Vivek A. Sujan
Steven Dubowsky

Department of Mechanical Engineering
Massachusetts Institute of Technology
Cambridge, MA 02139, USA
vasujan@mit.edu

Efficient Information-based Visual Robotic Mapping in Unstructured Environments

Abstract

In field environments it is often not possible to provide robot teams with detailed a priori environment and task models. In such unstructured environments, robots will need to create a dimensionally accurate three-dimensional geometric model of its surroundings by performing appropriate sensor actions. However, uncertainties in robot locations and sensing limitations/occlusions make this difficult. A new algorithm, based on iterative sensor planning and sensor redundancy, is proposed to build a geometrically consistent dimensional map of the environment for mobile robots that have articulated sensors. The aim is to acquire new information that leads to more detailed and complete knowledge of the environment. The robot(s) is controlled to maximize geometric knowledge gained of its environment using an evaluation function based on Shannon's information theory. Using the measured and Markovian predictions of the unknown environment, an information theory based metric is maximized to determine a robotic agent's next best view (NBV) of the environment. Data collected at this NBV pose are fused using a Kalman filter statistical uncertainty model to the measured environment map. The process continues until the environment mapping process is complete. The work is unique in the application of information theory to enhance the performance of environment sensing robot agents. It may be used by multiple distributed and decentralized sensing agents for efficient and accurate cooperative environment modeling. The algorithm makes no assumptions of the environment structure. Hence, it is robust to robot failure since the environment model being built is not dependent on any single agent frame, but is set in an absolute reference frame. It accounts for sensing uncertainty, robot motion uncertainty, environment model uncertainty and other critical parameters. It allows for regions of higher interest receiving greater attention by the agents. This algorithm is particularly well suited to unstructured environments, where sensor uncertainty and occlusions are significant. Simulations and experiments show the effectiveness of this algorithm.

KEY WORDS—visual mapping, information theory, unstructured environments

1. Introduction

An important goal of robotics research is to develop mobile robot teams that can work cooperatively in unstructured field environments (Baumgartner et al. 1998; Huntsberger, Rodriguez, and Schenker 2000); see Figure 1. Potential tasks include explosive ordinance removal, de-mining and hazardous waste handling, exploration/development of space, environment restoration, and construction (Shaffer and Stentz 1992; Baumgartner et al. 1998; Huntsberger, Rodriguez, and Schenker 2000). For example, space and planetary robotic missions will require robot scouts to lead the way, by exploring, mapping, seeking or extracting soil and rock samples, and eventually constructing facilities in complex terrains. Multiple cooperating robots will be required to set up surface facilities in challenging terrain for *in situ* measurements, communications, and to pave the way for human exploration of planetary surfaces. This will require the handling of relatively large objects, such as deploying of solar panels and sensor arrays, anchoring of deployed structures, movement of rocks, and clearing of terrain.

The control of such systems typically requires models of the environment and task. In unstructured field environments it is often not possible to have such a priori models. In such cases, the robot needs to construct these from sensory information usually from vision systems. A number of problems can make this non-trivial. These include the uncertainty of the

The International Journal of Robotics Research
Vol. 24, No. 4, April 2005, pp. 275-293,
DOI: 10.1177/0278364905051774

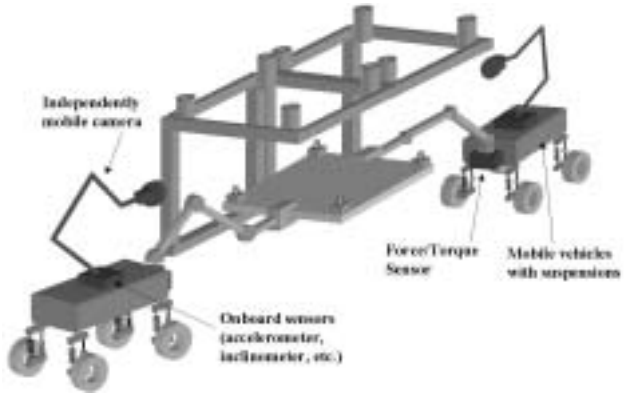


Fig. 1. Representative physical system.

task in the environment, location and orientation uncertainty in the individual robots, and occlusions (due to obstacles, work piece, other robots). If the systems are equipped with cameras mounted on articulated mounts, intelligent planning of the camera motion can alleviate problems of the occlusions, providing an accurate geometrical model of the task and environment. If the system consists of more than one robot, planning the behavior of these multi-information sharing systems can further improve the system performance.

Environment mapping by mobile robots falls into the category of simultaneous localization and mapping (SLAM). In SLAM a robot is localizing itself as it maps the environment. Researchers have addressed this problem for well-structured (indoor) environments and have obtained important results (Asada 1990; Kuipers and Byun 1991; Leonard and Durrant-Whyte 1991; Pito and Bajcsy 1995; Kruse, Gutsche, and Wahl 1996; Burschka, Eberst, and Robl 1997; Castellanos et al. 1998; Thrun, Fox, and Burgard 1998; Yamauchi, Schultz, and Adams 1998; Anousaki and Kyriakopoulos 1999; Victorino, Rives, and Borrelly 2000; Choset and Nagatani 2001; Tomatis, Nourbakhsh, and Siegwar 2001). These algorithms have been implemented for several different sensing methods, such as camera vision systems (Pito and Bajcsy 1995; Hager et al. 1997; Castellanos et al. 1998; Park, Jiang, and Neumann 1999), laser range sensors (Yamauchi, Schultz, and Adams 1998; Tomatis, Nourbakhsh, and Siegwar 2001), and ultrasonic sensors (Leonard and Durrant-Whyte 1991; Yamauchi 1998; Anousaki and Kyriakopoulos 1999; Choset and Nagatani 2001). Sensor movement/placement is usually performed sequentially (raster scan type approach), by following topological graphs or using a variety of "greedy" algorithms that explore regions only on the extreme edges of the known environment (Kuipers and Byun 1991; Leonard and Durrant-Whyte 1991; Yamauchi, Schultz, and Adams 1998; Anousaki and Kyriakopoulos 1999; Rekleitis, Dudek, and Milios 2000; Victorino, Rives, and Borrelly 2000; Choset and Nagatani 2001). Geometric descriptions of the environment are modeled in several ways, including generalized cones, graph models and Voronoi diagrams, occupancy grid models, segment models, vertex models, convex polygon models (Kuipers and Byun 1991; Choset and Nagatani 2001). The focus of these works is accurate mapping. They do not address mapping efficiency. Researchers have addressed mapping efficiency to a limited amount (Kruse, Gutsche, and Wahl 1996). However, sensing and motion uncertainties are not taken into account. They also generally assume that the environment is effectively flat (e.g., the floor of an office or a corridor) and readily traversable (i.e., obstacles always have a route around them; Lumelsky, Mukhopadhyay, and Sun 1989; Kuipers and Byun 1991; Yamauchi, Schultz, and Adams 1998; Anousaki and Kyriakopoulos 1999; Thrun, Burgard, and Fox 2000; Choset and Nagatani 2001) and have not been applied to robot teams working in rough planetary environments. Also, prior work has not addressed optimizing the communication between agents for both multi-agent planning and cooperative map building.

To achieve the localization function, landmarks and their relative motions are monitored with respect to the vision systems. Several localization schemes have been implemented, including topological methods such as generalized Voronoi graphs and global topological maps (Kuipers and Byun 1991; Victorino, Rives, and Borrelly 2000; Choset and Nagatani 2001; Tomatis, Nourbakhsh, and Siegwar 2001), extended Kalman filters (Leonard and Durrant-Whyte 1991; Anousaki and Kyriakopoulos 1999; Park, Jiang, and Neumann 1999), Bayesian filters (Howard, Matarić, and Sukhatme 2003), and robust averages (Park, Jiang, and Neumann 1999). Although novel natural landmark selection methods have been proposed (Yeh and Kriegman 1995; Hager et al. 1997; Simhon and Dudek 1998), most SLAM architectures rely on identifying landmarks as corners or edges in the environment (Kuipers and Byun 1991; Leonard and Durrant-Whyte 1991; Castellanos et al. 1998; Anousaki and Kyriakopoulos 1999; Victorino, Rives, and Borrelly 2000; Choset and Nagatani 2001). This often limits the algorithms to structured indoor-type environments. Others have used human intervention to identify landmarks (Thrun, Burgard, and Fox 2000).

Some studies have considered cooperative robot mapping of the environment (Thrun, Fox, and Burgard 1998; Yamauchi 1998; Jennings, Murray, and Little 1999; Rekleitis, Dudek, and Milios 2000). Novel methods of establishing/identifying landmarks, identifying relative robot poses and dealing with cyclic environments have been introduced for indoor environments (Thrun, Fox, and Burgard 1998; Jennings, Murray, and Little 1999; Howard, Matarić, and Sukhatme 2003). In some cases, observing robot team members as references to develop accurate maps are required (Rekleitis, Dudek, and Milios 2000). While the work done in this field has had significant impact on robot control architectures, these results largely do not address the problem of cooperative sensing in

the context of mobile robots in unknown, unstructured environments. In simple terms, the conventional approach for the control of robotic systems is to use sensory information as input to the control algorithms. System models are then used to determine control commands. The methods developed to date generally rely on assumptions that include simple well-known terrains, accurate knowledge of the environment, little or no task uncertainty, sufficient sensing capability, and sufficient communication capabilities. For real field environments, a number of these assumptions are often not valid. In general, current research has not solved the problem of controlling multiple mobile robots performing cooperative tasks in unknown environments, where uncertainty, limited sensing capabilities, and incomplete physical models of the system(s)/environment dominate the problem.

In this paper we propose an environment and task model building algorithm based on information theory. The objective is to efficiently build a geometrically consistent dimensional model of the environment and target, available to the robot(s), to allow it (them) to perform its (their) tasks, while overcoming the uncertainties in robot and camera location and orientation, for robots or robot teams cooperatively working in an unstructured field environment. It is assumed that the system consists of one (or more) mobile robots working in an unknown environment (such as constructing a planetary structure; see Figure 2). There are no physical interactions between the robots. The target is static. Each has a three-dimensional (3D) vision system mounted on an articulated arm. Sensing and sensor placement are limited, resulting in occlusions and uncertainties. It is assumed that dimensional geometric information is relevant and required for robots to perform their tasks, such as the construction of field facilities. The robot(s) is controlled to maximize geometric knowledge gained of its environment using an evaluation function based on Shannon's information theory. Using the measured and Markovian predictions of the unknown environment, an information theory based metric is maximized to determine the robotic agent's next best view (NBV) of the environment. Data collected at this NBV pose are fused using a Kalman filter statistical uncertainty model to the measured environment map. The process continues until the environment mapping process is complete. For multi-agent teams, each robot may share its model with the other agents providing knowledge of regions not visible by all robots (Sujan et al. 2003). For such scenarios, the NBV is determined using the entire fused map available to each agent. Each agent must account for the position of all the other agents in determining its NBV to avoid collisions. The details of the algorithm are presented next.

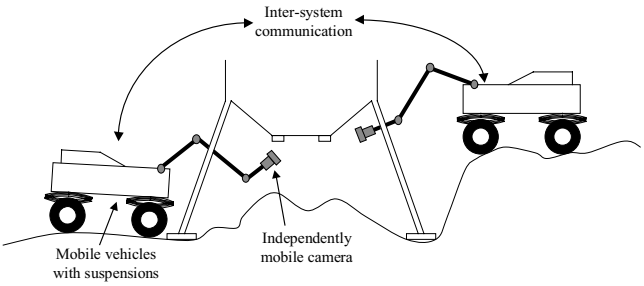


Fig. 2. Cooperative mapping by robots.

2. Algorithm Description

2.1. Overview

The first step in environment map building is to visually construct a model of the local environment, including the locations of the task elements and other robots (if necessary). The 3D map is modeled as a probabilistic discretized occupancy grid. Every voxel in the map has a value for probability-of-occupancy that ranges from 0 (empty) to 1 (occupied). A value of 0.5 indicates maximum uncertainty in occupancy of the voxel. We assume that only the geometry of the task elements (such as parts of a solar panel that is to be assembled; Huntsberger, Rodriguez, and Schenker 2000) and the robot(s) are well known. Obstacles and robot positions are unknown until measured/mapped.

The algorithm may be broken down as follows. First, the articulated vision sensors cooperatively scan the region around a target, generating a local 3D geometric model. This allows the robots to locate themselves and the obstacles in the target reference frame. This is achieved by "looking around" and matching the known target element geometric CAD model with visual data (Lara, Althoefer, and Seneviratne 1998). Next, these models are used to find optimum environment viewing poses for the multiple vision systems (NBV), by defining and optimizing a rating function (RF) over the possible sensor positions, subject to kinematic constraints of the sensor placement mechanisms for the individual robots. This RF aims to acquire as much new information (NI) about the environment as possible with every sensing cycle, while maintaining or improving the map accuracy, and minimizing the exploration time. The process is constrained by selecting goal points that are not occluded and that can be reached by a collision-free traversable path. The sensors then move to their new poses and acquire 3D data. Based on the sensor mount kinematics, the motion of the sensor is known. However, small motions of the robot base (due to suspension compliance) and errors in sensor mounts lead to additional uncertainties. These are accounted for by measuring common features during the vision sensor motion. Finally, the new data and their associated uncertainties are fused with the current environment map, resulting in an updated probabilistic environment map, which may then be shared with the other sensing agents. Each agent uses the updated probabilistic environment map to generate its NBV. The process continues until the environment is mapped to the predefined extent necessary. In general, for each sens-

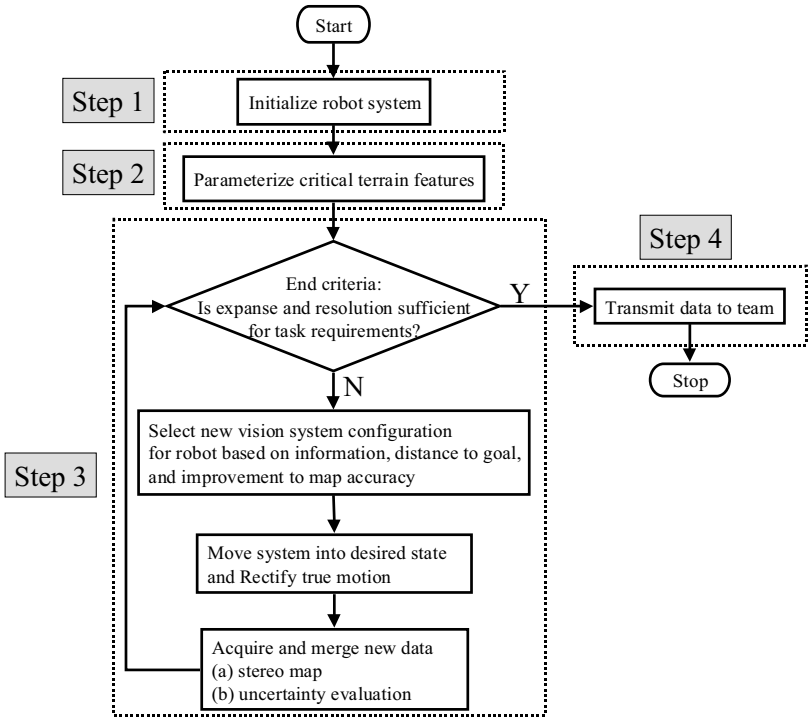


Fig. 3. Outline of model building and placement algorithm.

ing agent the algorithm consists of four steps, described as follows (see Figure 3; Sujan et al. 2003).

2.2. System Initialization

As described above, a common target must be located to establish a common inertial reference frame between the robots and the environment. All robots contributing to or requiring use of the map are localized with respect to the initial map. Localization may be achieved by either of the following:

(a) absolute localization, which is achieved by mapping a common environment landmark that is visible by all robots;

(b) relative localization, which is performed by mapping fiducials on all robots by other robot team members where one robot is selected as the origin (Howard, Matarić, and Sukhatme 2003).

Absolute localization is used in this application, although relative localization has been applied in the past (Howard, Matarić, and Sukhatme 2003; Sujan et al. 2003). Searching for the target (which forms the absolute origin) by moving the robot sensors can be carried out in many ways (Luo and Kay 1989; Tarabanis, Allen, and Tsai 1995). In this study, sensor placement for target searching is performed in the same way as sensor placement for environment model building. The absolute origin target is located by matching the known target element geometric CAD model with visual data (Lara, Althoefer, and Seneviratne 1998). Template matching for key fiducials in matching a CAD model would take $O(n^6)$ time for each fiducial (where n is the number of mapped points). However, this operation is only performed once during the entire process resulting in a one-time computational load. At this stage, the environment model is considered empty (i.e., no points are known). The first stereo range map (including the common target and all objects within the field of view) is taken by each agent. We assume that only the geometry of the task elements (such as parts of a solar panel that is to be assembled; Huntsberger, Rodriguez, and Schenker 2000) and the robot(s) are well known. The first stereo range map (including the common target and all objects within the field of view) is taken and fused to the model. This is described next.

2.3. Data Modeling and Fusion

At any time, the cameras on each mobile robot can only observe a small part of their environment. However, measurements obtained from multiple viewpoints can provide reduced uncertainty, improved accuracy, and increased tolerance in estimating the location of observed objects (Smith and Cheeseman 1986). To fuse multiple range measurements of a

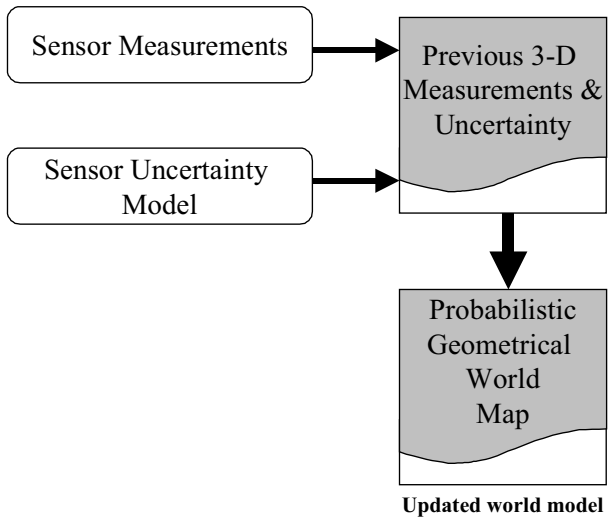


Fig. 4. 3D range measurement fusion with sensor uncertainty.

feature by sensors, a statistical model of sensor uncertainty is employed (see Figure 4). Current and previous range sensor measurements and their uncertainty models can be integrated to give an updated probabilistic geometric model of the environment.

A single range observation of a point ($\bar{x}$) is modeled as a 3D Gaussian probability distribution centered at $\bar{x}$, based on two important observations. First, the use of the mean and covariance of a probability distribution function is a reasonable form to model sensor data and is a second-order linear approximation (Smith and Cheeseman 1986). This linear approximation corresponds to the use of a Gaussian (having all higher moments of zero). Secondly, from the central limit theorem, the sum of a number of independent variables has a Gaussian distribution regardless of their individual distributions. The standard deviations along the three axes of the distribution correspond to estimates of the uncertainty in the range observation along these axes. These standard deviations are a function of intrinsic sensor parameters (such as camera lens shape accuracy) as well as extrinsic sensor parameters (such as the distance to the observed point or feature).

For most range sensing systems, this model can be approximated as (Sujan and Dubowsky 2002)

$$\begin{aligned}\sigma_{x,y,z} &= f\ (\text{extrinsic parameters, intrinsic parameters}) \\ &\approx S \cdot T_{x,y,z} \cdot L^n\end{aligned} \tag{1}$$

where S is an intrinsic parameter uncertainty constant, $T_{x,y,z}$ is an extrinsic parameter uncertainty constant, L is the distance to the feature/environment point, and n is a constant (typically 2). Provided two observations are drawn from a normal distribution, the observations can be merged into an improved estimate by multiplying the distributions. Because the result of multiplying two Gaussian distributions is another Gaussian distribution, the operation is symmetric, associative, and can be used to combine any number of distributions in any order. The canonical form of the Gaussian distribution in n dimensions depends on the standard distributions, $\sigma_{x,y,z}$, a covariance matrix (C) and the mean ($\bar{x}$) (Smith and Cheeseman 1986; Stroupe, Martin, and Balch 2000):

$$p(\bar{x}'|\bar{y}) = \frac{1}{(2\pi)^{n/2}\sqrt{|C|}} \exp\left(-\frac{1}{2}\left(\bar{y}-\bar{x}'\right)^{\mathrm{T}} C^{-1}\left(\bar{y}-\bar{x}'\right)\right)$$

$$\text{where } C = \begin{bmatrix} \sigma_x^2 & \rho_{xy}\sigma_{xy}\sigma_{xy} & \rho_{zx}\sigma_{zx}\sigma_{zx} \\ \rho_{xy}\sigma_{xy}\sigma_{xy} & \sigma_y^2 & \rho_{yz}\sigma_{yz}\sigma_{yz} \\ \rho_{zx}\sigma_{zx}\sigma_{zx} & \rho_{yz}\sigma_{yz}\sigma_{yz} & \sigma_z^2 \end{bmatrix} . \tag{2}$$

Here, the exponent is called the Mahalanobis distance. For uncorrelated measured data $\rho = 0$. The formulation in eq. (2) is in the spatial coordinate frame. However, all measurements are made in the camera (or sensor) coordinate frame. This problem is addressed through a transformation of parameters from the observation frame to the spatial reference frame as follows

$$C_{transformed} = R(-\bar{\theta})^{\mathrm{T}} \cdot C \cdot R(-\bar{\theta}) \tag{3}$$

where $R(\theta)$ is the rotation matrix between the two coordinate frames. The angle of the resulting principal axis can be obtained from the merged covariance matrix

$$C_{merged} = C_1\left(I - C_1\left(C_1 + C_2\right)^{-1}\right) \tag{4}$$

where C_i is the covariance matrix associated with the ith measurement. Additionally, a translation operation is applied to the result from eq. (2), to bring the result into the spatial reference frame. Determining this transformation matrix is described in Section 2.5. To contribute to the probabilistic occupancy environment model, all measured points corresponding to obstacles are merged. That is, all measured points falling in a particular grid voxel contribute to the error analysis associated with that voxel. Note that grid voxels falling within the field of view of the vision system that correspond to empty space result in no contribution to the uncertainty model (since these are not measured). However, these points are tabulated as measured (or known) points. This will be used to select the next pose for the vision system. The data fusion process is summarized in Figure 5.

2.4. Vision System Pose Selection

A RF is used to determine the next pose of the camera from which to look at the unknown environment. The aim is to acquire as much NI about the environment as possible with every sensing cycle, while maintaining or improving the map accuracy, and minimizing the exploration time. The process is constrained by selecting goal points that are not occluded and

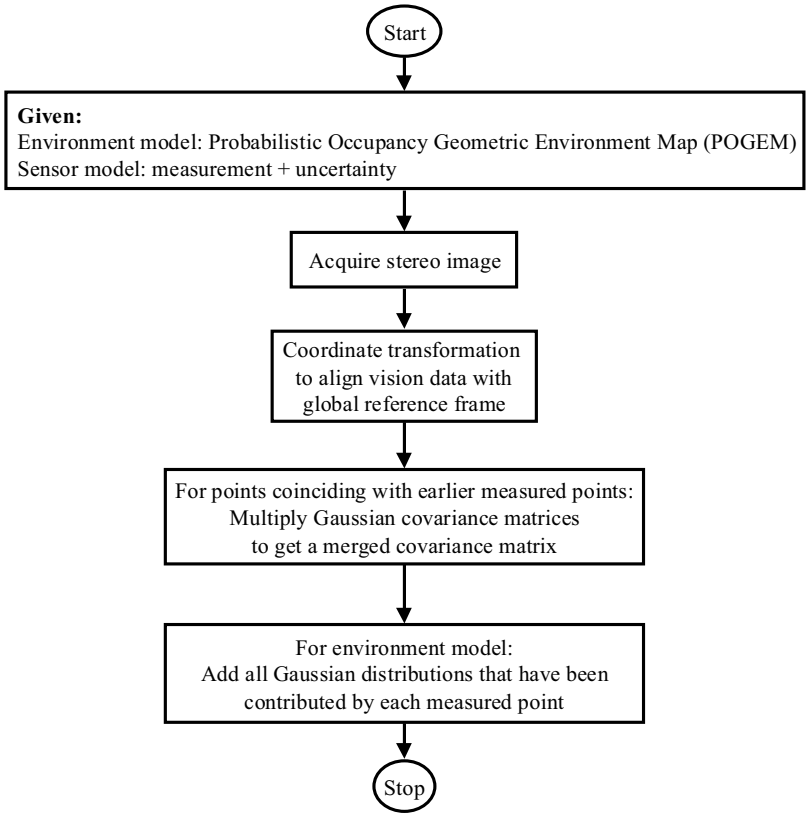


Fig. 5. Flowchart for data fusion using known vision system motion.

that can be reached by a collision-free traversable path. From the probabilistic geometric environment model, (x, y, z) locations with $P_{x,y,z} < P_{empty} = 0.05$ (2σ) are considered as unoccupied. Such points form candidate configuration space camera pose coordinates.

The NI is equal to the expected information of the unknown/partially known region viewed from the camera pose under consideration. This is based on the known obstacles from the current environment model, the field of view of the camera (see Figure 6) and a framework for quantifying information. Shannon (1948) showed that the information gained by observing a specific event among an ensemble of possible events may be described by the following function

$$H\,(q_1, q_2, ..., q_n) = -\sum_{k=1}^{n} q_k \log_2 q_k \tag{5}$$

where q_k represents the probability of occurrence for the kth event. This definition of information may also be interpreted as the minimum number of states (bits) needed to fully describe a piece of data. Shannon's emphasis was in describing the information content of one-dimensional signals. In two dimensions the gray-level histogram of an ergodic image can be used to define a probability distribution

$$q_i = f_i/N \text{ for } i = 1...N_{gray} \tag{6}$$

where f_i is the number of pixels in the image with gray-level i, N is the total number of pixels in the image, and N_{gray} is the number of possible gray levels. With this definition, the information of an image for which all the q_i are the same—corresponding to a uniform gray-level distribution or maximum contrast—is a maximum. The less uniform the histogram, the lower the information.

It is possible to extend this idea of information to a 3D signal—the environment model. In such an instance, the scene probability distribution for information analysis is still given by eq. (6). However, N is the maximum number of voxels visible by the vision system (limited by the depth of field and the field of view), and f_i is the number of voxels in the scene with gray-level i. The equation is evaluated separately for mapped verses unmapped regions:

$$H\,(q) = -\left(\left(\sum_{k=1}^{n} q_k \log_2 q_k\right)_{known} + \left(\sum_{k=1}^{n} q_k \log_2 q_k\right)_{unknown}\right). \tag{7}$$

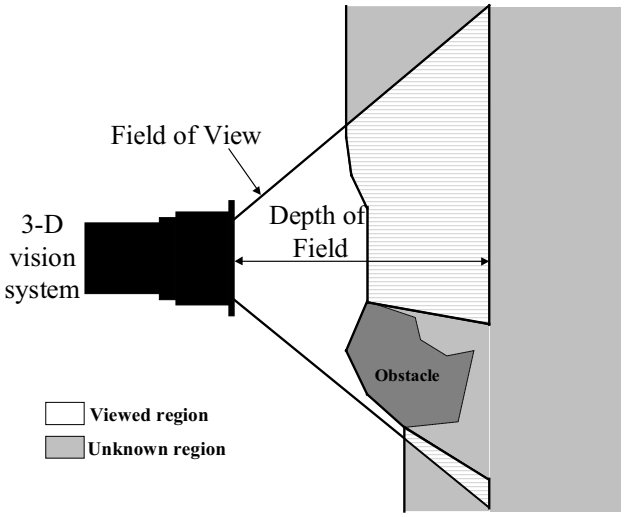


Fig. 6. Evaluation of expected NI.

The possible gray-level values are defined as follows. For all unknown/unsampled voxels, an occupancy value $p(\bar{x})_{unknown}$ may be defined in the form of a Markovian chain, i.e., $p(\bar{x})$ of a particular voxel is the average value of $p(\bar{x})$ of the neighboring voxels. Intuitively, this results in unknown regions that are mapped as averages of closest known regions. Thus, for all spatial voxels, a gray (probabilistic) occupancy value between 0 and 1 is found. Next, the values for $p(\bar{x})$ are modified as follows

$$\text{stretching: } p'(\bar{x}) = \begin{cases} \frac{1}{1-p(\bar{x})} \cdot \frac{1}{d_{voxel}} \ \forall\ p(\bar{x}) < 0.5 \\ \frac{1}{p(\bar{x})} \cdot \frac{1}{d_{voxel}} \ \forall\ p(\bar{x}) \geqslant 0.5 \end{cases}$$

$$+ \text{ scaling: } p''(\bar{x}) = \begin{cases} \frac{p'(\bar{x})-1}{2} \ \forall\ p(\bar{x}) < 0.5 \\ 1 - \frac{p'(\bar{x})-1}{2} \ \forall\ p(\bar{x}) \geqslant 0.5 \end{cases} \quad (8)$$

where d_{voxel} is the Euclidean distance of the voxel from the camera coordinate frame. This process causes regions with probability densities closer to 0 or 1 (regions of most certainty) to have a reduced effect on the NI expected. Regions that have a probability density closer to 0.5 (regions of least certainty of occupancy) are "stretched out" in the scene probability distribution, thus increasing the new expected information associated with those regions. A uniform discretization of this range of $p''(\bar{x})$ values may be performed to define the gray-level values i, f_i, N_{gray} and N (eq. (7)). With these definitions, q_k (eq. (6)) is evaluated and the results applied to eq. (8) resulting in a metric for NI. Alternatively, and possibly a better choice is a uniform discretization of $p(\bar{x})$ to define the gray-level values i, f_i, N_{gray} and N. To increase the contribution of regions with higher occupancy uncertainty to the information metric, the term $q_k \log_2 q_k$ of eq. (7) is premultiplied by $-\left(p_k \log_2 p_k + (1-p_k)\log_2(1-p_k)\right)$ reflecting the greater expected information available in such regions:

$$H(q) = -\left(\begin{array}{l} \left(\sum_{k=1}^{n} \left(-\left(p_k \log_2 p_k + (1-p_k) \right. \right. \right. \\ \left. \left. \left. \log_2 (1-p_k) \right) \right) \left(q_k \log_2 q_k \right) \right)_{known} \\ + \left(\sum_{k=1}^{n} \left(-\left(p_k \log_2 p_k + (1-p_k) \right. \right. \right. \\ \left. \left. \left. \log_2 (1-p_k) \right) \right) \left(q_k \log_2 q_k \right) \right)_{unknown} \end{array} \right). \quad (9)$$

This definition for NI does behave in an intuitively correct form. For example, for a given camera pose, if the field of view is occluded, then NI decreases. If every point in the field of view is known and is empty then NI = 0. NI increases as the number of unknowns in the field of view increases. Further, eq. (9) increases the NI expected with regions that are known with median probabilistic values (i.e., values that indicate with least amount of certainty whether a voxel is occupied or not). On the other hand, regions with high probabilistic values for occupancy result in reduced associated information.

In addition to maximizing information acquisition, it is also desirable to minimize travel distance and maintain/improve the map accuracy, while being constrained to move along paths that are not occluded. A Euclidean metric in configuration space, with individual weights α_i on each degree of freedom of the camera pose $\bar{c}$, is used to define the distance moved by the camera

$$d = \left(\sum_{i=1}^{n} \alpha_i (c_i - c_i')^2 \right)^{1/2}, \quad (10)$$

where $\bar{c}$ and $\bar{c}'$ are vectors of the new and current camera poses, respectively. Although in this paper α_i is set to unity, for general systems this parameter reflects the ease/difficulty in moving the vision system in the respective axis. Map accuracy is based on the accuracy of localization of each sensing agent. This may be obtained by adding the localization error of the agent along the path to the target (LE). Paths containing more promising fiducials for localization result in higher utility in determining both the goal location and the path to the goal. The NI, the travel distance, and the net improvement of map accuracy are combined into a single utility function that may be optimized to select the next view pose

$$\text{RF} = (w_{NI} \cdot NI - w_d \cdot d - w_{LE} \cdot LE) \cdot (1 - P_{x,y,z}), \quad (11)$$

where w_{NI}, w_d, and w_{LE} are scaling constants. This RF can be evaluated and optimized to find the next camera configuration from which to make future measurements of the environment. This choice for RF is a simple standard weighted linear combination. Other constraints (application-dependent) may be readily integrated. For example, when multiple robots are

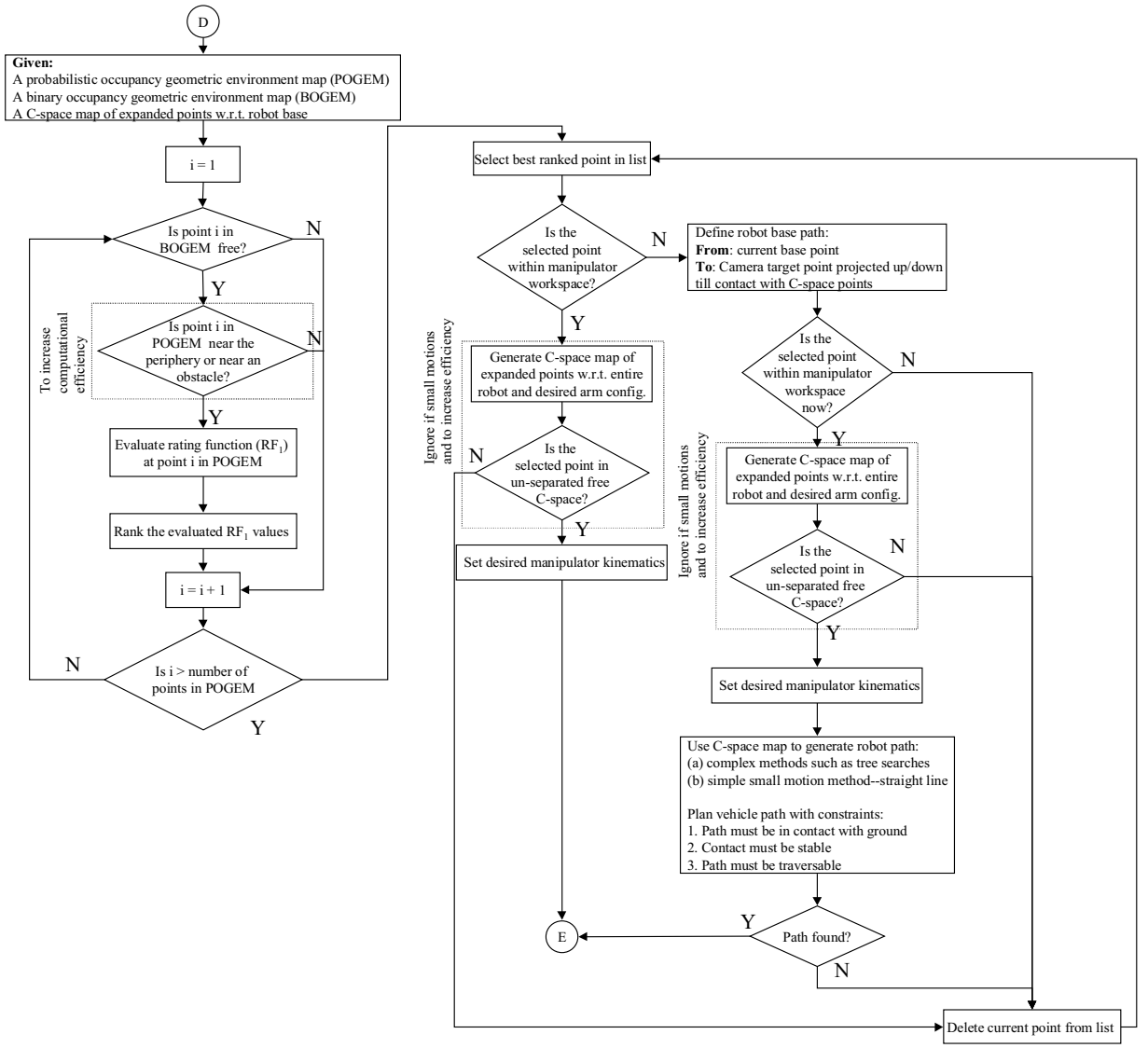


Fig. 7. Flowchart for vision system pose selection of environment mapping algorithm.

cooperatively mapping the environment, a weighted dependence on proximity to another robot can be added to prevent collisions. Selecting an optimum form for RF may not be possible, as an infinite array of more complex linear and nonlinear combinations can be made. However, the standard linear combination yields promising results, with significant improvements in mapping efficiency over conventional methods. The vision system pose selection algorithm is outlined in Figure 7. Note that the movement of the vision system may require motions by the mobile robot (in addition to manipulator motions). The flowchart in Figure 7 includes a simple path planning approach based on the principle of convex hulls (Sujan et al. 2003).

2.5. Camera Motion Correction

A final step in environment map building is to identify the motion of the camera. This process eliminates manipulator end-point positioning errors and vehicle suspension motions, and allows for accurate data fusion. A single spatial point in the reference frame, $\bar{r}_i$, is related to the image point (u_i, v_i) by the 4×4 transformation matrix $\mathbf{g}_{01}$ (see Figure 8).

For motion calibration, we need to identify $\mathbf{g}_{01}$:

$$\begin{bmatrix} k_i u_i \\ k_i v_i \\ k_i f \\ 1 \end{bmatrix} = \mathbf{g}_{01} \cdot \bar{r}_i = \begin{bmatrix} [\mathbf{R_{01}}]_{3x3} & \bar{X}_{3x1} \\ \bar{0} & 1 \end{bmatrix} \cdot \begin{bmatrix} r_i^x \\ r_i^y \\ r_i^z \\ 1 \end{bmatrix}. \quad (12)$$

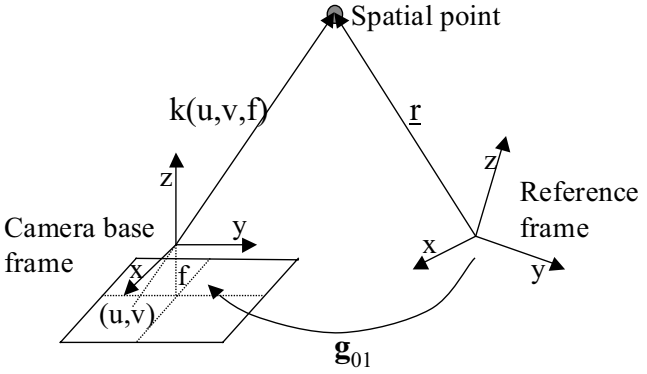


Fig. 8. Relationship of camera and target frames.

Here, $\mathbf{R}_{01}$ is the rotational matrix, $\bar{X}$ is the translation vector, f is the camera focal length, and k_i is a scaling constant. For computational reasons it is more convenient to treat the nine rotational components of $\mathbf{R}_{01}$ as independent, rather than a trignometric relation of three independent parameters. Each spatial point gives three algebraic equations, but also introduces a new variable, k_i—a multiplicative constant to extend the ith image point vector $(u, v, f)_i$ to the ith spatial point in the camera coordinate frame. k_i may be found from the disparity pair of the stereo images. For n points we have

$$\mathbf{u} = \mathbf{g}_{01}\mathbf{r} \Rightarrow \begin{bmatrix} k_1u_1 & k_2u_2 & \cdots & k_nu_n \\ k_1v_1 & k_2v_2 & \cdots & k_nv_n \\ k_1f & k_2f & \cdots & k_nf \\ 1 & 1 & \cdots & 1 \end{bmatrix} = \mathbf{g}_{01}\begin{bmatrix} r_1^x & r_2^x & \cdots & r_n^x \\ r_1^y & r_2^y & \cdots & r_n^y \\ r_1^z & r_2^z & \cdots & r_n^z \\ 1 & 1 & \cdots & 1 \end{bmatrix}. \quad (13)$$

This set of linear equations can be readily solved using the least-mean-square error solution:

$$\mathbf{g_{01}} = \mathbf{u}\left(\mathbf{r}^{\mathrm{T}}\mathbf{r}\right)^{-1}\mathbf{r}^{\mathrm{T}}. \quad (14)$$

The rotation matrix, $\mathbf{R}_{01}$, and the translation vector, $\bar{X}$, of the camera frame with respect to the base frame are extracted directly from this solution of $\mathbf{g}_{01}$. However, for real measured data and associated uncertainty, a larger number of spatial points are required to more correctly identify the geometric transformation matrix, $\mathbf{g}_{01}$. Given the $(i+1)$ spatial and image point, from eq. (14), $\boldsymbol{R}_{i+1}$ and $\bar{X}_{i+1}$ can be obtained. A recursive method can be used to determine the mean and covariance of $\bar{X}$ and $\mathbf{R}_{01}$ based on the previous i measurements as follows:

$$\begin{aligned} \hat{\bar{X}}_{i+1} &= \frac{\left(i\hat{\bar{X}}_i + \bar{X}_{i+1}\right)}{i+1} \\ C_{i+1}^{\bar{X}} &= \frac{iC_i^{\bar{X}} + \left[\bar{X}_{i+1} - \hat{\bar{X}}_{i+1}\right]\left[\bar{X}_{i+1} - \hat{\bar{X}}_{i+1}\right]^{\mathrm{T}}}{i+1} \\ \hat{R}_{i+1}^{(l,m)} &= \frac{\left(i\hat{R}_i^{(l,m)} + R_{i+1}^{(l,m)}\right)}{i+1} \\ C_{i+1}^{R(l,m)} &= \frac{iC_i^{R(l,m)} + \left[R_{i+1}^{(l,m)} - \hat{R}_{i+1}^{(l,m)}\right]\left[R_{i+1}^{(l,m)} - \hat{R}_{i+1}^{(l,m)}\right]^{\mathrm{T}}}{i+1}. \end{aligned} \quad (15)$$

This essentially maintains a measure on how certain the camera motion is with respect to its original configuration (assuming the original configuration is known very precisely with respect to the common reference frame). This camera pose uncertainty must be accounted for to obtain an estimate on the position uncertainty of a measured point in the environment. Let the measurement $\bar{z}$ be related to the state vector (actual point position) $\bar{x}$ by a non-linear function, $h(\bar{x})$. The measurement vector is corrupted by a sensor noise vector $\bar{v}$ of known covariance matrix, $\mathbf{R}$:

$$\bar{z} = h(\bar{x}) + \bar{v}. \quad (16)$$

Assume that the measurement of the state vector $\bar{x}$ is performed multiple times. In terms of the current measurement, a Jacobian matrix of the measurement relationship evaluated at the current state estimate is defined as

$$H_k = \left.\frac{\partial h(\bar{x})}{\partial \bar{x}}\right|_{\bar{x}=\bar{x}_k}. \quad (17)$$

The state (or positition) may then be estimated as follows:

$$\begin{aligned} K_k &= P_kH_k^{\mathrm{T}}\left[H_kP_kH_k^{\mathrm{T}} + R_k\right]^{-1} \\ \hat{\bar{x}}_{k+1} &= \hat{\bar{x}}_k + K_k\left[\bar{z}_k - h(\bar{x}_k)\right] \\ P_{k+1} &= \left[1 - K_kH_k\right]P_k. \end{aligned} \quad (18)$$

This estimate is known as the extended Kalman filter (Gelb 1974). Using this updated value for both the measured point $\bar{x}$ and the absolute uncertainty $\mathbf{P}$, the measured point may then be merged with the current envrionment model using eqs. (2) and (4). Note that combining noisy measurements leads to a noisier result. For example, the camera pose uncertainty increases as the number of camera steps increase. With every new step, the current uncertainty is merged with the previous uncertainty to obtain an absolute uncertainty in camera pose. However, merging redundant measurements (filtering) leads to a less noisy result (e.g., the environment point measurements).

Obtaining appropriate spatial points is now addressed. Spatial points are a visible set of fiducials that are tracked during camera motion. As the camera moves, the fiducials move relative to the camera, eventually moving out of the camera

view. This requires methods to identify and track new fiducials. Fiducials are selected from the probabilistic environment model based on three criteria: the degree of certainty with which a sampled point is known, the visual contrast of the sampled point with its surroundings, and depth contrast of the sampled point with its surroundings. These are combined into a single fiducial evaluation function:

$$\text{F.E.F.} = f\,(P\,(x)) + g\,(C\,(u, v)) + h\,(H\,(x))\,. \qquad (19)$$

Fiducial certainty: $f\,(P(\boldsymbol{x})) \sim P(\boldsymbol{x})/r$, where r is the radius of a sphere centered at the potential fiducial within which neighboring voxels have decending certainty levels. Outside this sphere, voxel certainty levels increase. Lower values for r suggest that the region surrounding a potential fiducial is well known, which is a desirable property.

Fiducial visual contrast: $g(C(u,v)) \sim$ contrast (C) * window size (w). Contrast is defined as

$$C(u, v) = \frac{I(x) - \bar{I}_w}{\bar{I}_w} \qquad (20)$$

where $I(\boldsymbol{x})$ is the two-dimensional image intensity value of the potential fiducial at $\boldsymbol{x}$, $\bar{I}_w$ is the average intensity of a window centered at the potential fiducial in the two-dimensional image, and w is the maximum window size after which the contrast starts to decrease.

Fiducial depth contrast: $h(H(\boldsymbol{x})) \sim H(\boldsymbol{x})$ * window size (w), where $H(\boldsymbol{x})$ is the maximum spatial frequency (from a 3D Fourier transform) at the potential fiducial within a window, and w is the maximum window size after which the power spectrum (of the 3D Fourier transform) starts shifting to higher frequencies. To simplify computation, this may be approximated with some heuristics.

Additionally, a penalty is added if a potential fiducial is too close to other identified fiducials. Using the identified fiducials, camera motion can be identified. Fiducials can be tracked with simple methods such as region growing or image disparity correspondence. This algorithm is summarized in Figure 9.

3. Results

3.1. Simulation Studies

Here we give the results using the RF for vision system pose selection to develop a probabilistic model of a planar environment. Two simulation results are presented: single camera/robot modeling of an unstructured environment and two cooperative cameras/robots modeling of an unstructured environment. Five camera pose selection methods are compared:

1. *Random pose selection*—the next camera pose is selected randomly within the known environment;
2. *Sequential/raster pose selection*—the next camera pose is selected as the next free location in the known environment from which measurements have not yet been made;
3. *Pose with maximum expected unmapped (new) region*—the next camera pose is selected as the location with the largest expected new region while accounting for known occlusions;
4. *Pose with minimum mapped (old) region (also known as the Frontier strategy)*—the next camera pose is selected as the location that will map the smallest previously mapped region;
5. *Pose with maximum expected information.*

The RF cannot be optimized analytically. In practice, finding an optimum value for the RF requires exhaustive searching although the entire n-point configuration space takes $0(n)$ time. This process to determine the true maxima may rapidly become intractable (as was evident in some of our initial tests). There are several methods around this. For example, a set of uniformly distributed randomly placed candidate points in the environment may be selected and evaluated which is representative of the "behavior" of points in its neighborhood. Clearly, end-effector positioning errors will prohibit us from truly reaching the exact point, and hence it would be more valuable to evaluate a representative point in a region rather than every point in the region. Determining region size or the number of points to evaluate is currently a topic of research. A small twist to this selection process is to employ a finite set (m-points) of goal configurations located uniformly but randomly near the current position. This process takes $0(m)$ time, where m is a constant. Thus, while the best goal configuration would be that maximizing the RF, any configuration chosen using the method above with a high value for the RF would be necessarily be in the neighborhood of the maxima. Such a configuration may be found with reasonable effort. Other ways to reduce search time include search space reduction using binning and decimation, finite neighborhood searches, etc.

Figure 10 shows an unknown environment ($100 \times 100\,\text{m}^2$) with occlusions (black) to be mapped/modeled. It is assumed that all mobile mapping sensing agents start at the center of this environment. Figure 11 shows the fraction of the envrionment mapped (Figure 11(a)) and the net distance moved by the vision system for the five mapping methods, using a single mobile vision system (with 90° field of view, 15 m depth of field). The energy consumption by the system is proportional to the net distance moved by the vision system. Hence, it is desirable to have a large fraction of the environment mapped with small net displacements. Figure 12 shows the fraction of the envrionment mapped in Figure 10 and the net distance moved by the vision system for the five mapping methods,

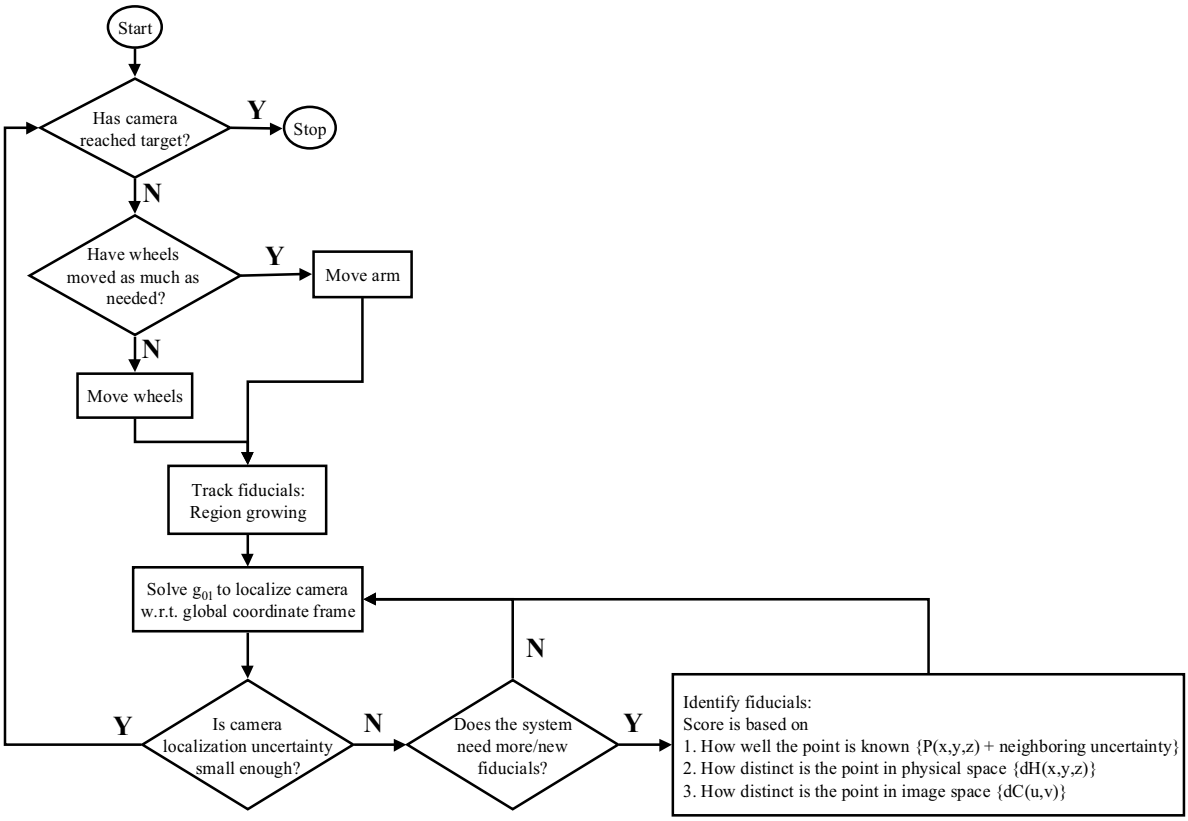


Fig. 9. Flowchart for vision system motion identification using scene fiducials.

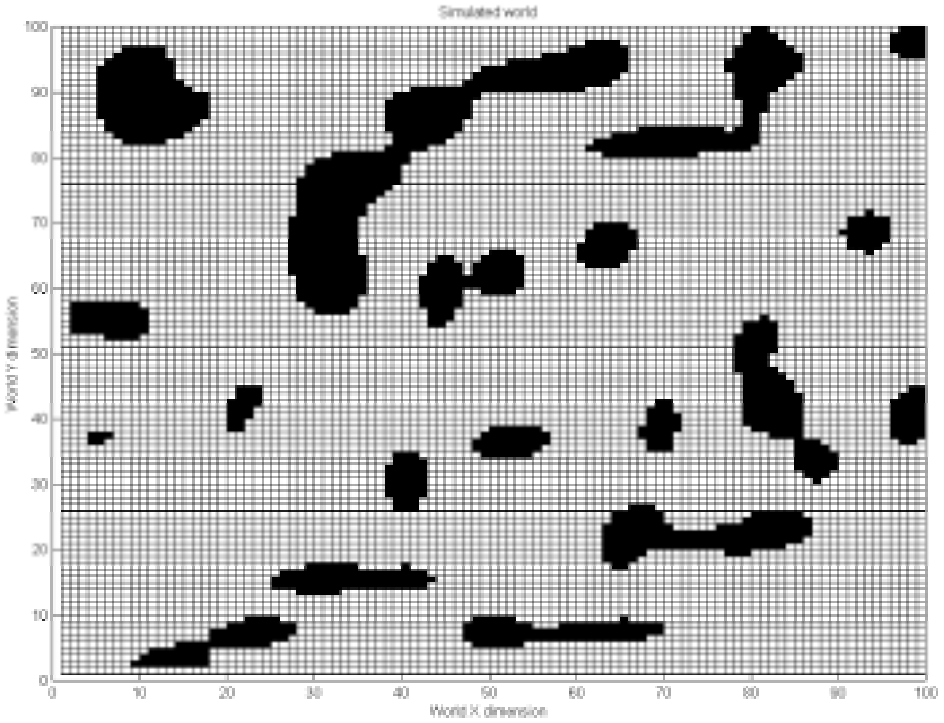
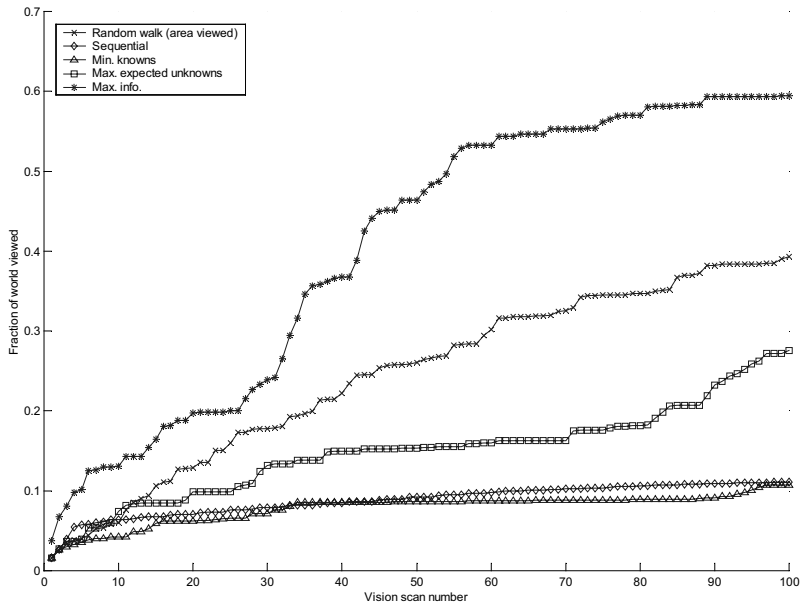
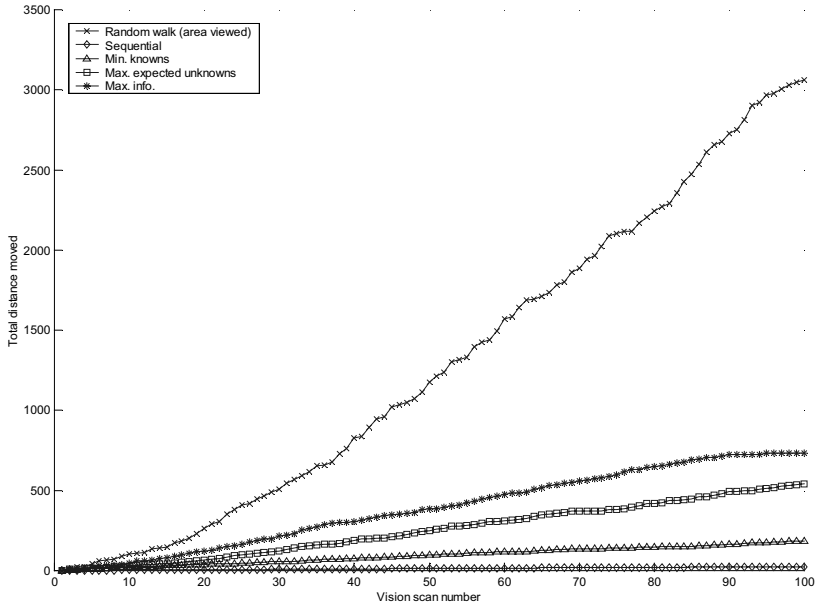


Fig. 10. Unknown planar environment.

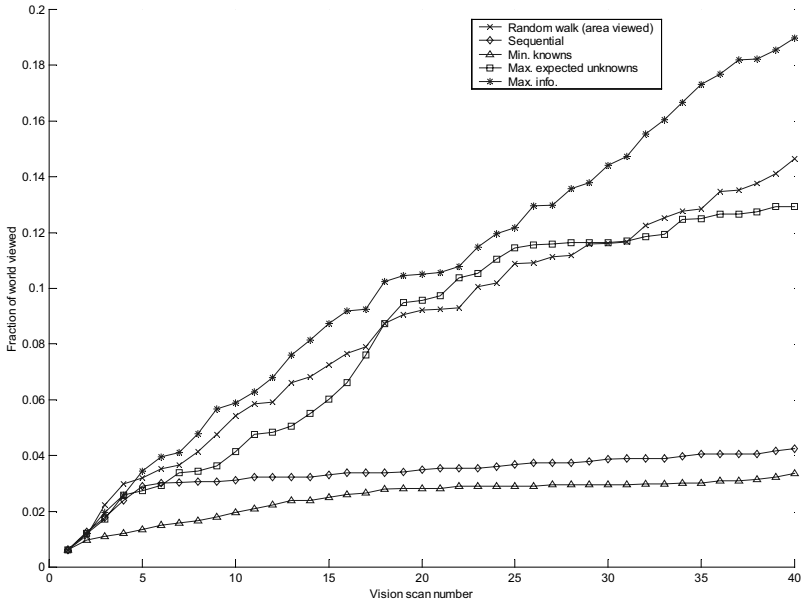


(a)

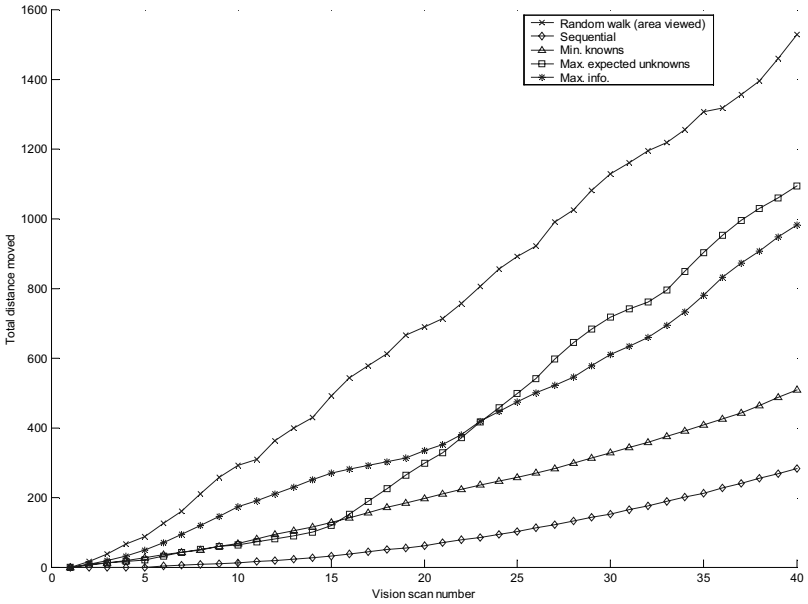


(b)

Fig. 11. Results of single vision system modeling an unknown environment: (a) fraction of environment modeled as a function of scan number; (b) distance moved by vision system as a function of scan number.



(a)



(b)

Fig. 12. Results of two vision systems modeling an unknown environment: (a) fraction of environment modeled as a function of scan number; (b) distance moved by vision system as a function of scan number.

using two cooperating mobile vision systems (with 75° field of view, 10 m depth of field). These results show the effectiveness of the information theroretic approach to vision system pose selection in environment modeling.

Figures 13–15 show the results of exploring this environment with a single robot (with a 90° field of view, 15 m depth of field). Figures 13 and 14 show examples of the area mapped and the path taken by the vision system using mapping/modeling methods (i) and (v). Figure 15 shows the average accumulated motion error of the vision system as it explores the environment as a function of traveled distance. Comparing Figures 15(a) and (b), it is seen that this error decreases substantially with redundancy of fiducials as well as with proximity of the fiducials. The energy consumption by the system is proportional to the net distance moved by the vision system. Hence, it is desirable to have a large fraction of the environment mapped with small net displacements. This is significantly more noticeable in Figure 14 where the robot path based on the amount of information obtained from its motion appears to criss-cross over itself far less frequently as compared to the random walk (with some heuristics). The former path (Figure 14(b)) is substantially shorter than the latter (Figure 14(a)) whereas the amount of the environment explored/mapped is greater in the former. The quantified result is seen in Figure 11.

3.2. Experimental Studies

As described above and demonstrated in simulation, the algorithm developed in this paper is equally applicable to a single robot mapping agent as it is to a team of multiple mapping agents. The difference is in the RF, where one must account for the presence of other agents to avoid collisions. Additionally, the presence of body parts of other robot agents in the maps taken must be accounted for and deleted from the map, as they are not natural occurrences in the environment. This is accomplished using CAD models of the agents combined with the current environment measurement along with an estimate of the agents' current pose. This process is not very computationally expensive, as it is being carried out on a small single sensor scan of the environment rather than on the whole environment map. Due to resource limitations, experimental validation was limited to a single mapping agent. Figure 16 shows a single vision system (stereo pair) mounted on a mobile manipulator. The experimental platform consists of a four-wheeled robot with a four-degrees-of-freedom (4-DOF) manipulator arm mounted on a six-axis force/torque sensor. Figure 16(a) shows the system's dimensions. On-board sensors also include a two-axis inclinometer. An off-board computer system (Pentium 166 MHz) is used for real-time control, data acquisition, and data processing. All programs are written in C++ operating on Windows NT. The robot was set in a rough unstructured Mars-type environment and is focused on mapping the environment around a mock solar panel array (representing a real system requiring service by a robot agent). Mapping is performed by breaking up the world into a grid of voxels of specified resolution. All measured points falling in a particular voxel contribute to the error analysis associated with that voxel. Voxels corresponding to empty space falling in the field of view of the vision system are tabulated as known, but with no contribution to the uncertainty model. The desired voxel resolution is a function of the task. For this test, the resolution is set at 1mm.

Figure 17 shows the number of points mapped in the environment for two pose selection methods: sequential/raster scan and maximum information based pose selection. The sequential/raster scan process is a common mapping process often applied by researchers and has been included here for comparison (Asada 1990; Kuipers and Byun 1991; Burschka, Eberst, and Robl 1997; Castellanos et al. 1998; Rekleitis, Dudek, and Milios 2000; Victorino, Rives, and Borrelly 2000; Choset and Nagatani 2001). Once again, the effectiveness of the information based pose selection process, seen as the rate of development of the environment map, is significantly greater using the algorithm developed in this work as compared to the sequential mapping process. Figure 18 shows the regions mapped using sequential camera pose selection and maximum information based camera pose selection. There is almost a factor of 6 increase in the region mapped for the same number of imaging steps when using the maximum information theory based approach as compared to the raster mapping method.

Due to the very rough nature of its surroundings, it was often difficult for the robot to find paths to its desired pose. As seen in Figure 16(b), the agent finds itself unable to move around to the rear of the target due to obstructing boulders and vertical "hill" sides. It was noted that if the desired pose to view the environment was a substantial distance away from the current pose, then the computational time for the mapping step was dominated by the process of determining a safe route to the target pose. To minimize this effect, greater weight was added to the term relating distance to target in the RF (eq. (11)). Future research will focus on determining the effect of a large number of team members and their influence on the mapping or scouting process of their environment. Cooperative robot control architectures for mapping may also exploit the added freedom associated with physical interactions to traverse more complex terrains.

4. Conclusions

In field environments it is often not possible to provide robotic teams with detailed a priori environment and task models. In such unstructured environments, cooperating robots will need to create a dimensionally accurate 3D geometric model by performing appropriate sensor actions. The performance of cooperative robots in field environments is limited by model

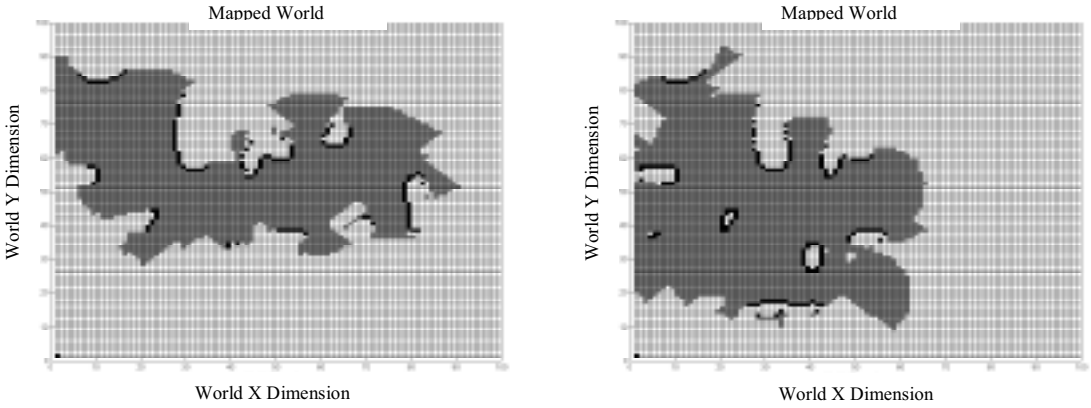
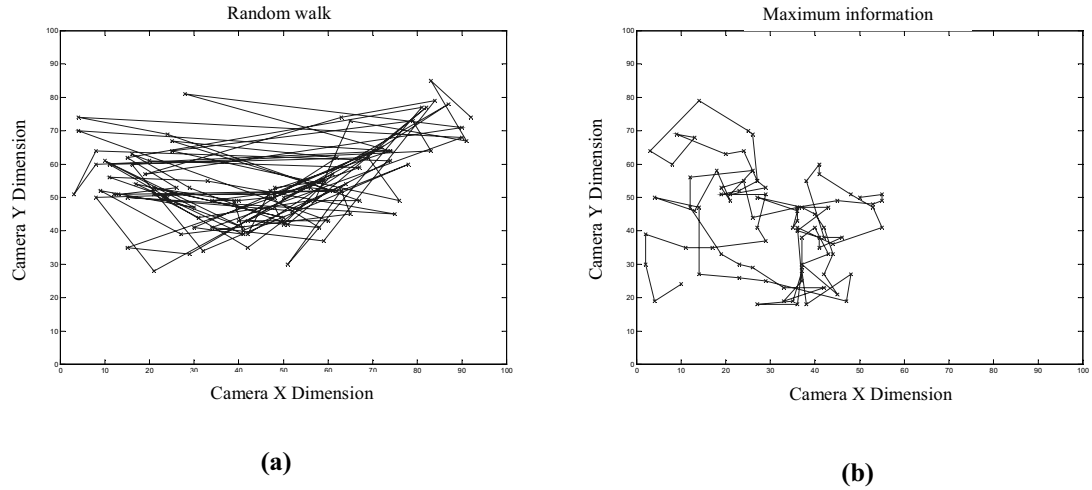
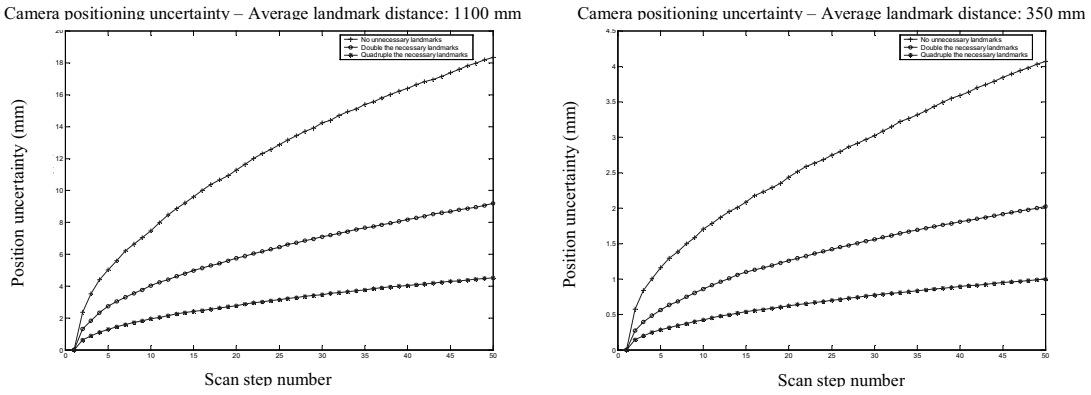


Fig. 13. Mapped area by a single vision system (gray = empty space; black = obstacle; white = unknown): (a) random walk pose selection; (b) maximum information pose selection.



(a) **(b)**

Fig. 14. Path taken by a single vision system: (a) random walk pose selection; (b) maximum information pose selection.



(a) **(b)**

Fig. 15. Accumulated root-mean-square translation error of vision system: (a) average fiducial distance 1100 mm; (b) average fiducial distance 350 mm.

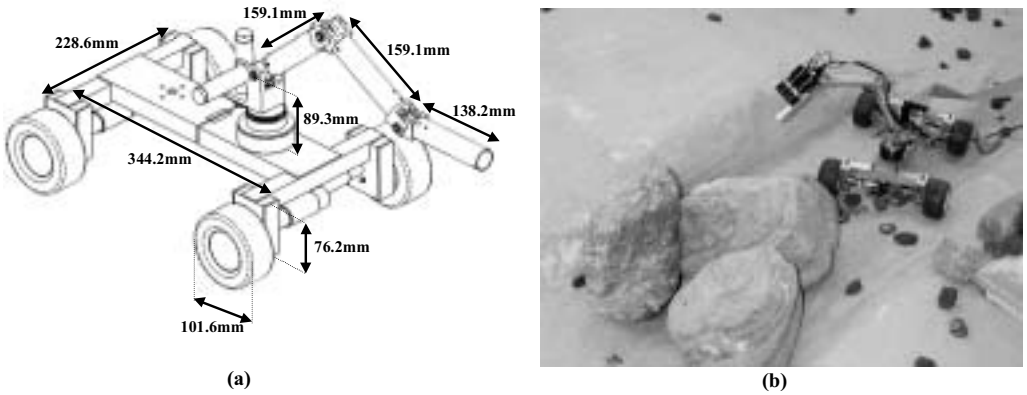


Fig. 16. Experimental mobile vision system modeling an unstructured environment: (a) schematic diagram of the physical system; (b) physical system implementation.

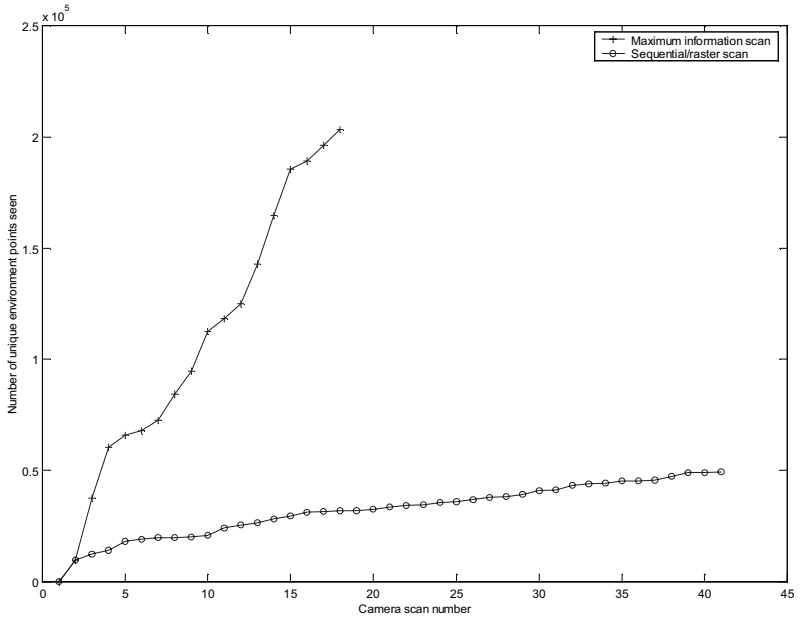
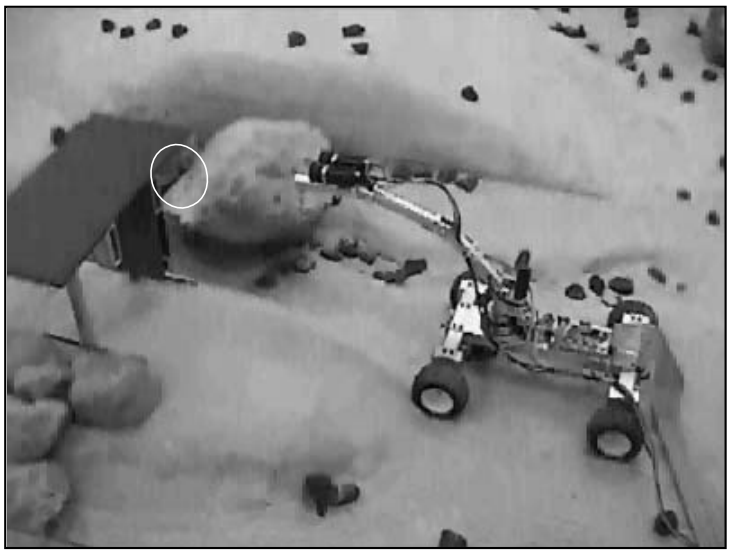
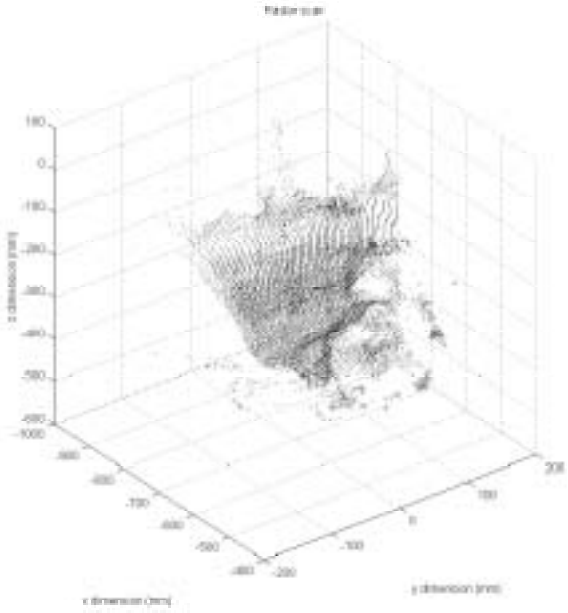


Fig. 17. Number of mapped environment points as a function of scan number.

uncertainties and on-board sensor limitations. In this paper we have developed sensing and estimation algorithms to enable multiple mobile robots to compensate for geometric model uncertainties and to successfully perform interacting cooperative manipulation tasks in highly unstructured/unknown field environments using optimal information gathering methods. A new algorithm based on iterative sensor planning and sensor redundancy has been presented to build a geometrically consistent dimensional map of the environment for mobile robots that have articulated sensors. This algorithm is unique in that it uses a metric of the quality of information previously obtained by the sensors to find new viewing positions for the cameras. The aim is to acquire NI that leads to more detailed and complete knowledge of the environment. Controlling robots to maximize knowledge is performed using Shannon's information theory based evaluation functions. The algorithm is initialized by a scan of the region around a target by all individual agents, to generate a local 3D geometric model of the origin reference frame. Next, these models are used to find optimum environment viewing poses for the multiple vision systems by defining and optimizing a RF over the possible sensor positions, subject to kinematic constraints of

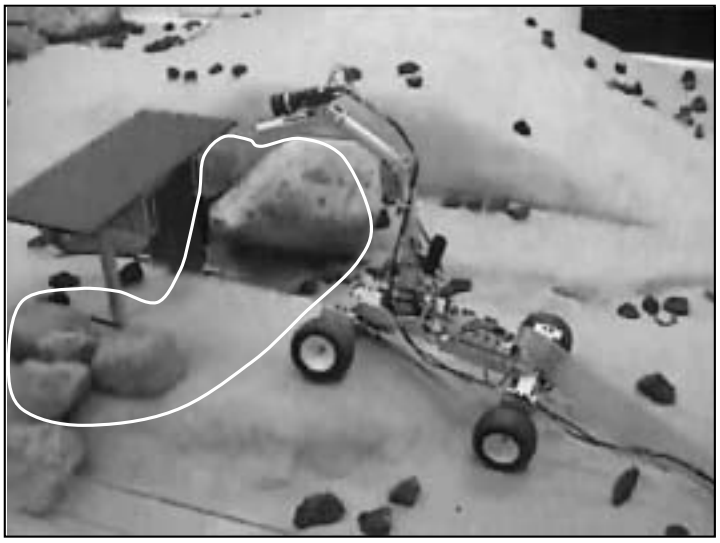


(i) region mapped in 41 steps

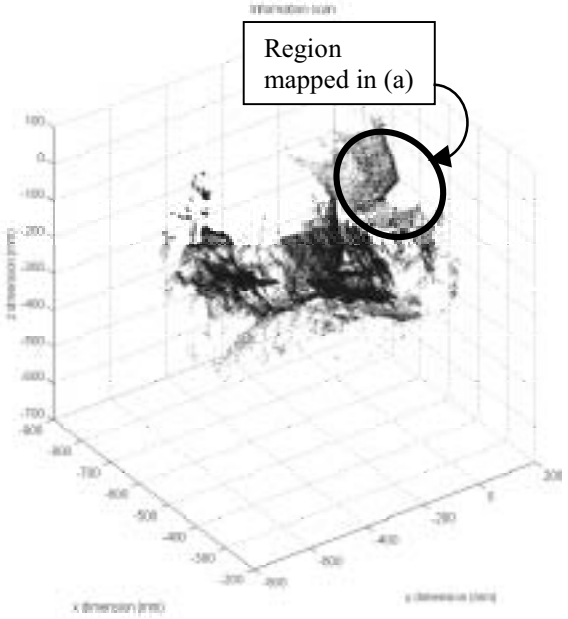


(ii) environment point cloud

(a)



(i) region mapped in 18 steps



(ii) environment point cloud

(b)

Fig. 18. Environment mapped/modeled: (a) sequential camera pose selection; (b) maximum information based camera pose selection.

the sensor placement mechanisms for the individual robots. This RF aims to acquire as much NI about the environment as possible with every sensing cycle, while maintaining or improving the map accuracy, and minimizing the exploration time. The process is constrained by selecting goal points that are not occluded and that can be reached by a collision-free traversable path. The sensors then move to their new poses and acquire 3D data. Based on the sensor mount kinematics, the motion of the sensor is known. However, small motions of the robot base (due to suspension compliance) and errors in sensor mounts lead to additional uncertainties. These are accounted for by measuring common features during the vision sensor motion. Finally, the new data and its associated uncertainty are fused with the current environment map, resulting in an updated probabilistic environment map, which may then be shared with the other sensing agents. Each agent uses the updated probabilistic environment map to generate its NBV. The process continues until the environment is mapped to the predefined extent necessary.

The algorithm may be used by multiple distributed and decentralized sensing agents for efficient and accurate environment modeling. The algorithm makes no assumptions of the environment structure. Hence, it is robust to robot failure since the environment model being built is not dependent on any single agent frame, but is set in an absolute reference frame. It accounts for sensing uncertainty, robot motion uncertainty, environment model uncertainty, and other critical parameters. It allows for regions of higher interest obtaining more attention by the agents. Simulations and experiments show promising results.

Acknowledgments

The authors would like to acknowledge the support of the Jet Propulsion Laboratory, NASA in this work (in particular Dr. Paul Schenker and Dr. Terry Huntsberger). The authors also acknowledge the help of Grant Kristofek, MIT, for the fabrication of the experimental system used in this work.

References

Anousaki, G. C., and Kyriakopoulos, K. J. 1999. Simultaneous localization and map building for mobile robot navigation. *IEEE Robotics and Automation Magazine* 6(3):42–53.

Asada, M. 1990. Map building for a mobile robot from sensory data. *IEEE Transactions on Systems, Man, and Cybernetics* 37(6):1326–1336.

Baumgartner, E. T., Schenker, P. S., Leger, C., and Huntsberger, T. L. 1998. Sensor-fused navigation and manipulation from a planetary rover. *Symposium on Sensor Fusion and Decentralized Control in Robotic Systems*, Boston, MA, November, *Proceedings of the SPIE* 3523:58–66.

Burschka, D., Eberst, C., and Robl, C. 1997. Vision-based model generation for indoor environments. *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, Albuquerque, NM, April.

Castellanos, J. A., Martinez, J. M., Neira, J., and Tardos, J. D. 1998. Simultaneous map building and localization for mobile robots: a multisensor fusion approach. *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, Leuven, Belgium, Vol. 2, pp. 1244–1249.

Choset, H., and Nagatani, K. 2001. Topological simultaneous localization and mapping (SLAM): toward exact localization without explicit localization. *IEEE Transactions on Robotics and Automation* 17(2):125–137.

Gelb, A. 1974. *Applied Optimal Estimation*, MIT Press, Cambridge, MA.

Hager, G. D., Kriegman, D., Teh, E., and Rasmussen, C. 1997. Image-based prediction of landmark features for mobile robot navigation. *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, Albuquerque, NM, Vol. 2, pp. 1040–1046.

Howard, A., Matarić, M., and Sukhatme, G. 2003. Putting the "I" in "team": an ego-centric approach to cooperative localization. *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, Taipei, Taiwan, September, pp. 868–892.

Huntsberger, T. L., Rodriguez, G., and Schenker, P.S. 2000. Robotics: challenges for robotic and human Mars exploration. *Proceedings of ROBOTICS2000*, Albuquerque, NM, March.

Jennings, C., Murray, D., and Little, J. 1999. Cooperative robot localization with vision-based mapping. *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, Detroit, MI, May.

Kruse, E., Gutsche, R., and Wahl, F. M. 1996. Efficient, iterative, sensor based 3-D map building using rating functions in configuration space. *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, Minneapolis, MN, April.

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

Lara, B, Althoefer, K., and Seneviratne, L. D. 1998. Automated robot-based screw insertion system. *Proceedings of the IEEE Conference of the 24th Annual Industrial Electronics Society (IECON '98)*, Aachen, Germany, September, Vol. 4, pp. 2440–2445.

Leonard, J. J., and Durrant-Whyte, H. F. 1991. Simultaneous map building and localization for an autonomous mobile robot. *Proceedings of the IEEE/RSJ International Workshop on Intelligent Robots and Systems (IROS)*, Osaka, Japan, November 3–5, Vol. 3, pp. 1442–1447.

Lumelsky, V., Mukhopadhyay, S., and Sun, K. 1989. Sensor-based terrain acquisition: the "sightseer" strategy. *Proceedings of the 28th Conference on Decision and Control*, Tampa, FL, December.

Luo, R. C., and Kay, M. G. 1989. Multisensor integration and fusion in intelligent systems. *IEEE Transactions on Systems, Man, and Cybernetics* 19(5):901–931.

Park, J., Jiang, B., and Neumann, U. 1999. Vision-based pose computation: robust and accurate augmented reality tracking. *Proceedings of the 2nd IEEE and ACM International Workshop on Augmented Reality (IWAR '99)*, San Francisco, CA, October, pp. 3–12.

Pito, R., and Bajcsy, R. 1995. A solution to the next best view problem for automated CAD model acquisition of free-form objects using range cameras. *Proceedings of the SPIE Symposium on Intelligent Systems and Advanced Manufacturing*, Philadelphia, PA, October.

Rekleitis, I., Dudek, G., and Milios, E. 2000. Multi-robot collaboration for robust exploration. *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, San Francisco, CA, April.

Shaffer, G., and Stentz, A. 1992. A robotic system for underground coal mining. *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, Nice, France, Vol. 1, pp. 633–638.

Shannon, C. E. 1948. A mathematical theory of communication. *Bell System Technical Journal* 27:379–423 and 623–656.

Simhon, S., and Dudek, G. 1998. Selecting targets for local reference frames. *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, Leuven, Belgium, Vol. 4, pp. 2840–2845.

Smith, R. C., and Cheeseman, P. 1986. On the representation and estimation of spatial uncertainty. *International Journal of Robotics Research* 5(4):56–68.

Stroupe, A. W., Martin, M., and Balch, T. 2000. Merging Gaussian distributions for object localization in multi-robot systems. *Proceedings of the 7th International Symposium on Experimental Robotics (ISER '00)*, Waikiki, HI, December 10–13.

Sujan, V. A., and Dubowsky, S. 2002. Visually built task models for robot teams in unstructured environments. *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, Washington, DC, May 11–15.

Sujan, V. A., Dubowsky, S., Huntsberger, T., Aghazarian, H., Cheng, Y., and Schenker, P. 2003. Multi-agent distributed sensing architecture with application to cliff surface mapping. *Proceedings of the 11th International Symposium of Robotics Research (ISRR)*, Siena, Italy, October 19–22.

Tarabanis, K. A., Allen, P. K., and Tsai, R. Y. 1995. A survey of sensor planning in computer vision. *IEEE Transactions on Robotics and Automation* 11(1):86–104.

Thrun, S., Fox, D., and Burgard, W. 1998. A probabilistic approach to concurrent mapping and localization for mobile robots. *Machine Learning* 31:29–53.

Thrun, S., Burgard, W., and Fox, D. 2000. A real-time algorithm for mobile robot mapping with applications to multi-robot and 3D mapping. *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, San Francisco, CA, April.

Tomatis, N., Nourbakhsh, I., and Siegwar, R. 2001. Simultaneous localization and map building: a global topological model with local metric maps. *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Maui, HI, October, Vol. 1, pp. 421–426.

Victorino, A. C., Rives, P., and Borrelly, J.-J. 2000. Localization and map building using a sensor-based control strategy. *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems* (*IROS 2000*), Takamatsu, Japan, October 30–November 5,Vol. 2, pp. 937–942.

Yamauchi, B. 1998. Frontier-based exploration using multiple robots. *Proceedings of the 2nd International Conference on Autonomous Agents (Agents '98)*, Minneapolis, MN, May, pp. 47–53.

Yamauchi, B., Schultz, A., and Adams, W. 1998. Mobile robot exploration and map-building with continuous localization. *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, Leuven, Belgium, Vol. 4, pp. 3715–3720.

Yeh, E., and Kriegman, D. J. 1995. Toward selecting and recognizing natural landmarks. *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Pittsburgh, PA, Vol. 1, pp. 47–53.