Learning Landmarks for Robot Localization

Robert Sim and Gregory Dudek
Centre for Intelligent Machines,
McGill University
3480 University St, Room 409
Montréal, QC
H2W 2B6
{simra,dudek}@cim.mcgill.ca

Abstract

Our work addresses the problem of *learning* a set of visual landmarks for mobile robot localization. The learning framework is designed to be applicable to a wide range of environments, and allows for different approaches to computing a pose estimate. Initially, each landmark is detected using a model of visual attention and is matched to observations from other poses using principal components analysis. Attributes of the observed landmarks can be parameterized using a generic parameterization method and then evaluated in terms of their utility for pose estimation. We discuss the status of the work to date, and future directions.

Problem Statement

Our goal is to develop a framework for a robotic system which can automatically acquire knowledge of its environment that is useful for the task of navigation. An important aspect of this system is the ability to localize. It is well known that odometry alone is not sufficient for a robot to maintain an accurate estimate of its position or pose. As a result, a robot requires other external cues in order to localize accurately. In previous work (Sim & Dudek 1999), we have developed a framework for the localization problem which employs supervised learning to infer the correlation between attributes of the observed landmarks and the pose of the robot. The landmarks themselves are initially extracted as the maximal responses of an attention operator.

Our prior work serves as a proof-of-concept of the approach. There remain several unanswered questions. First, what attention operators are best suited to the task of localization? Second, how should the correlation between landmark attributes and pose be calculated? Finally, what is the best approach to collecting training data? We intend to address these issues in our future work.

Previous Work

Prior work on the localization problem includes the Markov localization framework (Fox *et al.* 1998), which computes a probability density function over the pose space and updates this funtion over time by exploiting the Markov assumption and a prior map. There are two disadvantages to that approach – first, computational condiderations require that the pose space be discretized; and second, in the basic framework pose estimates are derived from the global sensor output – outliers in the field of view can pose problems for the estimator. Furthermore, the goal of the Markov localization framework is to minimize the positional uncertainty over the *trajectory* of the robot, and hence assumes a prior probability density estimate of the pose at each stage, requiring the robot to move in order to refine its pose estimate.

Our prior work addressed these issues in the context of computing a precise pose estimate without any *a priori* knowledge or motion on the part of the robot (Sim 1998). The mechanism operates by first *exploring* the environment, collecting images from a set of sample positions. From this training data a set of *tracked landmarks* are computed by first selecting a set of candidates on the basis of the output of an attention operator, and then tracking the candidates over the input images. Figure 1 depicts a set of candidates extracted from an image taken in our lab. A supervised learning scheme is employed to compute a set of parameterizations from the variations of visual attributes of the tracked landmarks with respect to changes in robot pose. For example, we consider the variation in appearance and position of a landmark as a function of the robot's position. The uncertainty of these parameterizations can then be estimated from the training data using a cross-validation scheme.

Once training is complete, the robot is equipped to generate a pose estimate from anywhere in its pose space. When an estimate is required, it acquires an input image and extracts a set of candidate landmarks. These candidates are matched to the learned tracked landmarks, and a separate pose estimate is generated for each match. The resulting estimates are then combined in a robust manner, taking into account the *a priori* uncertainty of the tracked landmark parameterizations. Figure 2 demonstrates the accuracy of the method. Each 'x'–'o' pair represents a generated pose estimate and the actual pose of the robot, over a 2m by 2m pose space in a laboratory environment.

Future Work

Currently, we are evaluating the suitability of different attention operators for the task of extracting candidate landmarks. The generic nature of our framework allows for a straight-

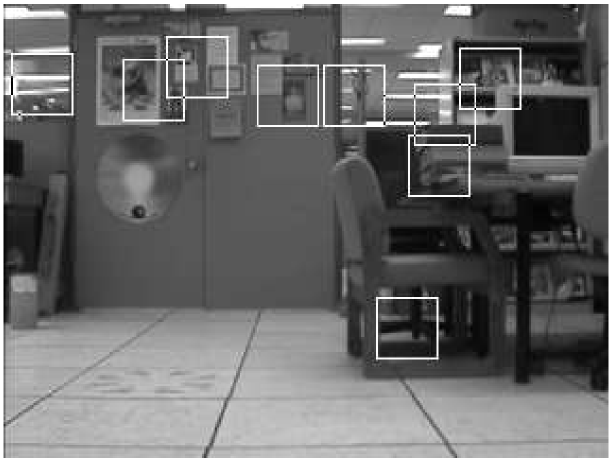

Figure 1: A scene with extracted landmark candidates.

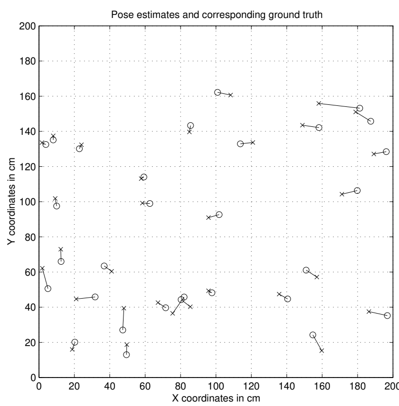

Figure 2: A set of pose estimates obtained from a 2m by 2m laboratory environment.

forward substitution of a new attention operator. The quality of the results will be measured in terms of the raw accuracy of the resulting pose estimation system, as well as its ability to measure the uncertainty of the pose estimates.

In order to address the problem of selecting an appropriate attention operator, we will consider the prior work of Shi and Tomasi (Shi & Tomasi 1994), which selects features which are well disposed to tracking, Reisfeld *et al.* (Reisfeld, Wolfson, & Yeshurun 1995), which selects features on the basis of symmetry, and Itti and Koch (Itti, Koch, & Niebur 1998), which employs a variety of criterion in order to select regions of high saliency.

Second, we are investigating alternative parameterization paradigms. Our initial parameterization method linearizes the joint data-pose space, and the results are very good compared to the standard linear least-squares estimation approaches which treat the pose and data spaces as duals. We are also considering non-linear parameterization approaches such as neural networks and regularization.

Finally, we are considering the important problem of deciding what are the best poses from which to collect training observations. In formulating a solution to such a problem, one must consider seriously the issue of what constitutes a good model of a landmark, as well as the costs associated with physically moving the robot to new vantage points.

Our consideration of these issues will rely heavily on the theoretical framework for inverse problem theory developed by Tarantola (Tarantola 1986), Whaite's approach to model parameter estimation using active vision (Whaite & Ferrie 1990) and the approach to map-building developed by Kuipers and Byun (Kuipers & Byun 1991), which constructs local regions of reliability in the pose space.

We expect that a theoretically rigorous approach to the problem of learning to estimate pose from visual landmarks will lead to an implementation that is successful both experimentally and in practical settings.

Acknowledgements

This work is funded by a PGS-B scholarship from the Canadian Natural Sciences and Engineering Research Council.

References

Fox, D.; Burgard, W.; Thrun, S.; and Cremers, A. B. 1998. Position estimation for mobile robots in dynamic environments. In *Proceedings of the 15th National Conference on Artificial Intelligence (AAAI-98) and of the 10th Conference on Innovative Applications of Artificial Intelligence (IAAI-98)*, 983–988. Menlo Park: AAAI Press.

Itti, L.; Koch, C.; and Niebur, E. 1998. A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 14:1254–1259.

Kuipers, B., and Byun, Y.-T. 1991. A robot exploration and mapping strategy based on a semantic hierarchy of spatial representations. *Robotics and Autonomous Systems* 8:47–63.

Reisfeld, D.; Wolfson, H.; and Yeshurun, Y. 1995. Context free attentional operators: the generalized symmetry transform. *International Journal Of Computer Vision* 14:119–130.

Shi, J., and Tomasi, C. 1994. Good features to track. In *Proceedings of the Conference on Computer Vision and Pattern Recognition*, 593–600. Los Alamitos, CA, USA: IEEE Computer Society Press.

Sim, R., and Dudek, G. 1999. Learning and evaluating visual features for pose estimation. In *Proceedings of the IEEE International Conference on Computer Vision(ICCV)*. Kerkyra, Greece: IEEE Press.

Sim, R. 1998. Mobile robot localization from learned landmarks. Master's thesis, McGill University, Montreal, Canada.

Tarantola, A. 1986. *Inverse Problem Theory*. Amsterdam, The Netherlands: Elsevier Science Publishers B. V.

Whaite, P., and Ferrie, F. P. 1990. From uncertainty to visual exploration. In *Proceedings of the 3rd International Conference on Computer Vision*.