Middle-East Journal of Scientific Research 16 (1): 88-94, 2013
ISSN 1990-9233

DOI: 10.5829/idosi.mejsr.2013.16.01.11624

Indoor Slam Using an Omnidirectional Camera

Khoa D. Phan and Aleksandr V. Ovchinnikov

Department of Automatic Control System, Tula State University, Tula, Russia

Abstract: This paper presents a Simultaneous Localization and Mapping (SLAM) algorithm for an indoor robot using bearing-only observations. An omnidirectional camera is used to observe indoor scene from which vertical lines are extracted to obtain bearing measurements. To track vertical lines through sequence of omnidirectional images, a matching algorithm based on histogram of oriented gradients technique is proposed. The Extended Kalman Filter (EKF) is used to estimate the 3-DoF motion of the robot along with two-dimensional positions of vertical lines in the environment. In order to overcome bearing-only initialization, the Unscented Transform is used to estimate the probability distribution function (PDF) of an initialized vertical line. Simulations have been carried out to validate the proposed algorithm.

Key words: SLAM · Omnidirectional Camera · EKF · Unscented Transform · Bearing-only

INTRODUCTION

Simultaneous Localization and Mapping (SLAM) is the key problem in mobile robotics research. Most implementations of SLAM is related to range-bearing SLAM which requires observations of range and bearing to features from the robot [1, 2]. However, range-bearing SLAM relies on accurate but expensive sensors, such as scanning laser rangefinder or radar. Therefore, inexpensive sensors are suitable for SLAM deployment in many practical applications. And bearing-only SLAM is an attractive solution as it enables the use of inexpensive vision sensors. A well-known bearing-only SLAM system consists of fusing measurements from proprioceptive (odometry) and exteroceptive sensors (vision sensor) by means of an EKF. EKF provides an estimate of the robot pose, as well as an estimate for feature location.

In contrast to range-bearing SLAM, only a single bearing measurement is not sufficient to determine the feature location and at least two measurements are required. However, location estimate may be ill-conditioned if the base-line between a pair of bearing measurements is insufficient [3]. Due to the difficulty of feature initialization, little work has been presented regarding bearing-only SLAM.

In this paper we highlight some of the specific issues faced in bearing-only SLAM using an omnidirectional camera as exteroceptive sensor. Due to the large field of view, features remain longer in sight of the omnidirectional camera that increase accuracy of robot pose estimate. We use the standard EKF based SLAM framework adapted to the visual case to solve the bearing-only SLAM. Vertical lines in the environment, such as doors and wall, are extracted to obtain bearing measurements which are incorporated in the update stage of the EKF. We will present an algorithm for extracting and matching vertical lines on consecutive omnidirectional images. In addition, we propose a feature initialization algorithm based on Unscented Transform allowing us to properly estimate PDF through non-linear function.

The format of the paper is as follows. The next section discusses the previous bearing-only SLAM algorithms. Section III presents proposed algorithms for extracting and matching feature. Section IV discusses our bearing-only SLAM based on EKF approach. Section V specifically describes the feature initialization algorithm. Finally, results of simulation are shown in Section VI and conclusions are made in Section VII.

Related Work: While range-bearing SLAM has received a lot of attention, little work has been presented regarding bearing-only SLAM. Similarly to range-bearing SLAM, bearing-only SLAM is essentially an estimation problem and can be solved using several stochastic techniques, such as Maximum Likelihood approaches, particle filter or EKF.

Corresponding Author: Khoa D. Phan, Department of Automatic Control System, Tula State University, Tula, Russia.

In [4] the author's use a multi-hypothesis filtering approach in which several hypotheses of the position of a landmark are created based along the direction of the first observation of a feature. The validity of the hypotheses is evaluated based on the sequential probability ratio test. Lemaire uses a Gaussian Sum Filter, but place Gaussians along the initial bearing to approximate a uniform uncertainty in depth [5].

Davison solves the initialization problem by assuming a uniform prior for the depth of a landmark [6]. Particle filter is then employed to recursively estimate the feature depth which is not correlated with the rest of the map. Each new observation is used to update the distribution of possible depths, until the variance range is small enough to consider a Gaussian estimation. In [7] the authors propose the algorithm which combines a bundle adjustment for feature initialization and a Kalman filter.

In [3] Bailey stores the robot pose and observation data in the state vector and uses constrained initialization to compute the feature position when robot is at a sufficient distance from the first observation. The Kullback-Leibler distance is used to determine whether feature initialization is well-conditioned. Computational cost of this method is high due to the calculation of the Kullback-Leibler distance.

There are some works that show solutions for SLAM problem using an omnidirectional camera. In [8] the authors integrate Spherical Camera Model for central omnidirectional systems into the EKF-based SLAM by linearizing the direct and the inverse projection. In [9] SLAM algorithm based on the FASTSLAM approach and the Hungarian algorithm for hierarchical data association is proposed.

Vertical Lines Extraction and Matching for Omnidirectional Images: The main advantage of an omnidirectional camera is that it provides a 360° filed of view which gives a very rich information. However, the mirror geometry provides radial distortion and non-uniform resolution on the image, so conventional image processing techniques are not directly applicable in omnidirectional images [10]. In order to apply image processing techniques for conventional cameras, omnidirectional images usually are unwrapped to perspective views which remove the radial distortion. But this procedure is computationally expensive.

In this section we propose an algorithm for vertical line extraction and matching without unwrapping omnidirectional images. Assuming that the axis of the omnidirectional camera is perpendicular to the floor of the environment, all world vertical lines project into radial lines on the image plane.

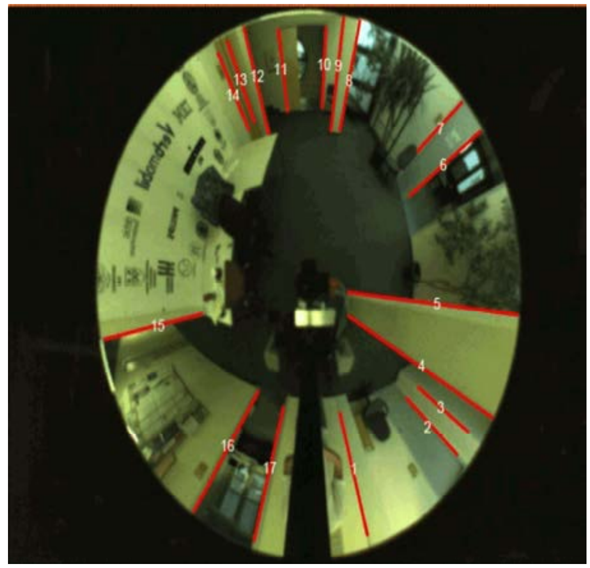

Fig. 1: Vertical line extraction on an omnidirectional image

To extract vertical lines, we first compute the image gradients and pixels whose gradient vector points to the image center are kept. Then the Hough transform is used to extract vertical lines. In order to reduce the number of features to be processed, an omnidirectional image is divided into two parts related to the center of the image. Line segments with the same Hough transform bin are merged if the distance between them is less than the specified value; too small line segments are discarded. Result of vertical line extraction algorithm is shown in Figure 1.

In order to match vertical lines between consecutive omnidirectional images, we use a descriptor, which is unique and invariant to the rotation and illumination change, to represent the local neighborhood of an vertical line. The descriptor is based on the histogram of oriented gradients, which counts occurrences of gradient orientation in interested portions of an image. Then matching can be done by finding the line with the closest descriptor.

To make the descriptor invariant to the rotation, vertical lines are rotated to a fixed direction, such as axis *Ox*. In order to reduce the computational cost, rotation is processed only for the pixels of the line and its neighborhood which forms a rectangular region of interest around the line. Then the region of interest is divided into sectors along the length of the line and symmetric about the line, as shown in Figure 2. In order to improve the

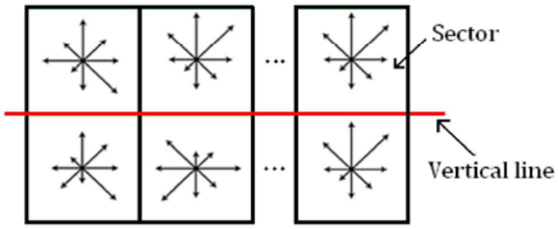

Fig. 2. Descriptor of a vertical line

performance of the descriptor, we apply a Gaussian spatial window, who's the standard deviation equal to one half the width of the region, within the region of interest. It allows us to weight pixels around the edge of the region less. For each sector, we compute a histogram of oriented gradients, each bin of which contains the sum of weighted magnitudes of gradients having the same orientation interval. The descriptor is formed from a vector containing the values of all the histogram of oriented gradients. To reduce the effects of illumination change, descriptor vector is normalized to unit length.

We use the Euclidean distance to determine a distance between two descriptor vectors. The correspondent of a vertical line in the consecutive images can be searched by finding the features with the closest descriptor. However, if a feature doesn't have a correspondent in another image, there is a closest feature anyway. Therefore, to improve the robustness of the comparison, we use the second criteria that the distance from the closest feature must be smaller than the distance from the second closest feature.

EKF-Slam Using Vertical Lines: The state vector of the system containing a robot pose $x_R = [x_r\ y_r\ \phi_r]^T$ and a set of 2D positions of vertical lines $x_L = [x_L\ \ y_L]^T$ at the time step $x[x] = [x_R^T[k] x_{L_1}^T[k]\ \cdots x_{L_n}^T[k]]^T$. The state vector is considered as Gaussian random variable with covariance P, which can be decomposed as:

$$P[k] = \begin{bmatrix} P_{rr} & P_{rl} \\ P_{lr} & P_{ll} \end{bmatrix} \quad (1)$$

where P_{rr} - covariance of the robot pose; P_u - covariance of the map of vertical lines; P_{rl}, P_{lr} – cross-covariance between the two.

At the beginning of time step k, a prior state vector $x^-[k]$ is predicted by a state-transition model which describes the evolution of the state vector from the previous state estimate $x^+(k-1)$ given a control vector $u[k]$:

$$x^-[k] = \begin{bmatrix} f(x_R^+[k-1], u[k], w[k]) \\ x_{L_1}^+[k-1] \\ x_{L_n}^+[k-1] \end{bmatrix} \quad (2)$$

where $u[k]$ is the control vector containing linear and angular velocity of the robot; $w[k]$ is the process noise vector which is assumed to be a Gaussian variable with covariance $Q[k]$.

The prior covariance of the state vector $P^-[k]$ propagated forward via:

$$P^-[k] = \begin{bmatrix} P_{rr}^-[k] & P_{rl}^-[k] \\ P_{lr}^-[k] & P_u^-[k] \end{bmatrix} == \begin{bmatrix} F_R P_{rr}^+[k-1] F_R^T + F_U U[k] F_U^T + Q[k] & F_R P_{rl}^+[k-1] \\ P_{lr}^+[k-1] F_R^T & P_{ll}^+[k-1] \end{bmatrix} \quad (3)$$

where $F_R = \nabla_{xR} f$ and $F_U = \nabla_u f$ are the Jacobian matrices of the state-transition function $f(x_R, u, w)$ w.r.t the robot pose and the control vector respectively.

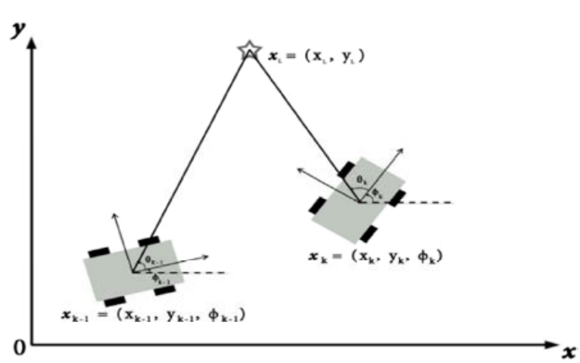

Fig. 3: The robot motion model

The fusion of the observation into the state estimate is accomplished by first calculating a predicted observation, using the observation model $h(x[k], v[k])$:

$$h = arctg\left(\frac{y_r^-[k] - y_L[k]}{x_r^-[k] - x_L[k]}\right) - \phi_r^-[k] + v[k] \tag{4}$$

where $v[k]$ is observation noise with covariance $R[k]$?

When vertical lines are obtained from the omnidirectional camera, they must be matched with initialized ones by the proposed algorithm at section 3. The difference between the actual observation $z[k]$ from the omnidirectional camera and the predicted observation $z^-[k]$ is known as the innovation $v[k]$.

$$v[k] = z[k] - z^-[k] = z[k] - h(x^-[k], v[k]) \tag{5}$$

The innovation covariance $S[k]$ is computed from current state covariance estimate $P^-[k]$, the Jacobian matrices of the observation model $H_R = \nabla_{xR}\, h$, $H_L = \nabla_{xL}\, h$, and the covariance of the observation model $R[k]$.

$$S[k] = H_R P_{rr}^-[k] H_R^T + H_R P_{rl}^-[k] H_L^T + + H_L P_{lr}^-[k] H_R^T + H_L P_{u}^-[k] H_L^T + R[k] \tag{6}$$

The state estimate and covariance are updated using optimal Kalman gain $K[k]$ which provides a weighted sum of the prediction and observation.

$$x^+[k] = x^-[k] + K[k]v[k] \tag{7}$$

$$P^+[k] = P^-[k] - K[k]S[k]K^T[k] \tag{8}$$

$$K[k] = P^-[k](\nabla_x\, h[k])^T\, S^T[k] \tag{9}$$

Unscented Transform Based Feature Initialization: The problem with bearing-only initialization is that a single observation is insufficient to determine the location of the feature and at least two observations from two different vehicle pose are required. The location of the feature is determined by triangulation method [3].

$$x_L = g\left(x_{R_i}, x_{R_j}, \theta_i, \theta_j\right) = \begin{bmatrix} \dfrac{x_{r_i} s_i c_i - x_{r_j} s_j c_i + \left(y_{r_j} - y_{r_i}\right) c_i c_j}{s_i c_j - s_j c_i} \\ \dfrac{y_{r_j} s_i c_i - y_{r_j} s_j c_i + \left(x_{r_i} - x_{r_j}\right) s_i s_j}{s_i c_j - s_j c_i} \end{bmatrix} \tag{10}$$

where $s_i = \sin(\phi_{r_i} + \theta_i), c_i = \cos(\phi_{r_i} + \theta_i)$.

Equation 8 may be extremely ill-conditioned depending on the uncertainty of the pose estimates, observations and the base-line between the two vehicle poses. Figure 4 shows PDF for the depth of a feature estimated from two noisy vehicle poses and bearing measurements at different base-lines. Estimates converge to a Gaussian shape when the base-line between two robot poses increases. Estimates at low base-lines are highly non-Gaussian and have "heavy-tailed" distributions which cannot be correctly processed by EKF. Feature initialization can be deferred until a base-line is sufficient. And a criterion, such as Kullback-Leibler distance [3], must be used to determine whether the estimate is well-conditioned. However, the complexity of methods used to compare two contributions is very high.

In order to overcome the problems of the bearing-only initialization, we propose Unscented Transform based algorithm to estimate PDF of a feature's location. The Unscented Transformation is a method for calculating the statistics of a random variable which undergoes a nonlinear transformation [11]. The state is represented by a Gaussian distribution and a set of points is used to sample the distribution of the state. These sample points

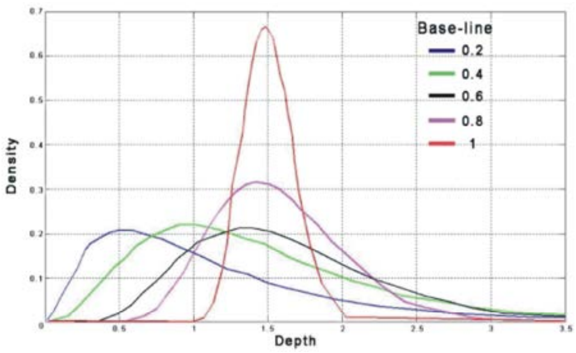

Fig. 4: PDF of the feature location with different base-line

are chosen such that they completely capture the true mean and covariance of the Gaussian distribution. After propagating these sample points through the non-linear system, we capture the posterior mean and covariance accurately to the 3-rd order.

We first form a Gaussian random vector μ which consists two robot poses, two correspondent bearing measurements. All these variables are needed to determine the location of a feature by the triangulation method. When a feature is first observed, the robot pose x_{r_1} and the bearing measurement of the feature θ_1 are stored in the random vector μ. Next observations of the feature and robot poses are alternately added to the random vector μ in order to determine the location of the feature using triangulation method from two bearings.

$$\mu = [x_{r_1}^T \ \ x_{r_i}^T \ \ \theta_1 \ \ \theta_i] \tag{11}$$

Covariance P of the random vector μ is diagonal matrix containing variances of the robot poses and observations. We form a set χ of $2n + 1$ sigma vectors χ_i (with corresponding weights W_i) which samples Gaussian distribution of the random vector μ.

$$X_0 = \mu \tag{12}$$

$$X_i = \mu + \left(\sqrt{(n+\rho)P}\right)_i \quad i = 1 \ldots n \tag{13}$$

$$X_i = \mu - \left(\sqrt{(n+\rho)P}\right)_{i-n} \quad i = n+1 \ldots 2n \tag{14}$$

$$W_0^{(m)} = \frac{\rho}{n+\rho} \tag{15}$$

$$W_0^{(c)} = \frac{\rho}{n+\rho} + (1 - \alpha^2 + \beta) \tag{16}$$

$$W_0^{(m)} = W_0^{(c)} = \frac{1}{2(n+\rho)} \quad i = 1 \ldots 2n \tag{17}$$

where $\rho = \alpha^2(n + \kappa)$ is a scaling parameter. α, κ determine how far the sample points are away from the mean μ. $\beta = 2$ Is optimal for Gaussian distribution. $\left(\sqrt{(n+\rho)P}\right)_i$ is the i-th row of the matrix square root $\sqrt{(n+\rho)P}$.

These sigma vectors are propagated through non-linear function given by Equation 8. The location and covariance estimates of the feature are approximated using weighted sample mean and covariance of the transformed sigma vectors.

$$\overline{X_{L_{new}}} = \sum_{i=0}^{2n} W_i^{(m)} x_{L_i} \tag{18}$$

$$P_{L_{new}} = \sum_{i=0}^{2n} W_i^{(c)} \left(x_{L_i} - \overline{x_{L_{new}}}\right)\left(x_{L_i} - \overline{x_{L_{new}}}\right)^T \tag{19}$$

The skewness of the estimate is calculated to determine whether the estimate is well-conditioned. For a well-conditioned estimate, its skewness must tend to zero.

$$\gamma_1 = \frac{\sum_{i=1}^{2n} W_i^{(c)} \left(x_{L_i} - \overline{x_{L_{new}}}\right)^3}{\sqrt{P_{L_{new}}^3}} \tag{20}$$

An ill-conditioned estimate is deleted and the next observation is obtained to process feature initialization. Once a new feature initialization is considered well-conditioned, its mean and covariance are augmented to the state vector and covariance of the system.

$$x_{aug}^+[k] = [x_r^+[k] \quad x_{L_1}^+[k] \quad \cdots \quad x_{L_n}^+[k] \quad \overline{x_{L_{new}}}]^T \tag{21}$$

$$P_{aug}^+[k] = \begin{bmatrix} P_{rr}^+ & P_{rl}^+ & P_{rr}^+ G_{r_R}^T \\ P_{lr}^+ & P_{ll}^+ & P_{lr}^+ G_{x_R}^T \\ G_{x_R} P_{rr}^+ & G_{x_R} P_{rl}^+ & G_{L_{new}} \end{bmatrix} \tag{22}$$

where G_{x_R} is the Jacobian matrix of the function $\nabla_{x_R} g$ w.r.t the current robot pose.

Experimental Result: Several simulations have been executed in order to test the performance of the proposed algorithm. Simulation provides ground truth and true parameters of the noise statistics. In this experiment,

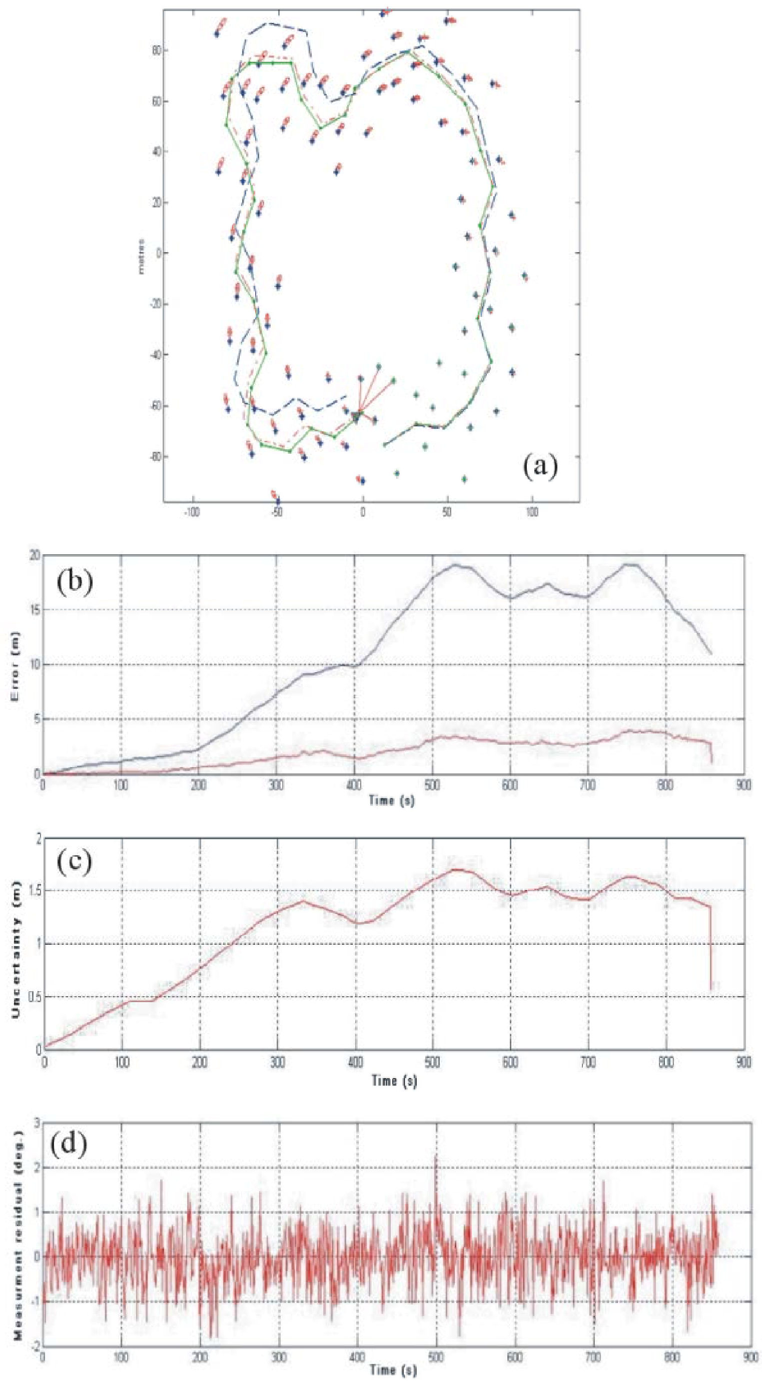

Fig. 5: (a) Result of simulation. (b) Error of estimate for odometry (blue) and proposed algorithm (red). (c) Uncertainty of estimate. (d) Measurement residual

the bearing measurement uncertainty is 1 degree. The simulated environment consists 80 features within a 80 by 80 meters region. The robot travels with average speed of 3 m/s. The standard deviations of the control values are 0.3 m/s and 3 degrees respectively.

The result of simulation is shown in Figure 5a. Here, the green trajectory depicts the true path, the red – the path estimated by the proposed algorithm and the blue – estimated by the odometry only. The blue points represent the true feature locations, the red – the estimated. The ellipses indicate the mean and uncertainty bounds of each feature estimate. The errors and uncertainty in the robot pose estimate by proposed algorithm and the odometry are shown in Figure 5b, 5c. Figure 5d represents the measurement residual between the actual observation from the omnidirectional camera and the predicted observation.

Obviously, the robot pose uncertainty is stable during the simulation which shows that the EKF process is consistent and new feature initializations are well-conditioned. Without the proposed feature initialization algorithm, SLAM diverges as soon as a new feature isn't properly initialized. At the end of the trajectory, the robot has revisited the place with the first observed features and the uncertainty of the estimation dramatically decreases.

CONCLUSION

In this paper we have presented an approach to bearing-only SLAM using the system based on an omnidirectional camera and an odometry. Azimuth of vertical lines extracted from omnidirectional images is used as exteroceptive measurements of the environment which the robot can use for navigation and localization. Conventional image processing techniques are not directly applicable in omnidirectional images, so algorithms for vertical lines extraction and matching are proposed. The bearing information from the omnidirectional camera is exploited in the update phase of an EKF. The feature initialization algorithm based on Unscented Transform is proposed. The Unscented Transform allows us to use the skewness of the estimate to determine whether the estimate is well-conditioned. Finally, we have shown results of simulation which validate the proposed algorithms.

REFERENCES

1. Guivant, J., E. Nebot and H.F. Durrant-Whyte, 2000. Simultaneous localization and map bulding using natural features in outdoor environments. In Sixth International Conference on Intelligent Autonomous Systems, 1: 581-588.
2. Dissanayake, M.W.M.G., P. Newman, S. Clark, H.F. Durrant-Whyte and M. Csorba, 2001. A solution to the simultaneous localization and map building (SLAM) problem. IEEE Transactions on Robotics and Automation, 17(3): 229-241.
3. Bailey, T., 2003. Constrained initialisation for bearing-only SLAM. In Proceedings of the 2003 IEEE International Conference on Robotics and Automation, pp: 1966-1971.
4. Kwok, N.M. and G. Dissanayake, 2004. An Efficient Multiple Hypothesis Filter for Bearing-Only SLAM. In Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems, pp: 736-741.
5. Lemaire, T., S. Lacroix and J. Sola, 2005. A practical 3D bearing-only SLAM algorithm. In Proceedings of IEEE Intelligent Robots and Systems, pp: 2449-2454.
6. Davison, A., 2003. Real-time simultaneous localization and mapping with a single camera. In Proceedings of the ICCV'03, 2: 1403-1410.
7. Deans, M. and M. Hebert, 2000. Experimental comparison of techniques for localization and mapping using a bearing-only sensor. In International Symposium on Experimental Robotics.
8. Rituerto L. Puig and J.J. Guerrero, 2010. Visual SLAM with an Omnidirectional Camera. In Proceedings of International Conference on Pattern Recognition, pp: 348-351.
9. Gamallo, M. Mucientes and C.V. Regueiro, 2013. A FastSLAM-based algorithm for omnidirectional cameras. Journal of Physical Agents, 7: 1.
10. Baker, S. and S.K. Nayar, 1999. A Theory of Single-Viewpoint Catadioptric Image Formation. International Journal on Computer Vision, 35(2): 175-196.
11. Julier, S. and J. Uhlmann, 1997. A new extension of the Kalman filter to nonlinear systems. In International Symposium on Aerospace/Defense Sensing, Simulation and Controls.