Detecting Useful Landmarks for Visual SLAM

Simone Frintrop, Patric Jensfelt, and Henrik Christensen

Computational Vision and Active Perception Laboratory (CVAP),
CSC, Kungliga Tekniska Högskolan (KTH), Stockholm, Sweden
Email: {frintrop/patric/hic}@csc.kth.se

Abstract. In this paper, we introduce a new method to automatically detect useful landmarks for visual SLAM. Landmarks are detected by a biologically motivated attention system. Various experimental results on real-world data show that the landmarks are useful with respect to be tracked in consecutive frames and to enable closing loops.

1 Introduction

An active area of research in mobile robotics is *simultaneous localization and mapping (SLAM)*, where a map is autonomously constructed of the environment. SLAM for indoor settings based on laser scanners is today considered a mature technology. However, the laser scanner is much too expensive for many applications. Therefore and because of the high amount of information offered by camera images, the focus within the community has shifted towards using visual information instead [1, 4, 6, 5]. One of the key problems in SLAM is *loop closing*, i.e. the ability to detect when the robot is revisiting some area that has been mapped earlier.

To perform SLAM, landmarks have to be detected in the environment. A key characteristic for a good landmark is that it can be reliably detected, this means that the robot is able to detect it over several frames. Additionally, it should be redetectable when the same area is visited again, this means the detection has to be stable under viewpoint changes. Often, the landmarks are selected by a human expert or the kind of landmark is determined in advance. Examples include localization based on ceiling lights [10]. As pointed out by [9], there is a need for methods which enable a robot to choose landmarks autonomously. A good method should pick the landmarks which are best suitable for the current situation.

In this paper, we suggest to use a computational visual attention system [2] to choose landmarks for visual SLAM. The advantage of this method is that it determines globally which regions in the image discriminate instead of locally detecting predefined properties like corners. We evaluate the usefulness of the attentional regions of interest (ROIs) and show that ROIs are easily tracked over many frames even without using position information and are well suited to be redetected in loop closing situations.

The application of attention systems to landmark selection has rarely been studied. Two existing approaches are [7], in which landmarks are detected in

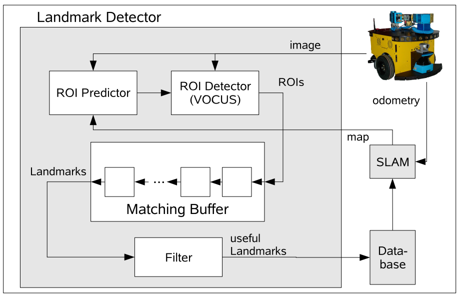

Fig. 1. Attentional landmark selection for visual SLAM

hand-coded maps, and [8], in which a topological map is build. The only approach we are aware of which uses an approach similar to a visual attention system for SLAM, is presented in [6]. They use a saliency measure based on entropy to define important regions in the environment primarily for the loop closing detection in SLAM. The map itself is built using a laser scanner though.

2 System Overview

In this section, we give an overview over the architecture in which the landmark detection is embedded (see Fig. 1). More details on the architecture can be found in [3]. The main components are the *robot* which provides camera images and odometry information, a *landmark detector* to create landmarks, a *database* to store the landmarks, and a *SLAM module* to build a map of the environment.

When a new frame from the camera is available, it is provided to the landmark detector. The landmark detector consists of a *ROI predictor* which predicts the position and appearance of landmarks depending on previous frames and on position estimations from the SLAM module, a *ROI detector* which redetects predicted ROIs and finds new ROIs based on the visual attention system VOCUS, a *matching buffer* which stores the last n frames, performs matching of the ROIs in these frames and creates landmarks from matched ROIs, and a *filter module* which filters out low quality landmarks.

The landmarks which pass the filter are stored in the database and provided to the SLAM module which performs the estimate of the position of landmarks and integrates the position into the environmental map. To detect old landmarks, the landmark position estimates from the SLAM module are used to narrow down the search space in the database.

The ROI detector consists mainly of the biologically motivated attention system VOCUS which detects regions of interest similar to the human visual system. The computations are based on detecting strong contrasts and uniqueness of features for the features intensity, orientation, and color (details in [2, 3]). The matching of ROIs between frames is based on the similarity of the ROIs (depending on the size of the ROI and a feature vector which describes the appearance of the ROIs) and on the prediction, i.e. the expected position of the

ROI considering the movement of the robot (details about the matching buffer in [3]). Details about the robot and the SLAM architecture can be found in [5].

3 Experiments and Results

The experiments were performed on a sequence of 658 images, obtained in a hallway by a robot driving 2.5 times in a circle. This setting is especially well suited to investigate loop closing situations, since most of the visited regions appear again after a while. The experiments consist of two parts: first, we show how the matching of ROIs can be used for tracking over consecutive frames and investigate the matching quality with and without position prediction. Second, we show how the matching of ROIs can be used in loop closing situations.

Tracking First, we investigate the matching of ROIs between consecutive frames. To investigate the matching by similarity independently from proximity, we first only consider matching of the ROIs based on the size of the ROIs and the similarity of the feature vector and second combine it with the position prediction.

We investigated how the number of landmarks, the length of the landmarks (i.e. the number of associated ROIs), and the quality of the matching (nb. false matches) depends on the choise of the matching threshold δ (Tab. 1, top). It shows that the number of landmarks increases when δ increases as well as the length of the landmarks, but on the other hand there are also significantly more false matches.

Finally, we performed the same experiments with additional position prediction (Tab. 1, bottom). It shows that the false matches are avoided and the matching is robust even for higher values of δ. This means that for a high threshold combined with position prediction, a higher amount of landmarks is achieved while preserving the tracking quality.

Loop closing In the next experiment, we investigate the matching quality in loop closing situations. We proceed as follows: for each ROI that is detected in a new frame, we match to all ROIs from all landmarks that occurred so far. We used a tight threshold for the matching ($\delta = 1.7$) and assumed that no position information is available for the ROIs. This case corresponds to the "kidnapped robot problem", the most difficult version of loop closing. If additional position information from odometry is added, the matching gets more precise, false matches are avoided and the threshold might be chosen higher.

In these experiments, the system attempted to perform 974 256 matches between ROIs (all current ROIs were matched to all ROIs of all landmarks that were detected so far). Out of these possible matches, 646 were matched, 575 (89%) of these matches were correct. Fig. 2 shows two examples of correct matches and one of a false match.

The experiments show that the attentional ROIs are useful landmarks which are both successfully tracked over consecutive frames and suitable to be redetected after visiting an area again from a different viewpoint.

matching without prediction:

Threshold	# landm.	# ROIs	# ROIs/Landmark (av)	# false m.
1.7	62	571	9.2	1
2.0	73	730	10.0	7
2.5	88	957	10.9	15
3.0	102	1109	10.9	42
5.0	130	1568	12.0	147

matching with prediction:

Threshold	# landm.	# ROIs	# ROIs/Landmark (av)	# false m.
1.7	61	566	9.3	0
2.0	73	724	9.9	0
2.5	88	955	10.9	0
3.0	98	1090	11.1	0
5.0	117	1415	12.1	0

Table 1. Matching of ROIs for different thresholds δ

Fig. 2. Matching results in loop closing situations: the first row shows a ROI in a current frame, the second row shows a previously seen ROI from the database. This situations usually occurred after the robot drove a circle and revisited the same area again. Two correct matches (left, middle) and one false match (right) are shown.

4 Conclusion

In this paper, we have presented a method for landmark selection based on a biologically motivated attention system which detects salient regions of interest. The matching of regions shows to be successful not only in short-term tracking but also in loop closing. The detection and matching method has been tested in a SLAM scenario to demonstrate the basic performance for in-door navigation.

Acknowledgment

The present research has been sponsored by the European Commission through the project NEUROBOTICS (EC Contract FP6-IST-001917) and the Swedish Foundation for Strategic Research through its Center for Autonomous Systems. The support is gratefully acknowledged.

References

1. Davison, A. J. Real-Time Simultaneous Localisation and Mapping with a Single Camera. In: Proc. of the ICCV (2003).
2. Frintrop, S. VOCUS: A Visual Attention System for Object Detection and Goal-directed Search. PhD thesis Bonn Germany (2005). Published 2006 in Lecture Notes in Artificial Intelligence (LNAI), Vol. 3899, Springer Verlag.
3. Frintrop, S., Jensfelt, P. and Christensen, H. Attentional Landmark Selection for Visual SLAM. In: Proc. of the International Conference on Intelligent Robots and Systems (IROS '06)(accepted) (2006).
4. Goncavles, L., di Bernardo, E., Benson, D., Svedman, M., Ostrovski, J., Karlsson, N. and Pirjanian, P. A visual front-end for simultaneous localization and mapping. In: Proc. of ICRA (2005) 44–49.
5. Jensfelt, P., Kragic, D., Folkesson, J. and Björkman, M. A Framework for Vision Based Bearing Only 3D SLAM. In: Proc. of ICRA'06 (2006).
6. Newman, P. and Ho, K. SLAM-Loop Closing with Visually Salient Features. In: Proc. of ICRA'05 (2005) 644–651.
7. Nickerson, S. B., Jasiobedzki, P., Wilkes, D., Jenkin, M., Milios, E., Tsotsos, J. K., Jepson, A. and Bains, O. N. The ARK project: Autonomous mobile robots for known industrial environments. *Robotics and Autonomous Systems* **25** (1-2, 1998) 83–104.
8. Ouerhani, N., Bur, A. and Hügli, H. Visual Attention-based Robot Self-Localization. In: Proc. of ECMR (2005) 8–13.
9. Thrun, S. Finding Landmarks for Mobile Robot Navigation. In: Proc. of the 1998 IEEE International Conf. on Robotics and Automation (ICRA '98) (1998).
10. Thrun, S., Beetz, M., Bennewitz, M., Burgard, W., Cremers, A., Dellaert, F., Fox, D., Hähnel, D., Rosenberg, C., Roy, N., Schulte, J. and Schulz, D. Probabilistic algorithms and the interactive museum tour-guide robot Minerva. *Int'l J. of Robotics Research* **19** (11, 2000).