STINT-KOSEF
Sweden/Korea workshop proposal:
Research on Visual SLAM

1 Introduction

SLAM (Simultaneous localization and mapping) has been a topic of significant interest in the robotic community over the last decade. While being considered widely solved for small indoor environments based on range sensors like laser scanners, current topics of active research include *visual SLAM* based only on camera data.

The use of cameras holds advantages as well as challenges and difficulties: on the one hand, cameras are low-cost and lightweight sensors which may be used in many applications where laser scanners are too expensive or too heavy. In addition, the rich visual information allows the use of more complex feature models for position estimation and recognition. On the other hand, visual processing holds difficulties: first, the high amount of data in images challenges real-time processing: choosing the relevant data for processing and storing is crucial. Second, depth estimation is difficult when performing bearing-only SLAM with a single camera without manual initialization. And third, different appearances of the same scene under illumination and viewpoint changes make tracking and matching a challenge.

2 Activities at KTH

This section gives an overview of some of the activities in the area of visual SLAM at KTH.

2.1 Exploiting SIFT for Point-based SLAM

The papers [1, 2] present a framework for 3D vision based bearing only SLAM using a single camera, an interesting setup for many real applications due to its low cost. The focus in is on the management of the features to achieve real-time performance in extraction, matching and loop detection. For matching image features to map landmarks a modified, rotationally variant SIFT descriptor is used in combination with a Harris-Laplace detector. To reduce the complexity in the map estimation while maintaining matching performance only a few, high quality, image features are used for map landmarks. The rest of the features are used for matching. The framework has been combined with an EKF implementation for SLAM.

2.2 Visual Scan Matching SLAM

Scan-matching based on data from a laser scanner is frequently used for mapping and localization. The paper [3] presents an scan-matching approach based instead on visual information from a stereo system. The Scale Invariant Feature Transform (SIFT) is used together with epipolar constraints to get high matching precision between the stereo images. Calculating the 3D position of the corresponding points in the world results in a visual scan where each point has a descriptor attached to it. These descriptors can be used when matching scans acquired from different positions.

Just like in the work with laser based scan matching a map can be defined as a set of reference scans and their corresponding acquisition point. In essence this reduces each visual scan that can consist of hundreds of points to a single entity for which only the corresponding robot pose has to be estimated in the map. This reduces the overall complexity of the map.

The SIFT descriptor attached to each of the points in the reference allows for robust matching and detection of loop closing situations.

One of the clear advantages of this method is that it extends nicely to an outdoor environment. The visual scan makes few assumptions about the structure and it would be very interesting to test it outdoors.

2.3 Active Camera Control for SLAM

When performing SLAM using a perspective camera it is typically quite difficult to get a large enough baseline to be able to accurately estimate the location of landmarks. The camera is often facing forward which coincides with the direction of motion. To create larger baselines and thus better estimates of the landmark locations one can actively control the direction the camera is facing. Some initial work has been performed in this area lately where a system that combines three different behaviors has been constructed. i) A tracking behavior tracks promising landmarks and prevents them from leaving the field of view. ii) A re-detection behavior directs the camera actively to regions where landmarks are expected and supports thereby loop closing. iii) Finally, an exploration behavior investigates regions where no landmarks have been found yet and enables thereby a more uniform distribution of landmarks.

2.4 Merging Object Recognition and SLAM

Linking semantic and spatial information has become an important research area in robotics since, for robots interacting with humans and performing tasks in natural environments, it is of foremost importance to be able to reason beyond simple geometrical and spatial levels. In [4] we considered this problem in a service robot scenario where a mobile robot autonomously navigates in a domestic environment, builds a map as it moves along, localizes its position in it, recognizes objects on its way and puts them in the map. For detecting objects we use the Canon VC-C4 pan/tilt/zoom camera. The zoom allows us to detect objects even if they are relatively far away. An interesting question for SLAM is how to design the representation of space such that geometric information from the laser scanner and the presence of object can both be included.

References

[1] P. Jensfelt, D. Kragic, J. Folkesson, and M. Björkman, "A framework for vision based bearing only 3D SLAM," in *Proc. of the IEEE International Conference on Robotics and Automation (ICRA'06)*, (Orlando, FL), May 2006.

[2] P. Jensfelt, J. Folkesson, D. Kragic, and H. I. Christensen, "Exploiting distinguishable image features in robotic mapping and localization," in *1st European Robotics Symposium (EUROS-06)* (H. I. Christensen, ed.), (Palermo, Italy), Mar. 2006.

[3] F. Bertolli, P. Jensfelt, and H. I. Christensen, "Slam using visual scan-matching with distinguishable 3d points," in *Proc. of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'05)*, 2006.

[4] S. Ekvall, P. Jensfelt, and D. Kragic, "Integrating active mobile robot object recognition and slam in natural environments," in *Proc. of the IEEE/RSJ International Conference on Robotics and Automation (IROS'06)*, (Beijing, China), 2006.