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

***Abstract*— The present work describes the implementation of a monocular visual SLAM applied to a wheeled mobile robot moving in an indoor environment. It will be described the whole system and details of each part of the current implementation. Parts comprising the system include 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 make use of the latest techniques for undelayed landmark initialization which are required given the partially observability of bearing only SLAM. Presented results show the performance of the implementation mainly for robot pose estimation, from which it can be observed a highly accurate result in robot orientation estimation.**

***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 exist an equivalent problem known as Structure from Motion (SfM). However, a significant difference between SLAM and SfM exist, the latter is used for off-line applications in batch processing whereas the former is used in on-line cases.

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 biggest contribution of this work was to demonstrate the high correlation between the uncertainties of the sensor pose and feature locations, and that these correlations grow with time as new observations are made. Solutions of 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 approach like the PF (Particle filter).

Earlier works of SLAM in robotics were based on using sensors like sonars, and laser range finders. Nowadays, the available computational power allows to use 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 until infinity), beside of being lightweight and cheap sensors. Moreover, visual SLAM (vSLAM) can use the accumulated theoretical knowledge and algorithmic solutions of the computer vision research community, which can be used to address two important issues in SLAM, namely feature extraction and data association.

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

The present work describes the implementation of a monocular visual SLAM system applied to a mobile robot moving on a plane surface for indoor applications. The main goal is to adapt the general implementation of vSLAM similar to the described in [6] for the special case of a mobile platforms. The paper is organized as follows: section 2 summarizes relevant works. 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 used models for monocular visual EKF-SLAM case. Results are presented in section 5 and finally section 6 remarks the conclusions and ideas for future works.

2 RELATED WORKS

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 classical EKF-SLAM solution, a recently observed landmark need to be fully described by a Gaussian

density in order to be added to the map. There exist two different approaches, namely Delayed Landmark Initialization (DLI) and Undelayed Landmark Initialization (ULI) [7]. DLI has the major disadvantage due to new features are not immediately used when they are seen for the first time in the correction of the camera pose. Moreover, it prevents the use of distant features (points at infinity) which are useful for orientation estimation [8].

The early work on monocular SLAM of Davison [9], which used an 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 a particle filter to approximate the depth coordinates until the distribution collapsed sufficiently in order to be represented by a Gaussian distribution. However, this approach has the main limitation that can only use nearby features which exhibit significant parallax during camera motion, limiting the application to room-scale scenes. Thought, Davison's implementation of monocular SLAM has been designed for a 3D motion (6DOF, Degrees of Freedom), it cannot deal with sudden change in motion direction given it uses a constant velocity motion model.

A widely used method for ULI in monocular SLAM is the Inverse Depth Parametrization (IDP) [10]. This parametrization allows efficient and accurate representation of uncertainties during feature initialization being 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 to infinity. This parametrization is an unified representation requiring no special feature initialization process, allowing an immediately 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 k denotes the discrete time step, $\mathbf{x}_k$ represents the unobserved state of the system, $\mathbf{u}_k$ is a known control action, and $\mathbf{z}_k$ is the observed measurement. The process noise $\mathbf{w}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_k)$ drives the dynamic system, and the observation noise is given by $\mathbf{v}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_k)$. The dynamic model $\boldsymbol{f}(\cdot)$ and measurement model $\boldsymbol{h}(\cdot)$ are assumed to be known.

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 $\mathbf{x}_{M,k} = \mathbf{x}_{M,k-1} = \mathbf{x}_M$.

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 which are: 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 eq. (1), and the SLAM measurement eq. (2), 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

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

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

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

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 pixels width and height, respectively. Euclidean points present significant non linearities in the observation function for monocular visual SLAM, being inappropriate for ULI process. The most adequate and widely used parametrization for ULI is the Anchored Modified Polar Point (AMPP) [11] also known as Inverse Depth Parametrization (IDP) [6]. The AMPP is represented by the camera position $\mathbf{t}_0$ (anchored point) when the feature is observed for the first time, 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 3D point, and the inverse depth to the point $1/\rho$ measured from $\mathbf{t}_0$. The resulting map feature 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 given by

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

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

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

where ρ_C is 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

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

IDP is described in details in [6] and [10].

4 IMPLEMENTATION OF VISUAL SLAM

4.1 General description

The presented implementation of monocular visual 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, which are obtained by calibration procedures previously to the SLAM running. The intrinsic camera parameters are obtained by calibration using standard techniques [12], and the rigid transformation or extrinsic parameters relating the CCS and robot coordinate system (RCS) can also being obtained experimentally [13].

The implemented monocular SLAM used an ULI approach based on inverse depth parametrization. As was previously mentioned, ULI has the main advantage of using new observed features immediately improving the estimation. Given that IDP is an overparametrization, 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 [14].

Figure 1 shows an schematic representation of the implemented monocular visual SLAM (vSLAM) system. The main parts of the implementation are the mobile robot along its sensor (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 adding and deleting map landmarks. Map landmarks addition is base on a minimum image features in each time step; and the deletion is determined by a rate between the amount of time a map landmark is predicted (in image plane) and use for filter correction.

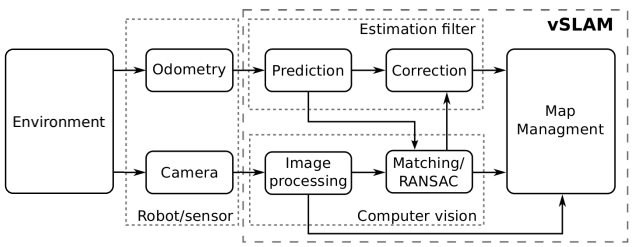

Figure 1: Monocular visual SLAM system

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 compose of

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

Assuming these variables are affected by Gaussian zero mean noise with

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

where $\alpha_i, i = 1, \ldots, 4$ are the motion parameters specific to the used robot. 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 accordingly to

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

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 (eq. (2)) is composed of two steps: first each map feature is projected to the image plane (in the actual camera pose) using eq. (6) or (9) depending on the parametrization; and second, a model for lens radial distortion is applied. Distorted image points are

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

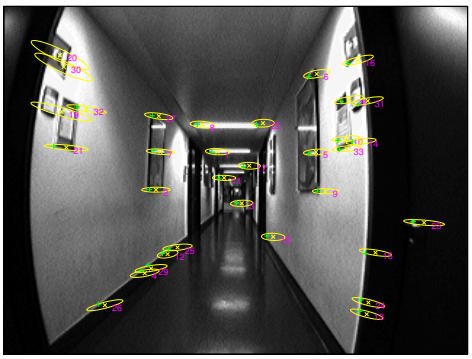

Figure 2: Measurement prediction and feature matching.

where $\mathbf{m} = \begin{bmatrix} u & v \end{bmatrix}^T$, κ_1 and κ_2 are the radial distortion parameters. The additive measurement noise considered to be Gaussian has a standard deviation of 1 pixel. Figure 2 shows predicted measurement 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 usage of the odometric motion model. The main difference in using an adequate motion model, like the odometric motion model presented here instead of constant velocity model, allows a better measurement prediction which also improve 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. Features extraction is based on a image interest point detector, being commonly used Harris, SIFT and SURF detectors. Data association process seek for correspondences between predicted measurements and feature descriptors. Moreover, it is needed an extra stage in order to reject wrong matches in data association, also known as outliers. The latter is typically done using RANSAC [15].

The proposed approach for feature extraction and data association is as follows: image features are extracted using the FAST interest point detector [16], and image patches are saved as descriptors. The matching process seek for correspondences between these patches and interest regions on the image defined by measurement predictions using prior information estimated by the filter. Finally, the 1-Point RANSAC algorithm [17] is used for outlier rejection. Interest regions for features matching are obtained similarly to [18], where the authors propose to define search or interest regions in image plane, aligned with image axes, by projecting four tangent planes to the 3D ellipsoids representing uncertainties of map landmarks. Instead, here is proposed to used directly 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

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

which can be represented in homogeneous coordinate [19] for the image projective plane as

$$\underline{\mathbf{m}}^T \mathbf{C} \underline{\mathbf{m}} = 0 \tag{13}$$

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

$$\underline{\mathbf{l}} = \mathbf{C} \underline{\mathbf{m}} \tag{14}$$

where $\underline{\mathbf{l}} = \begin{bmatrix} a & b & c \end{bmatrix}^T$ with $a = 0$ for horizontal lines and $b = 0$ for vertical lines. Points on both ellipse $\mathbf{C}$ and tangent lines are obtained solving the system of equation 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 and the bounding box defined by tangent (vertical and horizontal) lines to the ellipse. The bounding box defines the image interest region to seek for matching. Furthermore, it is also shown the feature patch (of 41x41 pixels) which is saved for each image features, and the warped patch (of 13x13 pixels). The latter is obtained applying a warping transformation based on the camera predicted pose given by the filter. The used similarity measurement for matching is the normalized cross-correlation. Results of image features matching can be seen in Fig. 2 where crosses (in green color) show the matching point, corresponding to maximum cross-correlation values.

5 RESULTS

The evaluation of the implemented monocular SLAM algorithm is performed using the RAWSEEDS dataset [20]. This dataset includes information from different sensors taken by a robot moving in an indoor environment, including a digital camera of 320 × 240 pixels. Furthermore, robot

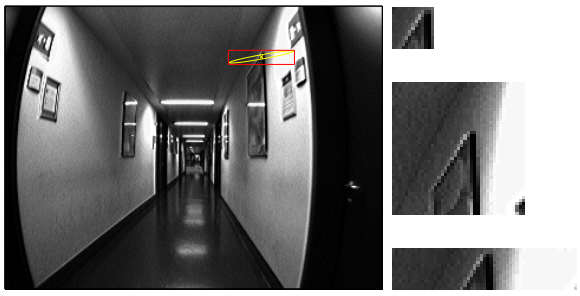

Figure 3: Image feature matching. *Left:* Measurement prediction, uncertainty and bounding box. *Right:* Interest region (bottom), feature patch (middle) and warped patch (top).

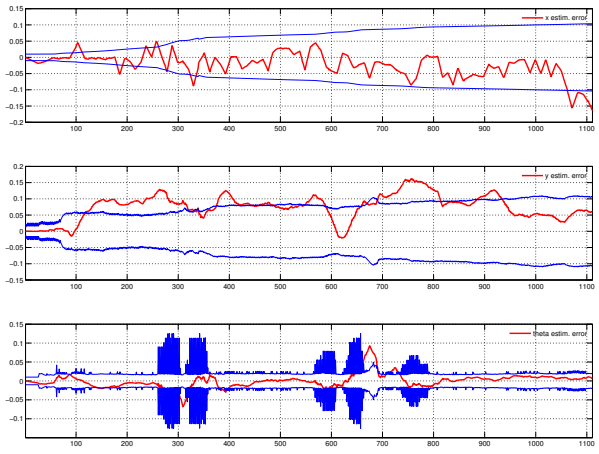

Figure 4: Estimation error in robot pose (x_r, y_r, θ_r) (from top to bottom respectively), together with 3σ bounds.

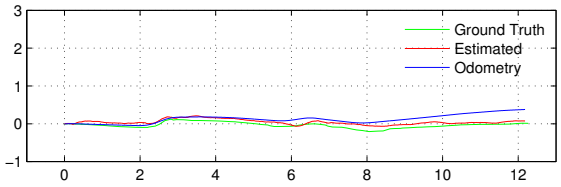

Figure 5: Ground truth (green), odometry (blue) and estimated (red) robot path for 12m corridor.

pose ground truth is also available which is aimed to evaluate the performance of SLAM and robot pose estimation algorithms.

Figure 4 shows the estimation error of robot pose together with the uncertainties in the estimation given by 3σ bounds, and Fig. 5 shows the robot path; both of the same SLAM running. These results are obtained with the robot moving along a corridor similar to Fig. 3. It can be observed how the uncertainty of the estimated robot y coordinate decrease after the robot deviate from the initial straight path. This is due to a better estimation of map landmarks after they are seen from different camera orientations. On the other hand, the estimation error in x coordinate at the end of the path (of approx. 12m) is near to 15cm (1.25%), given that a single camera cannot measure depth. However, even though the limited field of view of the camera the estimated y coordinate is rather accurate, being of less than 5cm at the end of the shown path. It is worthy to note the low estimation error in robot orientation showing that a monocular system is an accurate orientation sensor.

Figure 6 shows the robot pose estimation error together with the error using robot odometry, similar errors in robot x coordinate for both estimated and odometric 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 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. It was shown in details each part of the presented monocular visual SLAM system. The implemented algorithm was tested using a freely available dataset developed specifically for visual SLAM evaluation.

Presented results show the behavior of the monocular SLAM focused mainly for robot pose estimation. Even though a single camera as the only sensor 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. Moreover, it was verified with real data the precision of the implemented approach used for robot pose estimation for a straight path in typically indoor corridors, as well as for a longer path. Results demonstrated, as theory suggests that a monocular vision system acts as a very precise orientation sensor, mainly due to the usage of inverse depth parametrization for undelayed landmark initialization.

Future works includes the proposition of a new strategy

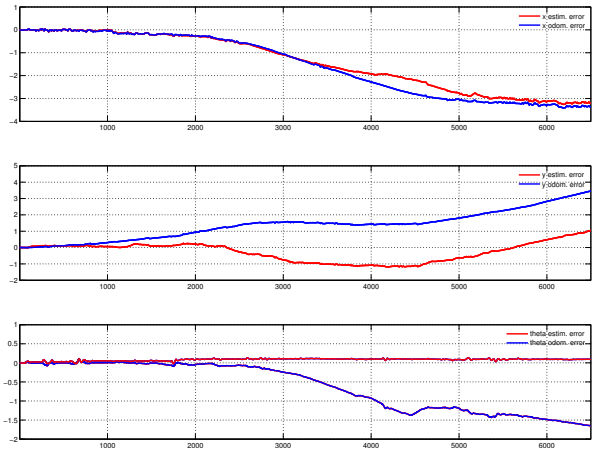

Figure 6: Robot pose odometry and estimation error, in (x_r, y_r, θ_r) (from top to bottom respectively).

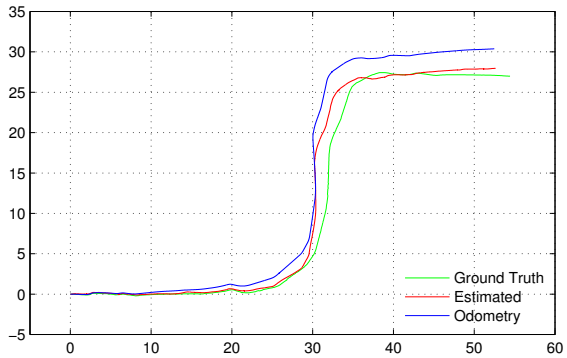

Figure 7: Ground truth (green), odometry (blue) and estimated (red) robot path of length more than 70m.

for image feature detection and matching, based on robust image point descriptors. This can improve data association and therefore the accuracy in the filter estimation.

Acknowledgments

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] 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.

[7] 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.

[8] J. M. M. Montiel and A. Davison, "A visual compass based on slam," in *Robotics and Automation, 2006. ICRA 2006. Proceedings 2006 IEEE International Conference on*, 2006, pp. 1917–1922.

[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. Montiel, J. Civera, and A. Davison, "Unified Inverse Depth Parametrization for Monocular SLAM," in *Robotics: Science&Systems*, July 2006.

[11] 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.

[12] 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.

[13] 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.

[14] 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.

[15] 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.

[16] 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.

[17] 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.

[18] 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.

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

[20] 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.