Mapping and Localization from a Panoramic Vision Sensor
Bunschoten, R.

Citation for published version (APA):
Summary

This thesis presents and discusses methods and algorithms that enable a mobile robot to learn an internal representation of its workspace. An internal representation is required to correct a pose estimate, which grows more uncertain when the robot navigates, based on new observations. We consider an application in which the robot is equipped with a panoramic vision sensor providing a 360° field of view.

First, we describe the panoramic vision sensor. The sensor consists of a conventional camera, which is aimed at a convex mirror. Although the camera is mounted below the mirror, the image observed by the camera can be regarded as though it is obtained from a single, fixed viewpoint residing inside the mirror. We use this property to construct virtual cameras. These virtual cameras share the same viewpoint, but yield a projection onto a different surface (for instance a cylinder). Throughout the thesis we show that images from such virtual cameras lend themselves better for certain tasks than the panoramic images from which they are derived. We also propose a quick and simple calibration method for the panoramic vision sensor.

Next, we discuss a method to estimate an internal model for localization (pose estimation) from a set of training samples (images and their associated pose). Our approach fits within a Bayesian framework for localization (Markov localization) where the estimate of the robot's pose at any time is represented as a probability density function. In order to adopt the Bayesian approach it is essential that low-dimensional features be extracted from the images. In our application, we use global features obtained by performing a principal component analysis (PCA) on the set of training images (appearance-based approach). We maximize the performance of the model by minimizing the expected Bayesian localization error. The same criterion is applied to determine which and how many PCA features are needed for global localization from a single observation.

Next, we focus on the problem of obtaining a 3-D reconstruction from a collection of images using stereo vision. We show that for (virtual) cylindrical panoramic images, the epipolar curves (along which image correspondences must be sought) are sinusoids. The first stereo vision method we present uses an angular parameterization of these epipolar curves. The chosen parameterization is suitable to obtain a reconstruction from a pair of images. A disadvantage when only using a pair of images is that the resulting depth image contains many errors. Furthermore, it is impossible to obtain reliable depth estimates in the direction of camera displacement. We propose a second method that addresses these
issues. We derive a parameterization of epipolar curves in terms of a quantity directly related to depth. This parameterization enables us to combine depth estimates obtained from different image pairs yielding a more reliable depth estimate.

The previously discussed methods assume that the input images are obtained at known camera poses. As a final contribution, we present a two-stage method to automatically estimate relative camera poses from pairs of images. Our method assumes that the camera motion is parallel to the ground plane. In the first stage, the relative orientation and direction of translation are estimated from image correspondences (established across virtual cylindrical panoramic images). In the second stage, the length of the translation vector is estimated based on images of the ground plane (obtained by a virtual planar perspective camera aimed straight at the ground plane). In practice, the first stage yields a robust estimate of the relative orientation and direction of translation. Simulation experiments show that the second stage gives a robust estimate of the length of the translation vector in the presence of specular reflections and objects extruding from the ground plane. We apply the proposed method to reconstruct a past trajectory from a sequence of images acquired during navigation. The relative pose estimates from successive images (neighboring in time) can be combined in a simple manner to obtain a reconstruction of the past trajectory. Applied in this fashion, a form of visual odometry is obtained. Like wheel odometry, this approach is subject to accumulation of (estimation) errors. We demonstrate that the effects of error propagation can be compensated for by considering images that are neighboring in space, but not necessarily in time.

The value of the methods and algorithms proposed in this thesis are endorsed by experiments performed on real-world images.