XV Reunión de Trabajo en Procesamiento de la Información y Control, 16 al 20 de septiembre de 2013

Monocular Simultaneous Localization and Mapping System for a wheeled mobile robot

Gonzalo F. Perez Paina and Eduardo A. Destéfanis

Research Center in Informatics for Engineering (CIII)
National Technological University, Córdoba Regional Faculty (UTN-FRC)
gperez@scdt.frc.utn.edu.ar

Abstract— **The present work describes the implementation of a monocular SLAM system applied to a wheeled mobile robot moving in an indoor environment. The whole system and details of each part of the current implementation will be described. The parts comprising the system are the estimation filter together with both motion and measurement models, as well as a set of computer vision algorithms for image processing and data association. The implemented visual SLAM makes use of the latest techniques for undelayed landmark initialization which are required given the partial observability of bearing only SLAM. Presented results show the performance of the implementation mainly for robot pose estimation, from which a highly accurate result in robot orientation estimation can be observed.**

Keywords— **Visual SLAM, Mobile robot, Monocular vision, Wheeled robot**

1 INTRODUCTION

Simultaneous Localization and Mapping (SLAM) has been an active research topic for the past decades, given that it solves two of the fundamental problems in order to build truly autonomous mobile robots. Recently, there is an increasing interest in using cameras as exteroceptive sensor for SLAM algorithms. On the other hand, in the computer vision community there is an equivalent problem known as Structure from Motion (SfM). A significant difference between SLAM and SfM exists, whereas the latter is used for off-line applications in batch processing, the former is used for on-line applications.

The early work by Smith *et.al.* [1] established the statistical basis in the treatment of uncertainties using Kalman filtering, which set the foundation for SLAM solutions. The most important contribution of this work was to demonstrate the high correlation between the uncertainties of the sensor pose and map landmark locations, and that these correlations grow with time as new observations are made. Solutions to the SLAM problems where proposed later applying different versions of Bayesian filtering techniques [2], being either Gaussian approaches like EKF (Extended Kalman Filter) [3], UKF (Unscented Kalman Filter) [4], EIF (Extended Information filter (EIF) [5]; or non Gaussian approaches like the PF (Particle filter) [6].

Earlier works on SLAM were based on using sensors like sonars and laser range finders. Nowadays, the available computational power allows the use of digital cameras as the main or even the only environmental sensor. The main advantage of using cameras for SLAM is that they provide a large amount of 3D information for potentially very large distances (ideally up to infinity), besides being lightweight and cheap sensors. Moreover, visual SLAM (vSLAM) can use the accumulated theoretical knowledge and algorithmic solutions of the computer vision research community. This can be used to address two important issues in SLAM, namely feature extraction and data association.

Two approaches related to environment perception using computer vision exist, depending on whether the robot is carrying a monocular or stereo vision system. The latter has significant implications on the observation function: stereo vision is able to obtain the 3D coordinates of scene landmarks, whereas in monocular vision only bearing measurements are observable. Given the partial observability nature of monocular vSLAM (also know as bearing-only SLAM), a feature initialization process is required in order to determine the complete landmark state, mandatory for the estimation filter.

The present work describes the implementation of a monocular SLAM system applied to a mobile robot moving on a planar surface for indoor applications. The main goal is to adapt the general implementation of vSLAM similar to the ones described in [7] for the special case of a mobile platform. The paper is organized as follows: section 2 summarizes the relevant work. Section 3 briefly describes the formulation of SLAM in general and the details for the particular case of vSLAM; whereas section 4 outlines the implemented whole system, and also shows the models used for monocular EKF-SLAM. Results are presented in section 5 and finally section 6 remarks the conclusions and future works.

2 RELATED WORK

An important issue in monocular or bearing-only SLAM is the landmark initialization process, due to the fact that depth is not measured using a single camera. Moreover,

in the context of the classical EKF-SLAM solution, a recently observed landmark needs to be fully described by a Gaussian density in order to be added to the map. Two different approaches exist, namely Delayed Landmark Initialization (DLI) and Undelayed Landmark Initialization (ULI) [8]. DLI has a major disadvantage which is that new landmarks are not immediately used when they are observed for the first time, in the correction of the camera pose. Furthermore, it prevents the use of distant landmarks (points at infinity) which are useful for orientation estimation [7].

The early work on monocular SLAM of Davison [9], which used a DLI process, shows that the standard EKF formulation for SLAM can work properly with a single camera as the only source of information. Davison's approach uses a kind of particle filter to approximate the landmark depth coordinate, until the distribution collapses sufficiently in order to be represented by a Gaussian distribution. However, this approach has the main limitation that can only deal with nearby landmarks that exhibit significant parallax during camera motion, limiting the application to room-scale scenes. Although Davison's monocular SLAM has been designed for a 3D camera motion estimation (6DOF, Degrees of Freedom), it cannot deal with sudden changes in motion direction given the constant velocity motion model used.

A widely used method for ULI in monocular SLAM is the Inverse Depth Parametrization (IDP) [7]. This parametrization allows efficient and accurate representation of the initial landmark depth uncertainty, and is able to work within the standard EKF. Explicit parametrization of the inverse depth can cope with depth uncertainties by means of a Gaussian distribution, spanning depth range from nearby up to infinity. IDP is a unified representation requiring no special landmark initialization process, allowing an immediate contribution to improve the camera pose estimation.

3 MONOCULAR VISUAL SLAM

3.1 EKF SLAM

Kalman filtering involves the estimation of the state of a discrete-time dynamic system defined by

$$\mathbf{x}_k = \boldsymbol{f}(\mathbf{x}_{k-1}, \mathbf{u}_{k-1}) + \mathbf{w}_k, \tag{1}$$

$$\mathbf{z}_k = \boldsymbol{h}(\mathbf{x}_k) + \mathbf{v}_k, \tag{2}$$

where $\mathbf{x}_k$ is the unobserved system state, $\mathbf{u}_k$ is the known control action, $\mathbf{z}_k$ is the observation, $\mathbf{w}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_k)$ is the process noise, and $\mathbf{v}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_k)$ is the observation noise, all in the time step k.

The state vector $\mathbf{x} = \begin{bmatrix}\mathbf{x}_R^T & \mathbf{x}_M^T\end{bmatrix}^T$ is composed of the robot state $\mathbf{x}_R = \begin{bmatrix}x_r & y_r & \theta_r\end{bmatrix}^T$, and the map state $\mathbf{x}_M = \begin{bmatrix}\mathbf{x}_1^T & \dots & \mathbf{x}_N^T\end{bmatrix}^T$, where $\mathbf{x}_i, i = 1, \dots, N$, are landmarks representing the environment map. As usually done in SLAM, the map is considered to be static, i.e. $\mathbf{x}_{M,k} = \mathbf{x}_{M,k-1} = \mathbf{x}_M$.

The goal of the Kalman filter is to optimally estimate the state $\mathbf{x}_k$ given the observations $\mathbf{z}_k$ up to time k. The EKF works in two stages: prediction and update. In the prediction stage the prior estimate is computed

$$\begin{aligned}\hat{\mathbf{x}}_k^- &= \boldsymbol{f}(\hat{\mathbf{x}}_{k-1}, \mathbf{u}_{k-1}) \\ \mathbf{P}_k^- &= \mathbf{F}_k \mathbf{P}_{k-1} \mathbf{F}_k^T + \mathbf{Q}_k,\end{aligned} \tag{3}$$

and the update stage corrects the actual estimation using new information from measurements

$$\begin{aligned}\mathbf{K}_k &= \mathbf{P}_k^- \mathbf{H}_k^T \left(\mathbf{H}_k \mathbf{P}_k^- \mathbf{H}_k^T + \mathbf{R}_k\right)^{-1} \\ \hat{\mathbf{x}}_k &= \hat{\mathbf{x}}_k^- + \mathbf{K}_k \left(\mathbf{z}_k - \boldsymbol{h}(\hat{\mathbf{x}}_k^-)\right) \\ \mathbf{P}_k &= \left(\mathbf{I} - \mathbf{K}_k \mathbf{H}_k\right) \mathbf{P}_k^-,\end{aligned} \tag{4}$$

where $\mathbf{F}_k$ and $\mathbf{H}_k$ are the Jacobian matrices of the SLAM process (1) and measurement (2) functions, respectively.

3.2 Landmark parametrizations

A 3D scene point in Euclidean representation is described by means of its three Cartesian coordinates as

$$\mathbf{x}_{\text{EU}} = \begin{bmatrix}X & Y & Z\end{bmatrix}^T \in \mathbb{R}^3, \tag{5}$$

which is projected to an image point on the image plane using the pin-hole model as

$$\underline{\mathbf{m}} = \mathbf{K}\mathbf{R}_W^C(\mathbf{x}_{\text{EU}} - \mathbf{t}_C^W) \in \mathbb{P}^2, \tag{6}$$

where $\underline{\bullet}$ stands for homogeneous coordinate in the projective space $\mathbb{P}^n$, $\mathbf{K}$ is the camera intrinsic parameters matrix given by

$$\mathbf{K} = \begin{bmatrix} f/h_u & 0 & u_0 \\ 0 & f/h_v & v_0 \\ 0 & 0 & 1 \end{bmatrix} \tag{7}$$

and $\{\mathbf{t}_C^W, \mathbf{R}_C^W\}$ represents the rigid transformation between the world (WCS) and camera (CCS) coordinate systems; f is the focal length, and h_u and h_v are the pixels width and height, respectively. Euclidean points present significant non linearities in the observation function for monocular SLAM, being inappropriate for the ULI process. The most adequate and widely used parametrization for ULI is the Anchored Modified Polar Point (AMPP) [10] also known as Inverse Depth Parametrization (IDP) [7]. The AMPP is represented by: the camera position $\mathbf{t}_0$ (anchored point) when the landmark is first observed, the azimuth and elevation angles (γ, ϕ) of the optical ray (expressed in WCS) joining $\mathbf{t}_0 = \begin{bmatrix}x_0 & y_0 & z_0\end{bmatrix}^T$ and the observed 3D point, and the inverse of the distance d from $\mathbf{t}_0$ to this point, $\rho = 1/d$. The resulting map landmark is

$$\mathbf{x}_{\text{ID}} = \begin{bmatrix}x_0 & y_0 & z_0 & \gamma & \phi & \rho\end{bmatrix}^T \in \mathbb{R}^6. \tag{8}$$

The projection to the image plane is

$$\underline{\mathbf{m}} = \mathbf{K}\mathbf{R}_W^C \left(\rho(\mathbf{t}_0 - \mathbf{t}_C^W) + \mathbf{d}(\gamma, \phi)\right) \in \mathbb{P}^2, \tag{9}$$

where $\mathbf{d}(\gamma, \phi)$ is the unit vector given by the azimuth and elevation angles, that is

$$\mathbf{d}(\gamma, \phi) = \begin{bmatrix} \cos\phi \sin\gamma \\ -\sin\phi \\ \cos\phi \cos\gamma \end{bmatrix}.$$

The back-projection which is used in the features initialization process is given by

$$\mathbf{x}_{\text{ID}} = \begin{bmatrix} \mathbf{t}_0 \\ (\gamma, \phi) \\ \rho \end{bmatrix} = \begin{bmatrix} \mathbf{t}_C^W \\ \boldsymbol{g}(\mathbf{R}_C^W \mathbf{K}^{-1} \underline{\mathbf{m}}) \\ \rho_C \end{bmatrix}, \quad (10)$$

where ρ_C is the initial inverse depth (prior information), and $\boldsymbol{g}(\cdot)$ gives azimuth and elevation angles from unit vector (pointing the optical ray) $\mathbf{r} = \begin{bmatrix} r_x & r_y & r_z \end{bmatrix}^T$ as

$$\boldsymbol{g}(\mathbf{r}) = \begin{bmatrix} \arctan(r_x/r_z) \\ \arctan\left(-r_y/\sqrt{r_x^2 + r_z^2}\right) \end{bmatrix} = \begin{bmatrix} \gamma \\ \phi \end{bmatrix}.$$

Further details on IDP can be found in [7].

4 IMPLEMENTATION OF VISUAL SLAM

4.1 General description

The presented implementation of monocular SLAM is based on a wheeled mobile robot carrying a single camera. All the parameters of the robot/camera system are considered to be known and are obtained by off-line calibration procedures. The intrinsic camera parameters are obtained by calibration using standard techniques [11], and the rigid transformation or extrinsic parameters relating the CCS and robot coordinate system (RCS) can also be obtained experimentally [12].

The implemented monocular SLAM used an ULI approach based on inverse depth parametrization. As was previously mentioned, ULI has the main advantage in that new observed landmarks are immediately used to improve the estimation. Given that IDP is an over-parametrization, map landmarks in IDP are converted to Euclidean representation as soon as they become adequately linear in the measurement function, reducing the dimension of the SLAM state vector. This conversion follows the approach presented in [13].

Figure 1 shows a schematic representation of the implemented monocular SLAM system. The main parts of the implementation are the mobile robot along with its sensors (odometry and camera), the estimation filter, and computer vision algorithms for image processing and data association. The implemented system uses a very simple map management approach, which is responsible of landmarks addition and deletion. New landmarks are added to the map based on a minimum threshold of image features (N_i), and they are deleted considering a threshold of the rate between the number of times a given map

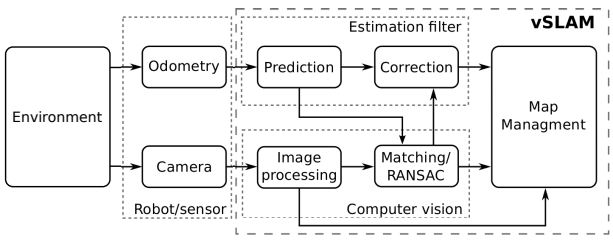

Figure 1: Monocular visual SLAM system.

landmark is predicted (N_p), and the number of times it is used for the filter correction (N_c). The values used are $N_i = 50$ and $N_c/N_p = 0.5$, resulting in a landmark deletion if it is predicted and not used by the filter 50% of the time.

4.2 Motion model

The model used corresponds to the probabilistic odometric motion model presented in [2]. This model uses relative motion readings from the robot odometry as the control action input, which is comprised of a first rotation, a translation and a last rotation, $\mathbf{u} = \begin{bmatrix} \delta_{rot1} & \delta_{trans} & \delta_{rot2} \end{bmatrix}^T$. Given the odometric reading at time step $k-1$, $\mathbf{x}_{k-1}^{odom} = \begin{bmatrix} x_{k-1}^{odom} & y_{k-1}^{odom} & \theta_{k-1}^{odom} \end{bmatrix}^T$ and at time step k, $\mathbf{x}_k^{odom} = \begin{bmatrix} x_k^{odom} & y_k^{odom} & \theta_k^{odom} \end{bmatrix}^T$, control action is composed of

$$\begin{aligned} \delta_{rot1} &= atan2(y_k^{odom} - y_{k-1}^{odom}, x_k^{odom} - x_{k-1}^{odom}) - \theta_{k-1}^{odom} \\ \delta_{trans} &= \sqrt{(x_{k-1}^{odom} - x_k^{odom})^2 + (y_{k-1}^{odom} - y_k^{odom})^2} \\ \delta_{rot2} &= \theta_k^{odom} - \theta_{k-1}^{odom} - \delta_{rot1}. \end{aligned}$$

Assuming these variables are affected by zero mean Gaussian noise with

$$\begin{aligned} \sigma_{rot1} &= \alpha_1 |\delta_{rot1}| + \alpha_2 |\delta_{trans}| \\ \sigma_{trans} &= \alpha_3 |\delta_{trans}| + \alpha_4 (|\delta_{rot1}| + |\delta_{rot2}|) \\ \sigma_{rot2} &= \alpha_1 |\delta_{rot2}| + \alpha_2 |\delta_{trans}|, \end{aligned}$$

where $\alpha_i, i = 1, \ldots, 4$ are the motion parameters specific to the robot used. The covariance matrix representing the uncertainty in motion action is a diagonal matrix $\mathbf{P}_{\mathbf{u}} = \text{diag}(\sigma_{rot1}^2, \sigma_{trans}^2, \sigma_{rot2}^2)$. The robot state evolves according to

$$\begin{aligned} \mathbf{x}_{R,k} &= \boldsymbol{f}(\mathbf{x}_{R,k-1}, \mathbf{u}_{k-1}, \mathbf{w}_{k-1}) \\ \begin{bmatrix} x_k \\ y_k \\ \theta_k \end{bmatrix} &= \begin{bmatrix} x_{k-1} \\ y_{k-1} \\ \theta_{k-1} \end{bmatrix} + \begin{bmatrix} \delta_{trans} \cos(\theta_{k-1} + \delta_{rot1}) \\ \delta_{trans} \sin(\theta_{k-1} + \delta_{rot1}) \\ \delta_{rot1} + \delta_{rot2} \end{bmatrix}. \end{aligned}$$

Putting together the robot motion model and the static map assumption results

$$\mathbf{x}_k = \begin{bmatrix} \mathbf{x}_{R,k} \\ \mathbf{x}_{M,k} \end{bmatrix} = \begin{bmatrix} \boldsymbol{f}(\mathbf{x}_{R,k-1}, \mathbf{u}_{k-1}, \mathbf{w}_{k-1}) \\ \mathbf{x}_{M,k-1} \end{bmatrix}, \quad (11)$$

which is the process equation for SLAM. A detailed description of this model can be found in [4].

4.3 Measurement model

Given the SLAM state vector $\mathbf{x}_k$ composed of the robot state vector $\mathbf{x}_R$ and the map state vector $\mathbf{x}_M = \begin{bmatrix} \mathbf{x}_{\square,1}^T & \cdots & \mathbf{x}_{\square,i}^T & \cdots & \mathbf{x}_{\square,N}^T \end{bmatrix}^T$, where the i-th map feature $\mathbf{x}_{\square,i}$ can be represented either in euclidean $\mathbf{x}_{EU,i}$ or inverse depth $\mathbf{x}_{ID,i}$ parametrization. The measurement function (2) is composed of two steps: first each map feature is projected to the image plane (in the actual camera pose) using (6) or (9) depending on the

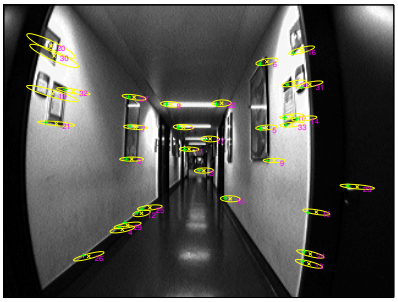

Figure 2: Measurement prediction and feature matching.

parametrization; and second, a model for lens radial distortion is applied. Distorted image points are

$$\mathbf{z}_i = \begin{bmatrix} u_i \\ v_i \end{bmatrix} = \begin{bmatrix} u_0 + \frac{u-u_0}{1+\kappa_1 r_d^2+\kappa_2 r_d^4} \\ v_0 + \frac{v-v_0}{1+\kappa_1 r_d^2+\kappa_2 r_d^4} \end{bmatrix}$$
$$r = r_d(1 + \kappa_1 r_d^2 + \kappa_2 r_d^4)$$
$$r = \sqrt{(h_x(u-u_0))^2 + (h_y(v-v_0))^2}$$

where $\mathbf{m} = \begin{bmatrix} u & v \end{bmatrix}^T$, κ_1 and κ_2 are the radial distortion parameters. The Gaussian additive measurement noise has a standard deviation of 1 pixel. Figure 2 shows predicted measurements in image plane where the uncertainties are represented by ellipses (in yellow color). All the shown uncertainty ellipses follow a specific direction given by the epipolar point, which is due to the use of the odometric motion model. Using an adequate motion model, like the odometric motion model presented here instead of a constant velocity model, allows a better measurement prediction which also improves the data association process.

4.4 Data association

When using a camera as exteroceptive sensor for SLAM, feature extraction and data association problems can be addressed by existing algorithms in the computer vision community. Feature extraction is based on an image interest point detector, being commonly used Harris, SIFT and SURF detectors. The data association process seeks for correspondences between predicted measurements and feature descriptors. Moreover, data association needs an extra stage in order to reject wrong matches also known as outliers. The latter is typically done using the RANSAC algorithm [14].

The proposed approach for feature extraction and data association is as follows: image features are extracted using the FAST interest point detector [15], and image patches are saved as descriptors. The matching process seeks for correspondences between these patches and interest regions on the image defined by measurement predictions using prior information given by the filter. Finally, the 1-Point RANSAC algorithm [16] is used for outlier rejection. Interest regions for feature matching are obtained similarly to [17], where the authors propose to define search regions or bounding boxes in the image plane, by projecting four tangent planes to the 3D ellipsoids representing uncertainties of map landmarks. Instead, we propose to directly use the uncertainties of predicted measurements (in image plane) which are also used by the filter. Interest regions are defined by bounding boxes which are determined by tangent lines to the ellipses representing a constant Mahalanobis distance of predicted measurements. Given a set of predicted measurements $\{\hat{\mathbf{z}}_i^-, \mathbf{R}_i^-\}$, a constant Mahalanobis distance is expressed by

$$(\mathbf{z} - \hat{\mathbf{z}}_i^-)^T \left(\mathbf{R}_i^-\right)^{-1} (\mathbf{z} - \hat{\mathbf{z}}_i^-) = k, \qquad (12)$$

which can be represented in homogeneous coordinate [18] in the projective plane as

$$\underline{\mathbf{m}}^T \mathbf{C} \underline{\mathbf{m}} = 0 \qquad (13)$$

where $\underline{\mathbf{m}} = \begin{bmatrix} u & v & w \end{bmatrix}^T$ is an homogeneous image point, and (the conic) $\mathbf{C}$ is an homogeneous matrix representing the uncertainty ellipse. Tangent lines to the conic are defined as

$$\underline{\mathbf{l}} = \mathbf{C} \underline{\mathbf{m}} \qquad (14)$$

where $\underline{\mathbf{l}} = \begin{bmatrix} a & b & c \end{bmatrix}^T$ with $a = 0$ for horizontal lines and $b = 0$ for vertical ones. Points on both ellipse $\mathbf{C}$ and tangent lines are obtained solving the system of equations given by (13) and (14) where $\underline{\mathbf{m}} = \begin{bmatrix} u & -c & b \end{bmatrix}^T$ and $\underline{\mathbf{m}} = \begin{bmatrix} -c & v & b \end{bmatrix}^T$ for horizontal and vertical tangent lines, respectively.

Figure 3 shows an indoor image with one predicted measurement represented by the uncertainty ellipse, together with the bounding box defined by tangent (vertical and horizontal) lines to the ellipse. This bounding box defines the image interest region to seek for matching. Moreover, the feature patch (of 41x41 pixels) which is saved for each image features, and the warped patch (of 13x13 pixels) are also shown. The latter is obtained applying a warping transformation based on the camera predicted pose given by the filter. The similarity measurement used for matching is the Normalized Cross-Correlation. Results of image feature matching can be seen in Fig. 2 where crosses (in green color) show the matching points, corresponding to maximum cross-correlation values.

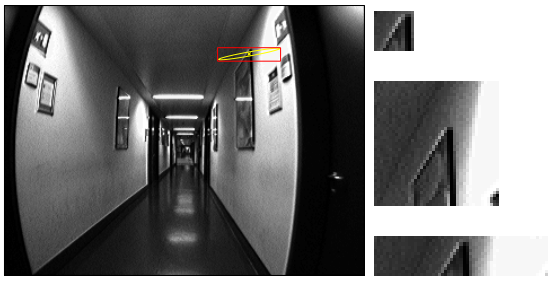

Figure 3: Image feature matching. *Left:* Measurement prediction, uncertainty and bounding box. *Right:* interest region (bottom), feature patch (middle) and warped patch (top).

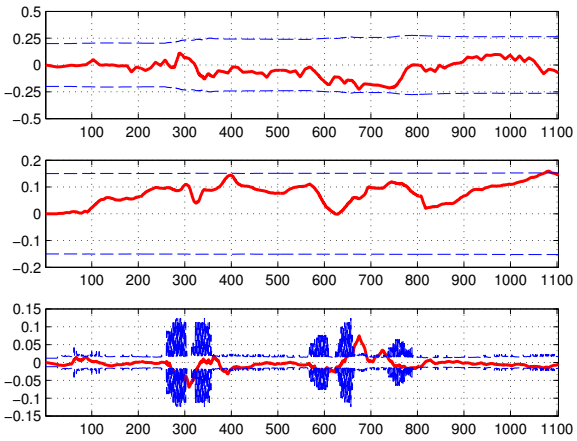

Figure 4: Estimation error in robot pose (x_r, y_r, θ_r) (from top to bottom respectively), together with $\pm 3\sigma$ bounds. Horizontal axis is the algorithm step.

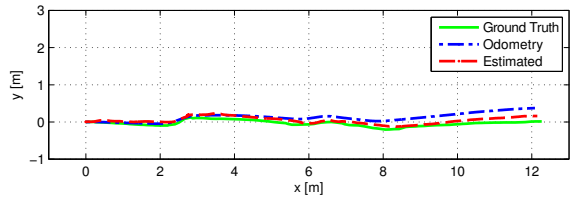

Figure 5: Ground truth (solid line), odometry (dash-dot line) and estimated (dashed line) robot path for 12m corridor.

5 RESULTS

The evaluation of the implemented monocular SLAM algorithm is performed using the RAWSEEDS dataset [19]. This dataset includes information from different sensors taken by a robot moving in an indoor environment, including a digital camera of 320×240 pixels. In addition, the robot pose ground truth is also available which is aimed at evaluating the performance of the SLAM and robot pose estimation algorithms. The motion model parameters used are: $\alpha_1 = 1$, $\alpha_2 = 35 \times 10^{-3}$, $\alpha_3 = 2 \times 10^{-2}$ y $\alpha_4 = 5 \times 10^{-2}$; and the measurement noise is $\sigma_z = 1$ pixel.

Figure 4 shows the estimation error of the robot pose together with the uncertainties in the estimation given by $\pm 3\sigma$ bounds, and Fig. 5 shows the robot path; both of the same SLAM running. These results are obtained with the robot moving forward (increasing its x coordinate) along a corridor similar to Fig. 3. A variation of the error in the robot's x coordinate can be observed due to the fact that a bearing only sensor, like a single camera, cannot measure the scene depth. However, despite the limited camera field of view this error can be reduced when the robot approximate to distant map landmarks. Similarly, the error in the robot's y coordinate presents a significant variation apart from a slight drift, due also to the limited camera field of view. On the other hand, even though the error in the robot orientation estimation is low, it grows significantly when the robot is rotating (near step 300 and 700). The latter can affect the data association process. At the end of the path of approx. 12m the position error is near 5cm in the robot x coordinate, and near 15cm in the y coordinate.

Figure 6 shows the robot pose estimation error together with the error using robot odometry, similar errors for both estimated and odometric robot x coordinate can be appreciated. However, estimated error in y coordinate is of 1m against 3.5m for odometry apart of being unbounded. As previously, a low estimation error in orientation is observed for the whole path. Lastly, Fig. 7 shows the odometric, estimated and ground truth path for the whole running corresponding to Fig. 6.

6 CONCLUSIONS

The implementation of an EKF-based monocular SLAM system applied to a wheeled mobile robot was presented. It is based on the state of the art of undelayed landmark initialization using the inverse depth parametrization. Each part of the presented monocular SLAM system was shown in details. The implemented algorithm was

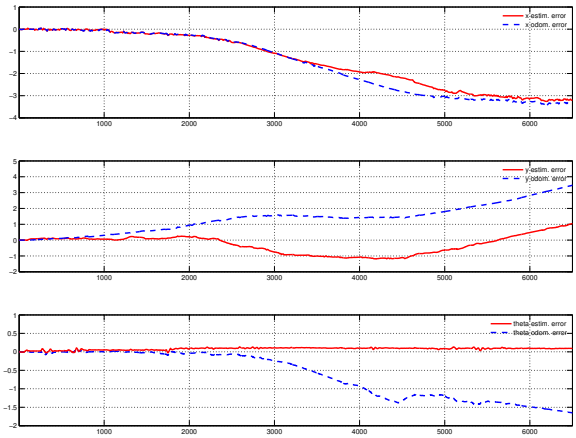

Figure 6: Robot pose odometry (dashed line) and estimation (solid line) error, in (x_r, y_r, θ_r) (from top to bottom respectively). Horizontal axis is the algorithm step.

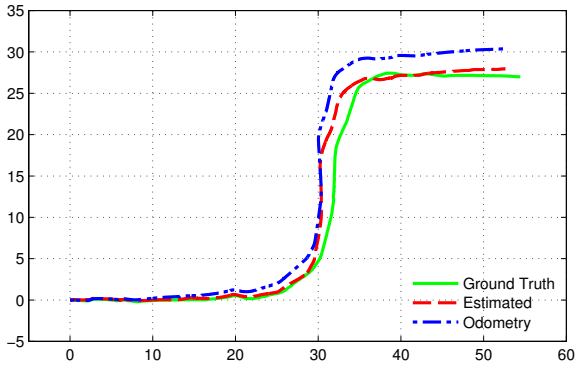

Figure 7: Ground truth (solid line), odometry (dash-dot line) and estimated (dashed line) robot path of length more than 70m.

tested using a freely available dataset developed specifically for visual SLAM evaluation.

Presented results show the behavior of monocular SLAM focused mainly on robot pose estimation. Even though a single camera is not able to perceive scene depth, besides of having a limited field of view, it was shown that it can be used in monocular SLAM for robot pose estimation. Furthermore, the precision of the implemented approach used for robot pose estimation was verified with real data, for both a straight path in a typical indoor corridor and for a longer paths. Results demonstrated, as theory suggests that a monocular vision system acts as a very precise orientation sensor, mainly due to the use of inverse depth parametrization for undelayed landmark initialization.

Future work includes the proposition of a new strategy for image feature detection and matching, based on robust image point descriptors. This can improve data association and therefore the accuracy of the filter estimation.

Acknowledgments

The authors thanks the anonymous referees for their comments and suggestions that led to an improvement of the present work.

This work is partially funded by the project "Autonomous Vehicle Guidance Fusing Low-cost GPS and other Sensors", PICT-PRH-2009-0136.

References

[1] R. Smith, M. Self, and P. Cheeseman, "Estimating uncertain spatial relationships in robotics," in *Robotics and Automation. Proc. 1987 IEEE Int. Conf. on*, vol. 4, Mar 1987, pp. 850–850.

[2] S. Thrun, W. Burgard, and D. Fox, *Probabilistic Robotics (Intelligent Robotics and Autonomous Agents)*. The MIT Press, 2005.

[3] M. Dissanayake, P. Newman, S. Clark, H. Durrant-Whyte, and M. Csorba, "A solution to the simultaneous localization and map building (slam) problem," *Robotics and Automation, IEEE Trans. on*, vol. 17, no. 3, pp. 229–241, Jun 2001.

[4] G. P. Paina, C. Paz, M. Baudino, A. Delfino, and E. Destéfanis, "Implementation and performance evaluation of ukf for simultaneous localization and mapping," in *VII Jornadas Arg. de Robótica*, 2012.

[5] F. Auat Cheein, G. Steiner, G. Perez Paina, and R. Carelli, "Optimized eif-slam algorithm for precision agriculture mapping based on stems detection," *Comput. Electron. Agric.*, vol. 78, pp. 195–207, Sep 2011.

[6] M. Montemerlo, S. Thrun, D. Koller, and B. Wegbreit, "FastSLAM 2.0: An improved particle filtering algorithm for simultaneous localization and mapping that provably converges," in *Proc. of the Sixteenth Int. Joint Conf. on Artificial Intelligence (IJCAI)*. Acapulco, Mexico: IJCAI, 2003.

[7] J. Civera, A. J. Davison, and J. M. M. Montiel, "Inverse depth parametrization for monocular SLAM," *IEEE Trans. on Robotics*, vol. 24, no. 5, pp. 932–945, Oct. 2008.

[8] J. Solà, T. Lemaire, M. Devy, S. Lacroix, and A. Monin, "Delayed vs undelayed landmark initialization for bearing-only SLAM," in *In Proceedings of the the IEEE Int. Conf. on Robotics and Automation workshop on SLAM - Workshops*, 2005.

[9] A. Davison, "Real-time simultaneous localisation and mapping with a single camera," in *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*, Oct 2003, pp. 1403 –1410 vol.2.

[10] J. Solà, T. Vidal-Calleja, J. Civera, and J. Montiel, "Impact of landmark parametrization on monocular EKF-SLAM with points and lines," *International Journal of Computer Vision*, pp. 1–30, Sep 2011.

[11] Z. Zhang and Z. Zhang, "A flexible new technique for camera calibration," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 22, pp. 1330–1334, 1998.

[12] G. Araguás, G. P. Paina, G. Steiner, and L. Canali, "Extrinsic calibration of a camera-robot system under non-holonomic constraints." AST, Argentine Symp. on Technology (JAIIO), 2011, pp. 157–167.

[13] J. Civera, A. Davison, and J. M. M. Montiel, "Inverse depth to depth conversion for monocular slam," in *Robotics and Automation, 2007 IEEE International Conference on*, 2007, pp. 2778–2783.

[14] M. A. Fischler and R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol. 24, no. 6, pp. 381–395, Jun 1981.

[15] E. Rosten and T. Drummond, "Fusing points and lines for high performance tracking." in *IEEE Int. Conf. on Computer Vision*, vol. 2, October 2005, pp. 1508–1511.

[16] J. Civera, O. G. Grasa, A. J. Davison, and J. M. M. Montiel, "1-point ransac for ekf-based structure from motion," in *Proc. of the 2009 IEEE/RSJ Int. Conf. on Intelligent robots and systems*, ser. IROS'09. IEEE Press, 2009, pp. 3498–3504.

[17] G. Bresson, T. Féraud, R. Aufrere, P. Checchin, and R. Chapuis, "A new strategy for feature initialization in visual slam," in *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, 2011, pp. 115–120.

[18] R. I. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*. Cambridge University Press, ISBN: 0521540518, 2004.

[19] S. Ceriani, G. Fontana, A. Giusti, D. Marzorati, M. Matteucci, D. Migliore, D. Rizzi, D. Sorrenti, and P. Taddei, "Rawseeds ground truth collection systems for indoor self-localization and mapping," *Autonomous Robots*, vol. 27, no. 4, pp. 353–371, 2009.