation (4.12). The transformation matrix ${}^{L1}R_C$ and the translation vector ${}^{L1}T_C$ are described by Equation (4.13).

$$a \cdot x + b \cdot y + c \cdot z = d \tag{4.11}$$

$${}^{L1}P = {}^{L1}R_C \cdot {}^{C}P + {}^{L1}T_C \tag{4.12}$$

$${}^{L1}R_C = [X \quad Y \quad Z] \tag{4.13}$$

where,

$$\begin{cases} X = \begin{bmatrix} 1-0 \\ 1-0 \\ \frac{d-a-b}{c} - \frac{d}{c} \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ -\frac{a+b}{c} \end{bmatrix} \\ Z = [a \quad b \quad c]^T \\ Y = Z \times X \end{cases} \tag{4.14}$$

and

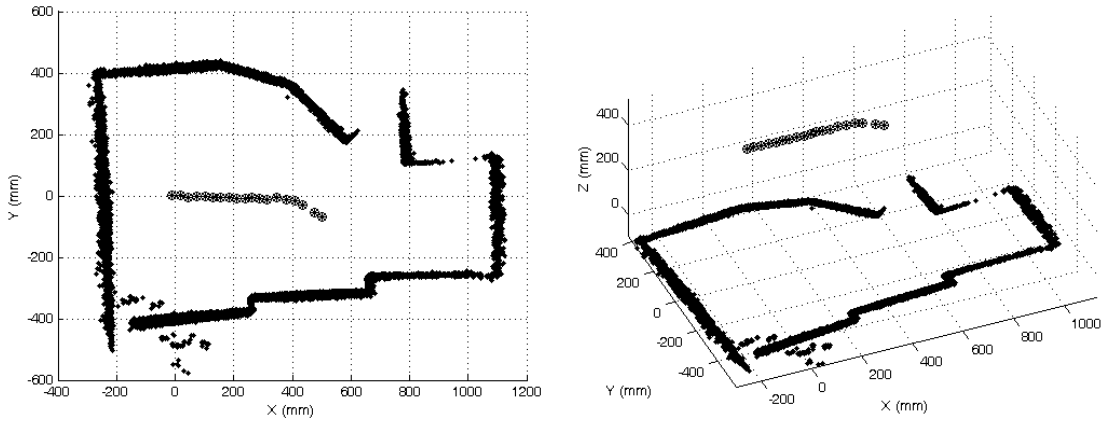
$${}^{L1}T_C = [0 \quad 0 \quad z_L]^T = [0 \quad 0 \quad -\frac{d}{c}]^T \tag{4.15}$$

Subsequently, the range scan points transformed in the 2D space are filtered through a *normal space sampling algorithm* [Rusinkiewicz and Levoy, 2001]. The algorithm consists in transforming the (x_L, y_L) space into a grid of $(n_x \times n_y)$ cells and it selects only one element in each cell. Therefore, the new space is formed by the sampled points $P_s(n_{xi}, n_{yi})$ and thus the weight of the points becomes inversely

proportional to their density.

The distances from the sampled points of the current view i to the planes (or lines in the 2D space) obtained from the un-sampled points of the previous view $i-1$ are calculated and the point-to-plane correspondences are established for all the points of view i. Considering a point in the current view, its pair in the previous view is the nearest point in the corresponding plane.

Finally, the 2D registration is performed by minimizing the distance between the two clouds of points using Equation (4.10). The result of the registration of the range scans presented in Figure 4.9 is shown in Figure 4.10. The base line can be seen in Figure 4.10 (b) where the distance between the camera position and the laser plane is visible.



(a) (b)

Figure 4.10: Cloud of points obtained by the registration of several partial range scans obtained placing the sensor at different locations in the scene. The camera view point is represented by the black sphere. (a) Projection on the XY plane. (b) Three dimensional view.

4.6 Line tracking and segments detection

The Hough transform is a well known image segmentation technique designed for isolating features of a particular shape within a given image. The classical Hough transform requires that the desired features be specified in some parametric form therefore it is commonly used for the detection of geometrical elements such as

lines, circles, ellipses, etc. The Hough technique is particularly useful for computing a global description of a given feature within a set of local measurements possibly affected by noise.

Most of the indoor scenes are formed by human-made shapes and the majority of these shapes contain vertical planes. Therefore, a mobile robot that carries our sensor and inspects an indoor scene should take advantage of the inherent structure of this environment. The Hough algorithm can be used to extract the lines in the cloud of points obtained from the range scan registration. Then, the extracted lines must be divided in segments representing the projections of the vertical planes onto the laser plane.

Let us describe briefly the Hough transform theory. Consider a line which can be expressed by the parametric Equation (4.16).

$$x \cdot \cos(\theta) + y \cdot \sin(\theta) = r \tag{4.16}$$

where r is the length of a normal from the origin to this line and θ is the orientation of r with respect to the X-axis, as shown in Figure 4.11. For any point on this line, r and θ are constant.

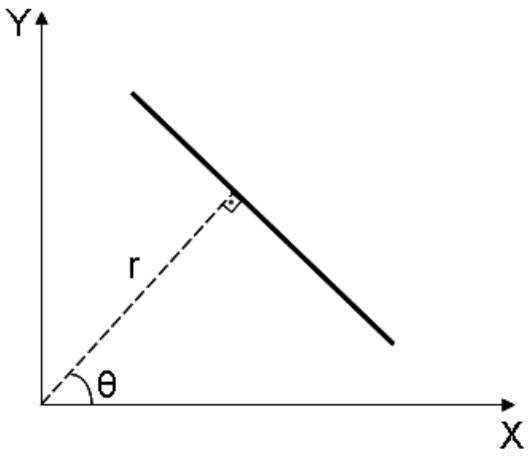


Figure 4.11: Parametric description of a straight line.

In an image analysis context, the coordinates of the points of interest (x_i, y_i) are known and can be plugged in the parametric line Equation (4.16). Therefore, we can define a cartesian space known as the Hough space in which each point (r, θ) corresponds to a line at angle θ and distance r from the origin in the original data

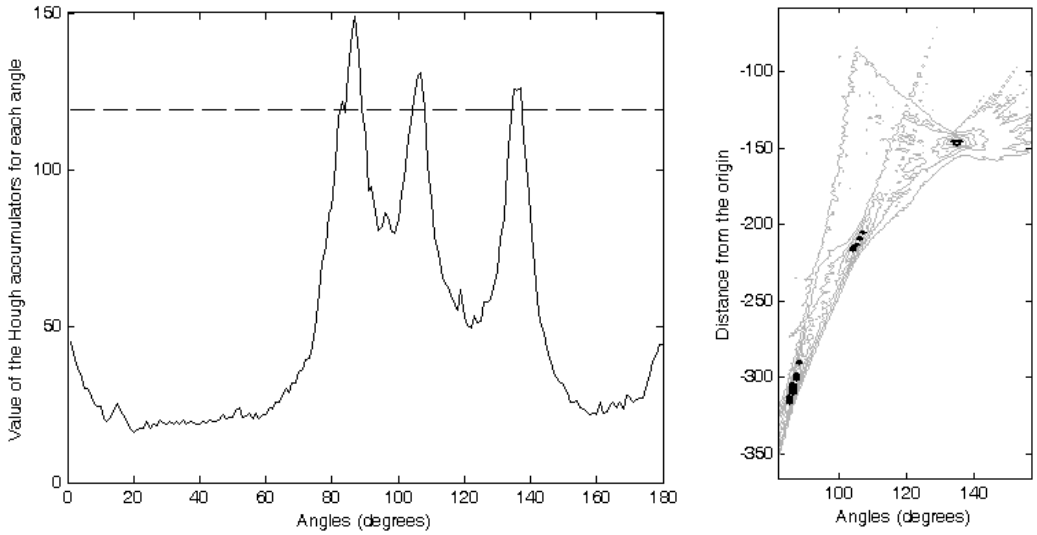
space. A discretisation of the Hough space will result in a set of cells called the Hough accumulators. Resulting peaks in the accumulator array represent strong evidence that a corresponding straight line exists in the image.

The major drawback of the classical Hough algorithm is the tradeoff between the feature search accuracy and the threshold value. A hight threshold would remove the weak peaks and thus the strongly defined lines are obtained. However, this method is based on the assumption that all the lines in the image are defined by a similar number of points and with a similar accuracy. This assumption is unacceptable in images that contain lines with different lengths or affected by different levels of noise. An apparent solution to this problem is to use a lower threshold. In practice this action yields more of the expected lines, but at the expense of many spurious lines. In order to avoid these problems we use a modified version of the Hough algorithm that iteratively changes the threshold until the maximum number of lines are detected.

Figure 4.12 (a) presents an example of the Hough accumulators for each angle θ in a range from 0 to 180 degrees with the step equal to the unity. For each line, we increment a counter (initialized at zero) in the Hough accumulator with coordinates (r, θ). The algorithm evaluates the resulting accumulator array and selects the regions containing a number of points higher than a threshold value tH represented by the dotted line. The centre of the selected regions approximates the lines determined by the points, see Figure 4.12 (b). The lines are memorized and the points that participated in their formation are excluded from the Hough space. The threshold is decreased and the process is repeated until all the lines are extracted from the point cloud.

One or several registered panoramic range scans provided by the sensor can be transformed in a 2D binary matrix which is analyzed using a recursive Hough algorithm with variable threshold adapted for the detection of the maximum number of lines. All the lines identified in the image are bounded and the segments corresponding to planar surfaces in the scene are detected. The incorrect segments are filtered and eliminated.

The 2D map obtained from the registered cloud of points of Figure 4.10 is shown in Figure 4.13.



(a) (b)

Figure 4.12: Line detection in the Hough space. (a) Each point votes for every possible line it could belong to. The lines are uniquely determined by an angle θ and distance d from the origin. (b) Detail of the Hough space containing the regions that approximate the lines determined by the points. The Hough accumulator with the value above the threshold tH are highlighted with a dark border. The center of the highlighted region approximates the coordinates of the most voted line.

4.7 Best pose evaluation and texture mapping

When different view points of the scene are available at the modelling stage we can chose between several images of each object. During the registration process, the transformation between the range scans taken at different locations is evaluated. Since the points of the range scans obtained in each shot are represented in the sensor coordinate system, the different poses of the sensor in relation to each segment can be calculated. Therefore, the best view pose of each segment is determined as being the sensor position from which the scene plane corresponding to the segment is observed under the most favorable angle. The image taken from the best pose is subsequently used for mapping the texture onto the scene objects.

The registration process provides the transformation matrix between the different range scans and thus the robot's position can be calculated with respect to any coordinate system. Figure 4.13 shows the locations through which the mobile platform moved and that were retrieved by registration. The transformation between two locations is computed by Equation (4.17) where R_L and t_L are the rotation and the translation matrices between the sensor coordinate system and the laser plane coordinate system and $t_{n(3x1)}$ is the translation between the n^{th} sensor position and the reference one.

$$\begin{bmatrix} x_{pos} \\ y_{pos} \\ z_{pos} \\ 1 \end{bmatrix} = \begin{pmatrix} R'_{L(3x3)} & (-R'_L \cdot t_L)_{(3x1)} \\ 0_{(3x3)} & 1 \end{pmatrix} \cdot \begin{bmatrix} t_{n(3x1)} \\ 1 \end{bmatrix} \tag{4.17}$$

Therefore, the sensor location and the segments detected in the scene map can be aligned with respect to the same reference coordinate. The optimal position for observing the texture of a given segment (i.e. planar object in the real scene) is the corner of an equilateral triangle containing that segment in the opposite side. The nearest robot location to the optimal position is chosen as the best pose, see Figure 4.13.

Assuming that the analyzed scene is composed mostly of vertical planar shapes placed on a flat floor surface, the 3D scene modelling is performed by generating the elevation map. The elevation map is directly obtained from the positions of the 2D segments considering the vertical field of view of the omnidirectional camera, see Figure 4.14.

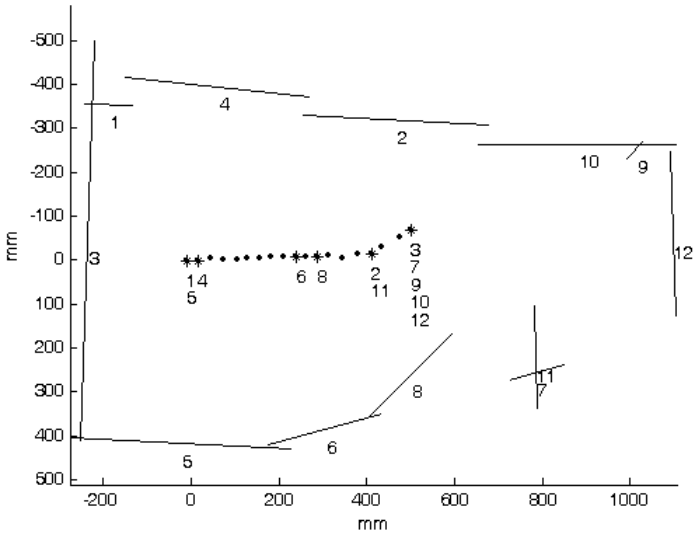


Figure 4.13: The *best view* for each segment in the scene. The segments in the scene are labelled with numbers from 1 to 12. The positions of the robot are marked with dots while the positions selected for mapping the texture on objects are marked with stars. Under each selected position are displayed the labels of the segments best observed from that position.

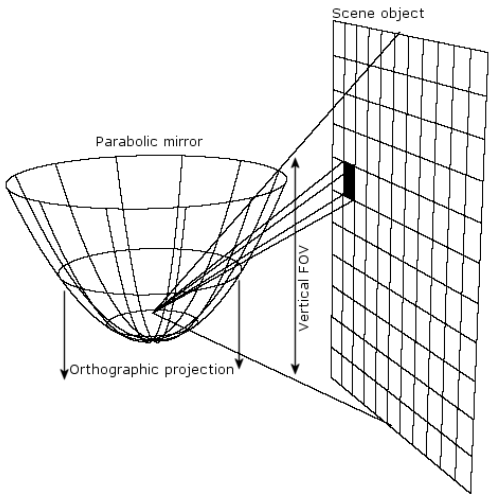


Figure 4.14: The 3D reconstruction of a plane in the scene. The height of the reconstructed plane depends on the distance from the sensor and on the vertical FOV.

4.8 Conclusions

The algorithms presented in this chapter were designed for processing the data provided by our sensor. We addressed two issues which involve the images and the depth measurements.

The images are processed by means of segmentation techniques or by applying geometrical transformations according to the camera model. The segmentation techniques were used for laser stripe detection, range scan computation and line tracking. The geometrical transformations were aimed to remove the distortion introduced by the mirror leading to the perspective and bird eye views and to the texture mapping.

The segmentation of the laser stripe in the image was our first priority as it is an essential condition for the operation of the proposed sensor. We developed a method that scans the omnidirectional image and retrieves the points that belong to the laser stripe. An accurate laser segmentation ensures an accurate depth reconstruction therefore a high accuracy method was implemented for the laser detection. The analysis of the raw light signal is realized in the frequency domain, which yields to the cut-off frequency and the transition band width for obtaining the coefficients of an optimized filter. Due to the filter design the method copes efficiently with the noise that affects the laser in the image and retrieves the coordinates of the points with sub-pixel accuracy.

The detected laser points together with the models for the laser and the camera enable the 3D perception by the triangulation. The range scan computation was successfully realized and the algorithm was described in detail. The depth measurements were involved in the range scan registration algorithms and in the evaluation of the best view. The range scan registration enhances the efficacy of the proposed sensor since it adds up the information obtained by the sensor at different instants and locations. The best view is another useful information because it enables the user to retrieve the best texture of a given object among all the available images of that object.

The lines in the scene are automatically extracted by a recursive Hough algorithm from the cloud of points formed by several aligned range scans. The algorithm described in section 4.6 overcomes the problems of the traditional Hough algorithm which works with a fixed threshold. The advantage of using a variable threshold for filtering the values of the accumulators in the Hough space is that we can accurately

detect the maximum number of lines in the scene independently of their length. Line detection is essential for obtaining a flexible parametric model of the scene. The parametric model provides an alternative to the point cloud and describes the scene in a cost-effective form regarding the computation and storage issues. Moreover, the scene map offers useful information such as size, shape, position and rotation of the objects.

The transformations applied to the omnidirectional images lead to new images similar with the ones taken by the conventional cameras. Consequently, such transformations are useful since the output images can be analyzed by means of traditional image segmentation techniques. Moreover, the resulting images are not distorted, offer new hints about the scene structure and are intuitively interpreted by human operators.

The algorithms presented in this chapter don't require user feed-back since the data is automatically processed. These algorithms represent the basic bricks that will be used to build the applications presented in the next chapter of the thesis.

Chapter 5

Experimental results

A technical view of the proposed sensor is given in this chapter. First, the components used to built the sensor prototype are described. Then, the accuracy of the sensor is determined by means of laboratory experiments. The chapter follows with the description of an application in which the sensor is mounted on a mobile platform and performs self localization and 3D map building. The resulting scene is obtained using the data collected by the sensor from different points of view and offers both 3D and texture information.

5.1 Introduction

Robots and humans as well need a model of their surroundings in order to be able to perform navigation tasks. Visual scene models are mostly used in virtual reality and in mobile robotics. While virtual reality focuses on the geometric and texture details, the scenes created for robot navigation put the emphasis on depth generation and mapping. Therefore one goal of the development of activities using mobile robots consists of achieving a scenario that allows the robot to model its unknown environment, in which it should operate in the future. During the modelling process the robot will learn its environment and the positions of objects in it, so that it is able to interact with them at a later point of time.

We want to combine the qualities of both approaches by generating a scene that contains depth information and texture. Our motivation comes from the necessity of having such a model in remote inspection tasks. We show that we can achieve this goal using the sensor described in the previous chapters of the thesis.

In this chapter, the sensor is firstly analyzed from the point of view of its structure and its components are presented in detail in section 5.2. The sensor specification also contains information about its accuracy presented in section 5.3. The proposed application uses the data obtained by the sensor for performing 3D map building and is suitable for robot navigation tasks or inspection of remote scenes. The effort was focused in modelling the scene geometry for an accurate environment reconstruction.

5.2 Sensor specifications

The sensor was built using off the shelf components. The optics and the mirror used for the omnidirectional camera were provided by Remote Reality, Figure 5.1 (a). The camera is a Sony SSC-DC198P with a 1/3 inch CCD and a resolution of 768x576 pixels. The camera with the mounted optics is shown in Figure 5.1 (b).

(a) (b)

Figure 5.1: Components used for the camera of the sensor. (a) Folded optics. (b) Camera with the optics.

The laser and its optics are produced by Lasiris, the diode power is 3mW and produces red light with a wavelength of 635nm, Figure 5.2 (a). The height of the conic mirror used to build the laboratory prototype is $h = 4.4$ cm and the cone aperture angle is $\beta = 52$ degrees. The laser projects a circular cone with a fan angle $\alpha = 11.4$ degrees. The laser is pointed at the conic mirror as shown in Figure 5.2 (b).

(c) (d)

Figure 5.2: Components of the structured light projector. (a) Laser diode presented with different diffraction heads. (b) Laser diode pointing to the conical mirror.

Depth is perceived due to the disparity generated by the two components of the sensor placed along a baseline, as explained in the previous chapters. At the design stage, the baseline has been implicitly chosen because the laser stripe had to be visible in the central part of the omnidirectional image. After calibration, the length of the baseline was calculated to be 95.26mm. Note that, due to the SVP property of the catadioptric camera, the baseline is constant and does not depend on the elevation of the observed scene point, which is not the case for non-SVP stereo catadioptric systems.

Several prototypes used during our experimental work are shown in Figure 5.3. The first two prototypes are not mobile because the camera is hold above the laser projector by means of a rigid arm fixed on the table. The third model can be moved since it is composed by a rigid structure that holds the two components.

With the sensor surrounded by four planes depth was calculated using a set of discrete points of the laser pattern. For a scene containing two cylinders the result is presented in Figure 5.4 with the two cylindrical shapes correctly identified. It is also noticeable that the points on the walls fall on the corresponding planes.

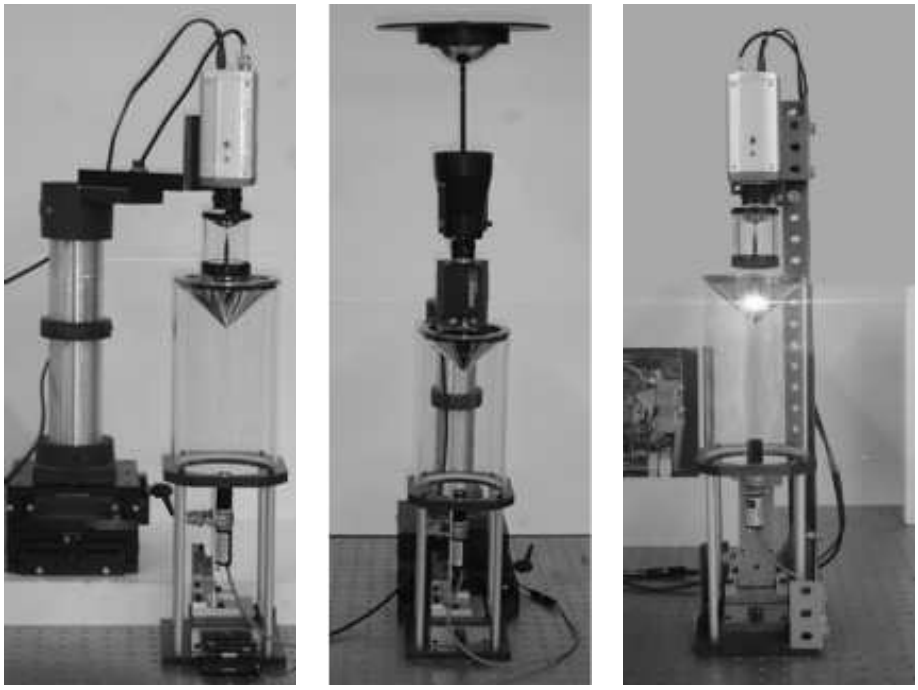


Figure 5.3: Different sensor prototypes.

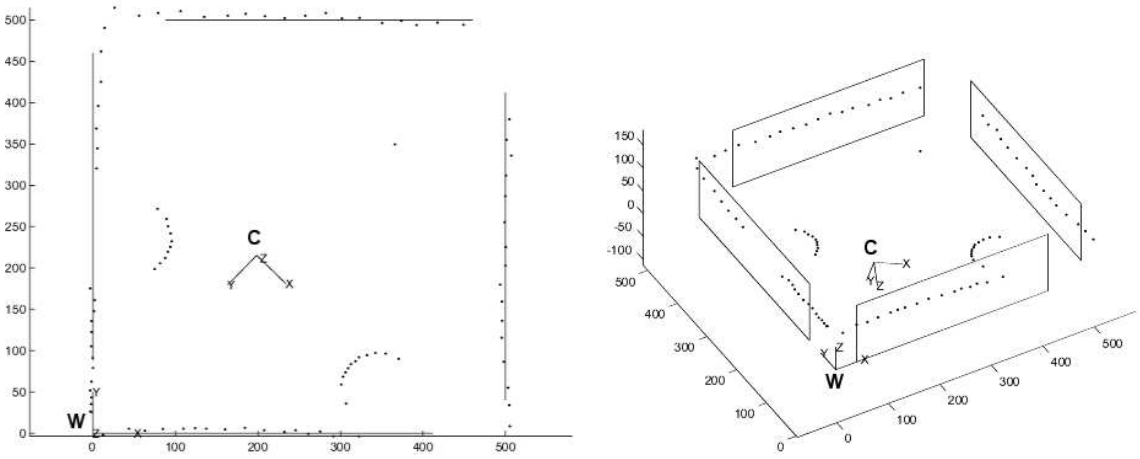


Figure 5.4: Omnidirectional 3D profile obtained along the laser stripe. The dots stand for the reconstructed 3D points. a. Lateral view b. Upper view.

5.3 Accuracy estimation

An important feature of the proposed sensor is its accuracy which was first estimated in [Orghidan *et al.*, 2005*d*]. In order to evaluate the sensor accuracy, we placed a planar surface at different known positions and its location was calculated, see Figure 5.5. The mean error is the average of the distances of the 3D points reconstructed by the sensor to the reference plane. The plane was placed at different distances from the sensor within a range from 40.64cm (16 inches) to 78.74cm (31 inches) and the error was calculated for each position.

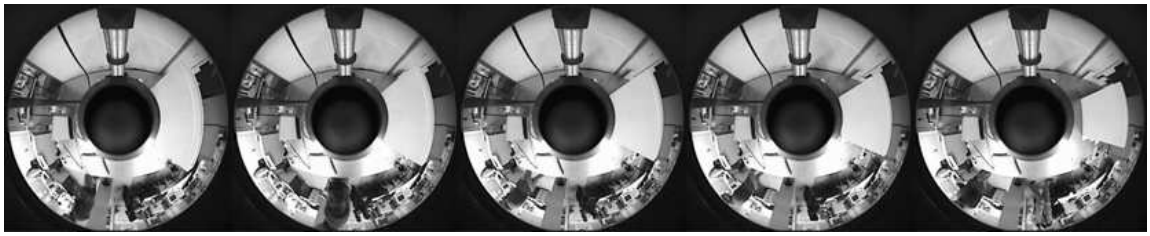


Figure 5.5: The planar surface, placed at different locations, imaged by the omnidirectional camera.

The sensor calibration was performed several times using calibration planes placed at different distances from the sensor. This technique is particularly useful for the proposed sensor since the laser stripe shows up in the omnidirectional image at unpredictable distances from the image centre. Consequently, we obtained five sets of parameters of the model. The first four are called partial calibrations and were obtained by calibrating the sensor using only the points from the planes located at 40.64, 48.26, 55.88 and 63.50 centimeters respectively, as shown in section 3.4.2. The fifth set of parameters was obtained by means of a global calibration that used all the points involved in the previous calibrations.

We observe a smooth variation of the 3D error when the all points calibration parameters are used as shown by the dashed line from Figure 5.6. A higher variation is obtained when the reconstruction is performed using the parameters obtained from the partial calibrations. Overall, we note that the average error of the sensor increases almost linearly and is less than 2.5 cm for obstacles placed in a range of 80 cm around it. In a small range the sensor provides reliable information about the scene depth since the reconstruction error is less than 10 millimeters for obstacles placed at approximately 450cm. Besides, in a larger range the sensor also detects the possible obstacles which give useful hints for navigation tasks.

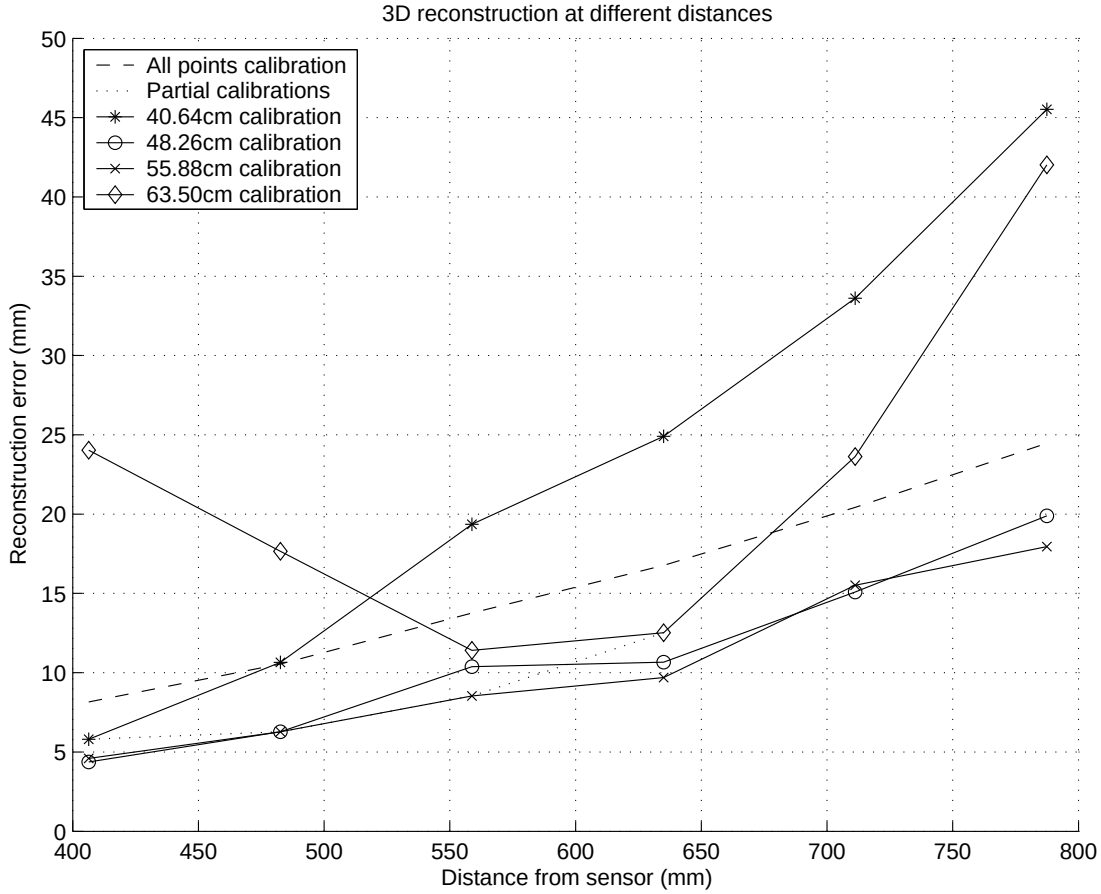


Figure 5.6: Accuracy of the proposed sensor with calibration patterns placed at different distances and with different parameters of the model. The dashed line represents the 3D reconstruction error when using the model obtained by all the calibration points. The dotted line stands for the error using the models obtained when the calibration planes are placed at the same distance as the reference one. The other lines follow the 3D reconstruction errors obtained for planes placed at different distances. The parameters of the four models have been obtained by placing the calibration patterns at 40.64, 48.26, 55.88 and 63.50 centimeters respectively.

Table 5.1: Sensor accuracy within different ranges. The presented error is the average distance between the 3D points reconstructed by the sensor and the reference plane. The values are given in millimeters.

Range	**Partial calibration models**		**All points calibration model**	
	Min. error	Max. error	Min. error	Max. error
400 - 450	4.7	24	8.2	9.1
450 - 500	5.1	21	9.1	11.02
500 - 550	6.5	18	11.5	14.1
550 - 600	8.2	22.5	14.1	15.6
600 - 650	9.1	26.8	15.6	17.3
650 - 700	11	32.5	17.3	19.9
700 - 750	14.2	39.8	19.9	22.5
750 - 800	16.3	47.1	22.5	25.3

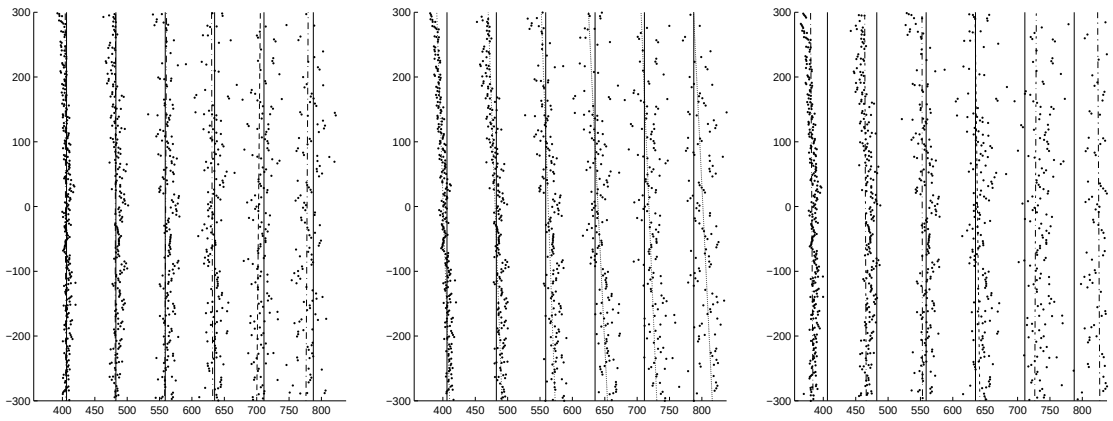
Table 5.1 presents the sensor accuracy within different ranges. The minimum and the maximum errors obtained using different model parameters are shown and can be compared with the error obtained from the all points calibration.

The reconstruction of three planes placed at different positions and using different parameter sets are presented in Figure 5.7 for a qualitative inspection. The measurable quantities represented on both axes are expressed in millimetres.

5.4 Self localization and 3D map building

The calibrated sensor was fixed on a mobile platform and placed in a man-made scene. The sensor is used for generating a complete 3D visual model of the surroundings of a mobile robot. The model is completed by mapping the imaged texture onto the objects through the central view point. This process is called *single shot scene modelling* because a visual model of the scene is obtained from a single omnidirectional image, see section 5.4.1. Several such reconstructions are then registered to build a 3D map of the scene while the robot is moving. We will call this process *scene modelling from range scan registration*, see section 5.4.2.

Generally, a 3D scene model can be represented either as a cloud of points or as a parametric model [Ahn, 2004]. The first mentioned technique is generally used for object visualization by means of polygonal mesh surfaces generated from the point cloud. The resulting models have a limited accuracy in describing the object and



(a) (b) (c)

Figure 5.7: Depth perception for planes at different positions. The continuous lines represent the real positions of the planes. The dotted lines are the planes fitted to the estimated points. (a) Calibration using only the calibration planes at 55.88cm. (b) Calibration using all the points. (c) Calibration using only the calibration planes at 63.50 cm.

require a high amount of computing time and memory space. The alternative to the point cloud is the parametric model which provides useful information such as size, shape, position and rotation of the objects. Moreover, the parametric model describes the object in a compact form alleviating the computation and storage issues. Because of the specific data acquisition of the proposed sensor the parametric model representation is an appropriate choice when using the sensor in structured human made environments. For these reasons, the parametric model was selected for the scene reconstruction.

The flowchart from image capture up to building the map is presented in Figure 5.8. The theory involved in the process was described in Chapter 4. The flowchart also reveals the data transfers between the data processing modules.

5.4.1 Single shot scene modelling

The sensor was placed in the experimental scene and the structured light projector produced the light stripe which was automatically detected in the image with subpixel accuracy using the optimized filter described in section 4.3. A detail of the laser stripe is shown in Figure 5.9 (a) in which the effects of the noise in the laser stripe are visible. The peaks detected are represented by dots in Figure 5.9 (b)

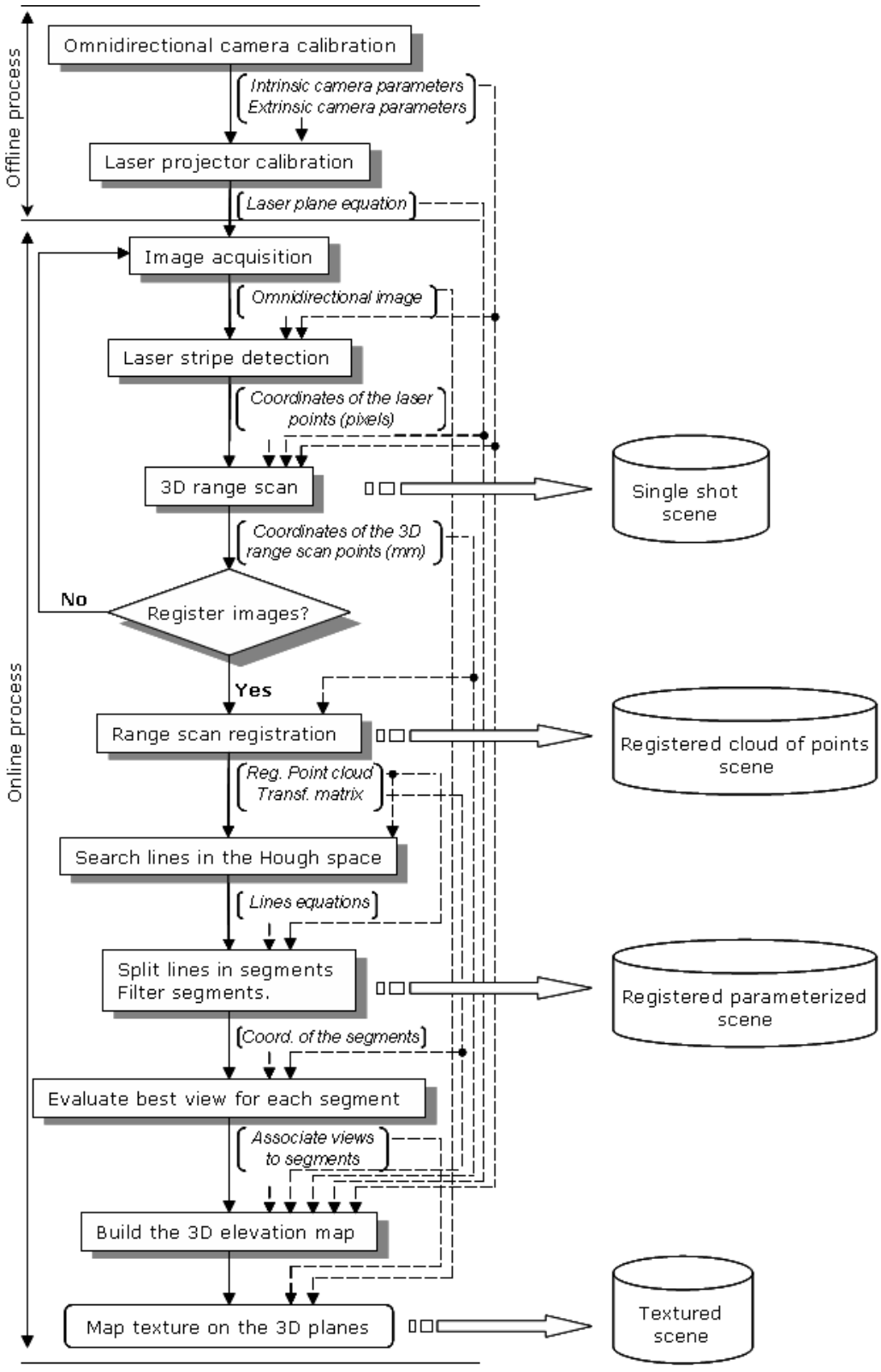


Figure 5.8: Image processing flowchart.

which are subsequently approximated to an ellipse in order to improve the process of detecting straight with subpixel accuracy. Note that lines in the world are seen as curves in the omnidirectional image.

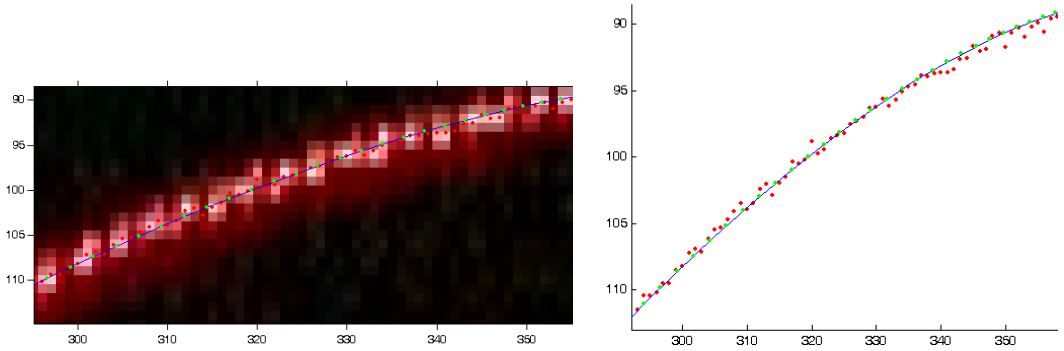
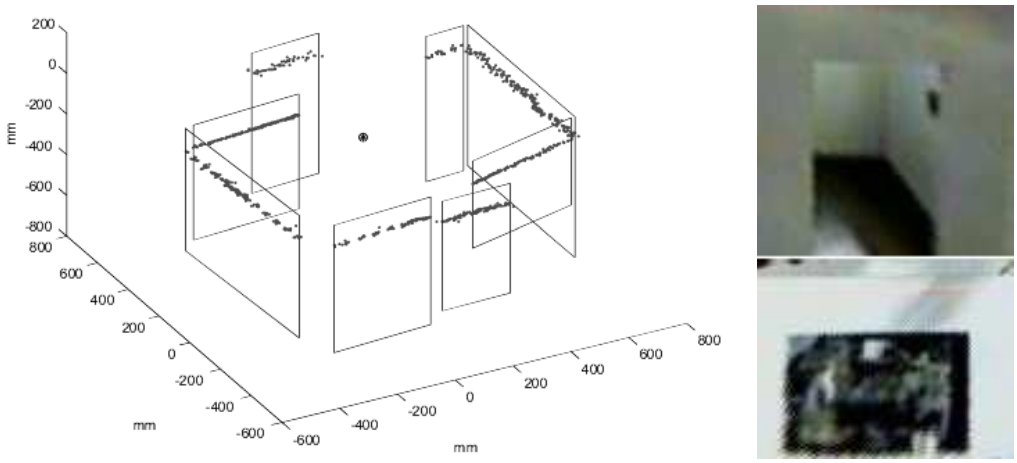


Figure 5.9: Detail of laser stripe detection in an omnidirectional image. (a) Detail of the real image. (b) Real and approximated points in the omnidirectional image.

The range scan provided by the sensor is analyzed using a recursive Hough algorithm and the vertical planes in the scene are detected. Since the FOV of the camera is known, the elevation map illustrated in Figure 5.10 (a) can be generated. The maximum vertical FOV was used due to the fact that all the scene objects were close enough. The parts of the image corresponding to each plane in the scene have been unwrapped and two examples of such patches are shown in Figure 5.10 (b).

The final model, obtained after mapping the texture on the corresponding objects is presented in Figure 5.11. One of the goals of the scene modelling is to provide a visual model for human inspection. Therefore, a VRML file has been built which permits virtual navigation in the scene.

The results obtained for another indoor scene are presented in Figure 5.12. Different points of view are obtained by changing the location of the sensor. The images captured by the omnidirectional camera at each location are shown in Figure 5.12 (a) while the range scan obtained is depicted in Figure 5.12 (b). Screen shots of the VRML scene models are shown in Figure 5.12 (c).



(a) (b)

Figure 5.10: Elevation map and texture unwrapping. (a) Elevation map of the experimental scene. The position of the camera is marked by a black sphere. (b) Examples of unwrapped image parts.

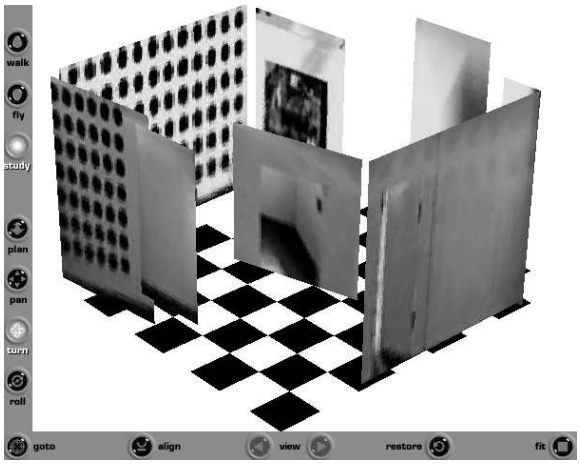


Figure 5.11: The texture is mapped on the corresponding objects which leads to a richer scene model. The checkered pattern, representing the floor, was introduced artificially.

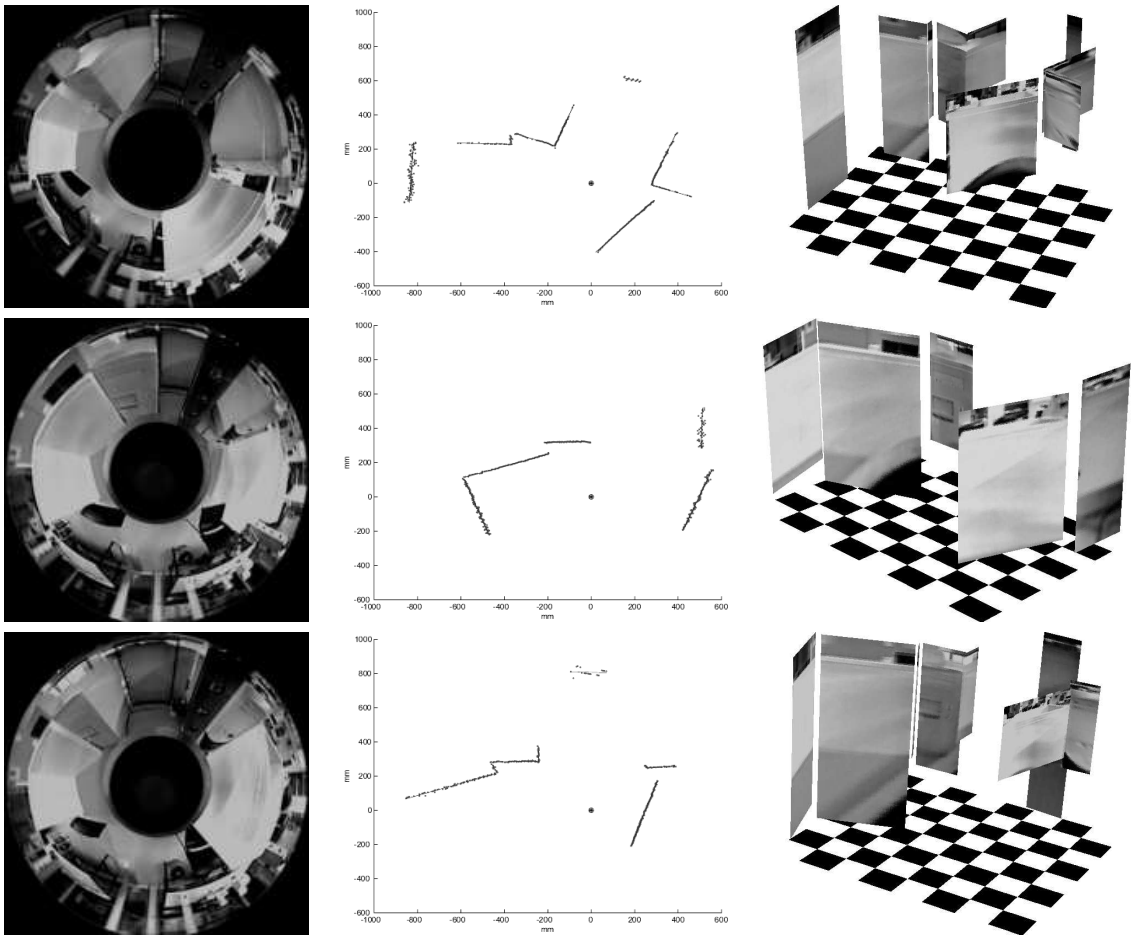


Figure 5.12: Scene modelling with the sensor placed at different locations. (a) The images captured by the omnidirectional camera. (b) The range scans provided by the sensor at each position. The position of the sensor is marked by a black dot. (c) Texture mapped on the surrounding planes.

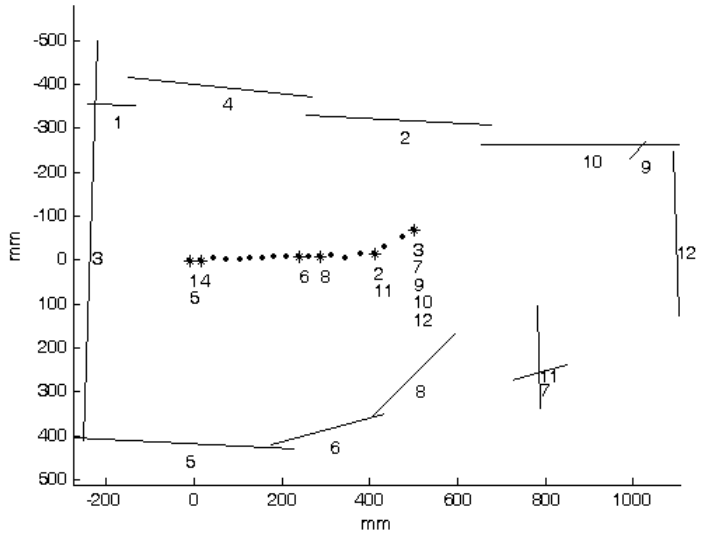


Figure 5.13: Scene map from the registration of range scans calculated at 19 positions.

5.4.2 Scene modelling from range scan registration

A more complex scene has been reconstructed using visual and range scan data from nineteen different points of view which permitted the reconstruction of planes that are not visible from all the view points. The scene map, i.e. the projection of the vertical planes in the scene on the laser plane, is shown in Figure 5.13.

The easiest way to build the scene is to use a single reference point and the maximum vertical field of view (FOV) available for each plane in the scene, as shown in Figure 5.14 (a). In the resulting 3D model the planes placed near the sensor will be lower than the planes placed at a larger distance. This model ensures an intuitive spatial view of the scene since high planes are immediately identified as being far from the robot. On the other hand, the large FOV of the parabolic mirror may create high planes artificially. A natural idea is to bound the height of the walls (Figure 5.14 (b)) since the robot is assumed to move on a flat surface. Even though different scene models can be reconstructed from a single reference point, these models suffer from lack of texture information since there are hidden planes that are not visible from the reference position. Figure 5.14 (c) shows the reconstruction of a scene from distinct points of view using the maximum vertical FOV at each position. This configuration ensures the best available texture for each object in the scene and an intuitive scene geometry. Therefore, the FOV was

bounded only for the planes that are "crossing the floor" on a vertical direction and the resulting model is presented in Figure 5.14 (d).

Finally, the texture is mapped on each plane of the 3D model using the information provided by the omnidirectional camera. The image taken from the location that ensures the best view of a plane is selected among all the available images and is used for extracting the texture of the object. Subsequently, the texture patch is stuck on the 3D model of the map. Two screen shots of the VRML model are presented in Figure 5.15.

5.5 Conclusions

This chapter presents experimental results obtained using our sensor. Firstly, the sensor accuracy is evaluated and the errors obtained for 3D reconstruction are presented. The 3D error which is the average 3D distance between the real points and the reconstructed ones depends on the model of the sensor obtained by calibration. Therefore, several calibrations have been performed. We explore the error obtained by the calibrated sensor model obtained using several calibration planes located at a given distance from the sensor. We show that the accuracy improves noticeably when the sensor is calibrated using points placed on planes located at different distances from the sensor. The reconstruction error obtained for a flat surface located at 80cm is less than 2.5cm and grows linearly. This is a good accuracy that indicates that the sensor can be used for a variety of applications with different accuracy requirements.

A 3D scene model is obtained by moving the sensor in a well structured scene containing planar surfaces. The sensor collects data at different locations. The virtual scene is built using two informational streams: the range-finder scan data and the texture of the scene objects. Both data sources were used to reconstruct the scene as they offer complementary hints about the environment. The position and orientation of the robot are also estimated at each step. The experimental results show that the shapes are properly retrieved and the textures are correctly mapped on the objects that surround the sensor. Mapping the texture on the corresponding objects enables the human operators to detect discontinuities that are not spotted by the laser pattern.

The results obtained from the experiments described in this chapter prove that

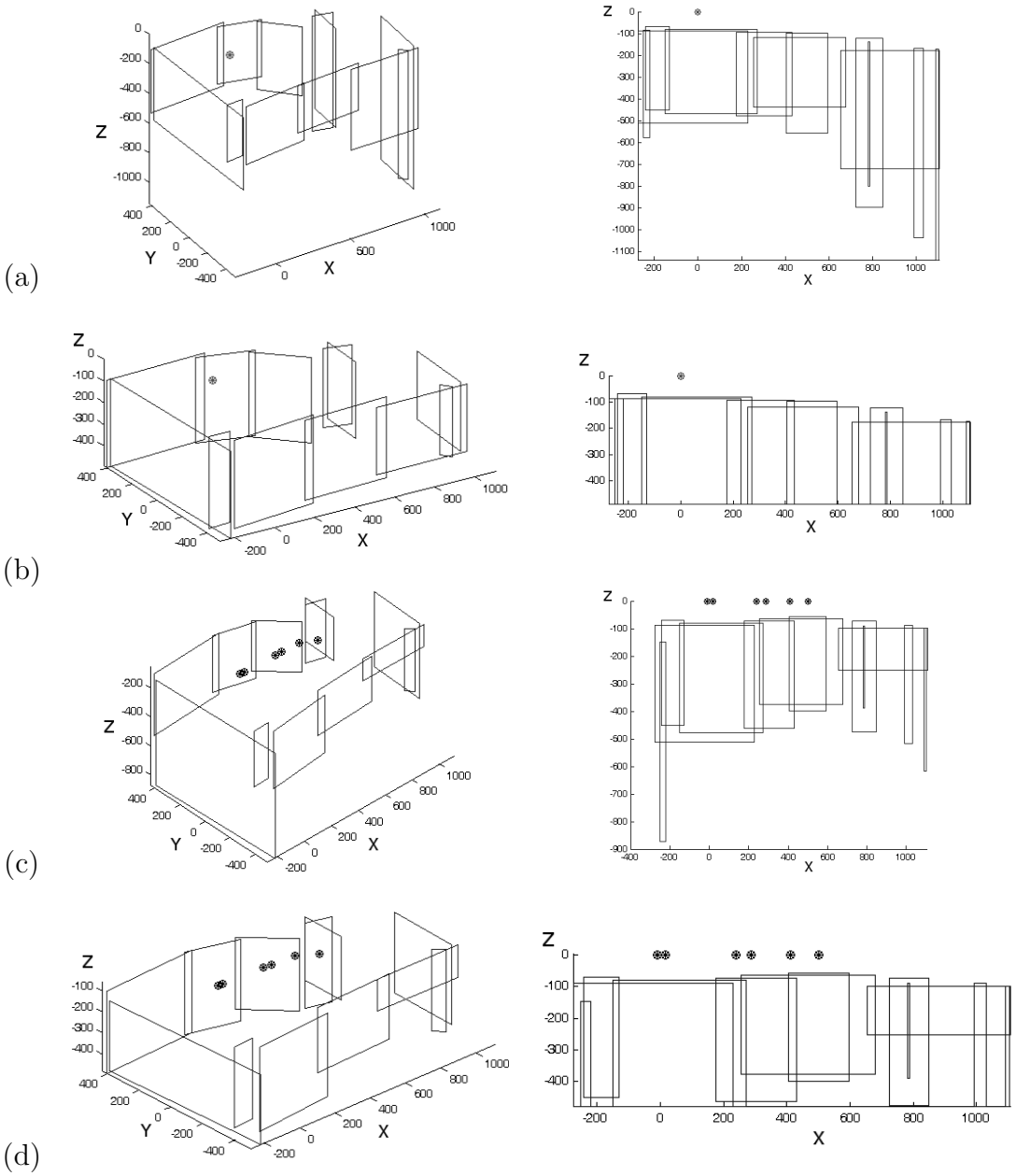


Figure 5.14: Different scene 3D models. (a) Scene from a single view point with maximum vertical FOV. (b) Scene from a single view point with constant vertical FOV. (c) Scene from multiple view points with maximum vertical FOV. (d) Scene from a multiple view points with thresholded vertical FOV.

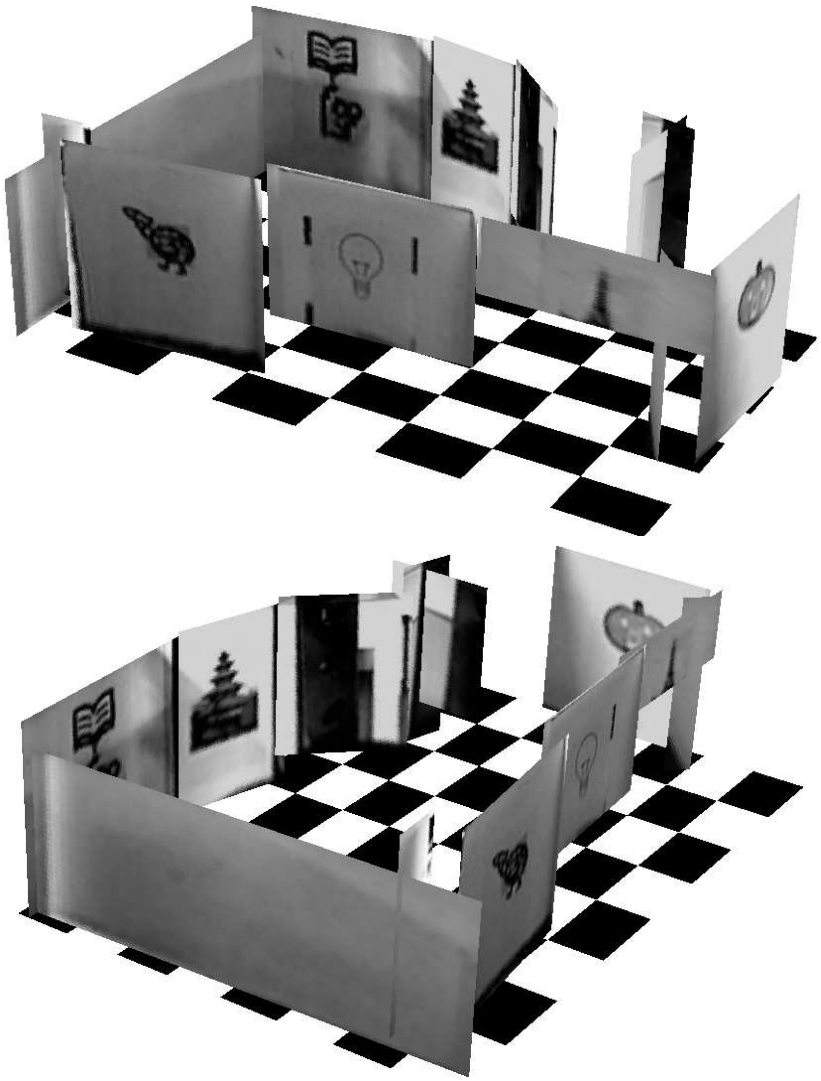


Figure 5.15: The 3D reconstruction of a scene from 19 omnidirectional views.

the sensor can be successfully used in a variety of real applications going from tasks such as autonomous pipe line inspection up to robotic navigation.

Chapter 6

Conclusions

This chapter concludes the dissertation. The conclusions of the contributions are discussed here and the real applications carried out during the research are highlighted since they prove the reliability of the proposed sensor. Moreover, we summarize the contributions of this research and its outcomes i.e. the publications in international conferences and journals. Further work is also analyzed and we suggest new paths for pursuing in the research.

6.1 Conclusions

This thesis is focused on the visual perception of depth by means of omnidirectional cameras. The 3D sensing is usually obtained by means of stereo configurations with the drawback of feature matching between images. Therefore, these techniques are reliable only when the observed scene contains textured objects or landmarks. Moreover, a good feature matching requires high resolution images and it is well known that the resolution of the omnidirectional cameras is much lower than that of the traditional cameras. The solution offered in this thesis uses structured light projection for solving the matching problem. Thus, the visual features are easy to identify in the scene and a more accurate triangulation can be performed.

The thesis starts with a generic introduction in omnidirectional vision including a discussion about the motivations and the techniques used for obtaining large fields of view. One of the contributions of this study is the comparative evaluation of om-

nidirectional cameras which are classified according to five criteria: the fabrication technology, the capability to measure depth, the resolution, the SVP property and the isotropy. Each category is explained and contains several examples of real sensors with detailed descriptions, references to articles and images. The classification parameters and the details of each type of sensors bring valuable guidelines for the choice of a camera depending on the application requirements. The wide range of literature references included in the survey offers an excellent starting points for any research line in the field of the omnidirectional vision.

The survey reveals the main trends for obtaining omnidirectional images and shows that catadioptric cameras provide an optimal trade off between price and performances. However, the search of correspondences between two images is a difficult problem in stereo configurations with catadioptric cameras even when the epipolar constraints are used. The main contribution of this thesis is the introduction of a formal model of a new omnidirectional sensor with depth perception capabilities. The sensor takes advantage of the qualities of the catadioptrics to obtain omnidirectional vision and to project a structured light pattern in the scene. As far as we are aware there are no other similar sensors that use omnidirectional light patterns without moving elements.

The sensor design is addressed in Chapter 3 that focuses on the particular stereo configuration that forms the proposed sensor. Several stereo configurations are evaluated and the optimal one is chosen for the two components: the omnidirectional camera and the laser projector. An accurate model of the sensor is obtained by a careful study of both components.

Two models of the projective geometry of the omnidirectional camera are described and implemented. The mathematical model of a catadioptric omnidirectional sensor is defined by a set of parameters that, basically, define the mirror shape, the camera and the relative location of the components. Since these parameters are difficult to measure directly, a set of calibration methods have been developed and validated by experimental results. The calibration methods presented are valid for any catadioptric set-up: the reflection model contemplates the case of catadioptric devices that use mirrors with known surfaces while the model based on the sphere of equivalence addresses the case of omnidirectional cameras with an unique center of projection. Therefore, the thorough study of the principles of catadioptrics presented in this dissertation offers valuable information and clear indications for modelling any type of catadioptric camera.

The laser projector was also studied and a general model was proposed. Since the laser stripe is observed by the camera after the reflection in two mirrors the noise and the distortion affecting the signal has to be modelled. Moreover, several noise sources affecting the laser have been identified. The analysis of the raw light signal is realized in the frequency domain, which yields to the cut-off frequency and the transition band width for obtaining the coefficients of an optimized filter. Due to the filter design the method copes efficiently with the noise that affects the laser in the image and retrieves the coordinates of the points with sub-pixel accuracy. A method for automatic detection of the peak of the signal was implemented.

Several higher level algorithms, presented in chapter 4, were carried out using the sensor parameters obtained by calibration. A qualitative proof of the correctness of the sensor model and of the calibration procedure are the transformations applied to the omnidirectional image. The Perspective and the Bird-eye projections are two methods used to remove the distortion from a catadioptric omnidirectional image. A visual inspection confirms that the curved lines in the image are correctly unwrapped by both methods. Other software tools for range scan computation and registration, line tracking, map building and texture mapping are described in detail in the same chapter.

The sensor was used in a map building application, presented in chapter 5. The application uses the data obtained by the sensor to perform 3D map building and is suitable for robot navigation tasks or inspection of remote scenes. The experimental results show that the reconstructed shapes are properly retrieved and the average error of the sensor is less than 2.5 cm for obstacles placed in a range of 80 cm in any direction. The results obtained are encouraging and prove that this sensor can be used with depth perception industrial applications such as scene modelling, pipe inspection, robot navigation, etc. The sensor is capable of finding obstacles placed vertically, common in man made environments. Moreover, when previous information about the scene geometry is available the accuracy can be increased by using this information at the segmentation stage for removing potential noise sources.

Another application that involves only the omnidirectional camera was designed for the interaction with a human operator charged with surveillance tasks, see Appendix B. This application highlights the potential of the sensor to detect movement within a wide range and to provide images easy to understand by human operators. In addition, our sensor adds a new dimension to the scene observed since it can

produce light spots that serve as landmarks with known positions. Therefore, a precise location of the ground features can be performed visually.

Finally, we would like to highlight some of the advantages of the proposed sensor: it is not intrusive since there are no modifications of the scene; unlike traditional stereo, the camera synchronization is not an issue since only a single camera is used; the sensor is more robust and easier to calibrate since it is compact and does not have mobile elements and, last but not least, it is cheaper than other similar structures since we only use two mirrors, one camera and a simple laser projector.

The research work also produced a Matlab toolbox containing user-friendly, commented code for sensor calibration and implementations of the algorithms for the applications described.

6.2 Contributions and grants

The contributions of the thesis are included in this section for a clear understanding of the achievements of our work. A detailed discussion about the contributions is presented in section 6.2.1. The financial support offered by different institutions played an essential role in the advance of this research as mentioned in the second part of this section.

6.2.1 Contributions

The contributions that this research brings to VICOROB group and CREA in particular and to the computer vision community in general are listed in the following.

- An exhaustive survey of omnidirectional cameras focused on catadioptric sensors. The survey is condensed in a table that provides more clues about the studied devices such as their resolution, the SVP property, their isotropy and the sensors that can provide 3D measurements are highlighted. The table also contains the corresponding literature references for each category.

- Study of the design and the geometry of a sensor that combines the qualities of the omnidirectional vision and structured light. As a result of this study, the prototype of the sensor was built.

- Analysis of the projective geometry and the mathematics involved in the catadioptric systems. The formulation and implementation of the mathematical models of both components of the proposed sensor were accomplished. Based on the proposed models, the calibration of the sensor was realized.

- 3D scene modelling oriented to robot navigation. The scene is reconstructed using depth and texture information provided by the sensor located on a mobile platform. We emphasize that the successful tests of this prototype open-up large perspectives for industrial applications.

- Surveillance using the sensor placed at a fixed location. Since our sensor provides both texture and depth information its use in surveillance tasks obviously enhances the performances of these applications.

6.2.2 Grants and projects

The work developed in this thesis has been partially funded by the Spanish government through the following projects:

- The MCYT[1] project TAP[2] 1999-0443-C05-01 from 31/12/99 until 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 for providing the robots the ability of exploring an unknown environment and building a dynamic map. This project took part of a bigger project coordinated by the Polytechnic University of Valencia (UPV) and integrated both the Polytechnic University of Catalonia (UPC) and the University of Girona (UdG).

- The MCYT project TIC[3] 2003-08106-C02-02 from 01/12/03 until 30/11/06. The aim of the global project in which it is integrated is to design and develop FPGA based applications with fault tolerance applied to active vision-based surveillance tasks in large surfaces like airports and train stations. Some of the tasks involved are automatic detection of dangerous situations or suspicious behaviours, and people tracking. The project is developed in collaboration with the UPV.

[1] Ministerio de Ciencia y Tecnología.
[2] Tecnologías Avanzadas de Produccíon.
[3] Tecnologías de la Informacíon y de las Comunicaciones.

The financial support needed for completing the thesis has been received mainly from the University of Girona (UdG)[4] and the Government of Catalunya as follows:

2001 - 2005 A four years PhD scholarship given by the UdG.

2005 Research stage of 1 month at the Centre of Robotics, Electrotechnics and Automation (CREA) [5], part of the University of Picardie Jules Verne (UPJV) [6], Amiens, France. The mobility grant was provided by the UdG.

2004 Research stage of four months at the CREA. The stage was financed by the AIRE[7] mobility grant provided by the Government of Catalunya.

2003 Participation during 3 days at the workshop: "SANKEN International Workshop on Intelligent Systems", Osaka, Japan. The participation was financed by the University of Osaka and partially by the MCYT project TIC 2003-08106-C02-02.

2003 Participation during 1 week at the "Doctorial days 2003", Barcelona, Spain. Funded by the UdG.

2002 Four months research stay at the CREA. The stay was partially financed by the UdG.

6.3 Publications

The results obtained from the research work of this thesis were published in international conferences and workshops. Moreover, two articles have been submitted to the "Pattern Recognition Letters" and "Robotics and Autonomous Systems" journals respectively. The list of all the publications is presented in the following.

6.3.1 Journals

2005 R. Orghidan, C. Matabosch, , E.M. Mouaddib, J. Serrano, J. Salvi.
Catadioptric Single Shot Rangefinder for Textured Map Building in

[4]Universitat de Girona.
[5]Centre de Robotique, d'Électrotechnique et d'Automatique.
[6]Université de Picardie Jules Verne.
[7]Ajuts per a la mobilitat interregional d'investigadors.

Robot Navigation
Robotics and Autonomous Systems, an Elsevier Journal
Submitted on the 8th of November 2005.
This article was focused on the automate reconstruction of a 3D scene using our sensor. Data processing techniques were presented and supported by experimental results.

2005 R. Orghidan, J. Salvi, E.M. Mouaddib.
Modelling and Accuracy Estimation of a New Omnidirectional Depth Computation Sensor,
Pattern Recognition Letters, an Elsevier Journal
Accepted with major revision changes on the 18th of August 2005. Submitted for the second revision on the 30th of September 2005.
Thorough calibration results and theory were presented in this article for both components of the sensor. The sensor accuracy was estimated. The precision to calculate depth improved due to the technique and pattern used for calibration.

6.3.2 Conferences

2005 R. Orghidan, J. Salvi, E. Mouaddib.
Accuracy Estimation of a New Omnidirectional 3D Vision Sensor
IEEE International Conference on Image Processing
Genoa, Italy, September 11-14, Pages: 365 - 368, Volume 3
This article was focused on the accuracy estimation of the proposed sensor by means of laboratory experiments. The calibration of the sensor was also described briefly.

2005 R. Orghidan, E. Mouaddib, J. Salvi.
A Computer Vision Sensor for Panoramic Depth Perception
2nd Iberian Conference on Pattern Recognition and Image Analysis,
Estoril , Portugal, June 7-9, Pages: 153-160
The proposed sensor was presented in this article. The geometry and the calibration of the sensor were described and a set of experimental results were presented in order to illustrate the sensor functionality.
The article was also published in: Lecture Notes in Computer Science, Volume 3522 / 2005, p. 153, Springer-Verlag GmbH, ISBN: 3-540-26153-2

2005 R. Orghidan, E. Mouaddib, J. Salvi.
Omnidirectional Depth Computation from a Single Image
IEEE International Conference on Robotics and Automation,
Barcelona, Spain, April 18-22, Pages: 1234-1239
One of the first prototypes of the sensor was presented in this article. The article contains two methods for the camera calibration. The laser projector was modelled by an elliptic cone shape and depth measurements were performed using the calibrated sensor.

2005 R. Orghidan, E. Mouaddib, J. Salvi.
Un Système de vision 3D omnidirectionnelle
9èmes journées ORASIS,
Puy-de-Dôme, France, May 24-27
The goal of this article was to introduce our sensor to the French-speaking community. The article presents the sensor, the calibration and experimental results related to depth evaluation accuracy.

2003 R. Orghidan, J. Salvi, E. Mouaddib. **Calibration of A Structured Light-Based Stereo Catadioptric Sensor**
Omnivis workshop held in conjunction with the Computer Vision and Pattern Recognition conference,
Madison, Wisconsin, U.S.A.
The geometry of the sensor was studied in this article. Several configurations were compared and the most appropriate one was chosen. Finally, a simulation of the sensor was presented. Note that the laser projector was initially designed to use a hyperbolic mirror. Our experiments and the recent published theoretical results about the conic mirrors determined us to use the conical shape instead of the hyperbolic one for the laser projector.

2003 R. Orghidan, E. Mouaddib, J. Salvi. **An Omnidirectional Sensor with Embedded Structured Light Projector**,
SANKEN International Workshop on Intelligent Systems,
Osaka, Japan, December 16-18
The theory and simulation of the sensor were presented in the workshop organized by the University of Osaka. This was an excellent opportunity to discuss about the problems of the new sensor with Japanese researchers.

2001 L. Miclea, S. Enyedi, R. Orghidan.

On line BIST Experiments for Distributed Systems
IEEE European Test Workshop,
Stockholm, Sweden, May 29th - June 1st, Pages: 37-39
This paper presents a set of experiments aimed at overcoming the problems during the on-line testing of distributed systems. The experiments used different programming languages such as Java and Visual Basic and CORBA as data exchange technique.

6.4 Future work

Single camera catadioptric systems offers a wide field of view but have lower resolution compared to common cameras. Therefore, such omnidirectional devices can be used for gathering information from the surrounding scene but is not suitable for precise detailed imaging. Following this logic, we conclude that the proposed sensor is appropriate for applications that need a large field of view and fairly low accuracy such as human made environment modelling that is useful for robot navigation and remote inspections. Further work can be done in adapting the proposed sensor for applications such as pipe inspection that deal with a structured environment in which the objects are placed in the proximity of the sensor.

An inherent disadvantage of the proposed sensor is that it can recover only one line of 3D dots at one position. The first prototype can be improved by the use of a more complex structured light pattern. For instance, a set of concentric circles or colour coded light patterns can increase the vertical FOV for stereo perception as much as the overlapping between the conical mirror and the parabolic mirror fields of view. Thus, full cylindrical surfaces of 3D dots could be recovered. However, it is not always necessary or worthwhile to assume the costs of building the entire 3D cylinder. If only a small region is required for a particular application or the sensor moves in a well structured environment then less computing time is necessary to build that region. Therefore, the use of more complex light patterns can be interesting for applications that need the reconstruction of a 3D cylinder.

Another shortcoming of the proposed sensor is its range. The laser used for the prototype presented here is not visible at distances greater than 80 cm. A stronger laser would overcome this problem. The only concern about increasing the laser strength is the eye safety restrictions that might apply in the environment where

the sensor is used. Moreover, the sensor detection range can be increased by using optical filters adapted to the laser frequency and a higher resolution camera in order to optimize the laser detection. Thus, we identified here three ways to improve the performance of the sensor prototype.

Interesting information can be retrieved from the variation of the range scan and color at a given position. Note that the range scan measured by the sensor can be seen as a discrete function of the azimuth angle. Therefore, the sharp function variations correspond to swift changes in the range distance which correspond to discontinuities in the scanned surface. The study of the derivative of the range scan function can be used for finding the corners in the scene surrounding the sensor. Moreover, the color of the objects is also available and a sudden color change along the azimuth scan line can also indicate a change in the orientation of the shapes exposed to the observer.

The deformation of the laser stripe in the omnidirectional image is strongly connected with the shapes of the objects in the scene. We suggest to study the relation between the laser stripe curvature and the shapes of the objects. Since the image formation model is known a direct relation between the parameters of the observed laser stripe and the shape of the object from the scene could be established. Helpful hints can be extracted from the study of the laser curve parameters corresponding to the detection of the same surface placed at different distances and orientations.

The 3D scene reconstruction presented in section 5.4.2 maps the textures obtained from the best view positions on the corresponding objects. The proposed definition of the best view offers the best image quality but does not use all the texture information available. There are several views that contain the object to reconstruct but only one is selected. Therefore we suggest to work on the improvement of the texture by fusing several views of the same object. A good starting point is the work of Nagahara et al. [Nagahara *et al.*, 2001] that propose a superresolution modelling method. The proposed method is remarkable because unlike other similar methods it does not rotate the camera. In addition, it generates a uniform texture directly on the 3D objects.

Appendix A

Notation

A.1 Mathematics Convention

This appendix contains the nomenclature used to express mathematical concepts and to denote the parameters of the models defined in the thesis.

$\{H\}$ defines a coordinate system H, which is composed of an origin O_H and either two $\{X_H, Y_H\}$ or three $\{X_H, Y_H, Z_H\}$ axis, depending on the number of dimensions defined.

The following coordinate systems are used in the thesis:

- $\{W\} = \{O_W, X_W, Y_W, Z_W\}$ defines the world coordinate system.
- $\{C\} = \{O_C, X_C, Y_C, Z_C\}$ defines the camera coordinate system located at the focal point which is the mirror focus in the case of SVP catadioptric cameras.
- $\{R\} = \{O_R, X_R, Y_R\}$ defines the retinal coordinate system located at the image principal point $O_R(u_0, v_0)$.
- $\{I\} = \{O_I, X_I, Y_I\}$ defines the computer image coordinate system located in the upper-left corner of the image plane.

A point P is always represented with respect to a coordinate system. Therefore, ${}^H P$ denotes the point P related to the system of coordinates $\{H\}$. A point ${}^H P$ of coordinates $({}^H X_h, {}^H Y_h, {}^H Z_h)$ is written in a compact form as $P({}^H X_h, {}^H Y_h, {}^H Z_h)$.

In order to distinguish a single point from a set, i.e. the set of test points, a second sub-index is used. Then, P_{wi} indicates the i^{th} point on a set, where $i = 1...n$.

A rigid transformation between two coordinate systems $\{H\}$ and $\{J\}$ is expressed by a transformation matrix denoted ${}^{J}K_{H}$ that expresses the points from $\{H\}$ with respect to $\{J\}$. Moreover,

$$ {}^{J}K_{H} = \begin{pmatrix} {}^{J}R_{H} & {}^{J}T_{H} \\ 0_{1x3} & 1 \end{pmatrix} \tag{A.1} $$

where $R = (r_1, r_3, r_3)$ expresses the orientation of $\{H\}$ with respect to the axis of $\{J\}$. Moreover, R can also be given considering the RPY angles i.e. α, β and γ that represent the rotation angles for the X, Y and Z axis respectively. Therefore, the following notations are used:

- ${}^{C}R_{XW}$ defines the rotation from $\{W\}$ to $\{C\}$ on the X axis.
- ${}^{C}R_{YW}$ defines the rotation from $\{W\}$ to $\{C\}$ on the Y axis.
- ${}^{C}R_{ZW}$ defines the rotation from $\{W\}$ to $\{C\}$ on the Z axis.

The translation of the origin of $\{H\}$ with respect to the origin of $\{J\}$ is given by ${}^{J}T_{H} = (t_x, t_y, t_z)^T$. Hence,

- ${}^{C}T_{W}$ defines the translation of the origin of $\{W\}$ with respect to $\{C\}$.

Due to the rotational symmetry of the mirrors used for the catadioptric configurations considered in this thesis we can reduce the study of the optical behavior of such a surface from three-dimensional Cartesian coordinates system (X, Y, Z) to a two dimensional one (r, z), where $r = \sqrt{X^2 + Y^2}$. A point represented in a two dimensional space $\{r, z\}$ is denoted $P(r, z)$.

A vector is formally defined as an element of a vector space. In the commonly encountered vector space $\mathbb{R}^n$ (i.e., Euclidean n-space), a vector is given by n coordinates and can be specified as $[A_1, A_2...A_n]$.

We denote a vector from a point A to a point B as $\overrightarrow{AB}$ or shortly $\overrightarrow{v}$. Therefore we can write,

$$ \overrightarrow{v} = \overrightarrow{AB} \tag{A.2} $$

The normal vector, often simply called the "normal", to a surface or to a vector is a vector perpendicular to it. We denote the normal to a vector $\overrightarrow{v}$ as $\overrightarrow{v}_{\perp}$.

The transpose of a vector is the vector obtained by turning the rows into columns and vice versa. For instance, the transpose of the line vector is a column vector and we can write,

$$[A_1, A_2 ... A_n]^T = \begin{bmatrix} A_1 \\ A_2 \\ ... \\ A_n \end{bmatrix} \tag{A.3}$$

The following parameters of the catadioptric models are used:

- u_0, v_0 are the two components of the principal point, i.e. the projection of O_C on the image plane.
- k_u, k_v are the two components which permit to transform a point from metric coordinates to pixels.
- α_u, α_v are the intrinsic camera parameters that describe the scale and depend on k_u, k_v and on the focal length.
- ξ which depends on the eccentricity.
- φ which is a function of both the eccentricity and the scale.

Appendix B

Application of omnidirectional vision to active surveillance

Surveillance systems are usually based on remote imaging sensors. However, conventional cameras have severely limited FOV so the use of an omnidirectional vision sensor is an attractive alternative. This appendix presents theoretical and practical issues related to the use of omnidirectional vision in surveillance applications. The endeavor falls on the transformations that have to be applied to the omnidirectional images in order to ease their interpretation by humans. The experimental results show the behavior of the sensor in a real scene.

B.1 Adaptive background model for motion detection

In order to detect movement in the scene we apply a background substraction technique. The image captured by the fixed camera at a time instant t is compared with a reference image of the background, estimated at the same temporal coordinate, in other words, the intensity of each pixel is compared with the corresponding value in the reference background image. The difference between the current image and the background image highlights the moving objects.

Motion detection can be solved in the two dimensional image space by considering the transformation of the foreground F_{2d} relative to the background B_{2d}. Along the

time, the background image is essentially influenced by light variations, namely by gradual light changes (due to the day stages or to partial or total sun occlusions) or by sudden light changes (due to artificial lightning). Other background alterations can be produced by objects that are added or removed from the scene, permanently or temporarily. Therefore, the background evolves with the time and needs to be continuously estimated. The moving objects can be seen as layers that float over the background.

A stack of images $I_t^{u,v}$, with the size of $u \times v$ pixels, taken at different time instants $t = 1..n$ is used for estimating the background B_{2d}. Therefore, the intensity value of a pixel of coordinates (i, j) at time t can be expressed as $I_t(i, j)$. The first n images are used for the learning process. Then, the stack is refreshed following the rule *First Input First Output (FIFO)*: at time t, with $t > n$, the image I_t enters the stack while the oldest image I_{t-n} is removed. For a small time interval, the light intensity of a pixel $I(i, j)$ is expected to have a gaussian variation around a mean value σ_{ij} as shown in Eq. (B.1).

$$\begin{cases} \sigma_{ij} = \frac{\Sigma_{t=1}^{n} I_t(i,j)}{n} \\ B_{2d}(i,j) = \{\sigma_{ij}\}_{i=1..u, j=1..v} \end{cases} \tag{B.1}$$

However, a quick light change, such as a flicker or an artificial light turned on and off, may alter the effective mean value σ_{ij} by producing high variations of the intensity values $I_t(i, j)$ in a time frame . This problem is solved by taking as reference the median value and removing the values considered as outlayers. If the light change persists the median value will gradually change as the oldest values are disregarded due to the $FIFO$ rule.

Consequently, the intensity value of each image pixel in the background image $B_{2d}(i, j)_{i=1..u, j=1..v}$ is calculated at each time instant t from a stack of n images as being the mean value over the time where the outlayers have been removed.

The image of the background, shown in Figure B.1, was extracted using the first 20 images of the surveillance sequence. Note that the moving silhouette disappeared from the background image.

Figure B.2 shows some samples of images captured by the surveillance camera. The person that moves is visible in the omnidirectional images. The lower part of Figure B.2 shows that the silhouette was correctly extracted. The image plane that was divided radially in small cells which were connected for weighting the detected

Figure B.1: Surveillance background image extracted from 20 omnidirectional images.

entity. A threshold was set in order to remove the entities with small size considered as outliers. The process of detecting movement in the scene and the connection of the regions of the image that belong to the same moving entity are explained in the next section of this Appendix.

B.2 Movement detection and blob building

The movement in the foreground is detected by comparing the current image with the background image. The foreground includes the pixels whose difference of intensity with respect to the background is above a threshold T_α^{2d}, as shown in Equation (B.2), and thus are considered as parts of a moving entity. Following this rule, the current image (Figure B.3 (a)) is filtered and the resulting image F is split in sectors, see Figure B.3 (b). The blobs are obtained by connecting the regions that contain foreground pixels, Figure B.3 (c). Regions containing fake foreground pixels, basically due to noise, may interfere in the blob building process so a threshold T_r must be set in order to reject the regions that contain less than T_r pixels.

$$F(i,j) = \{|I_t(i,j) - B2(i,j)| > T_\alpha^{2d}\} \tag{B.2}$$

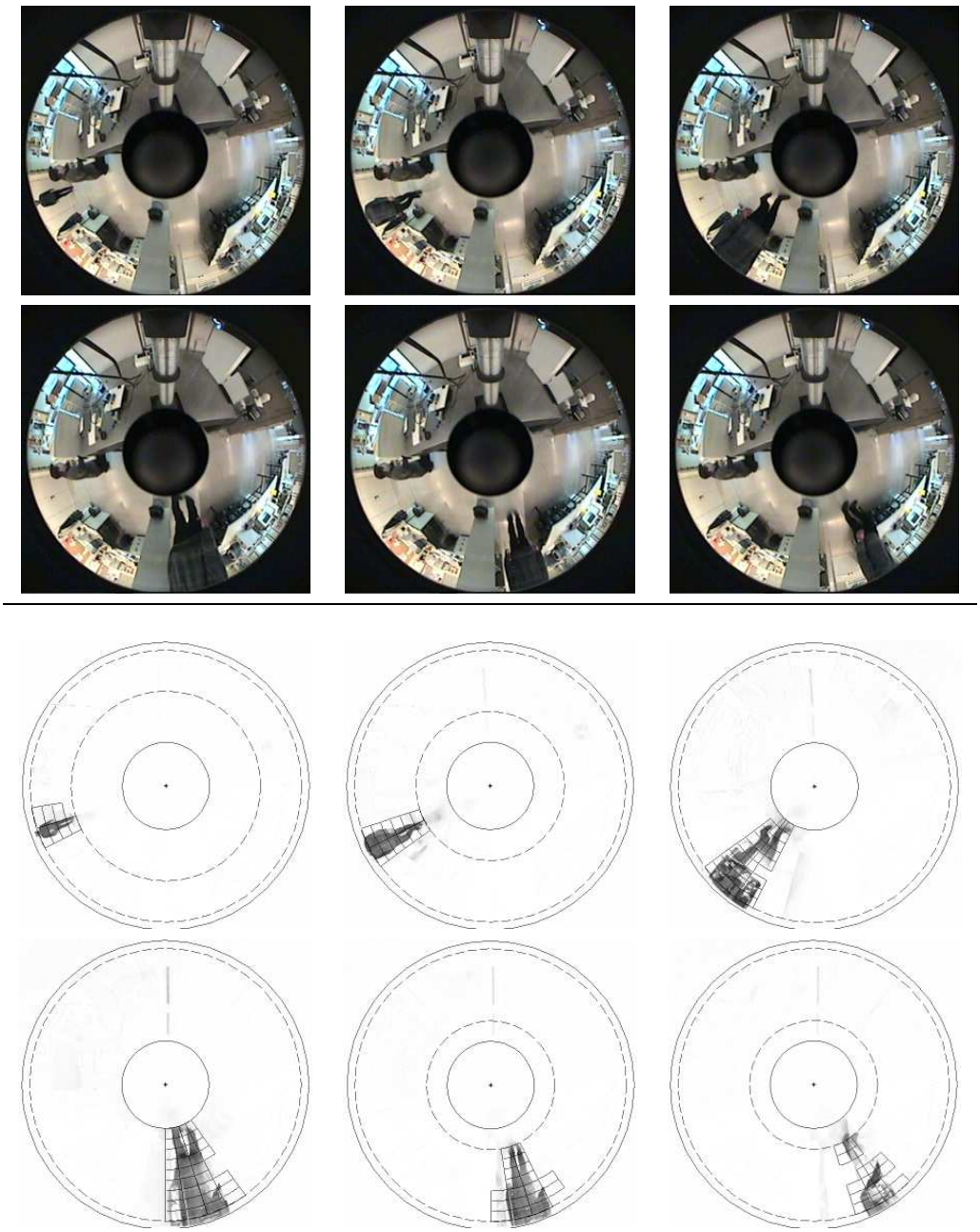
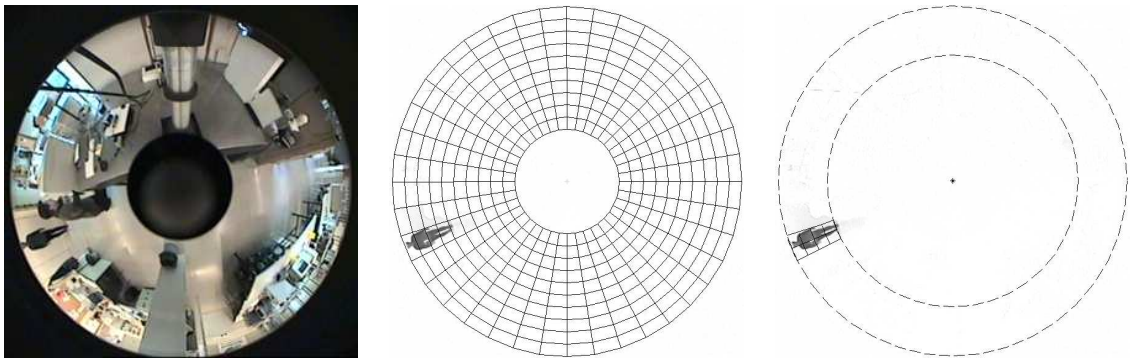


Figure B.2: (Upper part) A sequence of surveillance images captured by the surveillance camera. (Lower part) The detection of a moving silhouette in the observed scene.



(a) (b) (c)

Figure B.3: Blob building in an omnidirectional image. (a) One image of the sequence. (b) The sectors of the image. (c) The sectors where movement was detected.

The trajectory of the person moving in a scene can be calculated by estimating the gravity center of the moving blob. The trajectory of the person moving in the sequence of images presented in Figure B.2 was extracted and is shown in Figure B.4.

An interesting issue is the case when several entities appear in the foreground. In this case, the regions that belong to independent entities must not be connected. Figure B.5 (a) shows an image with several artificially introduced objects that are automatically detected and counted as different blobs as shown in Figure B.5 (b).

Finally, each blob in the image is unwrapped and a perspective image is generated for the user's convenience, as shown in Figure B.6.

In addition, since the camera is calibrated the bird eye view transform can be applied to the omnidirectional images. This transform gives a particularly intuitive view of the scene that enables the operator to visually locate the objects if the scene configuration is familiar. In the case presented in Figure B.7, the bird eye views give a visual clue about the approximative position of the person walking on the corridor since we know the relative distance between the optical table and the garbage can placed near the closet.

The application presented in this chapter uses the omnidirectional camera for surveillance tasks. The sensor is placed at a fixed location in a known scene and the movement in the scene is automatically detected. Two geometrical transformations are applied to the images provided by the omnidirectional camera in order to present understandable images for a human operator.

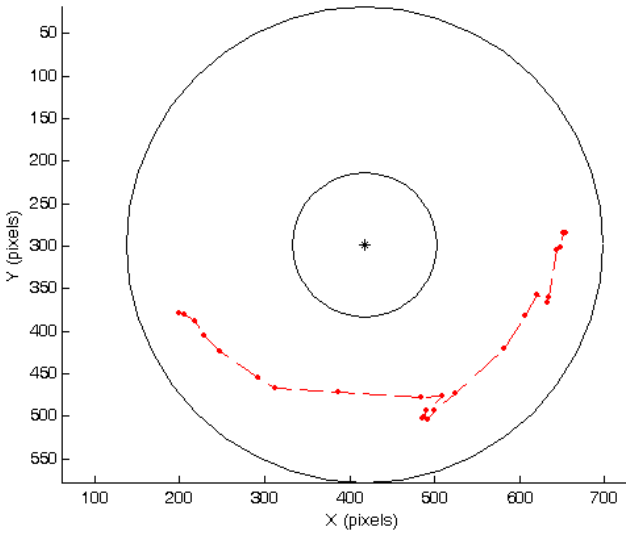
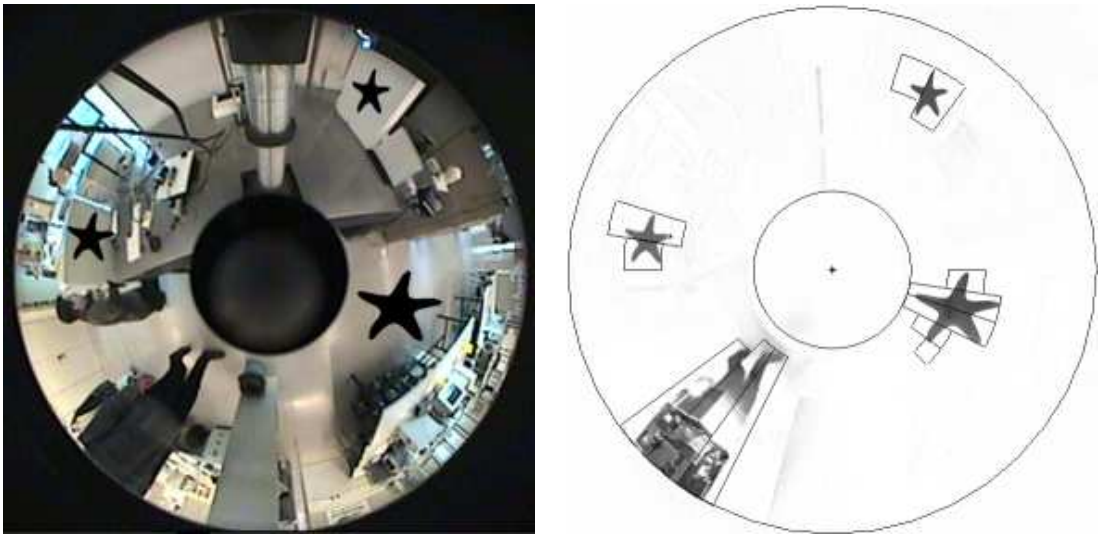


Figure B.4: Trajectory of a person moving in a scene surveyed by the sensor. The trajectory is represented by the dotted line and the positions at which the person was detected at each shot are marked with dots.



(a) (b)

Figure B.5: Multiple blob detection in an omnidirectional image. (a) Image with artificially introduced "objects". (b) The detected blobs formed by connected blocks.

Figure B.6: Perspective transform of the output images provided by the surveillance system. (Upper part) Detection of the slices containing moving entities. (Lower part) The slices are unwrapped and the perspective images are generated for a faster understanding of the scene.

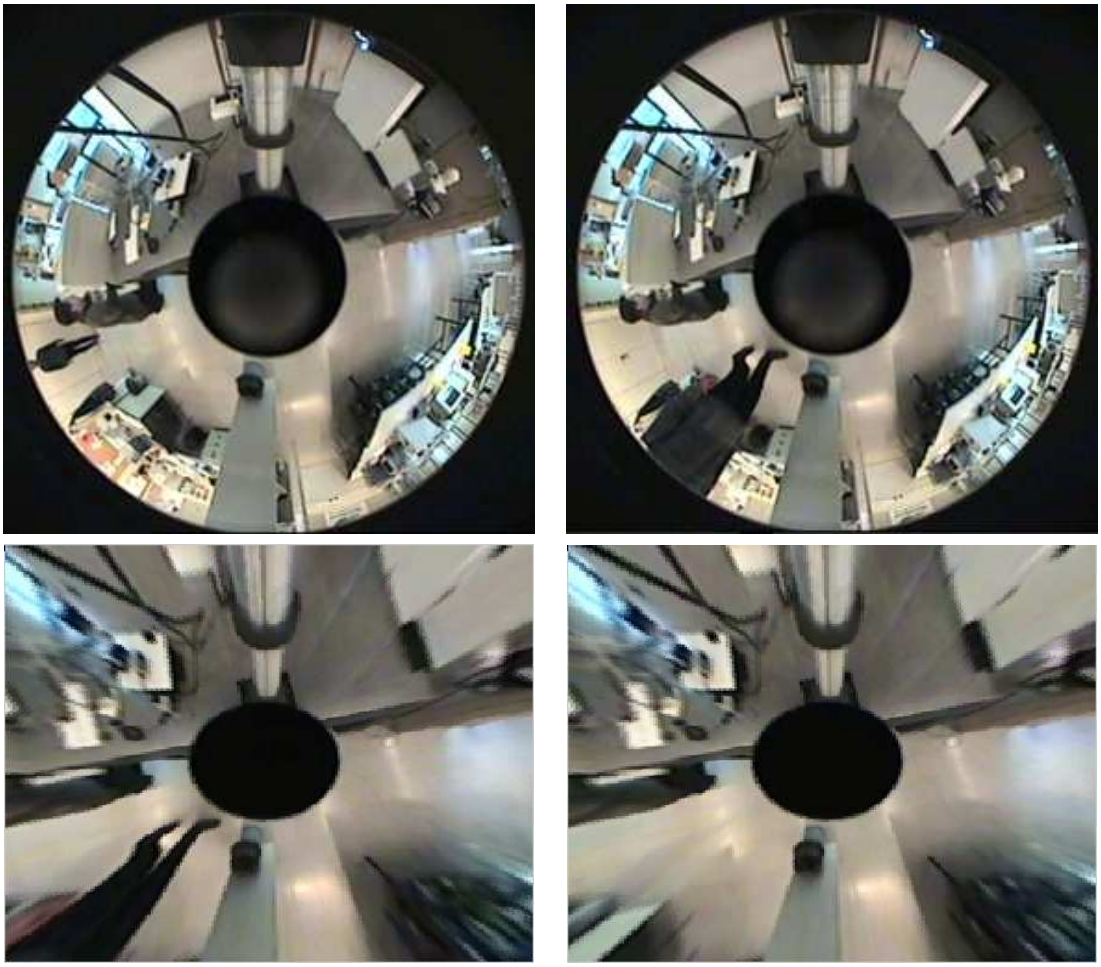


Figure B.7: Bird eye view transform of the output images provided by the surveillance system. (Upper part) Original omnidirectional images. (Lower part) Bird eye views extracted from the images.

Appendix C

Design of a hyperbolic mirror for a single view point configuration

During the study of the sensor prototype we first considered the use of a perspective camera in front of a hyperbolic mirror. Therefore, we studied the properties of this type of SVP catadioptrics. In this section we present some results obtained by this study such as the geometrical model of the configuration, the mirror design etc.

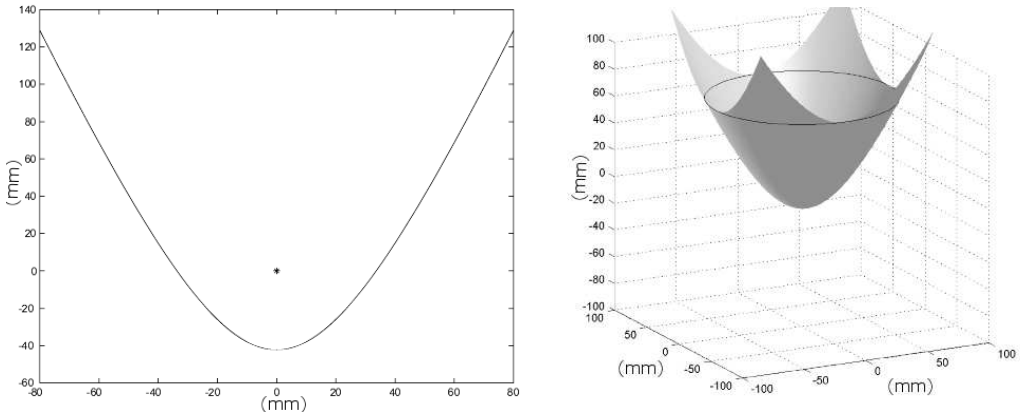
C.1 Hyperbolic mirror design

Let us consider the hyperbolic mirror shape given by Equation (C.1). An example of mirror profile obtained using this equation is shown in the Figure C.1

$$\frac{(z+\sqrt{a^2+b^2})^2}{a^2} - \frac{x^2+y^2}{b^2} = 1 \tag{C.1}$$

The hyperbola is a function of the two parameters a and b but it can also be described using the parameters c and k which determine the interfocus distance and the eccentricity, respectively. The relation between the pairs a, b and c, k is shown in Equation (C.2). Figure C.2 shows that the distance between the tips of the two hyperbolic napes is $2a$ while the distance between the two foci is $2c$.

$$\begin{aligned} a &= \tfrac{c}{2} \cdot \sqrt{\tfrac{k-2}{k}} \\ b &= \tfrac{c}{2} \cdot \sqrt{\tfrac{2}{k}} \end{aligned} \tag{C.2}$$



(a) (b)

Figure C.1: Hyperbolic mirror. (a) Hyperbolic profile with the parameters a=141 and b=57.3. The dot represents the focal point of the mirror, i.e., the SVP of the sensor. (b) The same hyperbolic mirror represented in 3D space.

By changing the values of these parameters the hyperbola changes its shape as well as the position of the focal point. In Figure C.3 are presented several hyperbolic shapes with the parameter a variable while in Figure C.4 varies b. The positions of the foci of the two napes of the hyperbola determine the size of the omnidirectional sensor.

The sensor is designed to be used in indoor environments and must be easy to handle so it has to be compact. Besides, image processing requires good resolution and a good vertical angle of view which means a larger mirror. A trade-off must be set up in order to obtain optimal performances.

It is obvious that the azimuth field of view is 360º since the mirror is a rotational surface upon the z axis. The vertical view angle is a function of the edge radius and the vertical distance between the focal point and the containing the rim of the mirror. This relation is expressed in Equation (C.3) where R_t is the radius of the mirror rim and α is the vertical view angle of the mirror.

$$\alpha = \arctan\left(\frac{h}{R_t}\right) + \frac{\pi}{2} \tag{C.3}$$

Therefore, R_t and h are the two parameters that bound the set of possible solutions.

The desired catadioptric sensor must possess a SVP, therefore, the pinhole of the

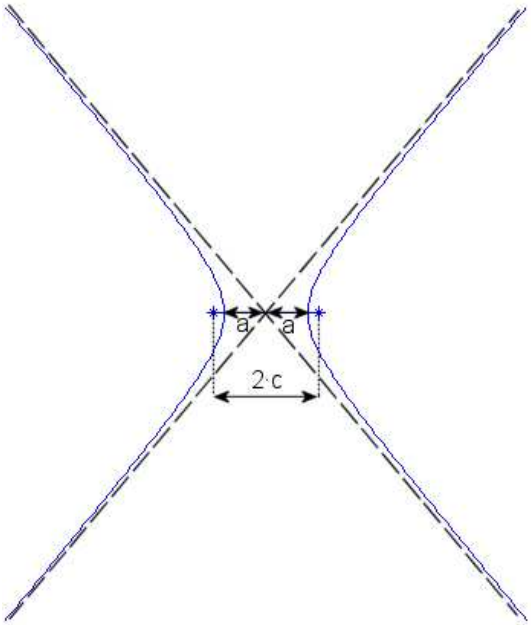
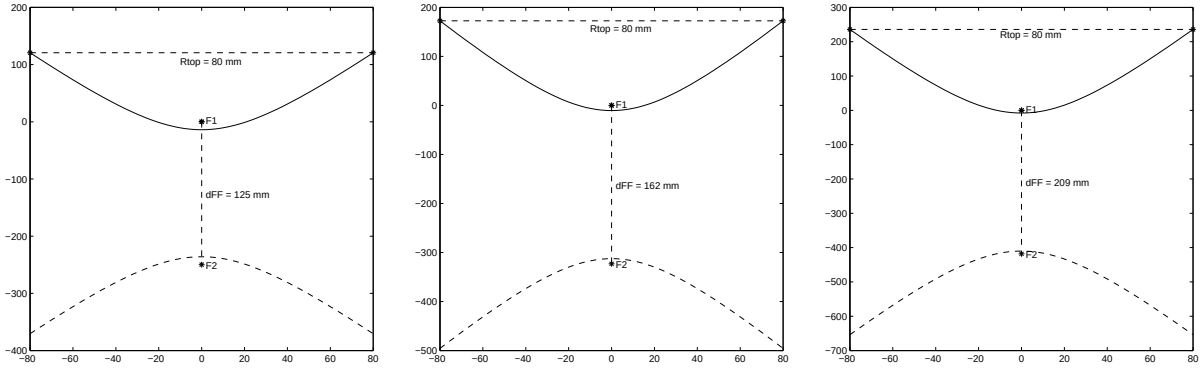
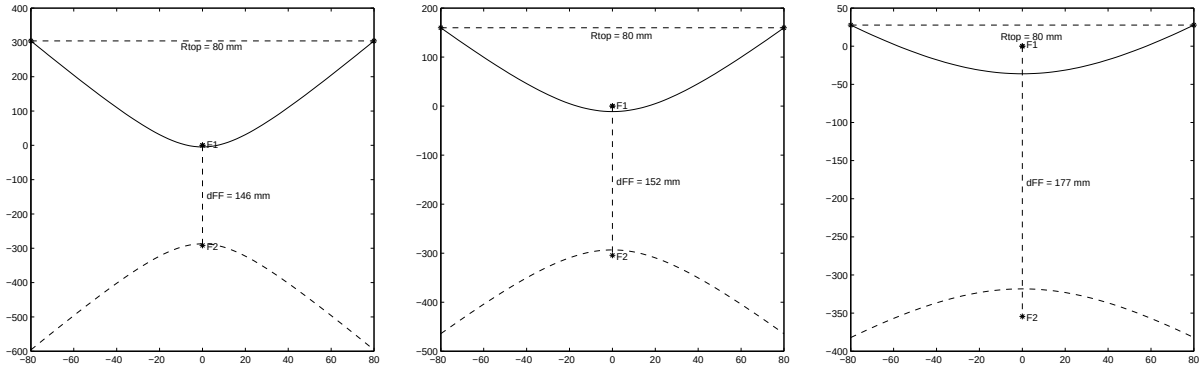


Figure C.2: The relation between the parameters a and c and the hyperbolic profile.



$a = 111, b = 57, k = 1.94$ $a = 151, b = 57, k = 2.64$ $a = 201, b = 57, k = 3.52$

Figure C.3: Hyperbolic profile change when varying the parameter **a** and keep b constant.



$a = 141, b = 37, k = 3.81$ $a = 141, b = 57, k = 2.47$ $a = 141, b = 107, k = 1.31$

Figure C.4: Hyperbolic profile change when keeping the parameter a constant and varying the parameter **b**.

camera model and the central projection point of the mirror have to be placed at the two foci of the hyperboloid. The relation between the profile of the mirror and the intrinsic parameters of the camera, namely the size of the CCD and the focal distance, is graphically represented in Figure C.5. Here, $P_{m1}(r_{rim}, z_{rim})$ is a point on the mirror rim, $P_{i1}(r_i, z_i)$ is the image of the point P_{m1} on the camera image plane, f is the focal distance of the camera and h is the vertical distance from the focal point of the mirror to its edge. Note that $z_{rim} = h$ and $z_i = -(c + f)$. Ideally, the mirror is imaged by the camera as a disc with circular rim tangent to the borders of the image plane.

Several constraints must be fulfilled during the design process of the hyperbolic mirror shape:

- The mirror rim must have the right shape so that the camera is able to see the point P_{m1}. In other words, the hyperbola should not be cut below the point P_{m1} which is the point that reflects the higher part of the desired field of view.
- The points of the camera mirror rim must be on the line $\overline{P_{i1}, P_{m1}}$ for an optimally sized image in the camera.
- From several suitable hyperbolic shapes for the camera mirror the one with the bigger "blind area" is chosen since it will provide a larger cone for placing the structured light projector. The "blind area" is the part of the mirror that back-reflects the image of the camera. This region of the image produced by the omnivision sensor is traditionally considered as being lost since it does not

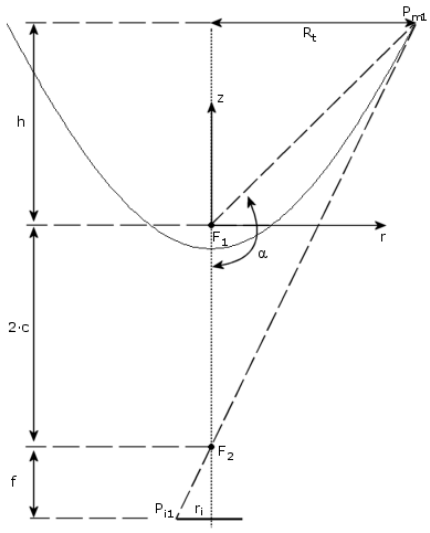


Figure C.5: The relation between the size of the sensor and the intrinsic parameters of the omnidirectional camera

provide any kind of information but the reflection of the camera itself.

A study about the influence of the parameters a and b upon the mirrors' profile was also realized by T. Svoboda et al. in [Svoboda *et al.*, 1997]. Svoboda underlined the influence that of the ratio $k = \frac{a}{b}$ upon the image formation when using a hyperbolic mirror:

- $k > \frac{b}{R_t}$ is the condition that the catadioptric configuration must satisfy in order to have a field of view higher than the horizon i.e. greater than the hemisphere.
- $k < \frac{h+2c}{R_t}$ is the condition for obtaining a realizable hyperbolic mirror. This requirement implies finding the right solution of the hyperbola equation.
- $k > \frac{h+2c}{4b} - \frac{b}{h+2c}$ prevents focusing problems by placing the mirror top far enough from the camera.

The first configuration of the proposed sensor [Orghidan *et al.*, 2003] used two hyperbolic mirrors with the laser projector embedded between the camera and the corresponding mirror, as shown in Figure C.6. The angle of view α of the sensor is bounded by the rims of the two mirrors. Figure C.6 also shows the focal points of the hyperbolic shape pointed by the camera: FC and F_{camera} which are the focal

points of the mirror and of the camera model respectively. The angle α and the range of view Rov determine the position of the highest PM and lowest Pm points of scene imaged by the sensor. These points are reflected on the laser mirror on points $P1L$ and $P2L$, respectively. The focal points of the laser mirror FL and the camera mirror FC are aligned along the optical axis and translated by a distance dFF.

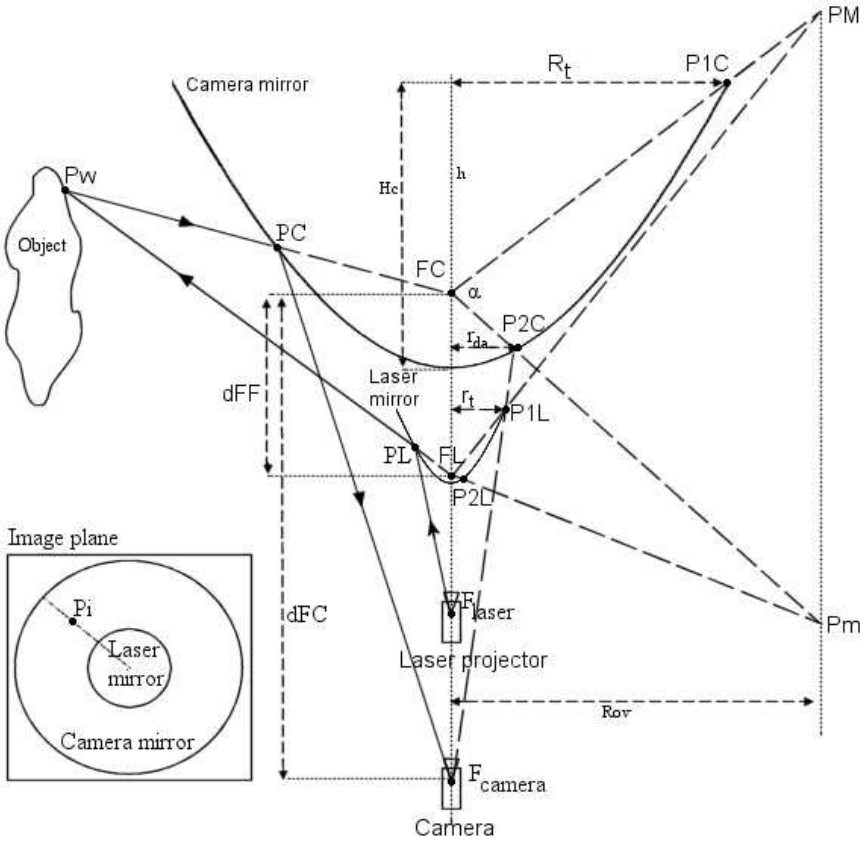


Figure C.6: The first sensor configuration using two hyperbolic mirrors.

An algorithm was developed in order to produce hyperbolic shapes according to the application requirements and taking into account the above considerations related to the mirror parameters. For instance, a surveillance application that uses a catadioptric camera fixed on the ceiling needs a good resolution on the part reflecting the ground rather than a high FOV. Therefore, a suitable mirror is the one presented in Figure C.7 (a) which has a FOV above the horizon of 6.2 degrees.

Besides, video-conference or pipe inspection applications need a high vertical FOV hence a sharper mirror must be built. Such a mirror is presented in Figure C.7 (b) and provides a vertical FOV above the horizon of 49.8 degrees.

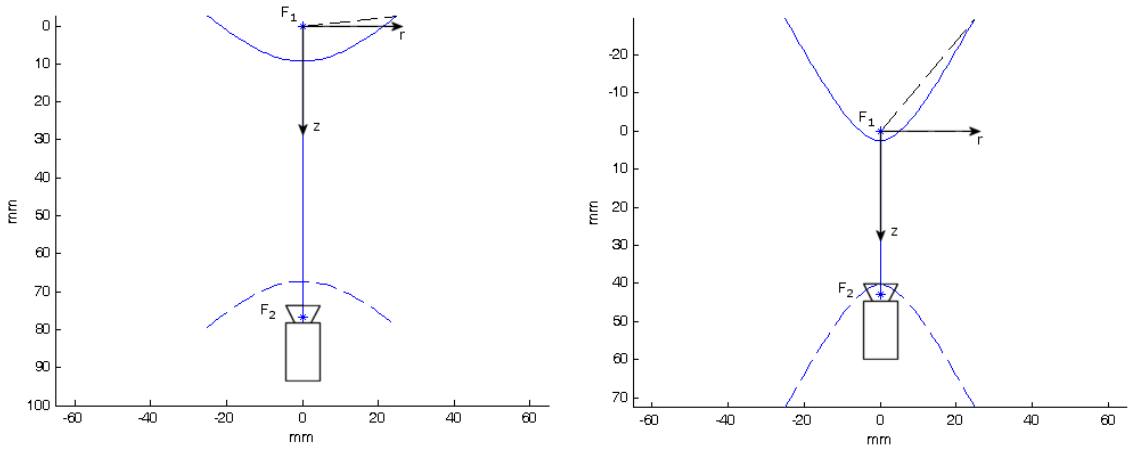


Figure C.7: Hyperbolic profiles designed according to the application requirements. (a) Low vertical FOV hyperbolic mirror suitable for surveillance applications. The parameters of the mirror are $a = 29$, $b = 25$ and $R_t = 25$. The vertical FOV above the horizon is of 6.2 degrees. (b) High vertical FOV hyperbolic mirror suitable for pipe inspection applications. The parameters of the mirror are $a = 19$, $b = 10$ and $R_t = 25$. The vertical FOV above the horizon is of 49.8 degrees.

However, due to the particular configuration of our sensor we decided to replace the laser mirror by a conic one and to place the laser projector under the catadioptric camera. Moreover, due to technical difficulties in manufacturing the mirror we bought an off the shelf catadioptric camera using folded mirrors that approximate a parabolic projection.

Bibliography

Ahn, S.J. (2004). Least-squares orthogonal distance fitting. *Lecture Notes in Computer Science*.

Aliaga, D.G. (2001). Accurate catadioptric calibration for real-time pose estimation of room-size environments. In: *IEEE International Conference on Computer Vision*. Vol. 1. pp. 127–134.

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

Baker, S. and S.K. Nayar (1998). A theory of catadioptric image formation. In: *IEEE Sixth International Conference on Computer Vision*. pp. 35–42.

Bakstein, H. and T. Pajdla (2003). Rendering novel views from a set of omnidirectional mosaic images. In: *Workshop on Omnidirectional Vision and Camera Networks*. p. CD ROM.

Barreto, J.P. (2003). General Central Projection Systems: Modeling, Calibration and Visual Servoing. PhD thesis. Univ. of Coimbra.

Barreto, J.P. and H. Araujo (2005). Geometric properties of central catadioptric line images and their application in calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **27**(8), 1327 – 1333.

Basu, A. and S. Licardie (1993). Multi-camera networks: Eyes from eyes. In: *IEEE/RSJ, Intelligent Robots and Systems*. Vol. 3. pp. 1822 –1828.

Benosman, R. and J. Devars (1998). Panoramic stereo vision sensor. In: *The Fourteenth International Conference on Pattern Recognition*. Vol. 1. pp. 767 – 769.

Benosman, R. and S.B. Kang (2001). *Panoramic Vision: Sensors, Theory, and Applications*. Springer-Verlag.

Benosman, R., T. Maniere and J. Devars (1996). Multidirectional stereovision sensor, calibration and scenes reconstruction. In: *The 13th International Conference on Pattern Recognition*. Vol. 1. pp. 161 – 165.

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

Brauer-Burchardt, C. and K. Voss (2001). A new algorithm to correct fish-eye and strong wide-angle-lens-distortion from single images. In: *International Conference on Image Processing*. Vol. 1. pp. 225 –228.

Bruckstein, A.M. and T.J. Richardson (2000). Omniview cameras with curved surface mirrors. In: *First Workshop of Omnidirectional Vision*. pp. 79 –84.

Cauchois, Cyril, Eric Brassart, Cyril Drocourt and Pascal Vasseur (1999). Calibration of the omnidirectionnal vision sensor : Syclop. In: *IEEE International Conference on Robotics and Automation*. Vol. 2. pp. 1287 – 1292.

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

Cutler, R., Y. Rui, A. Gupta, JJ Cadiz, I. Tashev, L. He, A. Colburn, Z. Zhang, Z. Liu and S. Silverberg (2002). Distributed meetings: A meeting capture and broadcasting system. In: *The tenth ACM international conference on Multimedia*. pp. 503 – 512.

Doubek, P. and T. Svoboda (2002). Reliable 3d reconstruction from a few catadioptric images. In: *Proceedings. of Omnidirectional Vision*. pp. 71–78.

Drocourt, C., L. Delahoche, B. Marhic and A. Clerentin (2002). Simultaneous localization and map construction method using omnidirectional stereoscopic information. In: *IEEE International Conference on Robotics and Automation*. Vol. 1. pp. 894 – 899.

Fermuller, C., Y. Aloimonos, P. Baker, R. Pless, J. Neumann and B. Stuart (2000). Multi-camera networks: Eyes from eyes. In: *Third Workshop on Omnidirectional Vision*. pp. 11–18.

Fiala, M. and A. Basu (2002*a*). Line segment extraction in panoramic images. *Journal of the Winter School of Computer Graphics* pp. 179–186.

Fiala, M. and A. Basu (2002*b*). Panoramic stereo reconstruction using non-svp optics. In: *16th International Conference on Pattern Recognition*. Vol. 4. pp. 27 – 30.

Fitzgibbon, Andrew W., Maurizio Pilu and Robert B. Fisher (1999). Direct least-squares fitting of ellipses,. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **21(5)**, 476–480.

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. pp. 646 – 649.

Franceschini, N., J.-M. Pichon and C. Blanes (1991). Real time visuomotor control: from flies to robots. In: *Fifth International Conference on Advanced Robotics, 'Robots in Unstructured Environments'*. Vol. 2. pp. 931 –935.

Gaspar, J., C. Decco, J. Jr. Okamoto and J. Santos-Victor (2002). Constant resolution omnidirectional cameras. In: *Third Workshop of Omnidirectional Vision*. pp. 27–34.

Gaspar, J., N. Winters and J. Santos-Victor (2000). Vision-based navigation and environmental representations with an omnidirectional camera. *IEEE Transactions on Robotics and Automation* **16**(6), 890 – 898.

Gaspar, Jose Antonio (2002). Omnidirectional Vision for Mobile Robot Navigation. PhD thesis. Technical Superior Institute, Technical University of Lisbon.

Geyer, C. and K. Daniilidis (1999). Catadioptric camera calibration. In: *IEEE International Conference on Computer Vision*. pp. 398–404.

Geyer, C. and K. Daniilidis (2000*a*). Equivalence of catadioptric projections and mappings of the sphere. In: *First Workshop of Omnidirectional Vision*. pp. 91 – 96.

Geyer, C. and K. Daniilidis (2000*b*). A unifying theory for central panoramic systems and practical applications. In: *IEEE Sixth European Conference on Computer Vision*. pp. 445–461.

Geyer, C. and K. Daniilidis (2001). Catadioptric projective geometry. *International Journal of Computer Vision* **45**(3), 223–243.

Geyer, C. and K. Daniilidis (2002). Paracatadioptric camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **24**(5), 687–695.

Gluckman, J. and S.K. Nayar (1998). Ego-motion and omnidirectional cameras. In: *IEEE International Conference on Computer Vision*. pp. 999–1005.

Gluckman, J. and S.K. Nayar (1999). Planar catadioptric stereo: Geometry and calibration. In: *IEEE Computer Vision and Pattern Recognition*. Vol. 1. pp. 22–28.

Gluckman, J. and S.K. Nayar (2002). Rectified catadioptric stereo sensors. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **24**(2), 224–236.

Gluckman, J., S. Nayar and K. Thorek (1998). Real-time omnidirectional and panoramic stereo. In: *DARPA Image Understanding Workshop*. pp. 299–303.

Greguss, P. (1984). *Panoramic imaging block for three dimensional space.*. United States Patent. No 4,566,763.

Grossberg, M.D. and S.K. Nayar (2001). A general imaging model and a method for finding its parameters. In: *IEEE International Conference on Computer Vision*. Vol. II. pp. 108–115.

Hicks, R. A. and K. R. Perline (2002). Equi-areal catadioptric sensors. In: *Workshop on Omnivision*. pp. 13–18.

Hicks, R.A. and R. Bajcsy (2000). Catadioptric sensors that approximate wide-angle perspective projections. In: *IEEE International Conference on Computer Vision and Pattern Recognition*. pp. 545–551.

Hicks, R.A. and R. Bajcsy (2001). Reflective surfaces as computational sensors. *IVC* **19**(11), 773–777.

Hicks, R.A. and R.K. Perline (2001). Geometric distributions for catadioptric sensor design. In: *IEEE Computer Society's Computer Vision and Pattern Recognition*. Vol. 1. pp. 584–589.

Hong, J., X. Tan, B. Pinette, R. Weiss and E.M. Riseman (1991). Image-based homing. In: *IEEE International Conference on Robotics and Automation*. pp. 620 –625, vol.1.

Hua, H. and N. Ahuja (2001). A high-resolution panoramic camera. In: *IEEE International Conference on Computer Vision and Pattern Recognition*. Vol. I. pp. 960–967.

Kang, S.B. (2000). Catadioptric self-calibration. In: *IEEE International Conference on Computer Vision and Pattern Recognition*. Vol. I. pp. 201–207.

Kar-Han, T., H. Hong Hua and N. Ahuja (2004). *IEEE Transactions on Pattern Analysis and Machine Intelligence* **26**(7), 941 – 946.

Kawanishi, T., K. Yamazawa, H. Iwasa, H. Takemura and N. Yokoya (1998). Generation of high-resolution stereo panoramic images by omnidirectional imaging sensor using hexagonal pyramidal mirrors. In: *Fourteenth International Conference on Pattern Recognition*. Vol. 1. pp. 485 – 489.

Kurita, T., H. Shimai, Y. Baba, T. Mishima, M. Tanaka, S. Akaho and S. Umeyama (2000). Gaze control on virtual active vision system with binocular fish-eye lenses. In: *IEEE International Conference on Systems, Man, and Cybernetics*. Vol. 3. pp. 1644 –1649.

Lin, S. S. and R. Bajcsy (2001*a*). The true single view point (svp) configuration for omni-directional view catadioptric system using cone mirror.. Technical report ms-cis-00-24. Computer and Information Science Department, University of Pennsylvania.. Philadelphia, USA.

Lin, S. S. and R. Bajcsy (2003). High resolution catadioptric omni-directional stereo sensor for robot vision. In: *IEEE International Conference on Robotics and Automation*. pp. 1694–1699.

Lin, S.S. and R. Bajcsy (2001*b*). True single view point cone mirror omni-directional catadioptric system. In: *IEEE International Conference on Computer Vision and Pattern Recognition*. Vol. II. pp. 102–107.

Matabosch, C., J. Salvi, D. Fofi and F. Meriaudeau (2005). Range image registration for industrial inspection. In: *Machine Vision Applications in Industrial Inspection XIII*. Vol. 5679. pp. 216–227.

Mouaddib, E. and B. Marhic (2000). Geometrical matching for mobile robot localisation. In: *IEEE Transactions On Robotics and Automation*. Vol. 16. pp. 542–552.

Mouaddib, E., R. Sagawa, T. Echigo and Y. Yagi (2005). Stereo vision with a single camera and multiple mirrors. In: *IEEE International Conference on Robotics and Automation*. pp. 812–817.

Nagahara, H., Y. Yagi and M. Yachida (2001). Resolution improving method from multi-focal omnidirectional images. In: *IEEE International Conference on Image Processing*. Vol. 1. pp. 654 – 657.

Nagahara, H., Y. Yagi and M. Yachida (2003). Wide field of view catadioptrical head-mounted display. In: *International Conference on Intelligent Robots and Systems*. pp. 3738–3743.

Nalwa, V. (1996). A true omnidirectional viewer. Technical report. Bell Laboratories.

Nayar, S. K. and V. Peri (1999). Folded catadioptric cameras. In: *IEEE International Conference on Computer Vision and Pattern Recognition*. pp. 217–223.

Nayar, S.K. (1997). Catadioptric omnidirectional camera. In: *IEEE International Conference on Computer Vision and Pattern Recognition*. pp. 482–488.

Nene, S.A. and S.K. Nayar (1998). Stereo with mirrors. In: *Sixth International Conference on Computer Vision*. pp. 1087–1094.

Neumann, J., C. Fermuller and Y. Aloimonos (2002). Eyes from eyes: new cameras for structure from motion. In: *Third Workshop on Omnidirectional Vision*. pp. 19 –26.

Ollis, M., H. Herman and S. Singh (1999). Analysis and design of panoramic stereo vision using equi-angular pixel cameras. Technical Report CMU-RI-TR-99-04. Robotics Institute, Carnegie Mellon University.

Onoe, Y., N. Yokoya, K. Yamazawa and H. Takemura (1998). Visual surveillance and monitoring system using an omnidirectional video camera. In: *Fourteenth International Conference on Pattern Recognition*. Vol. 1. pp. 588 –592.

Orghidan, R., E. Mouaddib and J. Salvi (2005*a*). A computer vision sensor for panoramic depth perception. In: *2nd Iberian Conference on Pattern Recognition and Image Analysis*. pp. 153–160.

Orghidan, R., E. Mouaddib and J. Salvi (2005*b*). Omnidirectional depth computation from a single image. In: *IEEE International Conference on Robotics and Automation*. pp. 1234–1239.

Orghidan, R., E. Mouaddib and J. Salvi (2005*c*). Un système de vision 3d omnidirectionnelle. In: *9èmes journées ORASIS*.

Orghidan, R., J. Salvi and E. Mouaddib (2003). Calibration of a structured light-based stereo catadioptric sensor. In: *Workshop on Omnidirectional Vision, IEEE Conf. on Computer Vision and Pattern Recognition*.

Orghidan, R., J. Salvi and E. Mouaddib (2005*d*). Accuracy estimation of a new omnidirectional 3d vision sensor. In: *IEEE International Conference on Image Processing*. Vol. 3. pp. 365 – 368.

Pagès, J. and J. Salvi (2004). A new optimised de bruijn coding strategy for structured light patterns. In: *17th International Conference on Pattern Recognition*. Vol. 4. pp. 284 – 287.

Peer, P. and F. Solina (2002). Panoramic depth imaging: Single standard camera approach. *International Journal of Computer Vision* **47**(1-3), 149–160.

Pegard, C. and E.M. Mouaddib (1996). A mobile robot using a panoramic view. In: *IEEE International Conference on Robotics and Automation*. Vol. 1. pp. 89–94.

Rees, D. (1970). Panoramic television viewing system. United States Patent (3,505,465).

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

Salvi, J., X. Armangue and J. Batlle (2002). A comparative review of camera calibrating methods with accuracy evaluation. *Pattern Recognition* **35**(7), 1617–1635.

Sandini, G., P. Questa, D. Scheffer and A. Mannucci (2000). A retina-like cmos sensor and its applications. In: *IEEE Sensor Array and Multichannel Signal Processing Workshop*.

Serrat, J., X. Varona, A. Lopez, X. Roca and J. Villanueva (2001). P3 : a three-dimensional digitizer prototype. In: *IX National Symposium on Pattern Recognition and Image Analysis (SNRFAI'01)*.

Shah, S. and J.K. Aggarwal (1994*a*). Depth estimation using stereo fish-eye lenses. In: *Proceedings of the IEEE International Conference on Image Processing*. Vol. 2. pp. 740 –744.

Shah, S. and J.K. Aggarwal (1994*b*). A simple calibration procedure for fish-eye (high distortion) lens camera. In: *IEEE International Conference on Robotics and Automation*. Vol. 4. pp. 3422 –3427.

Shih-Chieh, W., Y. Yagi and M. Yachida (1998). Building local floor map by use of ultrasonic and omni-directional vision sensor. In: *IEEE International Conference on Robotics and Automation*. pp. 2548 –2553 vol.3.

Southwell, D., A. Basu, M. Fiala and J. Reyda (1996*a*). Panoramic stereo. In: *13th International Conference on Pattern Recognition*. Vol. 1. pp. 378 – 382.

Southwell, D., B. Vandegriend and A. Basu (1996*b*). A conical mirror pipeline inspection system. In: *IEEE International Conference on Robotics and Automation*. Vol. 4. pp. 3253 – 3258.

Spacek, Libor (2005). A catadioptric sensor with multiple viewpoints. *Robotics and Autonomous Systems* **51**, 3–15.

Strelow, D., J. Mishler, D. Koes and S. Singh (2001). Precise omnidirectional camera calibration. In: *IEEE International Conference on Computer Vision and Pattern Recognition*. Vol. I. pp. 689–694.

Sturm, P. (2000). A method for 3d reconstruction of piecewise planar objects from single panoramic images. In: *IEEE Workshop on Omnidirectional Vision*. pp. 119 –126.

Svoboda, T. and T. Pajdla (2000). Panoramic cameras for 3d computation. In: *The Czeck Pattern Recognition Workshop*. pp. 63–70.

Svoboda, T., T. Pajdla and V. Hlavac (1997). Central panoramic cameras. Research report k335/97/147.

Svoboda, T., T. Pajdla and V. Hlavac (1998). Epipolar geometry for panoramic cameras. In: *Fifth European Conference on Computer Vision*. pp. 218–232.

Swaminathan, R., M.D. Grossberg and S.K. Nayar (2001). Caustics of catadioptric cameras. In: *IEEE International Conference on Computer Vision*. Vol. 2. pp. 2–9.

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**, 889–898.

Werner, T. and T. Pajdla (2001). Cheirality in epipolar geometry. In: *IEEE International Conference on Computer Vision*. Vol. I. pp. 548–553.

Winters, N., J. Gaspar, G. Lacey and J. Santos-Victor (2000). Omni-directional vision for robot navigation. In: *IEEE Workshop on Omnidirectional Vision*. pp. 21–28.

Xiong, Y. and K. Turkowski (1997). Creating image based vr using a self-calibrating fisheye lens. In: *IEEE Computer Vision and Pattern Recognition*. pp. 237–243.

Yachida, M. (1998). Omnidirectional sensing and combined multiple sensing. In: *IEEE Computer Vision for Virtual Reality Based Human Communications*. pp. 20 – 27.

Yagi, Y. (1999). Omnidirectional sensing and its aplications. In: *IEICE Transactions on Information and Systems*. Vol. E82-D. pp. 568–578.

Yagi, Y. and M. Yachida (2002). Omnidirectional sensing for human interaction. In: *Third Workshop on Omnidirectional Vision*. pp. 121–127.

Yagi, Y. and S. Kawato (1990). Panoramic scene analysis with conic projection. In: *IEEE/RSJ International Conference on Intelligent Robots and Systems*. Vol. 1. pp. 181 – 187.

Yamazawa, K., Y. Yagi and M. Yachida (1993). Omindirectional imaging with hyperboliodal projection. In: *IEEE/RSJ International Conference on Intelligent Robots and Systems*. pp. 1029–1034.

Yamazawa, K., Y. Yagi and M. Yachida (1995). Obstacle detection with omnidirectional image sensor hyperomni vision. In: *IEEE International Conference on Robotics and Automation*. pp. 1062–1067, vol.1.

Zhang, Z., R. Weiss and E.M. Riseman (1991). Feature matching in 360° waveforms for robot navigation. In: *IEEE Computer Vision and Pattern Recognition*. pp. 742 –743.

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