Selecting Targets for Local Reference Frames

Saul Simhon and Gregory Dudek
Centre for Intelligent Machines
McGill University
3480 University St, Montreal, Canada H3A 2A7

Abstract—**This paper addresses the problem of seeking out parts of the environment that provide adequate features in order to perform robot localization. The objective is to choose good regions in which local metric maps can be established. A distinctiveness measure is defined as a measure of how well the environment allows the robot to accomplish a task, in our case the task being localization. The distinctiveness measure is evaluated as a function of both the localization strategy and the environment. Areas in the environment are considered to have high distinctiveness measures if they exhibit both sufficient spatial structure and good sensor feedback. The problem is treated as defining an evaluation criterion based on the usefulness of gathered information.**

I. Introduction

This paper addresses the issue of how and when to generate local maps for a mobile robot. While a single metric coordinate system is a natural way to map space and is effective over small areas, over large extents of space it becomes problematic. In particular, over large regions of space incremental position errors can accrue to cause large errors in the global coordinate system. This can occur even when beacons or landmarks are used to reduce odometry error. Further, this type of error can causes inconsistencies in a map when updating is performed, since updated information may be put in the wrong place.

In many cases, it is sufficient to create local coordinate frames only in selected regions, where odometry error must be minimized. Mapping large scale environments can be accomplished using a collection of local coordinate frames [1]. Where should we attempt to create a local metric map so that it will be accurate and effective? We consider how to evaluate the local environment with respect to an arbitrary localization procedure so that candidate locations can be found where local geometric maps can be generated. The specific methods used to generate the local map are assumed to exist in advance. We consider two examples illustrating the use of two drastically different classes of approach. We show that for each method, we can develop a technique that predicts how appropriate a given region will be for localization (and hence for metric mapping).

II. Background

In order to perform accurate positioning, Dudek and Mackenzie [2] composed sonar based maps where explicit object models were constructed out of sonar reading distribution in space. The maps were used to determine robot pose by fitting new sensor data to the model. Dudek and Zhang [3] used a vision system to model the environment and extract positioning information. The model consisted of extracting appropriate features from images and correlating them to pose. Position calibration was attained by training a neural network for interpolation through the feature-pose space.

These traditional metric methods use a single reference frame. Although they provide accurate local correspondence, accumulated errors tend to warp the representation over larger scale areas. To compensate with the limitation of accumulated error, there are mapping techniques where ongoing localization is used. Leonard and Durrant-Whyte [4] and Feng Lu and Evangelos E. Milios [5] employ such methods. [4] provides a self consistent description of the environment by using Kalman filter techniques to compare predicted and perceived data.

These methods, like many others, can lead to two key problems if used indiscriminately. Firstly, time and energy may be wasted in attempting to accurately map regions irrelevant to the tasks of interest. Secondly, they may attempt to detect landmarks and establish a reference coordinate frame in regions where the local structure is ambiguous or unreliable. Thus map construction using a single absolute reference coordinate system can be problematic [6]. If attempted, certain positions must be tagged extremely unreliable.

It is not always beneficial to keep metric relations over large scales. An alternative is to provide topological or qualitative relations over such extents, while storing metric relations over local areas. Prior work in cognitive science suggests humans use a set of *local* reference frames topologically connected to model large scale environment. Yeap [7] shows that a module of the Cognitive Mapping Process can be represented with a Relative Absolute model. It consists of a global representation (referred to as Relative Space Representation, or RSR) that describes a qualitative composition of a sequence of local representations {S(1), S(2),...} called Absolute Space Representations or ASRs. That is, the global map can be considered as a set of clear and accurate patches of local information linked topologically by fuzzy, semi- unknown areas (Figure 1). This is easily depicted by a person travelling down a street. While walking on the uninteresting sidewalk, the persons' attention is often diverted from the environment and reallocated to other thoughts; the description of the environment is fuzzy. When reaching a point of interest or distinction, such as an intersection, the person redirects his attention to the environment in order to accurately localize to the sidewalk edge, check the street names etc. At this point, the environments' precise structure is re-acquired.

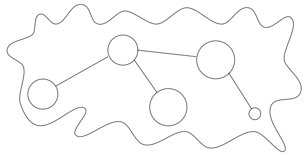


Fig. 1. Global map composed of a set of local maps. Circles represent metrically accurate local maps.

Kuipers and Byun [8] develop a mapping and exploration strategy based on both qualitative and quantitative components. Their method considers distinctiveness measures in terms of certain pre-defined sensory criteria. The map was composed by a set of edges (distinct paths) defined by 2-D distinctiveness measure criteria, and a set of nodes (distinct places) defined by 1-D distinctiveness measure criteria. A rehearsal procedure uses geometric information gathered along the nodes to distinguish new places from old ones. However, inappropriate sensory criteria can result in non-unique solutions. Furthermore, metric information is gathered in global correspondence, therefore accumulated dead reckoning error may distort true metric relations.

III. Distinctiveness Measure

In this work, the environment is represented by a set of accurate local frames. Each frame forms the node of a topological model of the world i.e., a graph [8], [9], [1]. The nodes can be referred to as *islands of reliability*. The edges of this graph correspond to control strategies that navigate the robot from one frame to another [8]. By using only separate local reference frames, we avoid the need to perform large-scale error integration [6].

A key issue in constructing such a representation, is selecting where to place the nodes of the map. That is, where are good candidate locations for local reference frames. This can be determined by evaluating the environment and selecting distinctive regions. The evaluation is derived with respect to a task, in our case localization. In the one dimension case, we define the general form of the distinctiveness measure R for a localization task as:

$$R \propto \frac{f(I, \Delta I)(1 + \sum_j \lambda_j Q_j)}{(1 + \sum_j \lambda_j)} \tag{1}$$

where I represents the strength of the response of a sensing technique and ΔI represents the amount of spatial change of that response (which may be expressed as spatial constraint). $f()$ is a function monotonically increasing with I and ΔI. Q_j is a quality measure specific to the properties of the localization technique and λ_j is a corresponding weight. That is, to successfully perform localization, there must be sufficient reliable information I subject to spatial variation ΔI along all degrees of freedom. The addition of ad-hoc quality measures Q_j, specific to the technique, can improve region selection. However, most of the emphasis is on searching for areas with rich information subject to spatial change (areas providing enough information and low structural ambiguity). A good choice for $f()$ is one of the form $I * \Delta I$, taken along the direction that results in the minimum value. Hence, R is large if both I and ΔI are large along all degrees of freedom.

IV. Local Map Perception

This section deals with the localization and modelling techniques used to build the islands. The methods relies on dead reckoning information for pose-data correspondence, therefore they are only locally consistent. For a full description refer to [2] and [3].

A. Sonar Based Environmental Model

In order to perform localization, a model is constructed of how sensory data varies as a function of the robots position. The model is built by fitting primitives to sensory data. Line segment primitives are efficient in modelling a collection of observations of the environment. (RCD's could also be used.) The line fitting method is done in several steps. First, a spatial clustering algorithm is employed to determine groups of neighbouring points that correspond to a potential line segment. Then, by using a line fitting procedure, a fitted line segment is used to model each cluster.

B. Sonar Based Pose estimation

The pose estimation problem is formulated as an optimization problem in terms of the extent to which map explains observed measurements. There are two phases involved in position calibration: 1) Classification of Data Points and 2) Weighted Voting of Correction Vectors. In the first phase, each measurement is associated to a line segment in the model. This allows to determine the Correction Vector relative to the line segment in the second phase. The second phase is that of a non-uniform weighting of Correction Vectors. Each point is given a weight in relation to the distance it lies from the associated line segment. The weighting factor is defined as a sigmoid function:

$$w(d) = 1 - \frac{d^m}{d^m + c^m} \tag{2}$$

where d is the distance from the line segment, m and c are constants. Points near their line segment are weighted more than far ones since far points may be outliers. The overall Correction Vector V is calculated as:

$$V = \frac{\sum_i w(\|v_i\|) v_i}{\sum_i w(\|v_i\|)} \tag{3}$$

where v_i is the perpendicular error vector for point i The position estimate is resolved after several iterations of translating about the Correction Vector. An important note is that only the perpendicular error of points are used to determine their Correction Vector. It is a one-dimensional position constraint provided by each measurement along the normal of the associated line segment. Ideally, the measurements would be distributed equally along all directions to allow equal localization confidence. These are some of the issues that differentiate good and bad candidate regions.

C. Image Based Environmental Modelling

The second localization technique we consider uses vision for position estimation. Rather than matching 3D models extracted from video, a notoriously difficult problem, it matches a statistical description of images to previous samples. This method has its own particular domain distinct from that of the sonar method.

For a camera mounted on a mobile robot, the dependency of the image and the pose $q = (x, y)$ is related by some function:

$$i = \Phi(q) \tag{4}$$

Where i is an N-dimensional vector of pixels. In order to solve the problem of computing the camera position we must invert the function:

$$q = \Phi^{-1}(i) \tag{5}$$

However, computing the inverse directly on images is computationally impractical. Φ in itself is not necessarily one-to-one and an inverse may not exist. To produce a computationally tractable solution the images are modelled by a set of M features:

$$G(i) = \{g_1 i, g_2(i), ..., g_M(i)\} \tag{6}$$

This produces a lower dimensional space that relates the features and the pose with a mapping:

$$f(q) = G(i) \tag{7}$$

Measurement features were derived from statistical properties of edge images (using the Canny-Deriche operator) to minimize the effects of illumination variation. The perceptual structure associated with a position in space consists of the following class of measurements:

- First and second moments of the edge distributions
- Mean edge orientation
- Densities of parallel lines at four orientations

These features compromise the first central moments of the edge distribution in space, and are the natural choices for efficient encoding.

D. Vision Based Pose Estimation

Since it is inefficient to sample the environment at every possible location and sensory data is often noisy, we must be able to interpolate within the feature space. For local areas this can be done by a linear interpolator:

$$q = \frac{\mid G(i) - G(i_1) \mid (q_2 - q_1)}{\mid G(i_2) - G(i_1) \mid} + q_1 \tag{8}$$

It was shown that the linear interpolator is only applicable to very restricted regions. In large regions or more complex areas, the linear interpolator fails. In practice, a three layer back propagation neural network is used. The network takes training examples and assigns appropriate weights to each network node by minimizing training set errors. When a new feature set is input to the network, the pose can be determined by taking a linear combination of the output units.

Generally, if the feature space is smooth, the interpolator produces good results. On the other hand, if the feature space consisted of many gaps and discontinuities, interpolation between these gaps may produce inaccurate results. We have a trade off between practical sampling resolution versus accuracy of the interpolator.

V. Distinctiveness Measure Criteria

Imagine human observers exploring a new territory, their notion of the environment can be described by a set of distinct landmarks. Once they veer off from the last known landmark, they set out to find the next distinct feature to be recorded in the cognitive map. What is distinct to human observers is associated with their goals and perception. Similarly, we seek out the best (most distinct) parts of the environment corresponding to the robots' perception (which also corresponds to a task). The task at hand is (x,y) position estimation and the sensing mechanisms considered are the sonar and vision systems described in the previous sections. The distinctiveness measure is derived accordingly

A. Sonar System Measure Criteria

A good distinctiveness measure for the sonar based technique is one that assigns high values to areas well constrained by near line segments of significant length. Furthermore, it is desirable that the line model shows similar orthogonal constraints in all directions, allowing equal localization confidence and keeping the error bound round. In the extreme case, parallel lines would result in ambiguities along one dimensions and will not provide enough information to adapt the full potential of the localization method. For the line model method, the distinctiveness measure R at a point $p = (x, y)$ over a square area $(2\epsilon)^2$ can be computed using:

$$R(p,\epsilon) = N(p,\epsilon) \int_{y-\epsilon}^{y+\epsilon} \int_{x-\epsilon}^{x+\epsilon} \frac{F(p)(1+Q(p))}{2} \delta x \delta y \tag{9}$$

where,

$$F(p) = Min[C_\perp(p), C_\parallel(p)] \tag{10}$$

$$N(p,\epsilon) = \frac{1}{\int_{y-\epsilon}^{y+\epsilon} \int_{x-\epsilon}^{x+\epsilon} \delta x \delta y}. \tag{11}$$

The functions $C_\perp$ and $C_\parallel$ represent $f()$ in equation 1 along two orthogonal directions. Q is a quality measure constraining the localization confidence to remain circular. It is based on the ratio of $C_\perp$ and $C_\parallel$, rating the equality of localization confidence about both dimensions. We integrate over a rectangular area defined by ϵ and normalize.

To describe $C_\parallel$ and $C_\perp$ we must first derive I and ΔI in terms of the line model. The strength of the response I is proportional to the length of the visible lines and their distance to the robot. Distant lines provide less reliable information than near ones due to sensing limitations an resolution. Furthermore, a line segment provides constraining information ΔI only along the normal. That is, an orthogonal position change with respect to the line guarantees a sensory measurement change. We define ΔI to be the orthogonal component of a sensed line segment.

For each line segment, we integrate the strength I and the orthogonal constraint ΔI to determine the *vector influence* along the normal to the line. We compute the vector influence for each visible line segment (in the form of $I * \Delta I$) by:

$$\mathbf{V_i}(\mathbf{p}) = \hat{N}_i \int_{\Theta} W(\mathbf{pp_i}) * (\hat{N}_i \bullet \mathbf{p}\hat{\mathbf{p}}_\mathbf{i}) \delta\Theta \tag{12}$$

Vi is the orthogonal vector influence for line segment i seen by point p and $\hat{N}_i$ is the unit normal of line segment i. Θ sweeps the visible viewing directions from point p to the line segment. Only angles within a reflectance threshold are taken to account in order to simulate specular reflection of real range signals. p_i is the intersection point of line segment i and a line emitted from point p along the viewing direction Θ. The constraining relation for $\mathbf{V_i}$ is in essence a projection of the vectors formed from point p to line segment points onto the normal of the line segment. $W(..)$ expresses the reduced probability of observing an object as a function of distance. W is described by an exponential decay function:

$$W(\mathbf{v}) = e^{-k\|\mathbf{v}\|} \tag{13}$$

k is the decay constant that is determined by the range of sensor confidence.

Once the vector influence is computed for all visible line segments, we choose a *reference vector* and determine the total number of components parallel and perpendicular to it. The reference vector defines the two orthogonal basis vectors for the projected components. The projected components of the vector influence are calculated as:

$$C_{\|}(p) = \sum_{i\ lines} \mid \hat{V}_{ref}(p) \bullet \mathbf{V_i}(\mathbf{p}) \mid \tag{14}$$

$$C_{\perp}(p) = \sum_{i\ lines} \|\hat{V}_{ref}(p) \times \mathbf{V_i}(\mathbf{p})\| \tag{15}$$

A good choice for the reference vector is that of largest magnitude vector, since it determines the dominating D.O.F. constraint. A bad choice may lead to inaccurate results, consider a long line segment of slope 1 relative to a basis composed of a minute line segment. The only constraint parallel to the long line segment is the projection from the minute line segment. However, the computed parallel and perpendicular constrains are almost equal (since the slope is 1) resulting in is a high distinctiveness measure. This is undesirable and to avoid it the dominating vector should always be the reference basis.

$C_{\perp}$ and $C_{\|}$ describe the total strength and constraint of all the line segments visible from point p. We then provide a *soft* classifier to discriminate good values from bad ones as a sigmoid re-mapping:

$$C = \frac{C^m}{c^m + C^m} \tag{16}$$

The cutoff threshold c and the decay rate m can be found empirically. The result is a measure ranging from 0-1 where values above .5 can be considered acceptable for localization.

We compute a quality measure Q as:

$$Q(p) = \begin{cases} \frac{C_{\perp}(p)}{C_{\|}(p)} & \text{if } C_{\|} > C_{\perp} \\ \frac{C_{\|}(p)}{C_{\perp}(p)} & \text{otherwise} \end{cases} \tag{17}$$

Q ranges form 0 to 1 where 1 represents equal orthonormal constraints and 0 represents that only one D.O.F. is constraint.

B. Vision System Measure Criteria

In terms of the vision based localization technique used, good areas are those that provide a strong, smooth and varying feature space. Weak features don't provide consistent reliability (indicated by a small value of I in equation 1). A near-constant feature space would lead to positioning ambiguities (indicated by a small value of ΔI in equation 1). A highly discontinuous space would reduce the interpolatory accuracy for position estimation due to lack of information. Furthermore, equal constraining information along both dimensions is desirable to keep the error bound circular (equal localization confidence for both D.O.F.) For the vision system, the distinctiveness measure R at point $p = (x, y)$ about a square neighbourhood $(2\epsilon)^2$ can be calculated as:

$$R(p,\epsilon) = N(p,\epsilon) \int_{y-\epsilon}^{y+\epsilon} \int_{x-\epsilon}^{x+\epsilon} \frac{F(p)(1 + Q_1(p) + Q_2(p))}{3} \delta x \delta y \tag{18}$$

where,

$$F(p) = Min(f_x(p), f_y(p)) \tag{19}$$

$$N(p,\epsilon) = \frac{1}{\int_{y-\epsilon}^{y+\epsilon} \int_{x-\epsilon}^{x+\epsilon} \delta x \delta y} \tag{20}$$

$f_x()$ and $f_y()$ play the role of $f()$ in equation 1 along two orthogonal directions. Q_1 expresses the uniformity of the position constraints in the alternative directions (same as previous quality measure). Q_2 expresses the smoothness of the feature space about a point. For the feature space G, we calculate $f_x()$ and $f_y()$ by the strength and variance of features (in the form of $I * \Delta I$) as:

$$f_j(p) = \sum_{i\ features} \mid G_i(p) * \frac{\delta G_i(p)}{\delta j} \mid \tag{21}$$

where j is any of the two orthogonal D.O.F. (x or y) and $G_i(p)$ is the feature space value at point p for feature i.

A high pass or band pass filter is used to reduce the weight of areas with low slopes or weak features while accepting larger slopes or stronger features. This is in the form of a sigmoid function:

$$f_j = \frac{f_j^m}{c^m + f_j^m} \tag{22}$$

The equality of constraint measure is calculated as follows:

$$Q_1(p) = \begin{cases} \frac{f_x(p)}{f_y(p)} & \text{if } f_y(p) > f_x(p) \\ \frac{f_y(p)}{f_x(p)} & \text{otherwise} \end{cases} \tag{23}$$

A result of 0 shows low equality and a result of 1 shows good equality. We then derive the feature space smoothness measure as:

$$Q_2(p) = \sum_{i\ features} \mid \nabla^2 G_i(p) \mid \tag{24}$$

A low pass filter is used to help evaluate the actual result. This would increase the weight for areas with low values of $\nabla^2 G$ (smooth ones) and reduce the weight for areas with large value of $\nabla^2 G$ (discontinuous ones). Again a sigmoid function is used to re-map the result:

$$Q_2 = 1 - \frac{Q_2^m}{c^m + Q_2^m} \tag{25}$$

VI. Results and Discussion

Testing was performed by comparing localization error and predictions from the distinctiveness measure. A good distinctiveness measure should show large values at areas with low error and low values at high error locations. Comparing the consistency of position confidence and distinctiveness measure confirms our ability estimate where the environment is suitable for localization.

A. Sonar System

For the sonar based method, a line segment model was manually constructed providing the simulated environment shown in figure 3. The distinctiveness measure for the model is plotted in figure 4. These results were obtained with the neighbouring area ϵ set to zero (such that measures consist of only a single point rather that an accumulation of a neighbourhood) and the reflectance threshold set to 30°. We can see how the long hallways show low measures while regions with good orthogonal constraints (such as intersections and bounding areas) lead to high measures. Furthermore, areas distant from line segments are of lower measure due to the exponential decay (the decay constant k was set to 1/200 cm). The sigmoid filter constant was determined by a minimum line segment threshold of 50 cm seen at distance $1/2k$ along the mid-line.

Figure 5 is a plot of the localization confidence. This plot was generated using a robot controller/simulator. At each position, simulated sonar data was collected, thereafter employing a position offset by a random value ranging 10-15 cm. The localization technique was then executed providing the position estimation. The error was the difference between the initial position and the estimated one. Confidence is simply $c - error$ where c is some constant. invis

We can seen how the confidence plot is consistent with the distinctiveness measure. Low confidence valleys match the low measure valleys, where there are not enough constraints. Figure 2 shows the accumulated residual plot between the distinctiveness. There are not many data points with residual greater than 0.4 (about 25% of the data).

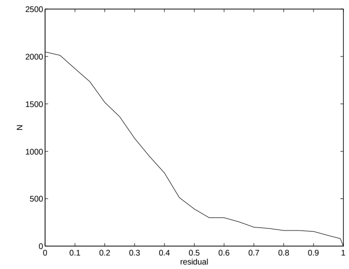


Fig. 2. Residual plot for figure 4 and 5, there are N data points with differences greater than residual.

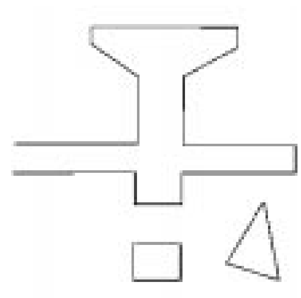


Fig. 3. Simulated environment used for measure evaluation.

B. Vision System

A sample scene was built that consisted of an interesting area, a non-interesting area and an intermediate region (Figure 6). The feature space model was built by extracting data from a pan and tilt camera where only one dimension was used (pan). The experiment was set up to capture an image, extract the features and pan the camera by .5° for the next iteration. An 18-D feature space was built using 17 features corresponding to each pan position. The total scene range was 30°. The distinctiveness measure was employed at each point with the ϵ window width of 10° (an offset of 5° at each end was included). The distinctiveness measure for this configuration is shown in figure 7(a).

The localization errors were determined by the confidence in the outputs of the vision based localizer. Data from the windows were used separately to train the three layer back propagation neural network interpolator. The training confidence for each window is shown in figure 7(b).

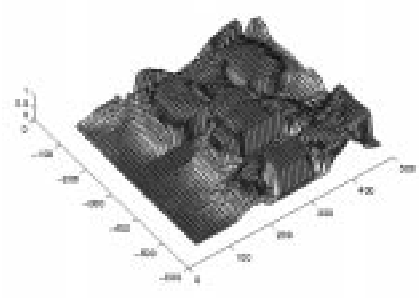


Fig. 4. Sonar based distinctiveness measure for figure 3.

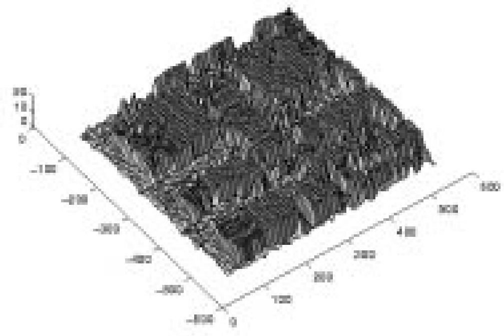


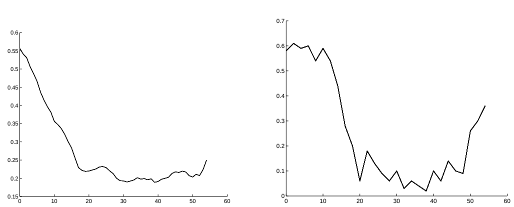
Fig. 5. Localization confidence for figure 3.

Fig. 6. Real environment with real object, room divider (shading is difficult to observe) and background clutter.

The distinctiveness measure and training confidence show similarity over the sample scene. Both the distinctiveness measure and confidence plot are high at areas that were deemed interesting and low at non-interesting areas. The residual plot is shown if figure 8.

A significant difference for both methods is where the localization confidence exhibits sharp drops while the distinctiveness measure undergoes smooth decays. The localization techniques are only accurate within a region of convergence, once the robot moves beyond that region the solution diverges; there is a narrow mid-ground. The analytic distinctiveness measure, on the other hand, is a smooth continuous function.

Keeping this in mind, a mapping threshold can be determined by the intersection point of the distinctiveness measure and the confidence cutoff region. Areas where the measure is less than the threshold are not reliable and should not be mapped but areas that give rise to larger values are good candidates. Furthermore, the filter parameters can be determined more accurately by forming an optimization criterion to minimize the residual.



(a) (b)

Fig. 7. Vision based distinctiveness measure 7(a) and localization confidence 7(b) for figure 6.

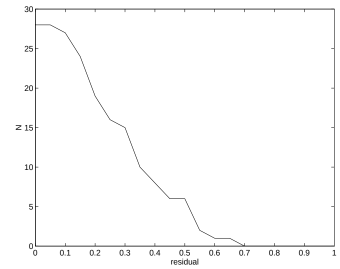


Fig. 8. Residual plot for figure 7(a) and 7(b), there are N data points with differences greater than residual.

VII. Conclusion

This paper describes an important step in the creation of large scale maps that combine both metric and topological knowledge. Specifically, we describe how the locations of individual localization regions or *islands of reliability* can be selected. Our general concept is illustrated using two specific yet very different types of localization procedure: a sonar system and a vision system. The distinctiveness measure for the placement of such islands showed consistency with localization confidence, making it a good measure of environment quality for localization. The paper was put in context to a higher level mapping goal that requires the framework of environment evaluation.

References

[1] Gregory Dudek, Michael Jenkin, Evangelos Milios, and David Wilkes, "A taxonomy for multi-agent robotics", *Autonomous Robots*, vol. 3, pp. 375–397, 1996.

[2] Paul MacKenzie and Gregory Dudek, "Precise positioning using model-based maps", in *IEEE International Conference on Robotics and Automation*, San Diego, California, May 1994, pp. 1615–1621.

[3] G. Dudek and C. Zhang, "Vision based robot localization without explicit object models", in *IEEE Conference on Robotics and Automation*, Minneapolis, MN, May 1996, pp. 76–82.

[4] John J. Leonard and Hugh F. Durrant-Whyte, "Mobile robot localization by tracking geometric beacons", *IEEE Transactions on Robotics and Automation*, vol. 7, no. 3, pp. 376–382, June 1991.

[5] F. Lu and E. Milios, "Optimal global pose estimation for consistent sensor data registration", in *IEEE International Conference on Robotics and Automation*, 1995, pp. 93–100.

[6] R. Smith and P. Cheeseman, "On the representation and estimation of spatial uncertainty", *The international journal of Robotics Research*, vol. 5(4), pp. 56–68, Winter 1986.

[7] W. Yeap, "Towards a computational theory of cognitive maps", *Artificial Intelligence*, vol. 32, pp. 297–360, 1988.

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

[9] Gregory Dudek, Michael Jenkin, Evangelos Milios, and David Wilkes, "Robotic exploration as graph construction", *Transactions on Robotics and Automation*, vol. 7, no. 6, pp. 859–865, December 1991.