Expectation-Based Vision for Self-Localization on a Legged Robot

Daniel Stronger and Peter Stone
Department of Computer Sciences, The University of Texas at Austin
{stronger,pstone}@cs.utexas.edu
http://www.cs.utexas.edu/~{stronger,pstone}

Introduction

This paper presents and empirically compares solutions to the problem of vision and self-localization on a legged robot. Specifically, given a series of visual images produced by a camera on-board the robot, how can the robot effectively use those images to determine its location over time? Legged robots, while generally more robust than wheeled robots to locomotion in various terrains (Wettergreen & Thorpe 1996), pose an additional challenge for vision, as the jagged motion caused by walking leads to unusually sharp motion in the camera image. This paper considers two main approaches to this vision and localization problem, which we refer to as the *object detection* approach and the *expectation-based* approach. In both cases, we assume that the robot has complete, a priori knowledge of the three-dimensional layout of its environment. These two approaches are described in the following section. They are implemented and compared on a popular legged robotic platform, the Sony Aibo ERS-7.

Many researchers have addressed the problem of vision and localization on a legged robot, in part because of the four-legged league of the annual RoboCup robot soccer competitions (Stone, Balch, & Kraetszchmar 2001). Nevertheless, of all of the approaches to this problem on a legged robot, we are not aware of any that take the expectation-based vision approach. However, we show that this approach can yield a higher overall localization accuracy than a state-of-the-art implementation of the object detection approach. This paper's contributions are an exposition of two competing approaches to vision and localization on a legged robot and an empirical comparison of the two methods.

The Object Detection Approach

The most common approach to vision and localization on a legged robot is the *object detection* approach. This approach relies on a set of landmarks: objects at fixed, known positions. Within a set of useful landmarks, distinct objects have recognizable, distinct appearances. If the agent can recognize these objects in the image based on their appearances, it can then use these sightings as landmarks for self-localization. For example, if the objects are a set of differently colored spheres, they will appear as differently colored solid circles in the image, and their size and color can be used to determine the robot's distance from the corresponding object in its environment.

Once the objects in the image have been identified, they can be used to inform the robot's localization process. An observed object is converted to a distance and horizontal angle from the known location of the given object. These observations are then accumulated over time, along with the robot's odometry information, into a running estimate of the robot's *pose*, its two-dimensional position and orientation. This final process can be accomplished by either particle filtering (Dellaert *et al.* 1999; Sridharan, Kuhlmann, & Stone 2005) or Kalman filtering (Kalman 1960; Welch & Bishop 2004).

The Expectation-Based Approach

The idea behind the expectation-based approach is that given a three-dimensional model of the robot's environment and an estimate of its pose, it can compute an *expected view*, a representation of the locations of the objects in the visual scene. In this paper, we focus on *line-model* expected views, which focus on the expected positions and orientations of the object boundaries in the image. This approach has been used previously in many domains, such as object tracking (Lowe 1991) and localization on a wheeled robot (Kosaka & Pan 1995).

Once the expected view is computed, the next step is to perform edge detection on the image and compare the observed and expected edges in the image. For each pair of an observed edge and an expected edge, if they are close enough in position and orientation, they are identified with one another. These line segment identifications are then used to update the robot's camera pose estimate, based on the robot's model mapping line segment positions to a pose. This process is repeated until the camera pose stabilizes. The details of this computation are given by Lowe (1991). Between each video frame and the next, the robot's pose is updated in accordance with its odometry estimate.

Empirical Validation

This section describes our test-bed robot and its environment, as well as our implementations of the two approaches to vision and localization described above. The methods are implemented on a Sony Aibo ERS-7.[1] The robot is roughly $280mm$ tall and $320mm$ long. It has 20 degrees of freedom: three in each of four legs, three in the neck, and five more in its ears, mouth, and tail. At the tip of its nose there is a CMOS color camera that captures images at 30 frames per second in YCbCr format. The images are 208×160 pixels giving the robot a field of view of $56.9°$ horizontally and $45.2°$ vertically. The robot's processing is performed entirely on-board on a 576 MHz processor. The robot's environment is a color-coded field measuring 4.4×2.9 meters, whose components are geometrical shapes: two colored goals composed of rectangular panes, inclined plane walls on the boundary, and four cylindrical beacons.

Our group has previously implemented the object detection approach to vision and localization in this domain. Our implementation has been highly optimized, and it is based on recent research contributions (Sridharan & Stone 2005; Sridharan, Kuhlmann, & Stone 2005).

Our implementation of the expectation-based approach, which is new to this paper, is demonstrated in Figures 1 and 2. Figure 1 shows the edges in the expected view given a prior pose estimate. Each line in the expected view also has two colors associated with it, based on the colors on each side of that line in the three-dimensional model. After the lines are identified, the robot's camera pose estimate is updated so that the lines are in the correct locations. This final situation is depicted in Figure 2. A video of the robot walking with superimposed model lines is available online.[2]

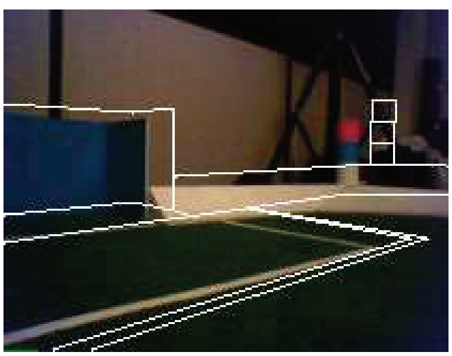

Figure 1: The white lines depict the expected view.

The two approaches were compared experimentally by running each one and manually measuring its localization accuracy. This was done by having the robot walk to a preset series of 14 poses on the field. For each pose that the robot attempted to visit, its position error and angle error were recorded. Each approach was evaluated over ten trials. The average position error achieved by the two methods were 11.81 ± 0.82 cm for the object detection method, and 7.55 ± 0.63 cm for the expectation-based method. The average angle errors were 6.49 ± 1.61 degrees for object detection and 7.18 ± 2.22 degrees for the expectation-based method.

[1] http://www.aibo.com

[2] www.cs.utexas.edu/~AustinVilla/?p=research/expectation-vision

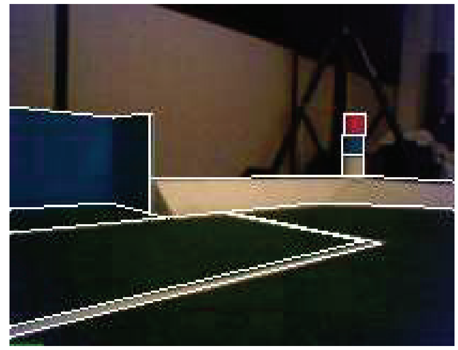

Figure 2: The white lines depict the final line placements.

The position errors attained by the object detection approach were, on average, 56% more than those attained by the expectation-based approach. This difference is statistically significant in a one-tailed t-test, with a p-value less than 10^{-9}. Although the particle filtering approach achieved a lower average angle error, this difference was not statistically significant, with $p = 0.44$ for a two-tailed t-test.

Conclusion

This paper presents two approaches to the problem of vision and localization on a mobile robot. One of them, the expectation-based method, has not previously been used on legged robots, to the best of the authors' knowledge. This method is implemented on a Sony Aibo ERS-7. We empirically compare these methods by using each one to localize the robot over a series of points and measuring its accuracy.

Acknowledgements

This research was supported in part by NSF CAREER award IIS-0237699, ONR YIP award N00014-04-1-0545, and DARPA grant HR0011-04-1-0035.

References

Dellaert, F.; Fox, D.; Burgard, W.; and Thrun, S. 1999. Monte carlo localization for mobile robots. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*.

Kalman, Rudolph, E. 1960. A new approach to linear filtering and prediction problems. *Transactions of the ASME–Journal of Basic Engineering* 82:35–45.

Kosaka, A., and Pan, J. 1995. Purdue experiments in model-based vision for hallway navigation. In *Proceedings of Workshop on Vision for Robots in IROS'95*.

Lowe, D. 1991. Fitting parameterized three-dimensional models to images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* PAMI-13(5):441–450.

Sridharan, M., and Stone, P. 2005. Real-time vision on a mobile robot platform. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*.

Sridharan, M.; Kuhlmann, G.; and Stone, P. 2005. Practical vision-based monte carlo localization on a legged robot. In *IEEE International Conference on Robotics and Automation*.

Stone, P.; Balch, T.; and Kraetszchmar, G., eds. 2001. *RoboCup-2000: Robot Soccer World Cup IV*. Lecture Notes in Artificial Intelligence 2019. Berlin: Springer Verlag.

Welch, G., and Bishop, G. 2004. An introduction to the kalman filter. Technical Report 95-041, University of North Carolina at Chapel Hill, Department of Computer Science.

Wettergreen, D., and Thorpe, C. 1996. Developing planning and reactive control for a hexapod robot. In *Proceedings of ICRA '96*, volume 3, 2718–2723.