Mapping and Localization from a Panoramic Vision Sensor

Bunschoten, R.

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
Chapter 1

Introduction

1.1 Mobile Service Robots

Mobile robots have a wide applicability and they are gradually making their way into our daily lives. Autonomous floor cleaning robots are already employed in shopping centers, airports, factories and other large buildings to perform a tedious cleaning task at nighttime. The “ST82 R” robot, produced by Hefer Cleantech, currently cleans floors in five Dutch supermarkets. Mobile robots perform security and surveillance tasks such as fire detection and intruder detection. The operational “MOSRO 1” surveillance robot, developed by Robowatch Technologies, uses a radar system that can identify contours of intruders even through walls. Mobile delivery robots, developed by Helpmate Robotics Inc., distribute medicines in over 100 American hospitals. Experience with this “Helpmate” robot shows that the robot performs the distribution tasks faster and more reliably than humans do. The above examples illustrate that mobile service robots can save both time and money, can perform tasks that humans consider dangerous or tedious, and may (unintentionally) entertain us in the process.

In order for a mobile service robot to perform its navigational tasks, it should be able to determine where it is. This localization problem, determining the pose (position and orientation) with respect to a global map of the workspace, has occasionally been referred to as “the most fundamental problem to providing a mobile robot with autonomous capabilities” [9].

Today’s commercial service robots often rely on an a priori provided map of their workspace. Such a map contains the locations of easily recognizable artificial landmarks, which have been mounted at strategically chosen locations. Localization is such a scenario may be considered a solved problem. Ideally, a mobile robot should be capable of learning and maintaining a map of its workspace while it performs its tasks. It should be able to do so without relying on the presence of artificial landmarks. The problem of mapping, creating a map for localization from sensor data collected during navigation in an initially unknown environment, is currently an active research topic.
1.2 Vision for Mobile Robots

Humans have evolved to rely primarily on vision for localization and navigation tasks (and many other tasks as well). Although we certainly use our other senses as well, these provide less informative clues as to where we are. Similarly, in robotics certain external sensors — sensors that measure aspects of the external world — may yield more informative clues than others may.

Early mobile robots were usually equipped with a ring of ultrasonic range sensors from which rough estimates of distances to surrounding objects can be derived. Nowadays, robots can be equipped with laser range finders providing precise distance information. Laser range finders can be used to construct accurate 2-D (or even 3-D) geometric maps of a robot’s workspace. Still, range information, however accurate it may be, may not be very informative for localization because many distinct places in a robot’s workspace can have a similar geometric structure. Moreover, today’s laser range finders are very expensive.

Vision sensors (i.e. cameras), on the other hand are cheap nowadays. A camera image provides abundant information about a scene observed and may thus yield stronger clues for localization. Employing a mobile robot with a camera furthermore opens up the possibility to perform a wide range of visual tasks, such as recognizing objects to grasp and recognizing people to interact with.

In spite of these advantages, range sensors are still very popular and commonly employed in studies concerning mapping and localization. This is likely because a range sensor provides both distance and bearing information, whereas a camera measures intensity of light reflected by objects and bearing only. The fundamental data association problem of establishing matches between a local sensor measurement and a (partially build) map is more complicated for a bearing only sensor because a match already established constrains other potential matches to a lesser extend. Furthermore, distance information — which is required for the construction of a metric map of the workspace — can only be recovered indirectly via triangulation of established point matches between at least two images.

In light of the above challenges, the large scope of potential applications and its cost effectiveness, in this thesis we explore the possibility of using a vision sensor for robot map building and localization.

1.3 Panoramic Vision using Mirrors

Conventional cameras have a relatively narrow field of view. In order to get an overall impression of its surrounding environment, a robot equipped with such a camera should actively look around. It could for instance use a pan-tilt-zoom mechanism to aim the camera in different directions, or it could rotate its body. It would be more practical if the robot were equipped with a panoramic vision sensor that provides a 360° impression of the surrounding environment instantaneously. One approach to achieve such panoramic
1.3 Panoramic Vision using Mirrors

Figure 1.1: Our experimental platform. The robot is a Nomad Super Scout II, manufactured by Nomadic Technologies Inc. The panoramic vision sensor, manufactured by Accowle Inc., is situated on top of the robot.

sight is to observe the world via a curved mirror. Vision sensors that combine mirrors and lenses are called **catadioptric** vision sensors. Dioptrics is the science of refracting elements. Catoptrics is the science of reflecting surfaces, or mirrors. Their combination is therefore called catadioptrics.

Mirrors, lenses and their useful properties have long been known to mankind. The earliest known mirrors were made out of polished volcanic glass and date back to c. 7000 BC. Lenses made out of crystal rock, dating back to c. 900 – 700 BC have been found at sites in Assyria. According to the Greek writer and philosopher Plutarchus (46 – 120 AD), the famous Greek mathematician Archimedes (287 – 212 BC) constructed concave mirrors during the siege of Syracuse by the Romans (214-212 AD). Supposedly, the mirrors were used to burn down Roman ships with reflected sunlight. Although this particular story remains unsubstantiated, ancient concave burning-mirrors made from polished metal have been found at sites in Egypt, China and Greece. One of the earliest known mathematical studies concerning the properties of mirrors and lenses was done by the Arabic mathematician, physicist and astronomer known as Alhazen (c. 965 – 1039 AD). He described results of experiments with spherical and parabolic mirrors (and many other optical phenomena) in his treatise “Kitab al-manazir”. In 1270, an influential Latin translation known as “Opticae thesaurus” was published in Europe. The burning mirrors from ancient times found a new application in astronomy. In 1672, Sir Isaac Newton presented his reflective telescope, featuring a concave parabolic mirror. His revolutionary design paved the way to magnification of object far beyond what could ever be obtained with a lens.

Nowadays, the use of curved mirrors to enable panoramic sight for mobile robots is
quickly gaining popularity. Although there exist alternative methods to obtain instantaneous panoramic vision (the interested reader is referred to the book edited by Benosman and Kang [2] and the proceedings of the workshops on omnidirectional vision [74, 75]), catadioptric systems appear particularly well suited for mobile robot applications. They are compact, relatively cheap, and have no moving parts so that they suffer little from wear and tear and consume little power. In this thesis, we present methods for robot mapping and localization in which we seek to take advantage of the large field of view offered by a catadioptric panoramic vision sensor. An image of our experimental platform, a Nomad Scout robot, with its catadioptric panoramic vision sensor is displayed in figure 1.1.

1.4 Original Contributions

In this thesis, we describe our research on visual mapping and localization. We have contributions in the areas of vision sensor design and calibration, robot localization, panoramic stereo vision and estimation of robot poses from images acquired during navigating in a previously unknown environment.

- We derive the parameters of a hyperboloid mirror so that the resulting catadioptric panoramic vision sensor meets a given view angle specification. We propose a simple method to calibrate the vision sensor. We show how various virtual cameras can be constructed. A virtual camera re-projects the catadioptric image onto a different surface. Throughout the thesis, we argue and demonstrate that images obtained by such virtual cameras are better suited for certain tasks than the panoramic images from which they are derived.

- We present a novel appearance-based model for probabilistic robot localization.

- We analyze the epipolar geometry for cylindrical panoramic images and propose a novel parameterization of (sinusoidal) epipolar curves that enables efficient stereo matching across multiple images obtained at (non co-linear) camera poses.

- We present a novel method to estimate a camera trajectory from a sequence of catadioptric images. Unlike many approaches presented in literature, our approach does not require many feature correspondences across many images. Instead, our method estimates the relative pose relationship between pairs of (catadioptric) images. For each pair, feature correspondences between two (virtual) cylindrical panoramic images are used to estimate the rotation and direction of translation. The length of the translation is subsequently estimated by registering two (virtual) planar perspective images of the ground plane.
1.5 Thesis Overview

Each of the contributions described in the previous section are presented in a separate chapter of this thesis.

Chapter 2 is not directly related to the problems of mapping and localization. It introduces the field of panoramic vision with a particular emphasis on vision sensor designs involving a curved mirror to obtain a panoramic field of view. We discuss the geometric properties of our particular sensor, present a calibration scheme, and show how the image obtained by the sensor can be mapped to into other valid perspective image representations.

Chapter 3 addresses the problem of mapping and localization based on images. We describe our approach, which models the relation between robot poses and the appearance of observed images directly. Modeling is done based on a set of images collected at known poses throughout the robot’s workspace. Such “appearance-based modeling” was first proposed in the field of object recognition. Since then, it has appeared in a variety of contexts, including robot mapping and localization. An important feature of our method is that it builds on a probabilistic framework for robot localization. Probabilistic approaches are generally less brittle than approaches that maintain a single pose estimate because uncertainty is explicitly represented and reasoned with. Appearance-based approaches require many training images to learn the map. In the resulting map, the concept of free versus occupied space is not explicitly represented. Collision free navigation can thus not solely be based on the appearance-based model.

Chapter 4 addresses both these issues. We present an efficient method to estimate depth from multiple panoramic images obtained at different poses. We show how the estimated depth information can be used to predict the appearance of images obtained at nearby poses. In principle, the methods discussed can be used to generate training images for the method presented in chapter 3. In order to estimate depth from images, the relative poses at which the images are acquired have to be known. In the chapter we derive the required pose information from wheel odometry and refine the pose information using vision.

Chapter 5 we present a method to estimate relative camera poses using visual information only. We apply our method to reconstruct a robot trajectory from consecutive images acquired during navigation in a previously unknown workspace. Like wheel odometry, small errors accumulate so that the estimated end pose in a trajectory may be far from the true end pose. Unlike wheel odometry, visual odometry is not “blind”. A previously visited place can be recognized. This provides a handle to correct the estimated past trajectory. The methods presented in chapter 3 and in chapter 4 require that the poses at which input images are obtained are known. The method presented in this chapter could be used to estimate the required pose information automatically.

Each chapter in this thesis has a section dedicated to discussion and conclusions. Chapter 6 draws general conclusions and indicates possible directions for future research.