Active Visual Localisation for Cooperating Inspection Robots

Andrew J. Davison **Nobuyuki Kita**

Intelligent Systems Division
Electrotechnical Laboratory
1-1-4 Umezono, Tsukuba
Ibaraki 305–8568, Japan
[davison,nkita]@etl.go.jp

Abstract

In the routine inspection of industrial or other areas, teams of robots with various sensors could operate together to great effect, but require reliable, accurate and flexible localisation capabilities to be able to move around safely. We demonstrate accurate localisation for an inspection team consisting of a robot with stereo active vision and its companion with an active lighting system, and show that in this case a single sensor can be used for measuring the position of known or unknown scene features, measuring the relative location of the two robots, and actually carrying out an inspection task.

1 Introduction

A task potentially well suited to autonomous mobile robots is the routine inspection of industrial or other areas: when tackled by humans, this can be repetitive and tedious, requiring a great deal of concentration, or even sometimes dangerous. For instance, in a nuclear power plant there are a mass of pipes, valves and other equipment which must be checked regularly for cracks or other flaws. Problems may have other non-visual symptoms, such as chemical leaks or tell-tale noises.

What type of robots are required for these inspection tasks? Potentially those with multiple sensing modalities to enable them to carry out a wide range of inspection tasks: cameras of different types, microphones, chemical sensors, etc. These different functionalities must be coordinated and sequenced in a unified framework.

While robots in many applications can operate without needing to know where they are (wandering lawn-mowing robots for instance), an inspection robot needs to follow a planned route to certain things and localisation is essential. The most easy way to achieve localisation is to place easy-to-find beacons in known locations in its environment: examples which have often been used are bar-code beacons detectable by a laser sensor, or of course GPS satellites for outdoor environments.

More generally, the robot can be supplied with a map of natural features in the environment which it can recognise with its sensors, or detect features itself and build a map as it moves. In robotic inspection, both of these modalities will be useful. While during normal operation the wisest course of action would be to supply the robot with a prior map of the environment to make reference to, there is the possibility that something unexpected may cause known features to become invisible or force the robot to move out of its normal area of operation: the ability to add its own features to the map would be invaluable here.

Most of the sensors that the robot carries can be used to aid with localisation: in general any sensor that is able to repeatably measure the relative location of the robot and something which is fixed in the scene will contribute information (the problem of combining many different sources of information has been referred to as *data fusion* [6]). Further, if the robot has a certain itinery of features it must inspect, these features themselves can be used for localisation. In our current implementation, we use point visual features, but this idea becomes more powerful once extended for instance to linear features such as pipes which must be inspected along their length [4]: this process takes some time, but if the measurements of the pipe are providing localisation information as well the robot could safely keep moving during the inspection.

There is a strong case for multiple robots to operate cooperatively in order to accomplish inspection tasks. One possibility is that similar robots could divide a large region between them in order to increase the speed of the work. Perhaps more powerfully, "specialist" robots, each equipped with different sensing equipment, could work together in an area, each using its particular talents as required. This is a common approach in human work teams of course, where each member will have his own function. As robots become more advanced and their capabilities extend to diagnosing or repairing problems in addition to simply flagging them, this team approach becomes increasingly valid.

1.1 In This Paper

In this paper we tackle the question of localisation for multiple robots with the specific example of a pair of cooperating inspection robots: one equipped with stereo active vision, and one a blind assistant carrying an active lighting system (see Figure 1(a). The key points of our approach are:

- Identification and measurement of the position of widely-spaced landmarks in 3D using active vision.
- The features being inspected by the robots can themselves be used as navigation landmarks.
- A fully automatic implementation of cooperative navigation for the robot pair, where measurements by the vision robot of a visual beacon mounted on its blind assistant aid the localisation of both.
- Localisation implemented using a general framework for map-building and localisation, which supports the simultaneous use of multiple robot, sensor and feature types and flexibility with respect to map-updating strategy [1]).

We will show that a sparse map of high-quality visual features can be used in combination with unreliable odometry to provide highly accurate localisation suitable for directing robots safely through a complex inspection environment. Accurate localisation is not a complete scheme for navigation and inspection — robots operating in the real world must be able to avoid unexpected obstacles and react to humans for instance — but is the most important and fundamental building block, providing a framework into which further capabilities can be built with confidence.

(a) (b)

Figure 1: (a) Robot inspection team: the robot on the right carries an active vision system, and that on the left active lighting. (b) ESCHeR, the ETL Stereo Compact Head for Robot Vision, features 4 degrees of rotational freedom and foveated lenses which provide high resolution in the image centre and low resolution in the periphery [5].

2 Localisation Using Active Vision

In this section we will describe the localisation process of a single robot equipped with a vision system, closely following the approach of [2], before moving on to the multiple robot case in Section 3.

2.1 Active Vision

In active approaches to sensing, sensor or information processing resources are directed purposively to regions of current interest in a scene, rather than being used to acquire and process data uniformly. In vision, active operation is achieved either by selective processing of the images acquired by fixed cameras, or in our case by physically directing the cameras as required using a motorised camera platform or "active head": see Figure 1(b).

Active vision is easily applied to short range "tactical" navigation tasks such as steering around a known obstacle: the obstacle is tracked and a simple law can be used to control the robot's steering with respect to the viewing angle [7]. However, there have only been a few attempts to apply it to more long-term navigation tasks such as map-building and using [2]. This is surprising since it is the main tool used in human navigation: as we move around our environment, our eyes constantly change their fixation point to look for of landmarks, check for obstacles or pick out headings. Attention must be divided between these important tasks as required, and this is what makes active sensing an interesting research area.

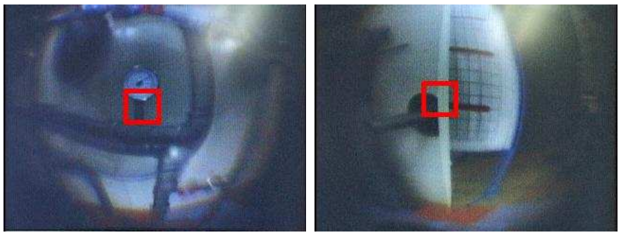

Figure 2: Fixated views of typical feature matches by correlation of 15×15 pixel image patches.

2.2 Visual Landmarks

The basis for localisation using vision is a map of features which are repeatably measurable using the robot's cameras. Within our general mapping framework, there is the potential for these features to have many different forms: points, lines [4] or planes for instance. In our current implementation, point features in 3D space are recognised using image correlation matching, which proves to be surprisingly robust to changes in viewpoint. The active head moves to fixate features for measurement with both of its cameras, acting as an accurate "pointing stick" which can measure the direction and depth of features over a very wide field of view. Figure 2 shows some typical features used. These features are either detected automatically by the robot as it navigates using an image interest operator, or can be supplied externally.

2.3 Storing and Updating Map Data

The current state of the robot and the scene features which are known about are stored in the system state vector $\hat{\mathbf{x}}$ and covariance matrix $\mathtt{P}$. These are partitioned as follows:

$$\hat{\mathbf{x}} = \begin{pmatrix} \hat{\mathbf{x}}_v \\ \hat{\mathbf{y}}_1 \\ \hat{\mathbf{y}}_2 \\ \vdots \end{pmatrix} \quad , \quad \mathtt{P} = \begin{bmatrix} \mathtt{P}_{xx} & \mathtt{P}_{xy_1} & \mathtt{P}_{xy_2} & \dots \\ \mathtt{P}_{y_1x} & \mathtt{P}_{y_1y_1} & \mathtt{P}_{y_1y_2} & \dots \\ \mathtt{P}_{y_2x} & \mathtt{P}_{y_2y_1} & \mathtt{P}_{y_2y_2} & \dots \\ \vdots & \vdots & \vdots & \end{bmatrix} . \tag{1}$$

$\hat{\mathbf{x}}_v$ is the robot state estimate, and $\hat{\mathbf{y}}_i$ the estimated state of the ith feature ($\hat{x}, \hat{y}, \hat{z}$ for the case of a 3D point). The state vector represents the robot's map of its environment and its place within it, and the covariance matrix how uncertain this information is.

Feature positions may be supplied as prior information when the robot starts navigating, or dynamically added to or removed from the map during navigation as required. A full SLAM approach [8] is used, propagating the covariances between the robot state and all feature estimates, and between the feature states themselves. This is essential in our system, since the small number of features generally used must have estimates which are of very high quality to provide accurate localisation information. The ability of active cameras to view features over a huge field of view is key to the quality of this map: the robot can really see the same features continuously as it goes through very large motions and rotations; thus fewer features need to be added to the map, and the uncertainties related to those present can be reduced succesively as the robot is able to measure them repeatably over long periods.

The data is updated sequentially as the robot moves around its environment and makes measurements of the features in its map following the rules of the Extended Kalman Filter: a prediction step when the robot moves, when a new position estimate is calculated based on odometry, and an update step when a measurement is made of a feature. Our map-building software system supports plug-in models which describe the specifics of robot motion and feature measurement [1].

2.4 Active Measurement

In an active scenario, it is necessary to decide at each instant which feature in the map to attempt to measure. This decision is made based on two criteria: expected visibility and the value of the measurement. The expected visibility (more precisely measurability) is something that depends on the sensor and feature type: for instance, with our point features matched by correlation, we do not expect to be successful with matching if the viewpoint is too different from that from which they were initially seen. Since we have an estimate of the current robot position, the predicted viewing direction can be evaluated in this respect.

Once a measurable subset of features in the map has been identified, the value of measuring each one is evaluated in terms of the uncertainty of their position relative to the robot (we choose a measurement which has a high innovation covariance [2]), the general principle being that there is little use in making a measurement of which we are sure of the result.

The selected feature is then measured by driving the active head to the angles predicted for fixation on that feature, and searching the images obtained for a match. Precise search regions are calculated from the uncertainty in the map, which maximise computational efficiency and reduce the chance of mismatches.

3 Multiple Robot Localisation

When the robot with active vision is accompanied by an assistant robot whose job is to provide lighting, this robot's position is another unknown quantity which must be sequentially estimated as it moves: its position state estimate and covariance are inserted next to those of the first robot in Equation 1.

Of course, the pair must coordinate their positions accurately to cooperate in inspection tasks. Lacking in this implementation sensors of its own, the assistant robot must rely purely on odometry during solitary movements. However, it is well known that the uncertainty of position estimates based solely on odometry grows without bound over time. To counter this, a visual beacon is attached to the second robot, of which the vision robot is able occasionally to make observations. This beacon is placed off-centre with respect to the second robot's rotation axis such that measuring it provides information on the robot's orientation as well as location. The meaurement takes place in exactly the same way as a normal feature measurement, with search regions generated and correlation matching, and is processed using the EKF in the same manner as a normal feature measurement.

Of course this approach would be equally valid should the second robot also possess sensors and the ability to measure its position with respect to a map. In that case, making an inter-robot measurement would be more of a mutually beneficial process of position estimation improvement, rather than the current situation where the vision robot transfers its rather good position information to the uncertain blind robot. Note that this occurs automatically due to the imbalance in the uncertainty of the two robots: we do not specify which robot is "helping" which.

An interesting recent approach to cooperative robot localisation by Fox *et al.* [3] used Monte Carlo representations of the probability distributions involved rather than the first-order approximations inherent in a Kalman Filter approach. This method is very powerful because it can deal transparently with the multi-modal distributions which stump the Kalman Filter. However, the computational expense involved means that it is not yet applicable to cases where the positions of features can be uncertain as well as that of the robots: in the work of [3] the robots navigated using a completely known map.

In our system, the inter-robot measurement is an implementation of the general *self-measurement* capability built into our localisation software framework. As well as being able to make measurements of arbitrary features in the world, a robot or robot group has the potential to make measurements of its own internal parameters; in the multiple robot case this includes measurements between the robots.

4 Experiments

4.1 Experimental Environment

In our laboratory we have constructed a mockup of a section of the interior of a nuclear power plant to use as an experimental environment. The mockup features authentic pipes, dials and valves, and has a reasonably large area for the robots to navigate around.

In our implementation, all visual and localisation processing is carried out by a Linux PC built into the vision robot. The lighting robot acts purely as a slave device, receiving control commands from the vision robot. Communications between the two robots are achieved via radio ethernet.

In the experiments below, ground-truth robot positions were measured by hand with respect to an accurate floor grid to compare with the output from the navigation algorithm.

4.2 A Cooperative Inspection Task

In this scenario, the active vision robot and its assistant carrying the active lighting system must collaborate to inspect a portion of pipe which lies towards the back of the arrangement of equipment in our plant mockup. In order to achieve a clear view, the vision robot must navigate through a narrow gap (only a few centimetres wider than itself) to get into position, while the lighting robot moves to a nearby position to supply illumination.

From starting points on the far side of the mockup, the robots navigate according to a pre-programmed route of known "waypoints": these are simply linked positions through which they must move. Of course though, a robot must know its location in order to know when it has reached a waypoint, so this is a pure test of the robots' localisation capabilities. The robots were initially manoeuvred by hand to known positions in the scene and these positions inserted as the initial state $\mathbf{x}_v$ — note though that the precision of these initial placements was also estimated and used as the initial covariance $\mathtt{P}_{xx}$ (for example the precision to which the robot's orientations could be aligned by hand was assessed as a couple of degrees).

During navigation, the vision robot made continuous measurements of features in the scene, which in this case had been supplied as a prior map so their

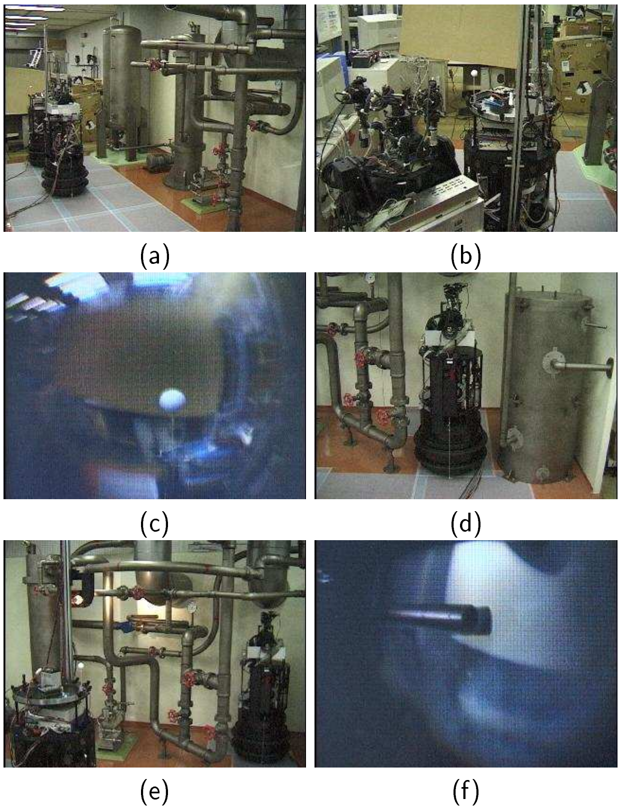

Figure 3: Collaborative inspection. (a) Start position. In (b) and (c) the vision robot makes an observation of the marker carried by the lighting robot to improve its position estimation (note that the vision robot rotates here to bring the lighting robot into view). At (d) the vision robot must pass through a narrow gap to reach the ideal inspection position (e), where the lighting robot illuminates the scene to provide view (f) of the pipe to be assessed.

positions were perfectly known. At points about 1m apart it turned to check on the progress of the lighting robot by measuring its beacon marker (although the active head carried by the vision robot provides an almost hemispherical field of view, when the lighting robot is behind it is necessary for the vision robot to rotate in order to bring it into view). After visual search has located the marker, the measurement of it's location relative to the vision robot is fed through the localisation filter to produce updated position estimates for both robots.

Figures 3 and 4 explain the inspection task in more detail.

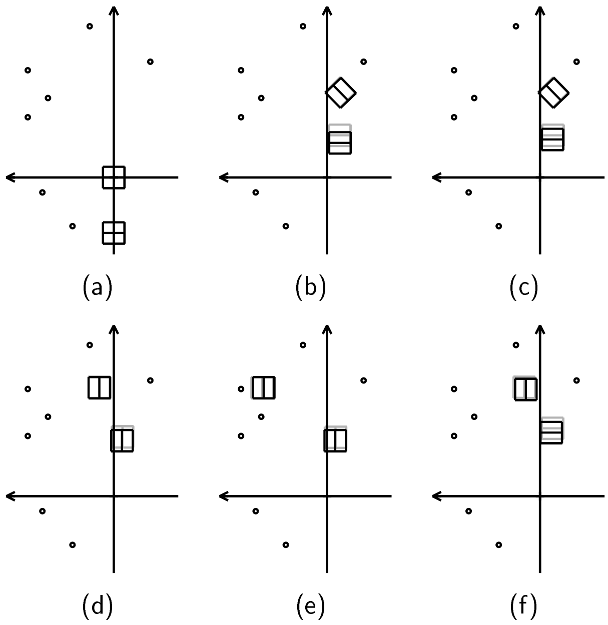

Figure 4: Collaborative navigation while performing an inspection task. Ground truth robot position measurements in black are superimposed on localisation estimates in grey. The point features referred to are also shown. From starting position (a), the robots move forward together in the upward direction; the vision robot is in front of the the lighting robot. Pair (b) and (c) are snapshots before and after an inter-robot measurement, and the improvement in the position estimate of the lighting robot can clearly be seen. The position estimate of the vision robot remains good throughout, such that at (d) it can enter a narrow gap to reach the inspection position (e). The vision robot exits through the gap again in (f) and the robots are ready to continue their inspection tour.

4.3 Observer-Aided Localisation

A more thorough experiment was carried out to assess the value of aiding the navigation of a blind robot with measurements from a robot with better localisation. The simple situation set up involved the vision robot remaining stationary while the lighting robot moved around under its occasional observation. Since the vision robot did not move, its position uncertainty remained very small and it was best able to assist the blind robot.

A "figure of 8" shaped path was planned for the lighting robot and two trials were carried out: first using odometry only (from which a sequence of snapshots are shown in Figure 5), and then with regular

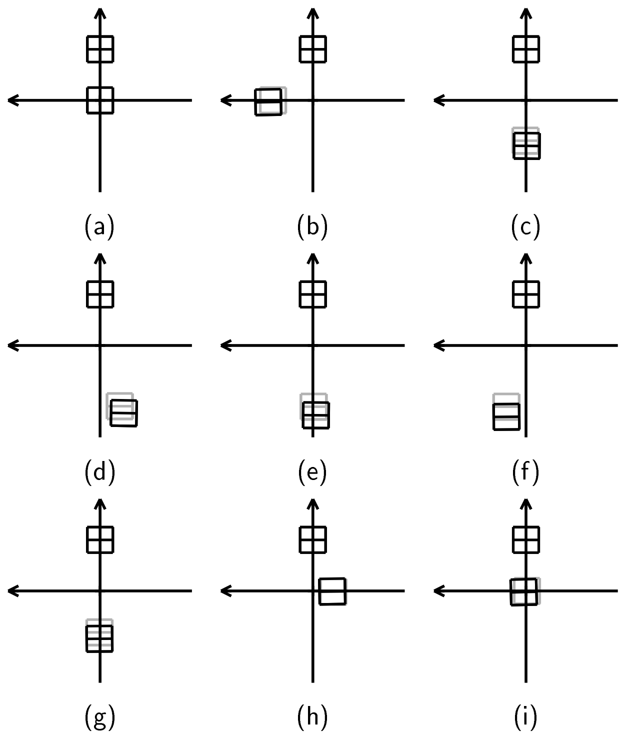

Figure 5: Unreliable odometry-based localisation for the blind lighting robot.

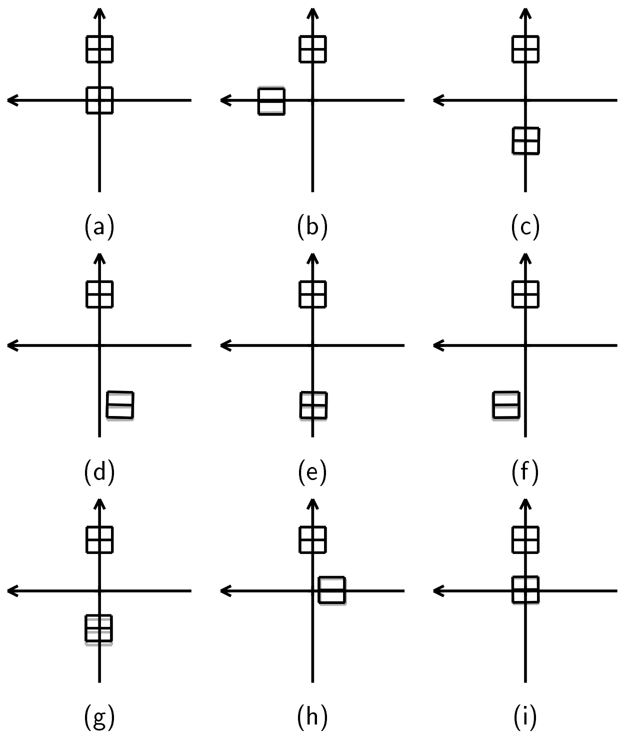

Figure 6: Observer-aided localisation: the vision robot made regular measurements to correct the localisation of the blind lighting robot as it moved.

measurements from the stationary vision robot (Figure 6). The real-world scale of these diagrams is about 3 metres. It can be seen that in the odometry-only case, after only a short motion the true and estimated positions began to differ significantly (on the order of 20cm), although perhaps a lucky coincidence saw the robot return quite accurately to its starting point in the last of the snapshots. When the robot was sent on another circuit of the route, however, built-up errors, particularly in the steering direction of the robot which is not directly visible in these diagrams, meant that the robot was soon dangerously off-course and had to be stopped before a collision occurred with the scenery.

In second case where regular observations took place, good localisation was maintained throughout with consistent accuracy of within 3–4cm. After returning to its starting point in the last snapshot, the robot was sent off on a further two circuits of the course, which it completed with consistent accuracy in the same range.

4.4 Evaluation

We have shown in these experiments how a very sparse map of visual features can be used for extremely accurate and repeatable localisation. Rather than as in many approaches to robot navigation where very dense feature maps are made, our approach concentrates on just a few *high quality*, *widely spaced* features, and uses intelligent active measurement selection to switch attention between them as necessary. By making occasional measurements of a carried beacon, it can also greatly aid the localisation of an assistant robot with poorer sensor capabilities.

In this style of multiple robot navigation, where an inter-robot measurement is only made quite infrequently, it is essential that the reliability with which the beacon placed on the second robot can be matched is high: since the uncertainty in the robots' relative location will be large, making this measurement will involve a large search region, and there is the potential for making mismatches.

5 Conclusions

Cooperative localisation for multiple robots can simply and rigorously be incorporated into the localisation and map-building framework for a single robot.

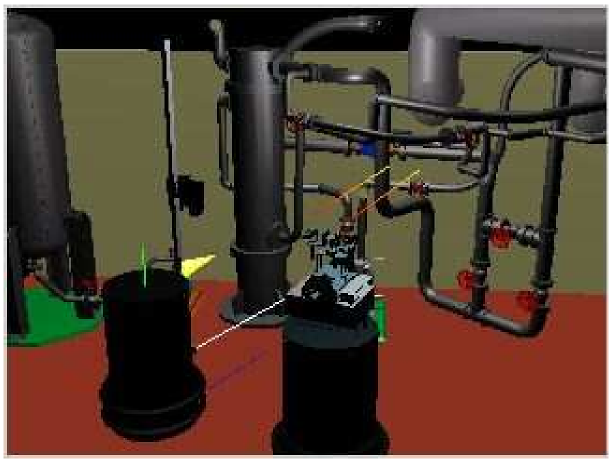

Figure 7: A graphical model of our nuclear plant mockup.

We have shown how inter-robot measurement can significantly improve localisation estimates in task-oriented navigation.

A very appealing direction for future research arises in conjunction with a full 3D graphical model we have constructed of our experimental plant environment and robots (see Figure 7). Interfacing this model with our navigation system in real time will permit more intelligent active choice of measurements for instance, since it will be possible to predict whether a feature will be occluded from a particular viewpoint. Further, with sufficient graphical quality it will be feasible to generate feature representations from the model itself rather than extracting those directly from the real world.

Against the increased capabilities of robots with multiple sensing modalities must be balanced the problems of added complexity and increased risk of hardware failure.

Acknowledgments

The authors would like to thank Isao Hara for implementational assistance with inter-robot communications.

References

[1] A. J. Davison and N. Kita. Sequential localisation and map-building in computer vision and robotics. In *Proceedings of the 2nd Workshop on Structure from Multiple Images of Large Scale Environments (SMILE), in conjunction with ECCV 2000, Dublin, Ireland*, 2000.

[2] A. J. Davison and D. W. Murray. Mobile robot localisation using active vision. In *Proceedings of the 5th European Conference on Computer Vision, Freiburg*, pages 809–825, 1998.

[3] D. Fox, W. Burgard, H. Kruppa, and S. Thrun. Efficient multi-robot localization based on monte carlo approximation. In *Proceedings of the 9th International Symposium of Robotics Research, Snowbird, Utah*, pages 113–120, 1999.

[4] N. Kita, D. W. Murray, and A. J. Davison. Mobile robot localization using active fixation on a line. In *IEEE Workshop on Perception for Mobile Agents*, pages 37–42, 1998.

[5] Y. Kuniyoshi, N. Kita, S. Rougeaux, and T. Suehiro. Active stereo vision system with foveated wide angle lenses. In *ACCV*, pages 359–363, 1995.

[6] J. Manyika and H. F. Durrant-Whyte. *Data Fusion and Sensor Management: A Decentralized Information-Theoretic Approach.* Prentice Hall, 1993.

[7] D. W. Murray, I. D. Reid, and A. J. Davison. Steering and navigation behaviours using fixation. In *Proceedings of the 7th British Machine Vision Conference, Edinburgh*, pages 635–644, 1996.

[8] R. Smith, M. Self, and P. Cheeseman. A stochastic map for uncertain spatial relationships. In *4th International Symposium on Robotics Research*, 1987.