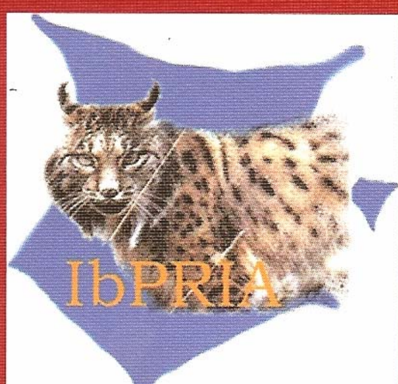
Joan Martí José Miguel Benedí
Ana Maria Mendonça Joan Serrat (Eds.)

LNCS 4477

Pattern Recognition and Image Analysis

Third Iberian Conference, IbPRIA 2007
Girona, Spain, June 2007
Proceedings, Part I

1 Part I



Springer

Lecture Notes in Computer Science 4477

Commenced Publication in 1973
Founding and Former Series Editors:
Gerhard Goos, Juris Hartmanis, and Jan van Leeuwen

Joan Martí   José Miguel Benedí
Ana Maria Mendonça   Joan Serrat (Eds.)

Pattern Recognition and Image Analysis

Third Iberian Conference, IbPRIA 2007
Girona, Spain, June 6-8, 2007
Proceedings, Part I

Springer

Volume Editors

Joan Martí
University of Girona
Campus Montilivi, s/n., 17071 Girona, Spain
E-mail: joanm@eia.udg.es

José Miguel Benedí
Polytechnical University of Valencia
Camino de Vera, s/n., 46022 Valencia, Spain
E-mail: jbenedi@dsic.upv.es

Ana Maria Mendonça
University of Porto
Rua Dr. Roberto Frias, s/n, 4200-465 Porto, Portugal
E-mail: amendon@fe.up.pt

Joan Serrat
Centre de Visió per Computador-UAB
Campus UAB, 08193 Belaterra, (Cerdanyola), Barcelona, Spain
E-mail: joan.serrat@cvc.uab.es

Library of Congress Control Number: 2007927717

CR Subject Classification (1998): I.4, I.5, I.7, I.2.7, I.2.10

LNCS Sublibrary: SL 6 – Image Processing, Computer Vision, Pattern Recognition, and Graphics

ISSN 0302-9743
ISBN-10 3-540-72846-5 Springer Berlin Heidelberg New York
ISBN-13 978-3-540-72846-7 Springer Berlin Heidelberg New York

Springer is a part of Springer Science+Business Media

springer.com

Printed in Germany

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper SPIN: 12070350 06/3180 5 4 3 2 1 0

A Probabilistic Observation Model for Stereo Vision Systems: Application to Particle Filter-Based Mapping and Localization

Francisco Angel Moreno, Jose Luis Blanco, and Javier Gonzalez

System Engineering and Automation Department
University of Malaga, Spain

Abstract. In this paper we propose a probabilistic observation model for stereo vision systems which avoids explicit data association between observations and the map by marginalizing the observation likelihood over all the possible associations. We define observations as sets of landmarks composed of their 3D locations, assumed to be normally distributed, and their SIFT descriptors. Our model has been integrated into a particle filter to test its performance in map building and global localization, as illustrated by experiments with a real robot.

1 Introduction

Due to the rich information cameras provide and their low cost in comparison with traditional robotics sensors, like laser scanners, vision systems have acquired more and more importance in mobile robotics during the last years. In particular, a large number of vision-based localization approaches have been reported in the literature either using single ([2],[3]), stereo ([13],[14]), or omnidirectional [8] cameras. Errors in the formation of the images (e.g. discretization) and in detecting features may lead to large inaccuracies in the robot localization. The resulting uncertainty can be managed by probabilistic Bayesian filters, extensively discussed elsewhere ([12], [16]). The underlying principle of those filters is the Bayes' theorem, which states how to update a prior belief about a variable x given a new observation z and an observation model:

$$\underbrace{p(x|z)}_{\text{posterior}} \propto \underbrace{p(x)}_{\text{prior}} \underbrace{p(z|x)}_{\text{obs. model}} \quad (1)$$

In mobile robot localization, the filter can be implemented by iteratively executing a prediction and an update step. In the former, the system state (the robot pose) is propagated in time according to an evolution model (the motion model), giving the *prior estimation* of the robot pose. In the second step, the prior is refined according to a given observation model, obtaining the *posterior distribution*.

Two widely extended implementations of Bayesian filters are the Extended Kalman Filter (EKF) [7], and the family of sequential Monte Carlo (SMC) methods (also named *particle filters* (PF)) [1]. EKF has been successfully used in mobile robot localization [4], but it is limited by the assumption of Gaussian models in both, the

J. Martí et al. (Eds.): IbPRIA 2007, Part I, LNCS 4477, pp. 346–353, 2007.

robot pose and the observation model. On the other hand, a PF can cope with complex, even multi-modal distributions, providing a unified method for map building and global localization. Due to these advantages, in this work we focus on observation models for PF. Although an observation model is required in both EKFs and PFs, notice that, in the former, the observation model has a parametrical form whereas in PFs it is necessary only to evaluate it pointwise.

This paper addresses the derivation of a suitable observation model for stereo vision systems. We assume that observations are sets of landmarks defined by their three-dimensional positions and a distinctive feature descriptor. In particular, we use SIFT [9] descriptors due to their invariance to image translation, scaling, and rotation. The main contribution of this work is the avoidance of explicit data association between observations and the map, which is achieved by marginalizing the observation model over all the possible associations. This model has been validated by experiments with a real robot: first a map of the environment has been successfully built applying vision-based SLAM, next, global localization is performed using the so built map.

The rest of this paper is organized as follows. In Section 2, we state the problem and define the involved variables. Section 3 describes our proposal for the observation model, and experimental results are presented in Section 4. Finally, we provide some conclusions and discuss possible future works in Section 5.

2 Problem Statement

2.1 Preliminary Definitions

Let m be the map of the environment, x_t the robot pose, u_t the robot action (which typically consists of odometry readings), and z_t the observation, all of them for the time step t. Then, the recursive Bayesian filter for the robot pose reads [16]:

$$\underbrace{p\left(x_t \mid z_{1:t}, u_{1:t}, m\right)}_{\text{current pose estimation}} \propto \underbrace{p\left(z_t \mid x_t, m\right)}_{\text{observation model}} \int \underbrace{p\left(x_t \mid x_{t-1}, u_t\right)}_{\text{motion model}} \underbrace{p\left(x_{t-1} \mid z_{1:t-1}, u_{1:t-1}, m\right)}_{\text{previous pose estimation}} dx_{t-1} \tag{2}$$

where we have employed the notation $z_{1:t}=\{z_1,\ldots,z_t\}$ for clarity. In this work we define observations z_t and the map m as sets of three-dimensional landmarks:

$$\begin{aligned} z_t &= \left\{z_t^i\right\}_{\mathrm{i}=\{1,\ldots,\mathrm{N}\}} \text{ where } z_t^i = \left\langle \mathbf{X}_t^i, \mathbf{F}_t^i \right\rangle, \;\; \mathbf{X}_t^i \sim N\left(\mu_t^i, \Sigma_t^i\right) \\ m &= \left\{m_j\right\}_{\mathrm{j}=\{1,\ldots,\mathrm{M}\}} \text{ where } m_j = \left\langle \mathbf{X}_m^j, \mathbf{F}_m^j \right\rangle, \;\; \mathbf{X}_m^j \sim N\left(\mu_m^j, \Sigma_m^j\right), \;\; \mathbf{F}_m^j \sim N\left(\mu_{\mathbf{F}_m^j}, \Sigma_{\mathbf{F}_m^j}\right) \end{aligned} \tag{3}$$

Notice that a landmark, either in the map or in an observation, comprises of its 3D location $\mathbf{X}$ (which we assume to be normally distributed), and its associated SIFT descriptor $\mathbf{F}$. For each landmark in the map, the descriptor $\mathbf{F}_m^j$ is also assumed to be normally distributed whereas the descriptor of an observation landmark $\mathbf{F}_t^i$ is just a sample vector. The process for extracting observations from the stereo images is described next.

2.2 Extraction of Reliable Landmarks for Observations

To extract the set of 3D landmarks (i.e. the observation) from a pair of stereo images we need to find feature points in both images, to match them, and to estimate their corresponding 3D locations. Next we describe the whole process in more detail.

Many methods have been proposed in the literature for extracting interest points from images, as the well known detectors of Kitchen & Rosenfeld [8] and Harris [5], based on the first and the second-order derivatives of images, respectively. More recently, the SIFT detector proposed by Lowe [9] deals with the detection process by identifying local extrema in a pyramid of Difference of Gaussians (DoG). It also provides the detected features with a descriptor that exhibits invariance to rotation and scale, and partial invariance to lighting changes and affine distortions. In our work, the detection of interest points in the images is carried out by the method proposed by Shi and Tomasi [15]. In addition, their corresponding SIFT descriptor is also computed to make them sufficiently distinguishable and improve the matching process robustness.

Once a set of keypoints has been detected in each image they are robustly matched according to both the similarity of their descriptors and the restriction imposed by the epipolar geometry. More precisely, in the former restriction, for each keypoint in the left image, it is computed the Euclidean distance between its descriptor and those of the keypoints in the right image. For a pair of keypoints to be considered as a candidate match, the minimum distance must be below a fixed threshold and the second lowest distance must be sufficiently apart from the minimum (see Fig. 1(b)). Moreover, the points must fulfill the epipolar constraint: they have to lay on the conjugate epipolar lines. In a stereo vision system with parallel optical axis as the one shown in Fig. 1(a), epipolar lines are parallel and horizontal, thus, the epipolar constraint reduces to checking that both features are in the same row.

Once matching have been robustly established, it is straightforward to estimate the most likely 3D coordinates of the landmark by using well-known methods [6],[14]. However, we also consider here the uncertainty in the 3D landmark position due to errors in the image quantization and in feature detection methods. Assuming a stereo system with parallel optical axes and a pinhole camera model (see Fig. 1(a)), the 3D coordinates (X, Y, Z) of a landmark can be computed from two matched points in the left and right images as [14]:

$$X = (c - c_0)\, b/d \quad Y = (r - r_0)\, b/d \quad Z = f\, b/d \tag{4}$$

where (r_0, c_0) are the coordinates of the reference image centre, (r, c) are the coordinates of the keypoint in the reference image (say, the left one), b is the baseline, d is the disparity, and f is the focal length of both cameras (please, refer to Fig. 1(a)).

The errors in obtaining the variables r, c, and d, are usually modelled as uncorrelated zero-mean Gaussian random variables [10]. Using a first-order error propagation to approximate the distribution of the variables in (4) as multivariate Gaussians, we obtain the following covariance matrix:

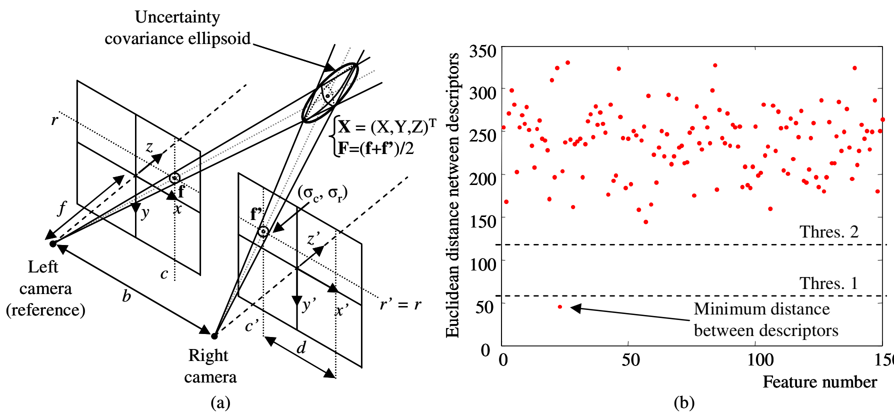


Fig. 1. (a) Configuration of a stereo vision system and schematic representation of uncertainty in the localization of the 3D landmarks. (b) Euclidean distance between the descriptors of a feature in the left image and all the features in the right one.

$$\Sigma = \mathbf{J}\, diag\left(\sigma_c^2, \sigma_r^2, \sigma_d^2\right)\mathbf{J}^{\mathrm{T}} \tag{5}$$

where $\mathbf{J}$ stands for the Jacobian matrix of the functions in (4), and $\sigma_X^2, \sigma_Y^2, \sigma_Z^2, \sigma_c^2, \sigma_r^2$, and σ_d^2 are the variances of the corresponding variables. Expanding (5) we come to the following expression for $\mathbf{\Sigma}$:

$$\mathbf{\Sigma} = \begin{pmatrix} \sigma_X^2 & \sigma_{XY} & \sigma_{XZ} \\ \sigma_{XY} & \sigma_Y^2 & \sigma_{YZ} \\ \sigma_{XZ} & \sigma_{YZ} & \sigma_Z^2 \end{pmatrix} = \begin{pmatrix} \frac{b^2\sigma_c^2}{d^2} + \frac{b^2(c-c_0)^2\sigma_d^2}{d^4} & \frac{(c-c_0)b^2\sigma_d^2(r-r_0)}{d^4} & \frac{(c-c_0)b^2\sigma_d^2 f}{d^4} \\ \frac{(c-c_0)b^2\sigma_d^2(r-r_0)}{d^4} & \frac{b^2\sigma_r^2}{d^2} + \frac{b^2(r-r_0)^2\sigma_d^2}{d^4} & \frac{(r-r_0)b^2\sigma_d^2 f}{d^4} \\ \frac{(c-c_0)b^2\sigma_d^2 f}{d^4} & \frac{(r-r_0)b^2\sigma_d^2 f}{d^4} & \frac{f^2 b^2\sigma_d^2}{d^4} \end{pmatrix} \tag{6}$$

which approximately models the uncertainty in the coordinates of landmarks computed from the noisy measurements of a stereo system. Finally, each landmark is assigned a SIFT descriptor which is simply computed as the mean value of the descriptors from each image (**F = (f+f'**)/2).

3 The Proposed Observation Model for Stereo Vision

In the following we introduce our proposal for the probabilistic observation model $p\left(z_t \middle| x_t, m\right)$, which models the likelihood of an observation at time *t*, given the robot pose (x_t) and a map (m). Firstly, assuming conditional independency between the errors in the detection of the individual landmarks z_t^i, the likelihood function can be factorized as follows:

$$p\left(z_t \middle| x_t, m\right) \overset{\text{cond.ind}}{=} \prod_i p\left(z_t^i \middle| x_t, m\right) \tag{7}$$

To avoid explicit data association between landmarks in the observation and in the map, we apply next the law of total probability to marginalize out the observation likelihood of individual landmarks by considering all the possible associations:

$$p\left(z_t^i \middle| x_t, m\right) = \sum_{j=\{1,\dots,M,\phi\}} p\left(z_t^i \middle| x_t, m, c_i = j\right) \underbrace{P\left(c_i = j \middle| x_t, m\right)}_{\eta} \tag{8}$$

where c_i is an unknown discrete variable that represents the correspondence of the i-th observed landmark. Its possible values are {1,…,M} for map landmarks, or ϕ for no correspondence with the map. Notice that the *a priori* probability of a given correspondence with the j-th landmark in the map, $P\left(c_i = j \middle| x_t, m\right)$, is constant since the actual observation z_t^i is not taken into account. Assuming the same probability for all the possible correspondences, including the null one, we have:

$$p\left(z_t^i \middle| x_t, m\right) = \eta \sum_{j=\{1,\dots,M,\phi\}} p\left(z_t^i \middle| x_t, m, c_i = j\right) \tag{9}$$

Next, if we expand the likelihood term conditioned to a given correspondence according to the definitions in (3), we obtain:

$$\begin{aligned} p\left(z_t^i \middle| x_t, m_j, c_i = j\right) &= p\left(\underbrace{\mathbf{X}_t^i, \mathbf{F}_t^i}_{z_t^i} \middle| x_t, \underbrace{\mathbf{X}_m^j, \mathbf{F}_m^j}_{m_j}, c_i = j \right) \\ &\overset{\text{cond.ind}}{=} \underbrace{p\left(\mathbf{X}_t^i \middle| x_t, \mathbf{X}_m^j, c_i = j\right)}_{\text{Localization term}} \underbrace{p\left(\mathbf{F}_t^i \middle| x_t, \mathbf{F}_m^j, c_i = j\right)}_{\text{Descriptor term}}, \end{aligned} \tag{10}$$

Here we have assumed conditional independence between the errors in localization and the descriptor of landmarks, which seems a plausible approximation. The descriptor term is easily computed by evaluating the probability density function associated to $\mathbf{F}_m^j$ at the vector $\mathbf{F}_t^i$. Regarding the localization term, we approximate it by the probability density of the pair of landmarks to coincide in the 3D space:

$$p\left(\mathbf{X}_t^i \middle| x_t, \mathbf{X}_m^j, c_i = j\right) = \int_{\mathbf{X} \in \mathbb{R}^3} p_{\mathbf{X}_m^j}\left(\mathbf{X}\right) p_{\mathbf{X}_t^i}\left(\mathbf{X}\right) d\mathbf{X} \equiv I \tag{11}$$

where $p_{\mathbf{X}_m^j}\left(\mathbf{X}\right)$ and $p_{\mathbf{X}_t^i}\left(\mathbf{X}\right)$ are the distributions of the random variables $\mathbf{X}_m^j$ and $\mathbf{X}_t^i$, respectively. Since both distributions are Gaussian their product is also a Gaussian and hence the integral I in (11) has a closed-form solution:

$$I = \left(2\pi \left|\Sigma_m^j + \Sigma_t^i\right|\right)^{-\frac{1}{2}} \exp\left\{ -\tfrac{1}{2} \left(\mu_m^j - \mu_t^i\right)^{\mathrm{T}} \left(\Sigma_m^j + \Sigma_t^i\right)^{-1} \left(\mu_m^j - \mu_t^i\right) \right\} \tag{12}$$

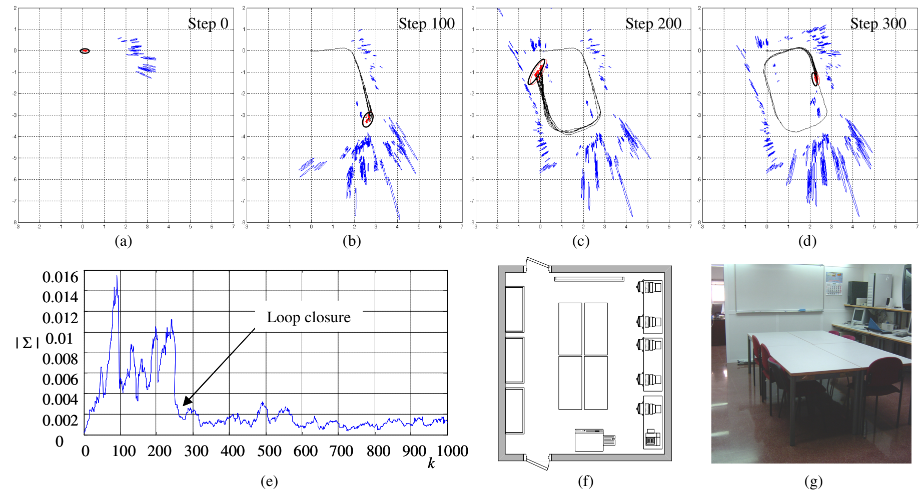


Fig. 2. Map building with a Rao-Blackwellised Particle Filter. (a)-(d) Map representation at different time steps. (e) Value of the covariance matrix determinant through time. (f) Plan and (g) a picture of the environment.

4 Experimental Results

Our proposed observation model has been tested within a particle filter framework for both Simultaneous Localization and Mapping (SLAM) and global localization. In robotics, particle filters represent the distribution of the robot pose by a set of samples which are propagated through the robot motion model and subsequently weighted according to the observation model. In the following experiments we additionally fit the particle set with a Gaussian to easily measure the degree of uncertainty in the robot pose estimation.

For the experiments, our Sancho mobile robot, equipped with a BumbleBee stereo vision system [18] (with a baseline of 11.9 cm and 2 mm of focal length), was manually driven following a circular trajectory of about 40 m inside one of our laboratories while taking images with the stereo camera at 3Hz (up to a grand total of 1000 stereo images). In the feature extraction process, it has been assumed that errors in the variables r, c and d have a variance of 1 pixel. A complete video showing the evolution of both experiments can be downloaded in [17].

4.1 Map Building

The sequence of stereo images is firstly used to build a map of the environment through a Rao-Blackwellised Particle Filter (RBPF). RBPFs are efficient solutions to the SLAM problem where each particle carries a hypothesis of the whole robot path and the associated map [16].

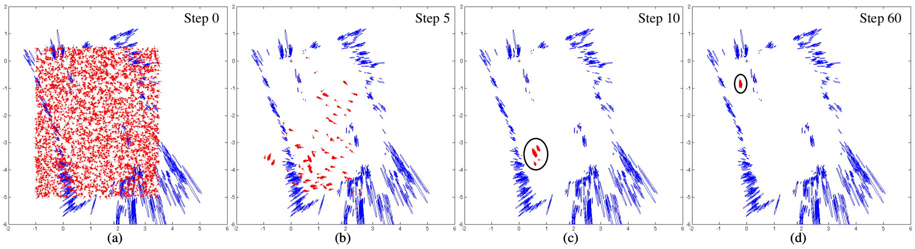


Fig. 3. (a)-(d) Global localization using the map built before. Initially, particles are uniformly scattered across the map and are subsequently concentrated around the robot real pose (*a circle surrounding the particles is shown in (c)-(d)*).

The evolution of the constructed map while the robot navigates is shown in Fig. 2 (a)-(d), where a top view of the 3D landmarks of the map being built (represented by 99% confidence intervals) is displayed at different time steps. Note that the uncertainty in the landmark positions decreases as they are detected in successive observations. Moreover, the covariance of the particles grows during the experiment since only new landmarks are being added to the map, until the point where the robot reaches an already navigated position, which is called the *loop closure*. Then, the estimation of the robot position is improved and particles converge towards the real robot location. This occurs in some point between Fig. 2 (c) and (d), and its effects in the uncertainty can be seen through the evolution of the determinant of the covariance matrix (Fig. 2 (e)). For this experiment, a sample size of 50 particles has been sufficient to yield a correct estimation of the map. The relatively small sample set is the reason of the noisy appearance of the fitted covariance in Fig. 2(e).

4.2 Global Localization

In this experiment we deal with the global localization problem. We take the map associated to the particle with the highest weight from the previous experiment as the map of the environment. Initially, a set of 3000 particles is uniformly distributed over the whole map (see Fig. 3 (a)) and, as the filter processes observations, they tend to converge towards the robot actual location (Fig. 3 (b)-(d)). Notice that in the early iterations particles are scattered over multiple possible positions since the available information is not enough to unambiguously localize the robot.

5 Conclusions

In this paper we have introduced a novel observation model for stereo vision systems suitable for particle filters, which considers observations as sets of landmarks determined by their 3D positions and their SIFT descriptors. As an important contribution, we avoid explicit data association by marginalizing out the observation likelihood over all the possible associations. Matching features in stereo image pairs is robustly solved by checking simple restrictions regarding their descriptors and epipolar geometry. The model takes into account the uncertainty both in the localization of landmarks and in their feature descriptors.

Two experiments have been performed in order to validate our proposal in the context of map building and global localization. The experimental results illustrate its adequate performance when coping with both problems and reveal the proposed observation model as a promising approach for stereo vision in robotics.

References

1. Arulampalam, M.S., Maskell, S., Gordon, N., Clapp, T.: A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. IEEE Transactions on Signal Processing 50, 174–188 (2002)
2. Davison, A.J.: Real-Time Simultaneous Localisation and Mapping with a Single Camera. In: Proc. International Conference on Computer Vision, vol. 2, pp. 1403–1410 (2003)
3. Davison, A.J., Cid, Y.G., Kita, N.: Real-Time 3D SLAM with Wide-Angle Vision. In: 5th Symposium on Intelligent Autonomous Vehicles. Lisbon Portugal (2004)
4. Dissanayake, M.W.M.G., Newman, P., Clark, S., Durrant-Whyte, H.F., Csorba, M.: A solution to the simultaneous localization and map building (SLAM) problem. IEEE Transactions on Robotics and Automation 17, 229–241 (2001)
5. Harris, C.J., Stephens, M.: A combined edge and corner detector. In: Proceedings of 4th Alvey Vision Conference, Manchester, pp.147–151 (1988)
6. Hartley, R., Zisserman, A.: Multiple View Geometry in Computer Vision, 2nd edn. Cambridge University Press, Cambridge (2003)
7. Julier, S.J., Ulhmann, J.K.: A New Extension of the Kalman Filter to Nonlinear Systems. Int. Symp. Aerospace/Defense Sensing, Simul. and Controls. Orlando (1997)
8. Kitchen, L., Rosenfeld, A.: Gray-level corner detection. Pattern Recognition Letters 1, 95–102 (1982)
9. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision 60(2), 91–110 (2004)
10. Matthies, L., Shafer, S.A.: Error modeling in Stereo Navigation. IEEE Journal of Robotics and Automation, vol. RA-3(3) (1987)
11. Menegatti, E., Pretto, A., Scarpa, A., Pagello, E.: Omnidirectional Vision Scan Matching for Robot Localization in Dynamic Environments. IEEE Trans. on Robotics 22(3), 523–535 (2006)
12. Montemerlo, M.: FastSLAM: A Factored Solution to the Simultaneous Localization and Mapping Problem With Unknown Data Association. PhD Thesis (2003)
13. Saeedi, P., Lawrence, P.D., Lowe, D.G.: Vision-Based 3-D Trajectory Tracking for Unknown Environments. IEEE Transactions on Robotics 22(1), 119–136 (2006)
14. Se, S., Lowe, D., Little, J.: Local and Global Localization for Mobile Robots using Visual Landmarks. In: Proc. International Conference on Intelligent Robots and Systems, pp. 414–420 (2001)
15. Shi, J., Tomasi, C.: Good features to track. In: Proc. Computer Vision and Pattern Recognition, pp. 593–600 (1994)
16. Thrun, S., Burgard, W., Fox, D.: Probabilistic Robotics. MIT Press, Cambridge (2006)
17. Website: http://www.isa.uma.es/C6/SLAM/default.aspx
18. Website: http://www.ptgrey.com