Mobile Robot Geometry Initialization from Single Camera

Daniel Pizarro, Manuel Mazo, Enrique Santiso, Hideki Hashimoto

Mobile Robot Geometry Initialization from Single Camera

Daniel Pizarro[1], Manuel Mazo[1], Enrique Santiso[1] and Hideki Hashimoto[2]

[1] Department of Electronics, University of Alcala,
Esc. Politecnica, 28871 Alcala de Henares, Spain
{pizarro,mazo,santiso}@depeca.uah.es

[2] Hashimoto Lab, University of Tokyo
4-6-1 Komaba, Meguro-ku, Tokyo 153-8505, Japan
hashimoto@iis.u-tokyo.ac.jp

Summary. Using external cameras to achieve robot localization has been widely proposed in the area of Intelligent Spaces. Recently, an online approach that simultaneously obtains robot's pose and its 3D structure using a single external camera has been developed [8]. Such proposal relies on a proper initialization of pose and structure information of the robot. The present paper proposes a solution to initialization which consists of retrieving 3D structure and motion of a rigid object from a set of point matches measured by the camera. A batch Structure from Motion (SFM) approach is proposed along a short path. By incorporating odometry information available in the robot, the ambiguity generated by a single view in the solution is solved. We propose to describe robot's motion and image detection as statistical processes in which the uncertainty is properly modelled. Using a Gaussian equivalence of the processes involved, the SFM cost function is expressed as a Maximum Likelihood optimization. The paper shows the improvements of the approach in the presence of the usual odometry drift noise, compared with those using Euclidean distance as a likelihood. The proposed method is assessed on synthetic and real data.

1 Introduction

The usage of external cameras to obtain the pose of a robot is inspired in the concept of 'Intelligent Space' [6]. The aim is to provide the environment with intelligence and to reduce its necessity in the robots. As a consequence the space is composed by a set of sensors (*e.g.* Cameras), a distributed intelligence, and a network layer, which allows communication with agents (*e.g. robots*). Under this framework, robot devices derive their localization and the high level navigation tasks from the space.

In this paper the usage of a single camera, external to the robot is proposed as a sensor device for robot localization. The ability of operating with a single sensor allows the system to work with few cameras, reducing heavily the cost and avoiding complex multiple camera configurations. There exist several approaches with the same aim that the present paper, by using artificial landmarks, either active [4] or passive ones [7]. In other works a geometrical model [10], is learnt previously to the tracking task in a supervised step. The common point of many of the proposals comes from the fact that high amount of prior knowledge is used for a posterior tracking.

In [8], a sequential algorithm that recovers the pose (with respect a global coordinate origin) and structure (set of 3D points from robot's structure) of a mobile robot is presented. The main novelty of this proposal is that it uses only as prior information the rigidity assumption in the geometry and the odometry readings present in the robot. The approach is based on retrieving the robot's pose by using a motion model and a sparse geometrical model which is learnt online. Such method depends heavily on a proper initialization of pose and structure before applying the online method.

In this work the initialization or self-calibration required in the online method is solved using one camera and odometry information. The initialization consists of a batch sequence of images of the robot taken during a short path which are synchronized with the odometry readings. In this paper a Structure from Motion (SFM) algorithm is proposed to solve the pose and structure by minimizing a likelihood function of the residuals in image frame. As the position estimation with odometry has a growing error the algorithm must tackle properly the uncertainty in the odometry metric reference.

In [2] the authors obtain the orientation and position of a camera with respect to a robot by using artificial markers and a set of predefined motions for calibration. The error in odometry position estimation and its influence in the solution is studied and the predefined motions are selected to minimize calibration error.

In the present paper we propose to describe robot's motion and image detection as statistical processes in which the uncertainty is properly modelled. Using a Gaussian equivalence of the processes involved, the SFM cost function is expressed as a Maximum Likelihood optimization. The main contribution of the paper is to show the improvements of this approach in the presence of the usual odometry drift noise, compared with those using Euclidean distance as a likelihood [9]. Besides, in this paper the natural structure of the robot is retrieved jointly with its pose.

We state the problem in §2. In §3 the likelihood distribution is analytically obtained from measurements and inputs. The final M.L cost function is presented in §4. Results on synthetic and real images are presented in §5. Finally, conclusions are presented in §6.

2 Problem Statement

As was commented before, the problem to solve is to achieve Structure from Motion by using one camera and the given odometry information which has growing error with robot's motion. To formalize the particularities of the problem, in this section it will be roughly commented on the notation and information involved.

Firstly, a set of discrete processes concerning pose, structure and measurements are presented. The number of time samples k_N is considered to be finite and the proposed method runs as a batch process, so all the measurements and inputs are available from the beginning.

Robot's pose at time k is described by a vector x_k. Usually for 3D motion (6 D.O.F) the vector is composed of 3 position components and 3 orientation angles $x_k = (r_k^x, r_k^y, r_k^z, \alpha, \beta, \gamma)$. For wheeled robots which its motion lie on a plane, the vector pose x_k is reduced to 3 components (r_k^x, r_k^y, ϕ). Motion model $x_k = f(x_{k-1}, u_k)$ obtains actual position with respect to previous time and some input u_k given by odometry (i.e. angular speed and linear speed of the robot).

The geometry of the robot is composed by a sparse set of N 3D points $\mathcal{M} = \{m^1, \cdots, m^M\}$ referred from a local coordinate origin described by robot's pose x_k. The rigidity in geometry involves that the points $\mathcal{M}$ are static in time and no temporal subindex is necessary. Function $m_x^i = t(x_k, m^i)$ uses actual pose x_k to express m^i in the global coordinate origin.

The camera sensor follows a perfect "pin-hole" described by its 3x4 projection matrix P, which encodes intrinsic and extrinsic parameters. The camera projection model is expressed by the non-homogeneous transformation $y = h(m_x, P)$ which converts a 3D point m_x expressed in a global coordinate origin into its 2D projection y in the image plane using camera parameters P. The set of measurements which correspond to structure points is encoded in vector $Y_k = \{y_k^i \quad i = 1, \cdots, M\}$ where $y_k^i = h(t(x_k, m^i), P)$. Each measurement in image plane is obtained by using a fiducial detector which also is able to establish correspondences between images.

The objective proposed in this paper is to retrieve the initial pose x_0 and the set $\mathcal{M}$ given a set of image measurements $Y_{1:N_K}$ and odometry input information $u_{1:N_k}$.

In the next sections a statistical model of all processes is proposed. A random process is defined by its P.D.F and it is expressed in bold typography. (i.e. $\mathbf{x_k} = N(\hat{x_k}, \Sigma_k)$ is the Gaussian random process and x_k is a realization of it).

3 Uncertainty Modeling

In this section, the objective is to obtain the statistical description of the processes stated in §2. By describing motion and image measurements as random processes, the uncertainty present in them is properly tackled and the odometry unbound error is directly modeled. In this paper it is shown how such representation improves the solution obtained in the reconstruction algorithm compared with a simple quadratic cost function.

The statistical models appear by including uncertainty noise variables into projection (h) and motion (f) models. We propose to approximate all statistical processes using Gaussian density distributions, which gives a meaningful description of the solution (mean) and uncertainty (covariance). The main difficulties come from the fact that the motion and projection models are in general non-linear equations, which transform the input Gaussian process and involves solving integral equations.

A first order approximation of the non linear function is used for propagating Gaussian processes. This approach has been widely used in Extended versions of Kalman Filtering [3]. It is known to suffer from bias [11], however it offers an analytical solution for the mean and covariance which will be necessary to minimize the likelihood cost function.

In the following sections the processes involved will be tackled in detail.

3.1 Motion uncertainty

Initial solution for pose x_0 is represented by a Gaussian process $\mathbf{x_0} = N(\hat{x}_0, \Sigma_0)$ with a mean vector $\hat{x}_0$ and a covariance matrix Σ_0. Between time samples, $\mathbf{x_k}$ is consider a Markov Process with a transition kernel $p(x_k|x_{k-1}, u_k)$ ruled by the motion model f.

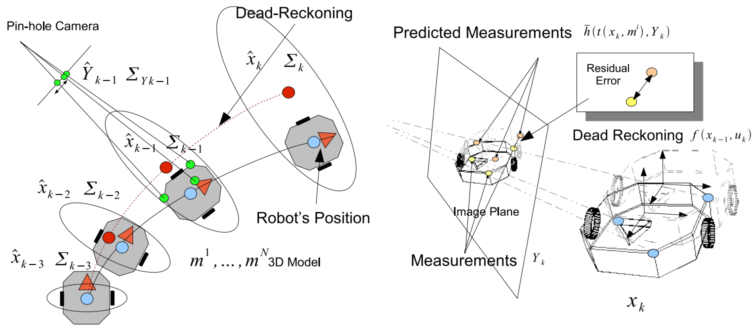
For simplicity, motion model used in this paper corresponds to a simple wheeled differential robot which moves over a ground plane. Pose parameters $x_k = (r_k, \phi_k)$ will be composed of position in the ground plane r_k with respect a global coordinate origin O and orientation ϕ_k:

$$x_k = x_{k-1} + \begin{pmatrix} (Vl_k + W_k^1)\cos(\phi_{k-1} + \Omega_k + W_k^2) \\ (Vl_k + W_k^1)\sin(\phi_{k-1} + \Omega_k + W_k^2) \\ \Omega_k + W_k^2 \end{pmatrix}, \tag{1}$$

Input from odometry is represented by the vector $u_k = (Vl_k, \Omega_k)$ with linear and angular speed components. Noise from odometry readings is modeled by a Gaussian process $\mathbf{W}_k = (\mathbf{W}_k^1, \mathbf{W}_k^2)$ which is statistically independent given x_{k-1}. Gaussian approximation of the distribution $p(x_k|x_0) = N(\hat{x_k}, \Sigma_k)$ is recovered by propagating the last pose $p(x_{k-1}|x_0)$ and the actual odometry u_k through the transition kernel $p(x_k|x_{k-1}, u_k)$. The result is a Gaussian process with continuously growing variance usually presented in Dead-Reckoning [1] processes (See Figure 1).

3.2 Measurement uncertainty

Measurement uncertainty depends on the method used to track fiducial points in the image. Measurements y_k^i are jointly attached to form a vector Y_k which



(a) Uncertainty Growing with odometry (b) Image residual information

Fig. 1.

will be described as a measurement process $\mathbf{Y_k}$. Each separate measurement y_k^i is contaminated by additive Gaussian noise $\mathbf{v_k^i} = N(0, \Sigma_v)$. The noise components are independent between each other and are included in the noise process $\mathbf{V}_k$.

Likelihood distribution $p(Y_{1:k_N}|x_0, \mathcal{M})$ represents the joint distribution of the whole set of measurements in the sequence. It is obtained by combining projection model from every measurement y_k^i into a global measurement equation $\bar{h}$:

$$\begin{aligned} Y_{1:k_N} &= \bar{h}(x_{1:k_N}, Y_{1:k_N}, V_{1:k_N}) \quad V_k = \{V_k^i \quad i = 1, \cdots, M\} \\ \bar{h} &= \left(h(t(x_1, m^1), P) + V_1^1, \cdots, h(t(x_{k_N}, m^M), P) + V_{k_N}^M\right) \in \mathbb{R}^{2 \cdot M \cdot k_N} \end{aligned} \quad (2)$$

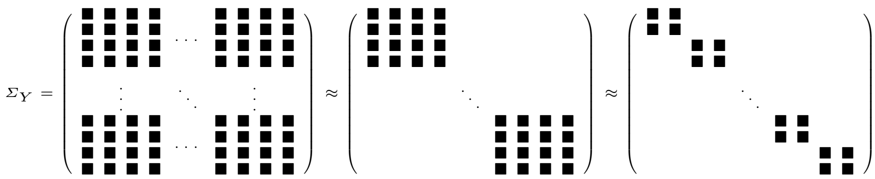
A Gaussian approximation of $p(Y_{1:k_N}|x_0, \mathcal{M}) = N(\hat{Y}_{1:k_N}, \Sigma_Y)$ is then obtained. It is used in this paper as the likelihood from which the solution to initial pose and structure is retrieved.

Usually, $\Sigma_Y \in \mathbb{R}^{n_Y \times n_Y}$ where $n_Y = 2Mk_N$, is a full correlated covariance matrix, in which all cross-correlation terms are present. The cross correlation terms will play an important role in the reconstruction step.

Assuming independence between each $\mathbf{x}_k$, the resulting covariance will be approximated by a box-diagonal one, where only terms from the same frame are correlated. In the same way, assuming that at time k there exists independence between different measurements, the resulting covariance matrix will be box-diagonal where only coordinates of the same measurement can be correlated. (See (3)).

In the next section, the differences and the importance of correlated terms will be explained. In §5, some synthetic experiments are run to test the approaches selected for Gaussian propagation.

a) Full Covariance b) Time Ind. c) Frame & Meas. Ind.



(3)

4 Maximum Likelihood Solution

Rigid structure from motion algorithms are at the moment in a mature state [5]. Thanks to the advances made in the last ten years, it can be considered a solved problem given that the correspondence between frames is correct. The usage of non-linear 'Bundle Adjustment' [9] with robust cost functions allows to obtain a solution for the SFM problem.

'Bundle Adjustment' makes use of reprojection errors to get robot initial pose x_0 and the set $\mathcal{M}$:

$$\min_{x_0,\mathcal{M}} \sum_{k}^{N_k} \sum_{i}^{M} \rho(y_k^i - h(t(x_k, m^i), P)), \tag{4}$$

where function ρ strictly depends on the reprojection error statistical properties. In this paper the cost function ρ is naturally obtained by maximizing the logarithm of the joint distribution $p(Y_{1:k_N}|x_0, \mathcal{M})$. By using Maximum Likelihood optimization, the cost function penalizes those residuals $y_k^i - h(t(x_k, m^i), P)$ which has more error due to odometry drift.

4.1 Maximum Likelihood weighted Least Squares

By taking logarithm, the likelihood function proposed $p(Y_{1:k_N}|x_0, \mathcal{M})$, and approximated by a Gaussian function, is maximized by minimizing the following quadratic cost function:

$$\min_{x_0,\mathcal{M}} (Y_{1:k_N} - \hat{Y}_{1:k_N})^T (\Sigma_Y)^{-1} (Y_{1:k_N} - \hat{Y}_{1:k_N}) \tag{5}$$

It must be noticed that both Σ_Y and $\hat{Y}_{1:k_N}$ are functions of the parameters to minimize x_0 and $\mathcal{M}$.

It is straightforward to see that when the covariance matrix Σ_Y is the identity, the resulting cost function is directly the one presented in (4) with square of residuals as ρ.

Expression (5) can be optimized by using any non-linear iterative optimization method such as Gauss-Newton or Region-Trust algorithms as Levenberg Mardquardt.

4.2 Sparse Hessian

Each basic step of optimization involve the calculation of the following operations: We define as $\Phi_n = \{x_0, \mathcal{M}\}_n$ as the value of the parameters to minimize at iteration number n.

Given Φ_{n-1} the solution for the increment $\Phi_n = \Phi_{n-1} + \Delta\Phi$ at iteration n is obtained by solving the following linear system:

$$J^T(\Sigma_Y)^{-1}J\Delta\Phi = J^T(\Sigma_Y)^{-1}(Y_{1:k_N} - \hat{Y}_{1:k_N}(\Phi_{n-1})), \tag{6}$$

where J represents the first derivative of $\hat{Y}_{1:k_N}$ evaluated at Φ_{n-1}.

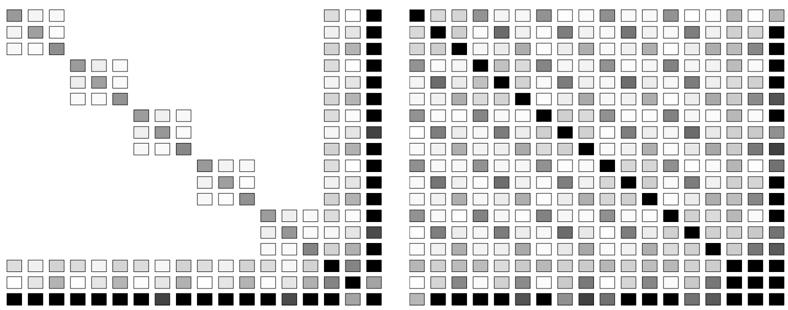
Matrix $H = J^T(\Sigma_Y)^{-1}J$ which is a square matrix of the size of Φ represents the approximated Hessian.

At each iteration system (6) is solved by inverting H.

$$\Delta\Phi = (H)^{-1}b, \tag{7}$$

If P_Y is a diagonal matrix, usually H have a sparse structure. This allows a very important improvement in the computational complexity for its inversion. The situation is specially critic when $\mathcal{M}$ contains a large number of 3D points.

As we can see in Figure 2, the inclusion of a full covariance Σ_Y or only a time correlated one, destroys the natural sparsity of the problem. The direct effect is that the systems become of higher complexity.



(a) Elements of H with diagonal or F.M.I Boxed diagonal Σ_Y

(b) Elements of H with full covariance or F.I Boxed diagonal Σ_Y

Fig. 2. Hessian Configurations

4.3 Initial guess

As a non-linear iterative algorithm, the proposed optimization requires an initial guess for vector Φ. The assumption is that if robot height is small compared to the distance to the camera the points from robot structure can be supposed to lie on the ground plane. Using the set of measurements from the first frame Y_0, and knowing the position of the ground plane, a set of 3D positions from global reference frame can be obtained, namely X_0. The set $\mathcal{M}$ is referred to the initial pose x_0. By using centroid of the set X_0 the position r_0 is easily obtained.

5 Results

In this section some results on synthetic generated data and real images are presented.

5.1 Synthetic data

The experiment conditions are the following:

- A set of M=30 points m_k^i initialized as a random Gaussian jointly with robot true initial position r_0. The geometry has a deviation around the center of rotation of $\sigma = 200(mm)$
- Camera calibration and ground position from a real configuration.
- Odometry noise variance Σ_W is varied from $\sigma^2 I_{2x2}$ $\sigma = 0.1, \cdots, 20$ (m/s,rad/s).
- Measurement noise Σ_V with standard deviation of around 0.1 pixels.

The following algorithms will be tested:

1. Element Diagonal covariance estimation (E.D): By supposing that Σ_Y is diagonal. No weighted optimization.
2. Frame Box Diagonal covariance matrix (F.D): See Figure 2(a).
3. Measurement Box Diagonal covariance (M.D): See Figure 2(b).

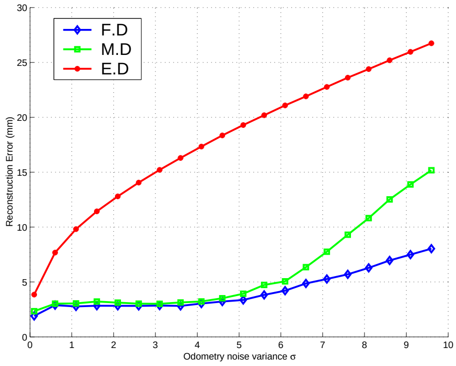
Experiments:

- Reconstruction mean error vs. Odometry noise variance. (Figure 3(a))
- Reconstruction mean error vs. % of outliers. (Figure 3 (b))

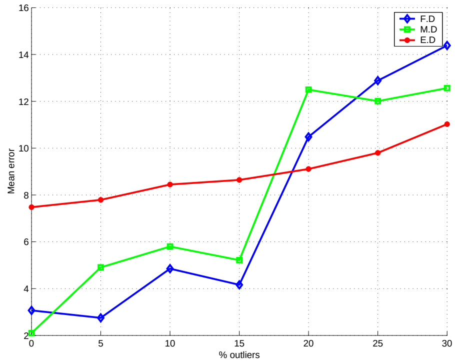
5.2 Real Data

A real system composed of a low cost camera and a real mobile robot is tested. The error initialization is the same used in the synthetic data.

The result of an initial calibration procedure is shown in Fig. 4 (b). As can be see five coplanar points are considered as initial state vector. The path required to obtain the estimation is shown in Fig. 4 (a).

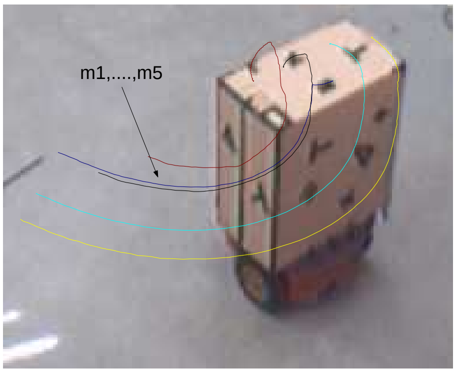


(a) Reconstruction error vs Odometry noise



(b) Real path and dead-reckon estimation

Fig. 3. Experiment 1



(a) Real robot initialization

m_x^1	55.6421	m_x^2	18.5460
m_y^1	-12.8574	m_y^2	117.9783
m_z^1	696.8408	m_z^2	700.4888
m_x^3	-63.2563	m_x^4	44.9386
m_y^3	-7.0530	m_y^4	194.9733
m_z^3	694.5169	m_z^4	681.7358
m_x^5	-125.9381	x_0	34.5282
m_y^5	-93.5649	y_0	756.0392
m_z^5	696.0449	α_0	111.2295

(b) Reconstructed table

Fig. 4. Real results

6 Conclusions

In this paper a method for computing robot geometry and pose from single camera is presented, with the aim of properly initializing a sequential approach. The inclusion of odometry allows metric reconstruction with single camera. Knowledge about the statistical properties of odometry readings and camera detection processes allows to propose a Maximum Likelihood approach which penalizes properly information which has more uncertainty as the last positions given by a dead-reckoning process. The equivalent approach is a weighted quadratic cost functions. The reconstruction accuracy is proved to be higher in the case of M.L versions. As can be seen in Figure 3 error is higher when no statistical information is used. Between the two approximations of the covariance matrix, the one which has more cross-correlated terms (F.D) achieves better performance when the error grows. The main drawback is that the more cross correlated terms the less sparse the Hessian becomes, which increases problem complexity. As a conclusion, for low error in odometry the Measurement Box Diagonal (M.D) approach will have the same behavior with the advantage that the sparsity of the problem remains equal to a Element Diagonal approach (E.D). Real results with coplanar points are also presented, achieving good accuracy with real odometry readings and image matching. This work was supported by the Ministry of Science and Technology under RESELAI project (reference TIN2006-14896-C02-01).

References

1. J. Borenstein and L. Feng. Measurement and correction of systematic odometry errors in mobile robots, 1996.
2. Huiying Chen, Kohsei Matsumoto, Jun Ota, and Tamio Arai. Self-calibration of environmental camera for mobile robot navigation. *Robotics and Autonomous Systems*, 55(3):177–190, March 2007.
3. A. J. Davison. SLAM with a single camera. In *Workshop on Concurrent Mapping and Localization for Autonomous Mobile Robots, in conjunction with ICRA*, 2002.
4. Y. Hada, E. Hemeldan, K. Takase, and H. Gakuhari. Trajectory tracking control of a nonholonomic mobile robot using igps and odometry. In *Multisensor Fusion and Integration for Intelligent Systems, MFI2003. Proceedings of IEEE International Conference on*, pages 51–57, 30 July-1 Aug. 2003.
5. Richard Hartley and Andrew Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, second edition edition, 2003.
6. Joo-Ho Lee, Kazuyuki Morioka, and Hideki Hashimoto. Intelligent space and mobile robots. In *The Industrial Information Technology Handbook*, pages 1–15. 2005.
7. K. Morioka, J.H. Lee, and H. Hashimoto. Human-following mobile robot in a distributed intelligent sensor network. *Industrial Electronics, IEEE Transactions on*, 51(1):229–237, 2004.
8. Daniel Pizarro, Enrique Santiso, and Manuel Mazo. Simultaneous localization and structure reconstruction of mobile robots with external cameras. In *ISIE05*, June 2005.
9. Bill Triggs, P. McLauchlan, Richard Hartley, and A. Fitzgibbon. Bundle adjustment – a modern synthesis. In *Vision Algorithms: Theory and Practice*, volume 1883, pages 298–372. Springer-Verlag, 2000.
10. T.Sogo, H.Ishiguro, and T.Ishida. Acquisition of qualitative spatial representation by visual observation. In *IJCAI99*, 1999.
11. Rudolph van der Merwe, Nando de Freitas, Arnaud Doucet, and Eric Wan. The unscented particle filter. Technical report, Cambridge University Engineering Department, aug 2000.