

# A Velocity-Based Rao-Blackwellized Particle Filter Approach to Monocular vSLAM

Morteza Farrokhsiar, Homayoun Najjaran

School of Engineering, University of British Columbia, Kelowna, Canada.  
Email: {morteza.farrokhsiar, homayoun.najjaran}@ubc.ca

Received January 6<sup>th</sup>, 2011; revised May 6<sup>th</sup>, 2011; accepted May 13<sup>th</sup>, 2011.

## ABSTRACT

*This paper presents a modified Rao-Blackwellized Particle Filter (RBPF) approach for the bearing-only monocular SLAM problem. While FastSLAM 2.0 is known to be one of the most computationally efficient SLAM approaches; it is not applicable to certain formulations of the SLAM problem in which some of the states are not explicitly expressed in the measurement equation. This constraint impacts the versatility of the FastSLAM 2.0 in dealing with partially observable systems, especially in dynamic environments where inclusion of higher order but unobservable states such as velocity and acceleration in the filtering process is highly desirable. In this paper, the formulation of an enhanced RBPF-based SLAM with proper sampling and importance weights calculation for resampling distributions is presented. As an example, the new formulation uses the higher order states of the pose of a monocular camera to carry out SLAM for a mobile robot. The results of the experiments on the robot verify the improved performance of the higher order RBPF under low parallax angles conditions.*

**Keywords:** Filtering, Higher Order Filter, Rao-Blackwellized Particle Filter, Bearing-Only Systems, Visual SLAM

## 1. Introduction

The basic idea of simultaneous localization and mapping (SLAM) was originally discussed for autonomous robots mainly because of the need to locate the robot in real time in a map which is incrementally built using the information obtained from the robot control and measurement systems. Smith *et al.* [1] established a probabilistic approach to model the uncertainty in the control and measurement signals. In essence, the SLAM provides a suitable framework for describing robot motion and landmark locations together with the underlying uncertainties associated with both of them. In this approach, motion model and observation model are described as the first-order and zero-order Markov processes, respectively. In the probabilistic approach, the SLAM problem is considered as a filtering problem in which the control input and robot observations are known and the robot pose and landmark locations are the desired unknown states. Although other approaches have been used to solve the SLAM problem, probabilistic approach has been remained as the mainstream of the SLAM literature [2].

In order to solve the SLAM problem, Csorba and Durrant-Whyte [3] proposed the use of the Extended Kalman Filter (EKF) which is capable of solving non-

linear filtering problems. Their work was further extended in [4,5] where convergence and performance of the EKF solution were discussed, and successful experimental results were reported.

A major limitation of the EKF solution is that the covariance matrix must maintain the information of all landmarks of the map, and hence the state vector is at least an  $N^2$  space, where  $N$  denotes the number of landmarks. Clearly, the covariance matrix has to be updated in each iteration at a cost proportional to  $N^2$ . This computational cost limits the real-time implementation of the EKF SLAM to simple environments and sparse mapping. As a result, reducing the computational load of the EKF-based SLAM solutions has been a major focus for the researchers working on the SLAM problem [6-8].

Liu and Thrun proposed a new solution for the SLAM problem using the Extended Information Filter (EIF) [9]. In their solution, the information matrix is sparsified and therefore, memory and computation load have been reduced significantly. Another solution for the SLAM problem, FastSLAM, based on the Rao-Blackwellized particle filtering (RBPF) was presented by [10]. FastSLAM is based on factorization of the robot path from landmark locations in which observation of each land-

mark is independent of any other landmark position [11,12]. The use of well-tuned particle filters instead of Kalman-type filters can result in less computational complexity and more robustness in data association [13]. Also, unlike the EKF, there is no need for the linearization of the motion model. The main shortcoming of FastSLAM was the large number of samples that are needed to avoid the degeneracy of the filter. This problem has been addressed in FastSLAM 2.0 in which both control inputs and observations are used to obtain the sampling distribution. The use of the observations in obtaining the sampling distribution not only reduces the number of samples required for the filtering process but also increases the overall efficiency of the filter [14]. However, it should be noted that in the original form of FastSLAM 2.0, the measurements are only useful for the sampling distribution of the states which are explicitly used in the measurement model. A goal of the algorithm proposed in this work is to eliminate this limitation.

Recent research on the SLAM problem focuses on finding more computationally efficient solutions using vision-based approaches, referred to as vSLAM. Typically, it is desirable to adopt a monocular vision system i.e., a single camera, for generating the map [15-18]. The use of a single camera, which is an example of a bearing only device when used as the only source of information, causes the problem of partial observability [17, 19] that is attributed to the lack information about the depth of the landmarks. However, the standard SLAM solutions are based on a fully observable system, and hence the main motivation of this work is to introduce an appropriate method for monocular vSLAM.

It is well-known that similar to any other filtering and estimation problem, the performance of the bearing-only SLAM systems is strongly influenced by the system order [20]. For example, higher order systems may be able to track more complicated maneuvers and abrupt changes of robot trajectories but they can also increase the estimation error unnecessarily for simple motion patterns. In the EKF-based SLAM framework, the motion model order can be modified easily and additional states can be added to the state vector and covariance matrix, but such modification for Rao-Blackwellized SLAM framework is not trivial since higher order states such as velocity and acceleration have no effect in the observation in the original formulation. Although FastSLAM 2.0 is superior to FastSLAM in terms of performance and convergence, it is not applicable to the bearing-only systems with higher order states.

The idea of the use of higher order state such as velocity in an RBPF-based SLAM was introduced and briefly discussed for bearing-only systems in [21]. Here, detailed formulation as well as a thorough analysis of the

estimation performance using experimental results is presented for the proposed approach. In this approach, first the robot pose is updated by sampling velocity from the proposal distribution. Similar to FastSLAM 2.0, the proposed approach incorporates the information obtained from both the control input and observations to obtain the proposal distribution. However, the can also adopt the information obtained the observations for the proposal distribution of a state which is of higher order than the observed state (*i.e.* velocity where the displacement is measured). In essence, this improves the versatility of the proposed approach with respect to the original formulation of FastSLAM 2.0. Considering the updated robot pose and last observation, the landmark locations are updated using the EKF. In the next stage, the importance weight of each particle is calculated. Finally, a set of particles is resampled with respect to the weight of the particles calculated in the previous step. In order to solve the problem of partial observability, each feature is initialized using a modified scheme in which the landmarks are defined using their inverse depth to decrease the nonlinearity of the measurement equation in low parallax angles [22].

Another implementation of the higher order Rao-Blackwellized SLAM has been introduced by Törnqvist *et al.* [23]. However, their approach is based on FastSLAM in which sampling distribution is independent of observations whereas the approach proposed in this paper proposes an extension of FastSLAM 2.0 in which sampling distribution relies on the control input as well as the observations. More precisely, they have focused on reducing the computational load of FastSLAM by introducing a new factorization technique, whereas in this paper the main concern is to extend FastSLAM 2.0 so that it incorporates observations in obtaining the sampling distribution of the states even if they are not explicitly used in the measurement model. Another difference between this work and the one presented in [23] is that their approach has been implemented into an unmanned aerial vehicle (AUV) equipped with multiple sensors including inertial navigation system (INS) and cameras whereas the main focus of this paper is on bearing-only systems and tackling the partial observability problem.

The detailed formulation of the proposed estimation approach including sampling and importance weight distributions are presented in Section 2. In Section 3, the motion and observation models of a bearing-only system are discussed. Section 4 presents another important feature of the proposed SLAM method that is the initialization scheme required to tackle the partial observability problem. Section 5 presents the results of a SLAM experiment carried out using a digital camera in a laboratory environment.

## 2. Velocity Based RBPF Approach

In the SLAM problem, the only available information for the observer is the measurements,  $z_{1:t}$  and the control input  $u_{1:t}$ , e.g. acceleration, as they become available. The index  $1:t$  refers to the time period between the first iteration and the iteration at time  $t$ . In the vision-based SLAM, it can be assumed that the data association problem [24] is solved using image processing methods such as the SIFT algorithm [25-27]. The bearing-only system is modeled using a constant velocity model which means that both position and velocity are included in the state vector. In this way, the accelerations are characterized using a Gaussian distribution. The approximated acceleration is assumed as the control input. The velocity dynamics can be presented as the first order Markov chain:

$$p(x_t | u_t, x_{t-1}) \approx N(x_t; f(x_{t-1}, u_t), R_t) \quad (1)$$

where  $x_t$  is the velocity in time  $t$ . The symbol  $N$  denotes the normal distribution with the mean of  $f(x_{t-1}, u_t)$  and covariance of  $R_t$ . The relationship between the velocity and pose is a deterministic function presented as:

$$y_t = g(y_{t-1}, x_t) \quad (2)$$

which means that the current pose,  $y_t$  is determined based on the previous pose,  $y_{t-1}$  and the current velocity,  $x_t$ . The observation equation is a zero order Markov chain:  $p(z_t | m_{c_t}, y_t, c_t) \approx N(z_t; h(m_{c_t}, y_t, c_t), Q_t)$  where  $z_t$  is the observation in time  $t$ . The symbol  $N$  stands for the normal distribution with the mean of  $h(m_{c_t}, y_t, c_t)$  and covariance of  $Q_t$ . The data association decision and landmark location are shown by  $c_t$  and  $m_{c_t}$  respectively. In the *full SLAM* problem formulation, the final objective is to find the posterior of the entire path and the map [13]:

$$p(s_{1:t}, m | u_{1:t}, z_{1:t}, c_{1:t}) \quad (3)$$

where  $s$  is the robot pose state vector,  $s = \{x, y\}$ ,  $c_{1:t}$  stands for known data association and  $m$  is the map which is the set of landmarks:

$$m = \{m_1, \dots, m_N\} \quad (4)$$

It is shown that this problem can be written in the factored form [28]:

$$p(s_{1:t}, m | u_{1:t}, z_{1:t}, c_{1:t}) = p(s_{1:t} | u_{1:t}, z_{1:t}, c_{1:t}) \prod_{i=1}^N p(m_i | z_{1:t}, s_{1:t}, c_{1:t}) \quad (5)$$

This factored form can be solved using one of the RBPF based approaches discussed in [10,14]. The localization problem is solved using the particle filter. Each

particle,  $y_t$ , in this approach consists of a robot trajectory estimate and a set of estimators of the individual landmark locations. In order to use the Rao-Blackwellized Particle Filter (RBPF) approach, proper proposal and importance weight distributions have to be determined

### 2.1. Proposal Distribution

Based on the idea discussed in [14], the proposal distribution is chosen as:

$$s_t^{[k]} \sim p(s_t | x_{1:t-1}^{[k]}, y_{1:t-1}^{[k]}, u_{1:t}, z_{1:t}, c_{1:t}) \quad (6)$$

This distribution includes all information related to the path,  $x_{1:t-1}^{[k]}$  and  $y_{1:t-1}^{[k]}$  controls,  $u_{1:t}$ , observations,  $z_{1:t}$ , and data associations,  $c_{1:t}$ . Considering the fact that  $s = \{x, y\}$ , this distribution can be expanded as:

$$\begin{aligned} & p(s_t | x_{1:t-1}^{[k]}, y_{1:t-1}^{[k]}, u_{1:t}, z_{1:t}, c_{1:t}) \\ &= p(x_t | x_{1:t-1}^{[k]}, y_{1:t-1}^{[k]}, u_{1:t}, z_{1:t}, c_{1:t}) \\ & p(y_t | x_t, x_{1:t-1}^{[k]}, y_{1:t-1}^{[k]}, u_{1:t}, z_{1:t}, c_{1:t}) \end{aligned} \quad (7)$$

Using Equation (2) and the Markov rule:

$$\begin{aligned} & p(y_t | x_t, x_{1:t-1}^{[k]}, y_{1:t-1}^{[k]}, u_{1:t}, z_{1:t}, c_{1:t}) \\ &= p(y_t | x_t, y_{1:t-1}^{[k]}) = \delta_{y_t, y_t^d} \end{aligned} \quad (8)$$

where  $\delta$  is Kronecker delta and  $y_t^d$  is the deterministic result of the Equation (3) when  $x_t$  and  $y_{1:t-1}^{[k]}$  are given. Therefore, the proposal distribution of Equation (6) is simplified as:

$$\begin{aligned} & p(s_t | x_{1:t-1}^{[k]}, y_{1:t-1}^{[k]}, u_{1:t}, z_{1:t}, c_{1:t}) \\ &= p(x_t | x_{1:t-1}^{[k]}, y_{1:t-1}^{[k]}, u_{1:t}, z_{1:t}, c_{1:t}) \delta_{y_t, y_t^d} \end{aligned} \quad (9)$$

Thus, sampling must occur in two separate stages. The velocity states,  $x_t$  has to be drawn from the following probability distribution:

$$x_t^{[k]} \sim p(x_t | x_{1:t-1}^{[k]}, y_{1:t-1}^{[k]}, u_{1:t}, z_{1:t}, c_{1:t}) \quad (10)$$

whereas the states related to the robot pose,  $x$  have to be calculated using Equation (2). Using the Bayesian formulation, the probability distribution of Equation (11) can be presented as:

$$\begin{aligned} & x_t^{[k]} \sim \eta^{[k]} p(z_t | x_t, x_{1:t-1}^{[k]}, y_{1:t-1}^{[k]}, u_{1:t}, z_{1:t-1}, c_{1:t}) \\ & p(x_t | x_{1:t-1}^{[k]}, y_{1:t-1}^{[k]}, u_{1:t}, z_{1:t-1}, c_{1:t}) \end{aligned} \quad (11)$$

and using Markov chain:

$$x_t^{[k]} \sim \eta^{[k]} p(z_t | x_t, x_{1:t-1}^{[k]}, y_{1:t-1}^{[k]}, u_{1:t}, z_{1:t-1}, c_{1:t}) p(x_t | x_{t-1}^{[k]}, u_t) \quad (12)$$

Considering the observation model, the posterior in Equation (12) can be presented as:

$$\eta^{[k]} \int p\left(z_t | x_t, m_{c_t}, x_{1:t-1}^{[k]}, y_{1:t-1}^{[k]}, u_{1:t}, z_{1:t-1}, c_{1:t}\right) \mathbf{d}m_{c_t} p\left(x_t | x_{t-1}^{[k]}, u_t\right) \quad (13)$$

A general closed form solution of Equation (13) is not available unless the linearizability assumption [14] is imposed on the system. Using these assumptions the closed form of the Equation (13) can be obtained as the:

$$\begin{aligned} & p\left(x_t | x_{t-1}^{[k]}, y_{t-1}^{[k]}, u_{1:t}, z_{1:t}, c_{1:t}\right) \\ &= \eta^{[k]} \exp\left\{-P_t^{[k]}\right\} \end{aligned} \quad (14)$$

where

$$\begin{aligned} P_t^{[k]} &= \frac{1}{2} \left( \left( z_t - \hat{z}_t^{[k]} - H_y \left( y_t - \hat{y}_t^{[k]} \right) \right)^T Q^{[k]-1} \right. \\ & \quad \left. \left( z_t - \hat{z}_t^{[k]} - H_y \left( y_t - \hat{y}_t^{[k]} \right) \right) \right. \\ & \quad \left. + \left( x_t - \hat{x}_t^{[k]} \right)^T R_t^{-1} \left( x_t - \hat{x}_t^{[k]} \right) \right) \end{aligned} \quad (15)$$

where  $\hat{z}_t^{[k]}$ ,  $\hat{y}_t^{[k]}$  and  $\hat{x}_t^{[k]}$  are prediction of the observation, velocity states and pose states, respectively. Also  $H_y$  is the Jacobian of the measurement model with respect to the pose states. In addition,  $Q_t^{[k]}$  is defined as:

$$Q_t^{[k]} = Q_t + H_m^T \Sigma_{c_t, t-1}^{[k]} H_m \quad (16)$$

where  $H_m$  is the Jacobian of the measurement model with respect to the landmark location and  $\Sigma_{c_t, t-1}^{[k]}$  is the covariance of the observed landmark location estimation. Considering the fact that the covariance of the Gaussian function is the inverse of the second derivative, and also invoking Equation (16), the sampling distribution covariance is obtained from:

$$\Sigma_{x_t}^{[k]} = \left( H_y Y_x \right)^T Q_t^{[k]-1} \left( H_y Y_x \right) + R_t^{-1} \quad (17)$$

where  $Y_x$  is the Jacobian of the Equation (2) with respect to the robot velocity states. Also, the mean value which is the root of the first order derivative equals to:

$$\Sigma_{x_t}^{[k]} = \left( H_y Y_x \right)^T Q_t^{[k]-1} \left( H_y Y_x \right) + R_t^{-1} \quad (18)$$

In summary, to sample from the trajectory, the first velocity states have to be drawn from the Gaussian distribution with the covariance and mean indicated in Equations (17) and (18), respectively. Then, the pose states have to be calculated using Equation (2).

## 2.2. Importance Weight

As it is discussed in [14] in the particle filtering and also RBPF, resampling is the crucial stage of the filtering process. In order to resample the particle, a proper importance weight has to be assigned to each particle. As the other particle filter approach, the importance weight is defined by the following quotient:

$$\omega_t^{[k]} = \frac{\text{target distribution}}{\text{proposal distribution}} \quad (19)$$

The target distribution is the robot trajectory posterior,  $p(s_{1:t} | u_{1:t}, z_{1:t}, c_{1:t})$ . The proposal distribution is  $p(s_t | s_{1:t-1}^{[k]}, u_{1:t}, z_{1:t}, c_{1:t})$  given the correct posterior from the previous stages,  $p(s_{1:t-1}^{[k]} | u_{1:t}, z_{1:t}, c_{1:t})$ . Therefore, Equation (19) can be written as:

$$\begin{aligned} \omega_t^{[k]} &= \\ &= \frac{p\left(s_t | u_{1:t}, z_{1:t}, c_{1:t}\right)}{p\left(s_t | s_{1:t-1}^{[k]}, u_{1:t}, z_{1:t}, c_{1:t}\right) p\left(s_{1:t-1}^{[k]} | u_{1:t}, z_{1:t}, c_{1:t}\right)} \end{aligned} \quad (20)$$

which can be presented as:

$$\omega_t^{[k]} = \eta p\left(z_t | s_t, s_{1:t-1}^{[k]}, u_{1:t}, z_{1:t}, c_{1:t}\right) \quad (21)$$

This equation can be transformed as:

$$\begin{aligned} \omega_t^{[k]} &= \eta \int \left( p\left(z_t | s_t, s_{1:t-1}^{[k]}, u_{1:t}, z_{1:t}, c_{1:t}\right) \right. \\ & \quad \left. p\left(s_t | s_{1:t-1}^{[k]}, u_{1:t}, z_{1:t-1}, c_{1:t}\right) \right) \mathbf{d}s_t \end{aligned} \quad (22)$$

Using Markov and a similar procedure to Equations (9)-(13):

$$\begin{aligned} \omega_t^{[k]} &= \eta \int \int p\left(z_t | s_t, s_{1:t-1}^{[k]}, m_{c_t}, u_{1:t}, z_{1:t}, c_{1:t}\right) \\ & \quad p\left(m_{c_t} | x_t, x_{1:t-1}^{[k]}, y_{1:t-1}^{[k]}, u_{1:t}, z_{1:t-1}, c_{1:t}\right) \\ & \quad p\left(x_t | s_{1:t-1}^{[k]}, u_{1:t}, z_{1:t-1}, c_{1:t}\right) \mathbf{d}m_{c_t} \mathbf{d}s \end{aligned} \quad (23)$$

This equation can be simplified further using Markov chain to:

$$\begin{aligned} \omega_t^{[k]} &= \eta \int p\left(x_t | x_{t-1}, u_t\right) \mathbf{d}x_t \\ & \quad \int p\left(z_t | y_t, m_{c_t}\right) p\left(m_{c_t} | s_{1:t-1}^{[k]}, u_{1:t}, z_{1:t-1}, c_{1:t}\right) \mathbf{d}m_{c_t} \end{aligned} \quad (24)$$

Using the system and measurement model, Equations (1)-(3), and linearizing Equations (2)-(3), the closed form of the Equation (24), can be obtained as the Gaussian distribution:

$$\begin{aligned} \omega_t^{[k]} &= \left| 2\pi L_t^{[k]} \right|^{-\frac{1}{2}} \\ & \quad \exp\left(-\frac{1}{2} \left( z_t - \hat{z}_t^{[k]} \right)^T L_t^{[k]-1} \left( z_t - \hat{z}_t^{[k]} \right)\right) \end{aligned} \quad (25)$$

where

$$\begin{aligned} L_t^{[k]} &= \left( H_y Y_x \right)^T Q_t^{[k]} \left( H_y Y_x \right) \\ & \quad + H_m^T \Sigma_{c_t, t-1}^{[k]} H_m + R \end{aligned} \quad (26)$$

The higher order RBPF for  $M$  particles can be summarized as it is shown in **Table 1**.

## 3. Motion and Measurements Model

### 3.1. Camera Motion Model

The camera motion is modeled using the constant veloc-

**Table 1. Pseudo-code of higher order RBPF approach.**

Do the following $M$ times
• Retrieve the previous trajectory information $s_{t-1}^{i^{k1}}$ from the $k^{\text{th}}$ particle set of the previous iteration, $Y_{t-1}^{i^{k1}}$ ( <i>Retrieval</i> ).
• Sample a new velocity $x_t^{i^{k1}}$ based on the proposal distribution ( <i>Prediction</i> ).
• Update the pose based on the sampled velocity ( <i>Prediction cont'd</i> )
• Update the mean and covariance of the landmark distribution using the predicted pose and the observation of the landmark ( <i>Measurement update</i> ).
• Calculate the importance weight $\omega_t^{i^{k1}}$ for the new particle ( <i>Importance weight</i> ).
• Sample, with replacement, $M$ particles, where each particle is sampled with a probability proportional to $\omega_t^{i^{k1}}$ ( <i>Resampling</i> )

ity input assumption in which the velocity update equation can be expressed as [29]:

$$\begin{pmatrix} \mathbf{v}_t^W \\ \mathbf{w}_t^W \end{pmatrix} = \begin{pmatrix} \mathbf{v}_t^W + \mathbf{V}^W \\ \mathbf{w}_t^W + \Omega^W \end{pmatrix} \quad (27)$$

where  $\mathbf{v}_t^W$  and  $\mathbf{w}_t^W$  are the velocity and rotational velocity of the camera in the world coordinates, respectively. Also

$$u = \begin{pmatrix} \mathbf{V}^W \\ \Omega^W \end{pmatrix} = \begin{pmatrix} a^W \Delta t \\ \alpha^W \Delta t \end{pmatrix} \quad (28)$$

where  $a^W$ ,  $\alpha^W$  and  $\Delta t$  are acceleration, angular acceleration and time interval, respectively. In the bearing-only situation acceleration and angular acceleration can be estimated by a Gaussian white noise. The camera pose can be defined by camera optical center,  $r^W$ , and the quaternion,  $q^W$ , defining the position and orientation of the camera in the global coordinates, respectively:

$$g = \begin{pmatrix} r^W \\ q^W \end{pmatrix}_t = \begin{pmatrix} r_{t-1}^W + \mathbf{v}_t^W \Delta t \\ q_{t-1}^W \times q(\mathbf{w}_t^W \Delta t) \end{pmatrix} \quad (29)$$

where  $q(\Omega_t^W \Delta t)$  is the quaternion defined by the rotation vector  $\Omega_t^W \Delta t$ .

### 3.2. Measurement Model

The measurement model has been adopted from [30]. In a 3D scene, the observation of every point  $i$  from a camera is defined by a vector consisting of six states:

$$m_i = (x_i, y_i, z_i, \theta_i, \varphi_i, \rho_i)^T \quad (30)$$

This vector determines the point location as:

$$Y = \begin{pmatrix} x_i \\ y_i \\ z_i \end{pmatrix} + \frac{1}{\rho_i} n(\theta_i, \varphi_i) \quad (31)$$

where  $(x_i, y_i, z_i)^T$  is the location of the optical center of

the camera at the first observation,  $\rho_i$  is the inverse of the point depth along the carrying ray,  $d_i$ ,  $\theta_i$  and  $\varphi_i$  are azimuth and elevation angles of ray directional vector at the global coordinate frame. The ray directional vector,  $n(\theta_i, \varphi_i)$  is defined as:

$$n(\theta_i, \varphi_i) = \cos(\varphi_i) \sin(\theta_i) - \sin(\theta_i) \cos(\varphi_i) \cos(\theta_i) \quad (32)$$

Each observation can be interpreted as a constraint between the camera pose and the corresponding landmark. Using the pin-hole model [31], the location of the landmark in the camera coordinate is given by:

$$h^C = R^{CW} (Y - r^W) \quad (33)$$

where  $R^{CW}$  is the rotation matrix from the global coordinate frame to the camera coordinate frame. The last equation can be expanded as:

$$h^C = R^{CW} \left( \rho_i \begin{pmatrix} x_i \\ y_i \\ z_i \end{pmatrix} - r^W \right) + n \leq (\theta_i, \varphi_i) \quad (34)$$

According to the pinhole model, the camera observes the projection of  $h^C$  into the image plane denoted by:

$$z = h(s, m) = \begin{pmatrix} u_0 - \frac{f}{d_x} v \\ v_0 - \frac{f}{d_y} v \end{pmatrix} \quad (35)$$

Where  $u_0$  and  $v_0$  are the camera center in pixels,  $f$  is the focal length, and  $d_x$  and  $d_y$  are the pixel sizes. Also,  $u$  and  $v$  are defined as:

$$\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} \frac{h_x^C}{h_z^C} \\ \frac{h_y^C}{h_z^C} \end{pmatrix} \quad (36)$$

### 4. Feature Initialization

The initialization process includes both initial estimation of the state values and assigning the related covariance matrix. Once a landmark is initialized, the estimation process is carried out using an EKF. The proposed initialization approach is similar to the initialization approach discussed by [30,32] but modified according to the RBPF approach used here. The initial state values for the landmark,  $\hat{m}_i$  can be expressed as:

$$m_i = (\hat{x}_i, \hat{y}_i, \hat{z}_i, \hat{\theta}_i, \hat{\varphi}_i, \hat{\rho}_i)^T \quad (37)$$

These values have to be determined using the location information, observation of new landmarks and the initial estimation of the depth inverse,  $\rho_0$ . The optic center

of the camera is directly taken from the current estimation of the camera location. The azimuth and elevation angles of the projection ray are obtained from inverse of the camera pin-hole model. In the range  $\left(0, \frac{1}{d_{\min}}\right)$ , the inverse depth,  $\rho_0$ , can be estimated as:

$$\hat{\rho}_0 = \frac{1}{2d_{\min}} \quad (38)$$

The difference between initialization scheme of the higher order RBPF approach and initialization discussed in [32] lies in the initial covariance matrix used. In the higher order RBPF, the covariance matrix after this initialization is:

$$\mathbf{P} = \mathbf{J} \begin{pmatrix} D_t & 0 & 0 \\ 0 & Q_t & 0 \\ 0 & 0 & \sigma_\rho^2 \end{pmatrix} \mathbf{J}^T \quad (39)$$

where  $Q_t$  is the noise covariance of camera, and the Jacobian is given by:

$$\mathbf{J} = \left( \frac{\partial m_i}{\partial r_k^w}, \frac{\partial m_i}{\partial q_k^w}, \frac{\partial m_i}{\partial h}, \frac{\partial m_i}{\partial \rho} \right) \quad (40)$$

Also,

$$D_t = H_x R_t H_x^T \quad (41)$$

where  $H_x$  is the Jacobian of the pose states with respect to the velocity states and  $R_t$  is the control input covariance. The depth covariance is defined as:

$$\sigma_\rho = \frac{1}{4d_{\min}} \quad (42)$$

## 5. Experimental Results and Discussion

An offline experiment using a video clip has been carried out to evaluate the proposed RBPF approach in combination with common feature extraction methods discussed

by [25,33].

A digital camera is installed on an Amigobot robot designed by Mobile Robotic Inc. The robot is programmed to move on a straight line with constant velocity of 0.6 m/sec while capturing a video from the laboratory environment. The robot navigation system equipped with wheel encoders indicates that the robot has been moved 2198 mm in  $x$ -direction and -121 mm in  $y$ -direction.

In this experiment, the RBPF consisting of 40 particles initializes and filters the features extracted using the SIFT algorithm from the video frames. The calibration parameters of the camera play a crucial role in initialization and filtering processes [29]. The camera has been calibrated based on the open access Matlab toolbox provided in [34]. Sample frame of the video and detected SIFT features are shown in **Figure 1**. In addition to the core RBPF algorithm used in the simulation program, additional subroutines have been added to the filtering algorithm to make it more robust in real world implementations. First, a probabilistic scheme has been added to the filter to evaluate the probability of the SIFT matching results and remove possible outliers. Specifically, RBPF will ignore the observation if the matching probability is smaller than a specific threshold value. The matching probability is computed based on the *correspondence likelihood* discussed in [13]. Also, the negative inverse depth which is an important issue in the inverse depth bearing-only SLAM [35] has been avoided using the ability of the particle filter to fuse the hierarchical knowledge with the positive information processed by the probabilistic approach [13]. The robot path and estimated localization results are shown in the **Figure 2**. The time evolutions of the robot pose states are shown in **Figures 3** and **4**. Also estimation of cosine of the robot heading angle has been shown in the **Figure 5**.

The RBPF estimated path which has been shown in **Figure 2** follows the true robot path with slight deviation in the  $y$ -direction. This deviation can be explained by basic experimental errors, such as the robot control sys-

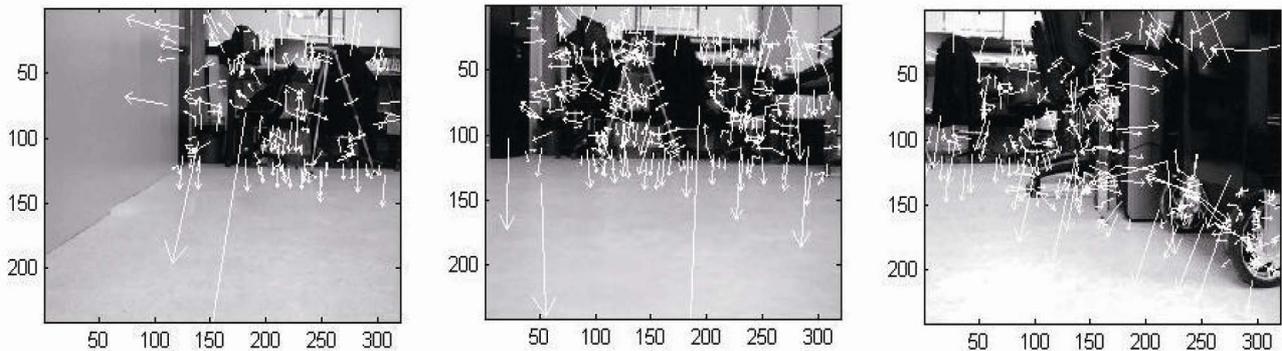


Figure 1. Robot path; true value (solid line) estimation (dashed line).

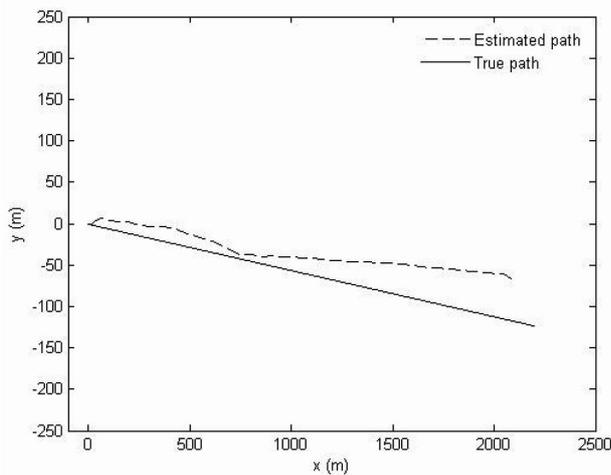


Figure 2. Robot path; true value (solid line), estimation (dashed line).

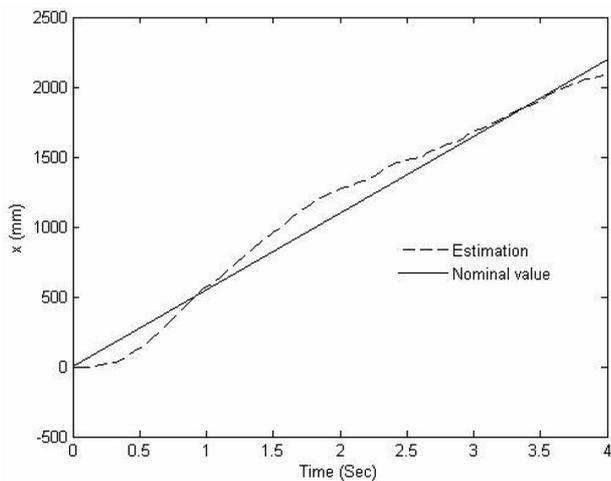


Figure 3. Estimation of the robot in the x-direction versus time.

tem error, the camera-robot alignment and also oversimplification of the measurement model e.g., the radial distortion of the camera lens has been totally ignored in this research. Robot motion estimation in the  $x$  and  $y$  directions are shown in **Figures 3** and **4**. The estimation of the heading angle cosine is shown in **Figure 5**.

Although the major objective in higher order monocular vSLAM is to solve the localization problem and produced sparse cannot be used directly, mapping results is able to provide better insight to the proposed approach. The Monte Carlo simulations of the spatial distributions of a sample landmark, after the first, fourth and seventh frame are shown in **Figure 6**. As it is shown in **Figure 6**, the spatial volume which represents the uncertainty in the location of landmark, reduces as more frames are processed. Also, the Euclidian norm of the landmark

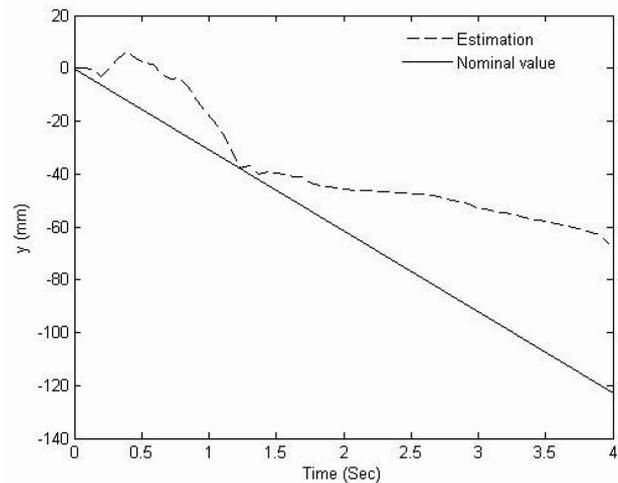


Figure 4. Estimation of the robot in the x-direction versus time.

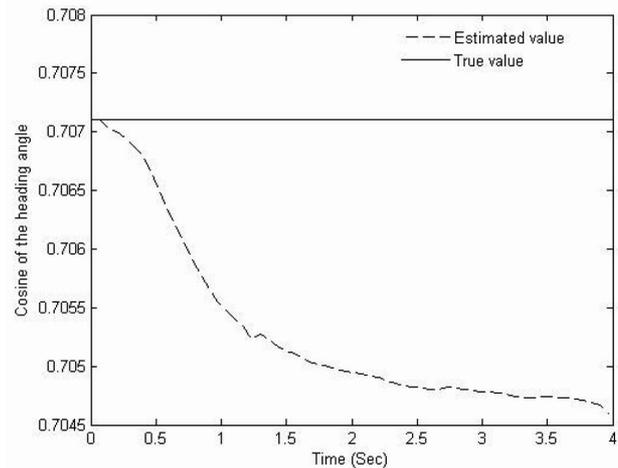
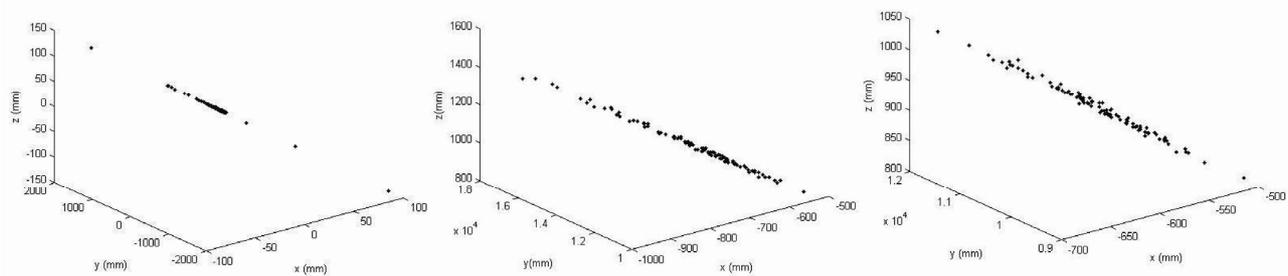


Figure 5. The robot velocity in the x-direction; nominal (solid line), estimated (dashed line).

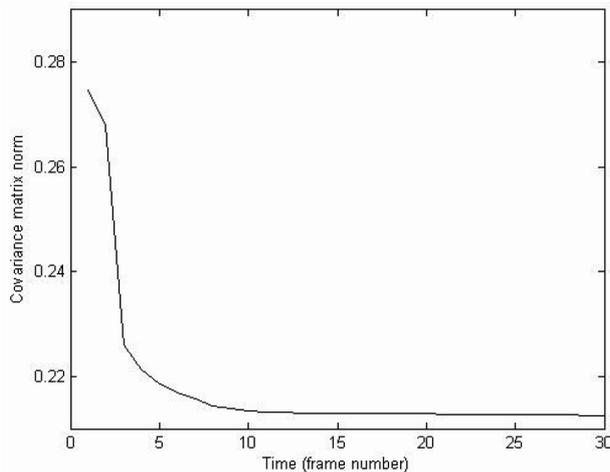
location covariance matrix has been calculated for different frames. The evolution of this norm versus number of frames in which the landmark has been observed is shown in **Figure 7**. As it is shown in **Figure 7**, the covariance norm decreases drastically after a few frames and then converges to a final value. The trend of the covariance norm is compatible with the Monte Carlo simulation results.

## 6. Conclusions

In this paper a new Rao-Blackwellized particle filter suitable for monocular vision (*i.e.*, bearing-only) systems is proposed. In the proposed approach, unlike the traditional formulation where the velocity is considered a constant, filtering is performed under the constant acceleration assumption meaning that the velocity is a state



**Figure 6.** The Monte Carlo simulations of the spatial distributions of a landmark; (from left to right), the first, fourth and seventh frame.



**Figure 7.** Euclidian norm of the landmark location covariance matrix versus time.

variable estimated by the filter. Although the inclusion of higher order terms in a parameterized filter such as an EKF may be considered straightforward, the entire formulation must be reworked for such modification of RBPF. New sampling and importance weight distribution based on the constant acceleration assumption have been derived. Also to solve the bearing only problem, a modified initialization scheme has been proposed. Numerical simulations have been carried out to examine the performance of the proposed system. The higher order RBPF approach has been compared to a lower order RBPF approach in similar conditions. Simulation results verify not only the feasibility of the proposed higher order RBPF approach but also its superiority to the lower order RBPF approach for high parallax angle. Also, an offline experiment has been conducted in which the higher order RBPF approach estimates the path successfully. It is also shown that the proposed filter requires large parallax angle to track the velocity accurately. It is also noted that the generalized higher order formulation is of great importance where there is a need to estimate the velocity as one of the states in the filtering process.

## 7. Acknowledgements

This work was supported in part by the Natural Sciences and Engineering Research Council of Canada (NSERC), Inuktun Services Ltd., Nanaimo, BC, Canada, and MITACS Inc., Canada.

## 8. References

- [1] R. Smith, M. Self and P. Cheeseman, "A Stochastic Map for Uncertain Spatial Relationships," In: R. Bolles and B. Roth, Eds., *The Fourth International Symposium of Robotics Research*, The MIT Press, Cambridge, 1988, pp. 467-474.
- [2] D. Micucci, D. G. Sorrenti, F. Tisato and F. M. Marchese, "Localisation and World Modelling: An Architectural Perspective," *International Journal of Advanced Robotic Systems*, Vol. 3, No. 1, 2006, pp. 79-84.
- [3] M. Csorba and H. F. Durrant-Whyte, "New Approach to Map Building Using Relative Position Estimates," *Navigation and Control Technologies for Unmanned Systems II*, SPIE, Bellingham, 1997, pp. 115-125.
- [4] M. W. M. G. Dissanayake, *et al.*, "A Solution to the Simultaneous Localization and Map Building (SLAM) Problem," *IEEE Transactions on Robotics and Automation*, Vol. 17, No. 3, 2001, pp. 229-241. [doi:10.1109/70.938381](https://doi.org/10.1109/70.938381)
- [5] G. Dissanayake, S. B. Williams, H. Durrant-Whyte and T. Bailey, "Map Management for Efficient Simultaneous Localization and Mapping (SLAM)," *Autonomous Robots*, Vol. 12, No. 3, 2002, pp. 267-286. [doi:10.1023/A:1015217631658](https://doi.org/10.1023/A:1015217631658)
- [6] S. Ahn, J. Choi, N. L. Doh and W. K. Chung, "A Practical Approach for EKF-SLAM in an Indoor Environment: Fusing Ultrasonic Sensors and Stereo Camera," *Autonomous Robots*, Vol. 24, No. 3, 2008, pp. 315-335. [doi:10.1007/s10514-007-9083-2](https://doi.org/10.1007/s10514-007-9083-2)
- [7] U. Frese, "Treemap: An  $O(\log n)$  Algorithm for Indoor Simultaneous Localization and Mapping," *Autonomous Robots*, Vol. 21, No. 2, 2006, pp. 103-122. [doi:10.1007/s10514-006-9043-2](https://doi.org/10.1007/s10514-006-9043-2)
- [8] S. Huang, Z. Wang, G. Dissanayake and U. Frese, "Iterated D-SLAM Map Joining: Evaluating Its Performance in Terms of Consistency, Accuracy and Efficiency," *Autonomous Robots*, Vol. 27, No. 4, 2009, pp. 409-429.

- [doi:10.1007/s10514-009-9153-8](https://doi.org/10.1007/s10514-009-9153-8)
- [9] Y. F. Liu and S. Thrun, "Results for Outdoor-SLAM Using Sparse Extended Information Filters," *Proceedings of the 2003 IEEE International Conference on Robotics and Automation*, Taipei, 2003, pp. 1227-1233.
- [10] M. Montemerlo, S. Thrun, D. Koller and B. Wegbreit, "Fast SLAM: A Factored Solution to the Simultaneous Localization and Mapping Problem," *Proceedings of 18<sup>th</sup> National Conference on Artificial Intelligence (AAAI-02)*, Edmonton, 2002, pp. 593-598.
- [11] K. P. Murphy, "Bayesian Map Learning in Dynamic Environments," *Advances in Neural Information Processing Systems*, Denver, 1999, pp. 1015-1021.
- [12] R. Kümmerle, B. Steder, C. Dornhege, M. Ruhnke, G. Grisetti, C. Stachniss and A. Kleiner, "On Measuring the Accuracy of SLAM Algorithms," *Autonomous Robots*, Vol. 27, No. 4, 2009, pp. 387-407.
- [13] S. Thrun, W. Burgard and D. Fox, "Probabilistic Robotics," The MIT press, Cambridge, 2006.
- [14] M. Montemerlo, S. Thrun, D. Koller and B. Wegbreit, "Fast SLAM 2.0: An Improved Particle Filtering Algorithm for Simultaneous Localization and Mapping that Provably Converges," *International Joint Conference on Artificial Intelligence*, Las Vegas, 2003, pp. 1151-1156.
- [15] A. J. Davison, "Real-Time Simultaneous Localisation and Mapping with a Single Camera," *Proceedings of Ninth IEEE International Conference on Computer Vision*, Nice, 2003, pp. 1403-1410. [doi:10.1109/ICCV.2003.1238654](https://doi.org/10.1109/ICCV.2003.1238654)
- [16] E. Eade, "Scalable Monocular SLAM," 2006 *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Vol. 1, New York, 2006, pp. 469-476.
- [17] N. M. Kwok and A. B. Rad, "A Modified Particle Filter for Simultaneous Localization and Mapping," *Journal of Intelligent and Robotic Systems*, Vol. 46, No. 4, 2006, pp. 365-382. [doi:10.1007/s10846-006-9066-0](https://doi.org/10.1007/s10846-006-9066-0)
- [18] A. Cesetti, E. Frontoni, A. Mancini, P. Zingaretti and S. Longhi, "A Vision-Based Guidance System for UAV Navigation and Safe Landing Using Natural Landmarks," *Journal of Intelligent and Robotic Systems*, Vol. 57, No. 1-4, 2010, pp. 233-257. [doi:10.1007/s10846-009-9373-3](https://doi.org/10.1007/s10846-009-9373-3)
- [19] T. Bailey and H. Durrant-Whyte, "Simultaneous Localization and Mapping (SLAM): Part II-State of the Art," *Robotics and Automation Magazine*, Vol. 13, 2006, pp. 108-117. [doi:10.1109/MRA.2006.1678144](https://doi.org/10.1109/MRA.2006.1678144)
- [20] Y. Bar-Shalom, X. R. Li, T. Kirubarajan and J. Wiley, "Estimation with Applications to Tracking and Navigation," Wiley, New York, 2001. [doi:10.1002/0471221279](https://doi.org/10.1002/0471221279)
- [21] M. Farrokhsiar and H. Najjaran, "A Higher Order Rao-Blackwellized Particle Filter for Monocular vSLAM," *American Control Conference 2010 (ACC-2010)*, Baltimore, 2010, pp. 6987-6992.
- [22] J. M. M. Montiel, J. Civera and A. J. Davison, "Unified Inverse Depth Parametrization for Monocular SLAM," *Proceedings of Robotics: Science and Systems*, 2006.
- [23] D. Törnqvist, T. B. Schön, R. Karlsson and F. Gustafsson, "Particle Filter SLAM with High Dimensional Vehicle Model," *Journal of Intelligent and Robotic Systems*, Vol. 55, No. 4-5, 2009, pp. 249-266. [doi:10.1007/s10846-008-9301-y](https://doi.org/10.1007/s10846-008-9301-y)
- [24] H. F. Durrant-Whyte and T. Bailey, "Simultaneous Localization and Mapping: Part I," *IEEE Robotics & Automation Magazine*, Vol. 13, No. 2, 2006, pp. 99-110. [doi:10.1109/MRA.2006.1638022](https://doi.org/10.1109/MRA.2006.1638022)
- [25] D. G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," *International Journal of Computer Vision*, Vol. 60, No. 2, 2004, pp. 91-110. [doi:10.1023/B:VISI.0000029664.99615.94](https://doi.org/10.1023/B:VISI.0000029664.99615.94)
- [26] S. Se, D. Lowe and J. Little, "Mobile Robot Localization and Mapping with Uncertainty Using Scale-Invariant Visual Landmarks," *The International Journal of Robotics Research*, Vol. 21, No. 8, 2002, p. 735. [doi:10.1177/027836402761412467](https://doi.org/10.1177/027836402761412467)
- [27] F. Caballero, L. Merino, J. Ferruz and A. Ollero, "Unmanned Aerial Vehicle Localization Based on Monocular Vision and Online Mosaicking," *Journal of Intelligent and Robotic Systems*, Vol. 55, No. 4-5, 2009, pp. 323-343. [doi:10.1007/s10846-008-9305-7](https://doi.org/10.1007/s10846-008-9305-7)
- [28] M. Montemerlo and S. Thrun, "FastSLAM, A Scalable Method for the Simultaneous Localization and Mapping Problem in Robotics," Springer, Berlin, 2007.
- [29] A. J. Davison, I. D. Reid, N. D. Molton and O. Stasse, "MonoSLAM: Real-Time Single Camera SLAM," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 29, No. 6, pp. 1052-1067. [doi:10.1109/TPAMI.2007.1049](https://doi.org/10.1109/TPAMI.2007.1049)
- [30] J. Civera, A. J. Davison and J. M. M. Montiel, "Inverse Depth Parametrization for Monocular SLAM," *IEEE Transactions on Robotics*, Vol. 24, No. 5, 2008, pp. 932-945. [doi:10.1109/TRO.2008.2003276](https://doi.org/10.1109/TRO.2008.2003276)
- [31] K. S. Fu, R. C. Gonzalez and C. S. G. Lee, "Robotics: Control, Sensing, Vision, and Intelligence," McGraw-Hill, New York, 1987.
- [32] M. Farrokhsiar and H. Najjaran, "Rao-Blackwellized Particle Filter Approach to Monocular vSLAM with a Modified Initialization Scheme," *Proceedings of 5<sup>th</sup> ASME/IEEE International Conference on Mechatronic and Embedded Systems and Applications*, San Diego, 2009, pp. 427-433.
- [33] J. Shi and C. Tomasi, "Good Features to Track," *Proceedings of 1994 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Seattle, 1994, pp. 593-600.
- [34] J. Y. Bouguet, "A Release of a Camera Calibration Toolbox for Matlab," 2008. [http://www.vision.caltech.edu/bouguetj/calib\\_doc/](http://www.vision.caltech.edu/bouguetj/calib_doc/)
- [35] M. P. Parsley and S. J. Julier, "Avoiding Negative Depth in Inverse Depth Bearing-Only SLAM," *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2008, pp. 2066-2071.