TITLE: Multi-robot visual SLAM using a Rao-Blackwellized Particle Filter

Arturo Gil *, Óscar Reinoso, Mónica Ballesta, Miguel Juliá

Universidad Miguel Hernández
Systems Engineering Department
Avda. de la Universidad s/n
03202 Elche (Alicante), Spain
e-mail: arturo.gil@umh.es

Preprint submitted to Robotics and Autonomous Systems *8 May 2008*

Multi-robot visual SLAM using a Rao-Blackwellized Particle Filter

Arturo Gil *, Óscar Reinoso, Mónica Ballesta, Miguel Juliá

Universidad Miguel Hernández
Systems Engineering Department
Avda. de la Universidad s/n
03202 Elche (Alicante), Spain

Abstract

This paper describes an approach to solve the Simultaneous Localization and Mapping (SLAM) problem with a team of cooperative autonomous vehicles. We consider that each robot is equipped with a stereo camera and is able to observe visual landmarks in the environment. The SLAM approach presented here is feature-based, thus the map is represented by a set of three dimensional landmarks each one defined by a global position in space and a visual descriptor. The robots move independently along different trajectories and make relative measurements to landmarks in the environment in order to jointly build a common map using a Rao-Blackwellized particle filter. We show results obtained in a simulated environment that validate the SLAM approach. The process of observing a visual landmark is simulated in the following way: first, the relative measurement obtained by the robot is corrupted with gaussian noise, using a noise model for a standard stereo camera. Second, the visual description of the landmark is altered by noise, simulating the changes in the descriptor which may occur when the robot observes the same landmark under different scales and viewpoints. In addition, the noise in the odometry of the robots also takes values obtained from real robots. We propose an approach to manage data associations in the context of visual features. Different experiments have been performed, with variations in the path followed by the robots and the parameters in the particle filter. Finally, the results obtained in simulation demonstrate that the approach is suitable for small robot teams.

Key words: SLAM, Cooperative robots, Particle filter, Visual landmarks

* Corresponding author:
Email address: arturo.gil@umh.es (Arturo Gil).
URL: www.isa.umh.es/arvc (Arturo Gil).

1 Introduction

Acquiring maps of the environment is an essential ability for autonomous mobile robots, since the maps are needed to perform higher level tasks. In consequence, in the last two decades the problem of simultaneous localization and mapping (SLAM) has received significant attention. The SLAM problem considers the situation in which an autonomous mobile robot moves through an unknown space and incrementally builds a map of this environment while simultaneously uses this map to compute its absolute location. This problem is considered inherently difficult, since noise introduced in the estimate of the robot pose leads to noise in the estimate of the map and *viceversa*.

To date, typical SLAM approaches have been using laser range sensors to build maps in two and three dimensions (e.g., [18,22]). Recently, the interest on using cameras as sensors in SLAM has increased and some authors have been concentrating on building three dimensional maps using visual information obtained from cameras. These approaches are usually denoted as visual SLAM. The reasons for this interest stem from:

i Stereo vision systems are typically less expensive that laser range systems.
ii Typical laser ranging systems provide 2D information from the environment whereas stereo vision systems are able to provide a more complete 3D representation of the space.
iii Vision systems provide a great quantity of information and allow to integrate in the robot other techniques, such as face or object recognition.

However, stereo systems are usually less precise than laser sensors and the information from the cameras needs normally to be processed in order to extract salient features. In common configurations, the camera is installed at a fixed height and orientation [5,17] with respect to the robot reference system and the movement of the camera is restricted to a plane.

Most approaches to visual SLAM are feature-based. In this case, a set of significant points in the environment are used as landmarks. Mainly, two steps must be distinguished in the selection of visual landmarks. The first step involves the detection of interest points in the images that can be used as reliable landmarks. The points should be detected at different distances and viewing angles, since they will be observed by the robot from different poses in the environment. At a second step the interest points are described by a feature vector which is computed using local image information. This descriptor is used in the data association problem, that is, when the robot has to decide whether the current observation corresponds to one of the landmarks in the map or to a new one. When the robot observes a visual landmark in the environment, it obtains a distance measurement and computes a visual descriptor.

Next, the descriptor and the measurement are used to find the landmark in the map that generated the observation. Typically, when the robot traverses previously explored places, it re-observes landmarks. In this case, if current observations are correctly associated with the visual landmarks, the robot will be able to find its location with respect to these landmarks, thus reducing the error in its pose. If the observations cannot be correctly associated to the landmarks in the map, the map will become inconsistent. To sum up, the data association is a fundamental part of the SLAM process, since wrong data associations will produce incorrect maps.

An important subfield within mobile robotics that requires accurate maps is the performance of collaborative tasks by multiple vehicles. Multiple vehicles can frequently perform tasks more quickly and robustly than a single one [15]. However, little effort has been done until now in the field of multi-robot visual SLAM, which considers the case where several robots move along the environment and build a map. In this paper we concentrate on this problem and propose a solution that allows to build a map using a set of visual observations obtained by a team of mobile robots. It is worth noting that SLAM algorithms focus on the incremental construction of a map, given a set of movements carried out by the robots and the set of observations obtained from different locations. However, SLAM algorithms do not consider the computation of the movements that need to be performed by the robots, since this is generally considered a different problem, denoted as exploration. Classical exploration strategies often try to cover unknown terrain as fast as possible and avoid to repeatedly visit known areas. This strategy, however, is suboptimal in the context of the SLAM problem because the robot typically needs to revisit places in order to localize itself and reduce the uncertainty [18]. If the exploration has to be accomplished using a team of robots, the problem becomes harder, since the exploration algorithm needs to avoid that different robots explore the same areas in the map. In this paper we concentrate on the visual SLAM problem and assume that the robots are able to explore the environment in an efficient way.

The major contribution of this paper is twofold. First we propose an approach to the multi-robot SLAM problem using a Rao-Blackwellized Particle Filter (RBPF). To the best of our knowledge, this is the first work that uses visual measurements provided by several robots to build a common 3D map of the environment. The results obtained with different number of robots are compared. Second, we simulate the process of visual SLAM using multiple robots. We consider that the robots perform observations on visual landmarks using stereo cameras, and, in addition, each observation is associated with a visual descriptor, that allows to partially differentiate between landmarks. In addition, we propose a solution to the data association in the context of visual SLAM and compare the results when the noise in the descriptor varies, which simulates the variation in the descriptor when a landmark is observed from

different viewpoints. Introducing noise in the descriptor makes the data association harder and affects negatively the results of the SLAM algorithm. The results presented demonstrate that the approach is suitable to build visual maps using small teams of robots.

The main reason that motivated to carry out a series of experiments using simulated data is that this enables to assess the validity of the presented algorithm under different conditions. In simulation, the map and the true path of the robots is known in advance, and this enables us to evaluate the precision of the approach when varying some of the parameters in the Rao-Blackwellized filter. Furthermore, we compare the results when different number of robots are used and evaluate the computational cost introduced by more robots. On the contrary, evaluating the results of visual SLAM using real data is complex. On the one hand the true path followed by the robots is not known. On the other hand, the map cannot be known precisely, since the exact location of the landmarks cannot be established in advance.

The remainder of the paper is structured as follows. First, Section 2 discusses related work. Section 3 presents the approach to multi-robot SLAM, our map representation is also detailed. Section 4 deals with visual landmarks and their utility in SLAM. Next, Section 5 exposes our approach to the data association problem in the context of visual features. The most relevant features of the simulated environment are explained in Section 6. Next, in Section 7 we present the results obtained. Finally, Section 8 summarizes the most important conclusions and proposes future extensions of the work.

2 Related work

Up to now, the approaches to multi-robot SLAM can be grouped in one of the two following solutions:

i Approaches in which each robot estimates its own individual map using its observations. At a later stage, a common map is formed by fusing the individual maps of the robot team.
ii Approaches where the estimation of all the trajectories and the map is made jointly. A single map is computed simultaneously using the observations of all the robots.

The work exposed in [19] can be classified in the first group. The idea here is that each robot builds an own map, and, at the same time continuously attempts to localize in the maps built by other robots using particle filters. The approach can cope with the situation where the initial locations of the robots are unknown, solving the data association problem with a *rendez-vous*

technique. However, the fusion of the individual maps is computationally expensive, since K^2 particle filters must be maintained for a team of K robots.

The approach presented in [4] can be classified in the second group. It uses an extended Kalman filter (EKF) to estimate a state vector formed by the poses of all the vehicles and a set of 2D landmarks. With this extension, the robots obtain observations and construct a single unified map using the update equations of the classical EKF [3]. The initial positions of the vehicles must be known in advance and the data association is assumed to be known. In addition, the gain in performance when using multiple vehicles is proved theoretically within this context. In this case, the main drawback stems from the fact that a single hypothesis over the robot pose is maintained. If false data associations are made the whole EKF may diverge [14].

In [16] stereo vision is used to extract 3D visual landmarks from the environment. During exploration, the robot extracts SIFT (Scale Invariant Feature Transform) features from stereo images and calculates relative measurements to them. Landmarks are then integrated in the map with an Kalman Filter associated to each one. However, this approach does not manage the uncertainty associated with robot motion, and only one hypothesis over the pose of the robot is maintained. Consequently it may fail in the presence of large odometric errors (e.g. while closing a loop). In [5] a Rao-Blackwellized particle filter is used to estimate simultaneously the map and the path of a single robot exploring the environment. In the mentioned work, SIFT features are used too as landmarks in the environment and extracted using a stereo pair of cameras.

3 Multi-robot visual 3D SLAM

In this section, we describe our approach to the SLAM problem in the case where a team of robots explore simultaneously the environment. SLAM is considered to be a complex task due to the mutual dependency between the map of the environment and the pose of the robot. If we make an error in the estimation of the pose, this induces an error in the estimation of the map and viceversa. The robots share the observations performed over the landmarks and create a common map of the environment. The solution presented here is based, in essence, on a Rao-Blackwellized particle filter (RBPF), proposed initially by Murphy [13] and commonly referred as FastSLAM in the SLAM community [11]. Basically, a Rao-Blackwellized particle filter combines a representation of the pose by means of particles with a closed estimation of some variables.

In order to build the map, we assume that the robots are equipped with a

stereo camera system, which enables them to obtain relative measurements from landmarks that are detected in the environment. In addition, we assume that a descriptor associated to the landmark can be obtained, based on its visual appearance. We also consider that the robots are able to communicate among themselves and with a central agent in the system. In addition, we consider that the relative starting position of the robots is approximately known in advance.

For simplicity, we first start our explanation assuming that a single robot explores the environment while it observes visual landmarks. At time t the robot obtains an observation z_t, constituted by $z_t = (v_t, d_t)$, where $v_t = (X_c, Y_c, Z_c)$ is a three dimensional vector relative to the left camera reference frame and d_t is the visual descriptor associated to the landmark. The map L is represented by a collection of N landmarks $L = \{l_1, l_2, ..., l_N\}$. Each landmark is described as: $l_k = \{\mu_k, \Sigma_k, d_k\}$, where $\mu_k = (X_k, Y_k, Z_k)$ is a vector describing the position of the landmark referred to a global reference frame, with associated covariance matrix Σ_k. In addition, each landmark l_k is associated with a descriptor d_k that partially differentiates it from others. This map representation is compact and has been used to effectively localize a robot in unmodified environments [6].

Following the usual nomenclature in Rao-Blackwellized SLAM, we denote the robot pose at time t as x_t and the movement of the robot at time t as u_t. On the other hand, the robot path until time t is referred as $x^t = \{x_1, x_2, \ldots, x_t\}$, the set of observations made by the robot until time t will be designated $z^t = \{z_1, z_2, \ldots, z_t\}$ and the set of actions $u^t = \{u_1, u_2, \ldots, u_t\}$. We formulate the SLAM problem as that of determining the location of all landmarks in the map L and robot poses x^t from a set of measurements z^t and robot actions u^t. Thus it can be stated as the estimation of the posterior:

$$p(x^t, L|z^t, u^t, c^t) \tag{1}$$

where c^t designates the set of data associations performed until time t, $c^t = \{c_1, c_2, \ldots, c_t\}$. While exploring a particular environment, the robot has to determine whether a particular observation $z_t = (v_t, d_t)$ corresponds to a previously mapped landmark or to a new one. Given that, at time t the map is formed by N landmarks, this correspondence is represented by c_t, where $c_t \in [1 \ldots N]$, meaning that the observation z_t corresponds to the landmark c_t in the map. When no correspondence is found we denote it as $c_t = N + 1$, indicating that it is a new landmark. For the moment, we consider this correspondence as known.

The SLAM problem can be separated into two parts:

i The estimation of the trajectory of the robot.
ii The estimation of the map by means of a series of measurements.

However, the two facets can be separated and, if somehow we could know the trajectory of the robot in the environment, then the estimation of the map would be trivial. This conditional independence property of the SLAM problem implies that the posterior (1) can be factored as [11]:

$$p(x^t, L|z^t, u^t, c^t) = p(x^t|z^t, u^t, c^t) \prod_{k=1}^{N} p(l_k|x^t, z^t, u^t, c^t) \tag{2}$$

This equation states that the full SLAM posterior is decomposed into two parts: one estimation over robot paths, and N independent estimators over landmark positions, each conditioned to the path estimate. We approximate $p(x^t|z^t, u^t, c^t)$ using a set of M particles, each particle having N independent landmark estimators (implemented as EKFs), one for each landmark in the map. Each particle is thus defined as:

$$S_t^{[m]} = \{x^{t,[m]}, \mu_{t,1}^{[m]}, \Sigma_{t,1}^{[m]}, d_1^{[m]}, \ldots, \mu_{t,N}^{[m]}, \Sigma_{t,N}^{[m]}, d_N^{[m]}\}, \tag{3}$$

where $\mu_{t,k}^{[m]}$ is the best estimation at time t for the position of landmark l_k based on the path of the particle m and $\Sigma_{t,k}^{[m]}$ is the associated covariance matrix. The visual descriptor associated to the landmark j is represented by $d_j^{[m]}$. The particle set $S_t = \{S_t^{[1]}, S_t^{[2]}, \ldots, S_t^{[M]}\}$ is calculated incrementally from the set S_{t-1} at time $t-1$ and the robot control u_t. Thus, each particle is sampled from a proposal distribution $x_t^{[m]} \sim p(x_t|x_{t-1}, u_t)$ that models the noise in the odometry of the robots. Next, and following the approach of [11], each particle is then assigned a weight according to:

$$\omega_t^{[m]} = \frac{1}{\sqrt{|2\pi Z_{c_t}|}} e^{\{-\frac{1}{2}(v_t - \hat{v}_{t,c_t})^T [Z_{c_t}]^{-1} (v_t - \hat{v}_{t,c_t})\}} \tag{4}$$

where v_t is the actual measurement and $\hat{v}_{t,c_t}$ is the predicted measurement for the landmark c_t based on the pose $x_t^{[m]}$. The matrix Z_{c_t} is the covariance matrix associated with the innovation $(v_t - \hat{v}_{t,c_t})$. Note that we implicitly assume that each measurement v_t has been associated to the landmark c_t of the map. This problem is, in general, hard to solve, since similar-looking landmarks may exist. In Section 5 we describe our approach to this problem. In the case that B observations from different landmarks exist at a time t, that is $z_t = \{z_{t,1}, z_{t,2}, \ldots, z_{t,B}\}$, we calculate the total weight assigned to the particle as: $\omega_t^{[m]} = \prod_{i=1}^{B} \omega_{t,i}^{[m]}$, where $w_{t,i}^{[m]}$ is the weight associated to the observation $z_{t,i}$, computed using Equation (4).

In the following, we assume that a team of K robots explores the environment. Our objective here is to jointly estimate the paths followed by the robots and the map. At a time step t the robot $\langle i \rangle$ is at pose $x_{t,\langle i \rangle}$ and performs a single observation $z_{t,\langle i \rangle} = \{v_{t,\langle i \rangle}, d_{t,\langle i \rangle}\}$. We denote the path of the robot $\langle i \rangle$ until time t as $x_{\langle i \rangle}^t = \{x_{1,\langle i \rangle}, x_{2,\langle i \rangle}, \ldots, x_{t,\langle i \rangle}\}$. For simplicity, we refer

to $x^t_{\langle 1:K\rangle} = \{x^t_{\langle 1\rangle}, x^t_{\langle 2\rangle}, \ldots, x^t_{\langle K\rangle}\}$ to the set of robot paths in the team until time t. Analogously, we denote $u^t_{\langle 1:K\rangle} = \{u^t_{\langle 1\rangle}, u^t_{\langle 2\rangle}, \ldots, u^t_{\langle K\rangle}\}$ the set of actions performed by the robots and $z^t_{\langle 1:K\rangle} = \{z^t_{\langle 1\rangle}, z^t_{\langle 2\rangle}, \ldots, z^t_{\langle K\rangle}\}$ refers to the set of observations performed by the robot team until time t. Similarly to the case of a single robot, the data association variable will be denoted as c_t, which indicates that the observation $z_{t,\langle i\rangle}$ is associated to the landmark c_t in the map. It is worth noting here that we are estimating a map common to all the robots, in consequence the observation is associated to a landmark in the map independently of the robot that observed it. The data association until time t is referred as $c^t = \{c_1, c_2, \ldots, c_t\}$. In consequence, the multi-robot SLAM problem can be stated as the estimation of the following probability function:

$$p(x^t_{\langle 1:K\rangle}, L | z^t_{\langle 1:K\rangle}, u^t_{\langle 1:K\rangle}, c^t) = \\ p(x^t_{\langle 1:K\rangle} | z^t_{\langle 1:K\rangle}, u^t_{\langle 1:K\rangle}, c^t) \prod_{k=1}^{N} p(l_k | x^t_{\langle 1:K\rangle}, z^t_{\langle 1:K\rangle}, u^t_{\langle 1:K\rangle}, c^t) \tag{5}$$

This equation proposes a manner to estimate a group of K paths $x^t_{\langle 1:K\rangle}$ and a map L conditioned to the case that the robots have performed a number of movements $u^t_{\langle 1:K\rangle}$ and a series of observations $z^t_{\langle 1:K\rangle}$ associated to landmarks in the map c^t. In consequence, analogously to Equation (2), Equation (5) expresses that we can separate the estimation of the map and the estimation of K different paths into two parts: The function $p(x^t_{\langle 1:K\rangle} | z^t_{\langle 1:K\rangle}, u^t_{\langle 1:K\rangle}, c^t)$ is estimated using a particle filter, while the map is estimated using N independent estimations conditioned to the paths $x^t_{\langle 1:K\rangle}$. As a result, we are decomposing the SLAM problem into two different parts: A localization problem of K robots in an environment and a series of individual landmark estimations conditioned to robot paths $x^t_{\langle 1:K\rangle}$. In order to achieve this, each one of the M particles in the filter is accompanied with N independent estimators for each one of the landmarks, implemented as EKFs. In our case, each of the Kalman Filters will be conditioned to the K paths of the robot team. In consequence, each particle is represented as:

$$S^{[m]}_t = \{x^{t,[m]}_{\langle 1:K\rangle}, \mu^{[m]}_{1,t}, \Sigma^{[m]}_{1,t}, d^{[m]}_1, \cdots, \mu^{[m]}_{N,t}, \Sigma^{[m]}_{N,t}, d^{[m]}_N\} \tag{6}$$

The fundamental difference compared to the particle defined in (3) is that, in this case, the state that we would like to estimate is composed by the pose (x, y, θ) of K robots, thus $x_{t,\langle 1:K\rangle} = \{x_{t,\langle 1\rangle}, x_{t,\langle 2\rangle}, \cdots, x_{t,\langle K\rangle}\}$. As a result, we propose a joint estimation over a path state of dimension $3K$. According to [21], the number of particles needed to obtain a good estimation increases exponentially with the dimension of the state. However, the results that we present here show that the approach works perfectly for robot teams of 2–3 members using a reasonable number of particles. In the case presented, the same map is shared by all the robots, which means that an observation performed by a particular robot affects the map of the whole robot team. In conse-

Table 1
Particle set S_t. Each particle is accompanied by N Kalman Filters.

Particle 1	$\{(x,\ y,\ \theta)_{\langle 1\rangle}, \cdots, (x,\ y,\ \theta)_{\langle K\rangle}\}^{[1]}$	$\mu_1^{[1]}\ \Sigma_1^{[1]}\ \mathrm{d}_1^{[1]}$	$\ldots$	$\mu_N^{[1]}\ \Sigma_N^{[1]}\ \mathrm{d}_N^{[1]}$
$\vdots$				
Particle M	$\{(x,\ y,\ \theta)_{\langle 1\rangle}, \cdots, (x,\ y,\ \theta)_{\langle K\rangle}\}^{[M]}$	$\mu_1^{[M]}\ \Sigma_1^{[M]}\ \mathrm{d}_1^{[M]}$	$\ldots$	$\mu_N^{[M]}\ \Sigma_N^{[M]}\ \mathrm{d}_N^{[M]}$

quence, one member of the team may observe a landmark previously mapped by a different robot and update its estimate. In addition, this means that a robot does not need to explicitly close a loop in order to reduce the uncertainty in its pose (i.e. return to a previously mapped area). A robot can reduce the uncertainty in its pose when it observes landmarks previously mapped by other robots. The structure of the particle defined in (6) is clearly explained in Table 1. To sum up, we propose a method based on a Rao-Blackwellized particle filter for the case where a robot team cooperates to build a map of a given environment. The algorithm can be decomposed in 4 basic steps:

- Generate a new particle set based on the prior set.
- Update the estimation of each landmark based on the observations.
- Calculate a weight for each particle.
- Perform a resampling based on the weight of each particle.

In order to clarify the ideas presented here we describe the whole process in Algorithm 1, which considers the case in which 3 robots explore simultaneously the environment.

3.1 Generating a new particle set

The first step is to generate a new set of hypothesis S_t based on the set S_{t-1}. That means that we obtain a new particle set over the robot poses $x_{t,\langle 1:K\rangle}$. That is, we obtain a new pose $x_{t,\langle i\rangle}^{[m]}$ for each of the robots by sampling from a motion model $p(x_t|x_{t-1}, u_t)$:

$$x_{t,\langle i\rangle}^{[m]} \sim p(x_{t,\langle i\rangle}|x_{t-1,\langle i\rangle}, u_{t,\langle i\rangle}) \tag{7}$$

The function $p(x_{t,\langle i\rangle}|x_{t-1,\langle i\rangle}, u_t)$ defines the movement model for the mobile agent. We apply this movement model to each of the K poses of the particle separately, based on the movement performed by the $\langle i\rangle$ robot. Figure 1 shows the application of Equation (7). The solid lines represent the true movements performed by the robots while the dotted lines correspond to the odometry readings. In the beginning of the movements, all the particles concentrate on the origin. After applying the movement model we obtain a set of particle clouds that represent the probability over the poses of the robots at each time step. In this way, non-linear motion models can be easily represented, as opposed to the linearization required by EKF-SLAM approaches [3].

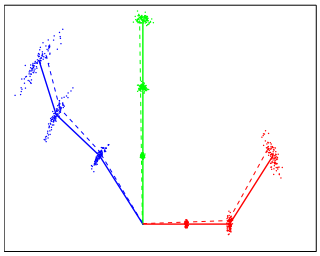


Fig. 1. The figure shows a new set of particles generated by sampling from the motion model.

3.2 Landmark estimation

Updating each landmark estimate is performed based on the pose of the $\langle i \rangle$ robot that performed the observation $z_{t,\langle i \rangle} = \{v_{t,\langle i \rangle}, d_{t,\langle i \rangle}\}$ with data association c_t. Updating each landmark l_k is performed independently, using the standard EKF equations detailed below:

$$\hat{v}_{t,\langle i \rangle} = g(x^{[m]}_{t,\langle i \rangle}, \mu^{[m]}_{c_t,t-1}) \tag{8}$$

$$G_{l_{c_t}} = \nabla_{l_{c_t}} g(x_t, l_{c_t})_{x_t = x^{[m]}_{t,\langle i \rangle}; l_{c_t} = \mu^{[m]}_{c_t,t-1}} \tag{9}$$

$$Z_{c_t,t} = G_{l_{c_t}} \Sigma^{[m]}_{c_t,t-1} G^T_{l_{c_t}} + R_t \tag{10}$$

$$K_t = \Sigma^{[m]}_{c_t,t-1} G^T_{l_{c_{t,\langle i \rangle}}} Z^{-1}_{c_t,t} \tag{11}$$

$$\mu^{[m]}_{c_t,t} = \mu^{[m]}_{c_t,t-1} + K_t(v_{t,\langle i \rangle} - \hat{v}_{t,\langle i \rangle}) \tag{12}$$

$$\Sigma^{[m]}_{c_t,,t} = (I - K_t G_{l_{c_t}}) \Sigma^{[m]}_{c_t,t-1} \tag{13}$$

where $\hat{v}_{t,\langle i \rangle}$ is the prediction for the current measurement $v_{t,\langle i \rangle}$ assuming that it has been associated with landmark c_t in the map . The observation model $g(x_t, l_{c_t})$ is linearly approximated by the Jacobian matrix $G_{l_{c_t}}$. It is assumed here that the noise in the observation is Gaussian and can be modeled with the covariance matrix R_t. Equation (12) represents the update of the estimation of the landmark c_t: $\mu^{[m]}_{c_t,t-1}$ in terms of the innovation $v = (v_{t,\langle i \rangle} - \hat{v}_{t,\langle i \rangle})$. Finally, Equation (13) updates the covariance matrix $\Sigma^{[m]}_{c_t,t}$, which is associated to the m particle and the landmark c_t. Updating each EKF requires a constant time per landmark, since the dimensions of each one are 3×3. Note that we implicitly assume that the observation $z_{t,\langle i \rangle}$ corresponds to the landmark l_{c_t} in the map. By now we assume this correspondence to be known. Later, in Section 5 we deal with this problem with more detail.

The matrix R_t associated with the noise in the observation is modeled using the equations of a standard stereo pair of cameras (assuming pin-hole cameras and parallel optical axis of both cameras). The 3D coordinates of a point in

the space relative to the left camera reference system can be calculated as:

$$X^r = \frac{I(c - C_x)}{d}, Y^r = \frac{I(C_y - r)}{d}, Z^r = \frac{fI}{d} \tag{14}$$

where, (r, c) are the row and column of the projected 3D point in the left image. Correspondingly, the same point is projected in $(r, c + d)$ in the right image, being d the disparity associated to that point. The parameter I is named baseline, and corresponds to the horizontal separation of both cameras. The parameters C_x and C_y refer to the intersection of the optical axis with the image plane in both cameras and f is the focal distance of the cameras. We can calculate the covariances associated to a relative measurement (X^r, Y^r, Z^r) supposing a linear error propagation:

$$\sigma^2_{X^r} = \frac{I^2\sigma_c^2}{d^2} + \frac{I^2(c - C_x)^2\sigma_d^2}{d^4},\ \sigma^2_{Y^r} = \frac{I^2\sigma_r^2}{d^2} + \frac{I^2(C_y - r)^2\sigma_d^2}{d^4},\ \sigma^2_{Z^r} = \frac{f^2I^2\sigma_d^2}{d^4} \tag{15}$$

This model has previously used at a visual SLAM context in [16]. In the experiments we have used $\sigma_r = \sigma_c = 10$ pixels and $\sigma_d = 0.5$, which produce typical errors in a common stereo camera. The noise matrix R_t is computed as $R_t = diag(\sigma^2_{X^r}, \sigma^2_{Y^r}, \sigma^2_{Z^r})$.

3.3 Assigning a weight to each particle

The set of particles generated by the movement model are distributed according to $p(x^t_{\langle 1:K\rangle}|z^{t-1}_{\langle 1:K\rangle}, u^t_{\langle 1:K\rangle}, c^{t-1}_{\langle 1:K\rangle})$, since the lasts observations $z_{t,\langle 1:K\rangle}$ obtained by the robots have not been included in the filter. This distribution is commonly known as proposal distribution. However, we would like to estimate the posterior $p(x^t_{\langle 1:K\rangle}, L|z^t_{\langle 1:K\rangle}, u^t_{\langle 1:K\rangle}, c^t_{\langle 1:K\rangle})$ which includes all the information from odometry and sensors until time t, generally stated as the target distribution. This difference is corrected with a process generally called sample importance resampling (SIR). Basically, a weight is assigned to each particle depending on the the quality that the current observation and the map matches. Following, a new set of particles S_t is created by sampling from S_{t-1}. Each particle is included in the new set with probability proportional to its weight. During the resampling process, particles with low weight are normally replaced with others with a higher weight. Assuming that a team of K robots exist and each one performs a single measurement $z_{t,\langle i\rangle}$ with data association c_t. We compute, the weight $\omega^{[m]}_{t,\langle i\rangle}$ associated with the particle m and the robot $\langle i\rangle$ as:

$$\omega^{[m]}_{t,\langle i\rangle} = \frac{1}{\sqrt{|2\pi Z_{c_t}|}} e^{\{-\frac{1}{2}(v_{t,\langle i\rangle} - \hat{v}_{t,c_t})^T[Z_{c_t}]^{-1}(v_{t,\langle i\rangle} - \hat{v}_{t,c_t})\}} \tag{16}$$

In consequence, for each particle $S_t^{[m]}$ defined in Equation (6) a number of K weights $\omega_{t,\langle i\rangle} = \{\omega_{t,\langle 1\rangle} \cdots \omega_{t,\langle K\rangle}\}$ are calculated. Since we are estimating the joint probability over the robot paths, the total weight associated to the particle $S_t^{[m]}$ must be calculated by multiplying the weights associated to the K robots: $\omega_t^{[m]} = \prod_{i=1}^{K} \omega_{t,\langle i\rangle}^{[m]}$. Next, the weights are normalized to approximate a probability density function $\sum_{i=1}^{M} \omega_t^{[i]} = 1$. The same procedure can be extended to the case where each of the vehicles obtains a set of B observations $z_{t,\langle i\rangle} = \{z_{t,\langle i\rangle,1}, z_{t,\langle i\rangle,2}, \ldots, z_{t,\langle i\rangle,B}\}$. This can be achieved by calculating a weight for each of the observations using Equation (16) and then multiplying the B results to calculate $\omega_{t,\langle i\rangle}$.

3.4 Path and map estimation

The algorithm we have just described, maintains a set of particles that represent a set of plausible paths that the robot team may follow. Conditioned to these paths, a set of three dimensional landmarks are estimated, each one represented by a Kalman filter. Finally, we need to choose the path and the map that best represent the true trajectories and map. In order to do this, we maintain a logarithmic sum over the global weights of the particles. In consequence, we choose the best particle as the one that maximizes the sum:

$$\hat{m} = \underset{m}{argmax} \sum_{t=1}^{A} \log(w_t^{[m]}) \tag{17}$$

4 Visual landmarks

We define visual landmark as a point in space that can be easily detected from different distances and viewing angles by means of a vision sensor. Corners in images are typically good candidates to be employed as landmarks for SLAM tasks, since they are stable and can typically be detected from different viewpoints.

Solving the SLAM problem with a landmark oriented approach involves two main processes: The detection, that enables the landmark to be detected from a set of poses in the environment, and the description of the landmark, that aims at representing the landmark based on its visual appearance. The descriptor enables the robot to recognize a particular landmark in the map. In the case of visual landmarks, we aim at detecting some points in the images that are highly salient, and can easily be detected from different distances and viewing angles. The description of the visual landmarks is based on a

Algorithm 1 Summary of the proposed algorithm.

$S = \emptyset$
$[z_{t,\langle 1\rangle}, z_{t,\langle 2\rangle}, z_{t,\langle 3\rangle}]$ = ObtainObservations()
InitialiseMap$(S, x_{0,\langle 1:3\rangle}, z_{t,\langle 1:3\rangle})$
for $t = 1$ to numMovements **do**
  $[z_{t,\langle 1\rangle}, z_{t,\langle 2\rangle}, z_{t,\langle 3\rangle}]$ = ObtainObservations()
  $[S, \omega_{t,\langle 1\rangle}]$ = FastSLAMMR$(S, z_{t,\langle 1\rangle}, R_t, u_{t,\langle 1\rangle})$
  $[S, \omega_{t,\langle 2\rangle}]$ = FastSLAMMR$(S, z_{t,\langle 2\rangle}, R_t, u_{t,\langle 2\rangle})$
  $[S, \omega_{t,\langle 3\rangle}]$ = FastSLAMMR$(S, z_{t,\langle 3\rangle}, R_t, u_{t,\langle 3\rangle})$
  $\omega_t = \omega_{t,\langle 1\rangle}\omega_{t,\langle 2\rangle}\omega_{t,\langle 3\rangle}$
  S = ImportanceResampling(S, ω_t) //Sample randomly from S according to $\omega_t^{[m]}$
end for

function $[S_t]$ **= FastSLAMMR**$(S_{t-1}, z_{t,\langle i\rangle}, R_t, u_{t,\langle i\rangle})$
$S_t = \emptyset$
for $m = 1$ to M {For every particle} **do**
  $x_{t,\langle i\rangle}^{[m]} \sim p(x_{t,\langle i\rangle} | x_{t-1,\langle i\rangle}, u_{t,\langle i\rangle})$
  for $n = 1$ to $N_{t-1}^{[m]}$ //Loop over all possible data associations **do**
    $\hat{v}_{t,\langle i\rangle} = g(x_{t,\langle i\rangle}^{[m]}, \mu_{n,t-1}^{[m]})$
    $G_{l_n} = \nabla_{l_{c_t}} g(x_t, l_n)_{x_{t,\langle i\rangle} = x_{t,\langle i\rangle}^{[m]}; l_{c_t} = \mu_{c_t,t-1}^{[m]}}$
    $Z_{n,t} = G_{l_n} \Sigma_{n,t-1}^{[m]} G_{l_n}^T + R_t$
    $D(n) = (v_{t,\langle i\rangle} - \hat{v}_{t,\langle i\rangle})^T [Z_{n,t}]^{-1} (v_{t,\langle i\rangle} - \hat{z}_{t,\langle i\rangle})$
    $E(n) = (d_{t,\langle i\rangle} - d_n)^T (d_{t,\langle i\rangle} - d_n)$
  end for
  $D(N_{t-1}^{[m]} + 1) = D_0$
  $j = find(D \leq D_0)$ {Find candidates below D_0}
  $c_t = argmin_j\, E(n)$ {Find minimum among candidates}
  if $E(c_t) > E_0$ // Create a new landmark? **then**
    $c_t = N_{t-1}^{[m]} + 1$
  end if
  if $c_t = N_{t-1}^{[m]} + 1$ //New landmark **then**
    $N_t^{[m]} = N_{t-1}^{[m]} + 1$
    $\mu_{c_t,t}^{[m]} = g^{-1}(x_{t,\langle i\rangle}^{[m]}, z_{t,\langle i\rangle})$
    $\Sigma_{c_t,t}^{[m]} = G_{l_{c_t}}^T R_t^{-1} G_{l_{c_t}}$
    $\omega_t^{[m]} = p_0$
  else
    $N_t^{[m]} = N_{t-1}^{[m]}$ //Old landmark
    $K_t = \Sigma_{c_t,t-1}^{[m]} G_{L_{c_t}}^T Z_{c_t,t}^{-1}$
    $\mu_{c_t,t}^{[m]} = \mu_{c_t,t-1}^{[m]} + K_t(v_{t,\langle i\rangle} - \hat{v}_{t,\langle i\rangle})$
    $\Sigma_{c_t,t}^{[m]} = (I - K_t G_{l_{c_t}}) \Sigma_{c_t,t-1}^{[m]}$
  end if
  $\omega_{t,\langle i\rangle}^{[m]} = \frac{1}{\sqrt{|2\pi Z_{c_t}|}} e^{\{-\frac{1}{2}(v_{t,\langle i\rangle} - \hat{v}_{t,c_t})^T [Z_{c_t}]^{-1} (v_{t,\langle i\rangle} - \hat{v}_{t,c_t})\}}$
  add $\{x_{t,\langle i\rangle}^{[m]}, N_t^{[m]}, \mu_{1,t}^{[m]}, \Sigma_{1,t}^{[m]}, d_{1,t}^{[m]}, \cdots, \mu_{N_t^{[m]},t}^{[m]}, \Sigma_{N_t^{[m]},t}^{[m]}, d_{N_t^{[m]},t}^{[m]}, \omega_t^{[m]}\}$ to S_t
end for
return S_t

sub-image centered at the detected point. The case of visual SLAM is particularly difficult, since:

i The landmarks cannot always be detected from different viewpoints. In consequence, it is difficult to re-detect previously mapped landmarks.

ii The description of the points must be invariant to changes in viewing distance (scale) and viewing angle. To achieve an invariant description is par-

ticularly complex in the case of visual landmarks, since the appearance of a point in space varies greatly with viewpoint changes. In consequence, the data association problem becomes hard to solve.

To date, different detectors and descriptors have been used for mapping and localization using monocular or stereo vision, such as SIFT [16,5], the Harris corner detector [2], Harris-Laplace [7] or SURF [12]. SIFT and SURF combine a detection and description method. In particular, SIFT (Scale Invariant Feature Transform) features were developed for image feature generation, and used initially in object recognition applications (see for example [8] and [9]). SIFT features combine a method to extract stable locations in images and a description that enables to identify those points. First, SIFT features are located at maxima and minima of a difference of Gaussian function applied in scale space. Next, each SIFT location is associated to a descriptor, computed using the image information at each location and scale. The descriptor provides invariance to image translation, scaling, rotation and is only partially invariant to illumination changes and affine projection. The resulting descriptor is of dimension 128 and enables the same points in the space to be recognized from different viewpoints, which may occur while the robot moves around its workplace. Lately, SIFT features have been used in robotic applications, showing its suitability for localization and SLAM tasks ([16,17,6,5]). In a prior work we performed an evaluation over different detection [10] and description [1] methods and their suitability for visual SLAM.

5 Data association

While a robot explores the environment it must decide whether the observation $z_{t,\langle i\rangle} = (v_{t,\langle i\rangle}, d_{t,\langle i\rangle})$ corresponds to a previously mapped landmark or to a different landmark. This concept is shown in Figure 2. Marked with stars is the position of the two landmarks, with the uncertainty depicted with an ellipse. Next, the robot obtains a measurement $v_{t,\langle i\rangle}$ shown with dashed line, with an uncertainty denoted by its associated ellipse. In addition, and since we are using a vision sensor, a descriptor $d_{t,\langle i\rangle}$ describes the visual appearance of the point in space. In this case, the robot must decide whether the measurement corresponds to landmark θ_1, landmark θ_2 or it is a new landmark. In [11] a Mahalanobis distance function is computed to associate each measurement to a landmark in the map. Thus, the distance D is computed for all the landmarks in the map:

$$D = (v_{t,\langle i\rangle} - \hat{v}_{t,c_t})^T [Z_{c_t}]^{-1} (v_{t,\langle i\rangle} - \hat{v}_{t,c_t}) \tag{18}$$

The measurement $v_{t,\langle i\rangle}$ is associated with the landmark c_t in the map that minimizes D. If the minimum value surpasses a pre-defined threshold, a new landmark is created.

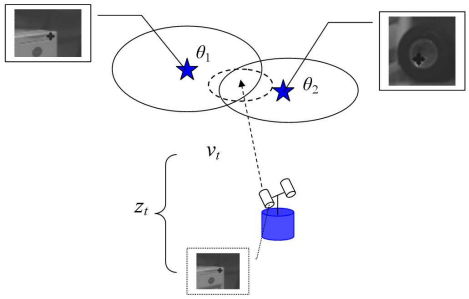


Fig. 2. The figure describes the data association problem in the context of visual landmarks.

To work properly, the previous solution needs the landmarks to be placed sufficiently apart from each other, in order to avoid false data associations. However, normally, this is not the case in visual SLAM, since the landmarks may be placed nearby. As stated in [14] the solution is not optimal, since wrong data associations will be made. As depicted in Figure 2, we propose here to improve the data association process by incorporating the visual description of the landmark. In our case, we will employ a SIFT descriptor associated to each landmark in the map. To do this, we first compute distance D for all the landmarks in the map and select those that are below a threshold D_0 as candidates. Next, given a SIFT descriptor as observed by the robot $d_{t,\langle i \rangle}$ and a SIFT descriptor d_j associated to a landmark in the map, we compute a squared Euclidean distance function:

$$E = (d_{t,\langle i \rangle} - d_j)(d_{t,\langle i \rangle} - d_j)^T \tag{19}$$

We compute the distance E within the group of candidates and select the landmark in the map that minimizes E. Whenever the distance E is below a threshold E_0 the observation is associated with the landmark. On the other hand, a new landmark is created whenever the distance E exceeds a pre-defined threshold (selected experimentally), since the current observation cannot be correctly explained by any of the landmarks in the map up to now.

The data association is performed independently for each particle, and this means that different particles would make different data associations. In addition, the method presented is simple, since it needs to be computed for each particle. In a previous work [5], we used this approach in the case of visual SLAM with a single robot. The experimental results showed that the high specificity of the SIFT descriptors allowed to build a precise map of the environment.

6 Simulation environment

The main reason that motivated to perform a series of experiments using simulated data is that this enables to assess the validity of the presented algorithm. In this case, the map and the true path of the robots is known in advance, and this enables us to evaluate the precision of the approach when varying some of the parameters in the Rao-Blackwellized filter. In this way, we can evaluate the algorithm with different number of particles and analyze the effect of adding more robots to the precision of the map and the estimated paths.

We simulate an indoor environment and a group of robots that navigate through it. The simulations were carried out using Matlab. We assume that the mobile agents move around the space simultaneously, they find 3D visual landmarks and obtain relative observations to them, consisting of a 3D relative measurement and a visual descriptor for each landmark. Figure 3 shows a 3D view of the simulated environment. The position of each landmark is indicated with an asterisk. In order to simulate a typical indoor environment, the landmarks have been placed over planar regions, that simulate walls which restrict the visibility of the points (i.e. a robot cannot observe a landmark through a wall). We consider that there is a SIFT descriptor associated to each landmark in the map. To do this, we previously captured a set of images from an office environment and obtained a set of SIFT descriptors, which were stored in a database. Next, we randomly associated each landmark to one of the SIFT descriptors. The size of the environment is $30 \times 30 \times 2$ meters, and there exist two loops inside.

The process of observing a visual landmark is simulated in the following way: we consider that a robot is able to observe a landmark in the environment when it lies in the field of view of the stereo cameras and the relative distance is below a threshold d_{max}. As previously said, an observation is constituted by $z_{t,\langle i\rangle} = (v_{t,\langle i\rangle}, d_{t,\langle i\rangle})$, where $v_{t,\langle i\rangle} = (X_r, Y_r, Z_r)$ is a relative measurement and $d_{t,\langle i\rangle}$ is a SIFT descriptor associated to the point. When a robot observes a landmark, we compute $v_{t,\langle i\rangle}$ using Equations (14) and add gaussian noise to the measurement as computed by Equations (15). To simulate the process of observing a landmark from different viewpoints, we add gaussian noise to the descriptor $d_{t,\langle i\rangle}$ and study the effects of this variations in the results.

Given an observation $z_{t,\langle i\rangle} = (v_{t,\langle i\rangle}, d_{t,\langle i\rangle})$ obtained by one of the robots, the data association is solved as explained in Section 5, thus assigning the observation $z_{t,\langle i\rangle} = (v_{t,\langle i\rangle}, d_{t,\langle i\rangle})$ to the landmark θ_j in the map, or creating a new landmark.

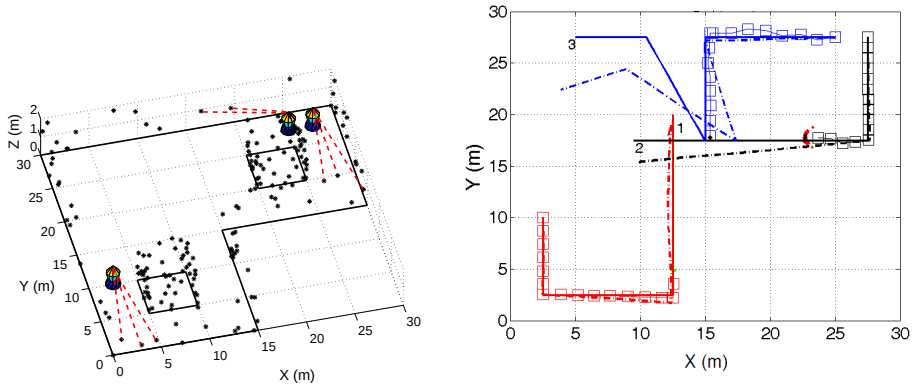
Once the simulation is finished, the path s^t estimated is compared with the

true path and an RMS error is computed with the differences along the path. Analogously, the true map is compared with the map of the most probable particle, as determined by Equation (17) and a RMS error is computed in base of the Euclidean distance between the true an estimated path. The generation of a new set of particles creates a random set of new poses by sampling from the motion model stated in Equation (7). In consequence, different results may be obtained when running the simulation repeatedly with the same data. In consequence, the simulations are repeated a number of times, computing a mean value and 2σ intervals.

7 Results

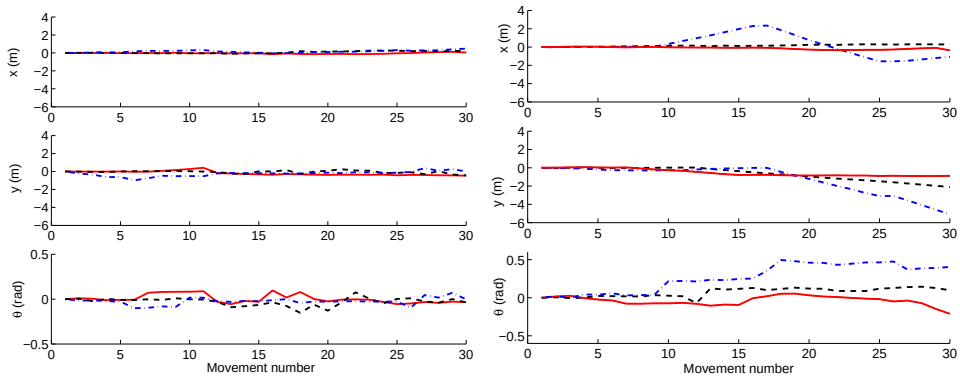
SLAM solutions aim at building a map, given a set of movements performed and observations obtained by the robots. However, SLAM approaches do not consider the problem of computing the trajectory of the robots, e.g. in order to minimize the time needed to explore the environment. In particular, this is normally considered a different problem, denoted as Exploration. In the simulated experiments we consider that the robots are able to explore the environment in an efficient manner. That means that, if a robot needs T movements to explore the whole environment, we consider that each one of the K robots require T/K movements to achieve the exploration of the environment. In other words, the total distance traveled with only one robot is divided along the group so that every robot traverses the same distance. The paths followed by the robots were set by hand, considering this requirement. In addition, different paths were used in the experiments, with variations in the initial position of the robots. During the simulation, the only information that the robots can use to derive its position is the noisy odometry and the measurements performed over the landmarks in the map.

In Figure 3 we show the results of a particular simulation conducted with a team of 3 robots. Figure 3(a) shows the environment and the position of the robots. The true position of each landmark is indicated with and asterisk, whereas the observations obtained by the robots are designated with dashed lines. Figure 3(b) shows the simulation at a particular step. The true path followed by the robots is shown with a solid line. Using Equation (7) a noisy odometry is simulated, which is shown as a dashed line. The noise parameters used in the odometry have values that resemble the noise in real robots from our laboratory. In the same figure, we present the best estimation over the paths followed by the robots, denoted with squares. In addition, the uncertainty over the pose of each one of the robots is presented with a set of particles. In Figure 4(a) we show the difference between the estimated and the true poses during all the movements performed by the robots. In Figure 4(b) we show the error between the odometry and the true poses for all



(a) (b)

Fig. 3. Figure (a) shows the simulated environment. Figure (b) shows the true path followed by the robots (continuous line) the odometry (dashed line) and the estimated paths (marked with squares).



(a) (b)

Fig. 4. Fig. (a) shows the error between the estimated and the true pose (x, y, θ) for each of the movements. Fig. (b) shows the error between the odometry and the true pose. In both figures, the first robot is indicated with a continuous line, the second with a dashed line, and the third is indicated with a dash-dot line.

the robots. We can clearly see that the error in odometry grows without bound and cannot be directly used to estimate the map. It can be clearly seen that the error in the estimated path is reduced significantly when compared with the odometry.

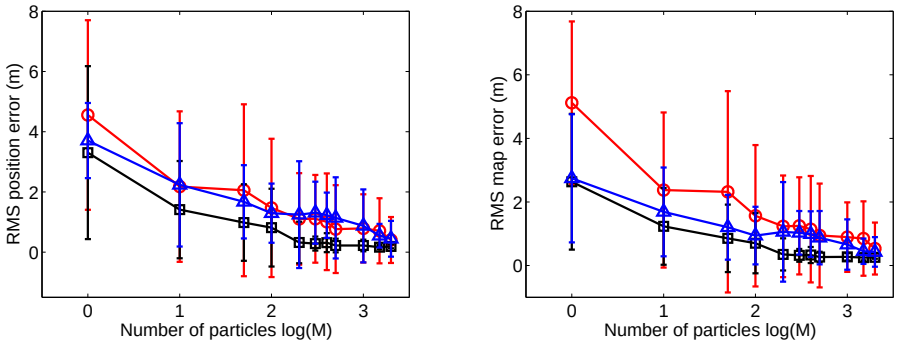
We conducted a series of experiments and evaluated the results when varying the parameters used by the SLAM algorithm to estimate the map and the paths. The most important parameters that were studied are gathered at Table 2. The same simulations were performed using different number of robots.

First, we performed a series of simulations with different number M of particles. We consider that each robot obtains B measurements at each simulation

Table 2
Most important parameters studied in the simulations

$M = 500$	number of particles used in the filter
$B = 15$	number of observations integrated at each simulation step.
$d_{max} = 10$ m	maximum distance at which the observations are made.
$\sigma_d = 0.1$ pix	standard deviation in the disparity computation.
$\sigma_{d_t} = 0.01$	standard deviation of the noise introduced in the descriptor.

step, which are integrated in the Rao-Blackwellized filter. As a result of each simulation, we obtain a RMS error over the robot paths and the map. The RMS error in position is computed with respect to the true path. The same experiment was repeated several times for each value of M, calculating a mean value and a standard deviation over the results for each M. Different trajectories were performed by the robots, considering different starting positions in the map. When varying M the rest of parameters remain constant, with the values indicated in Table 2. The values that appear in the table are reasonable and correspond to parameters used when building a visual map with a single robot using real data [5]. Figure 5(a) shows the RMS error in the estimated paths with error bars indicating 2σ intervals, whereas Figure 5(b) presents the error in the map. Both, Figure 5(a) and 5(b) present the results when 1, 2 or 3 robots are used and follow a similar trend. In general, for any number of robots deployed, we can clearly observe that the error in the estimation decreases when the number of particles increases. This result was expected, since the function $p(x^t_{\langle 1:K\rangle}, L|z^t_{\langle 1:K\rangle}, u^t_{\langle 1:K\rangle}, c^t)$ is approximated with more precision when the number of particles M is large. The state to estimate is of dimension $3K$, being K the number of robots. According to [21] in a general case, the number of particles needed to obtain an accurate estimation grows exponentially with the dimension of the state. As a result, we should expect that the RMS error when using 3 robots would be always greater than the error obtained with one robot, since more particles are needed to accurately estimate 3 robot paths. However, the results demonstrate that for the same number of particles, more accurate results are obtained when using more robots. For example, Figure 5(b) demonstrates that the map is more accurately estimated when the number of robots increases. This result can be explained in the following manner: when one robot explores the environment on its own, it must return to previously explored places in order to reduce the uncertainty over its pose. In a typical environment, as the one simulated, the robot needs to travel a long distance in order to re-visit previously explored places (i.e. close a loop) and accumulates a high uncertainty. This normally induces big errors in the estimation of the path and requires a big number of particles for an accurate estimation. In the same environment, when more robots are used, it occurs frequently that one of the robots observes landmarks observed previously by other robots. In the algorithm proposed, this has the same effect as if the robot observed landmarks detected before and the uncertainty is maintained low. As a consequence, since the uncertainty of each robot is low, it can be represented using less particles. To sum up, for a given number of



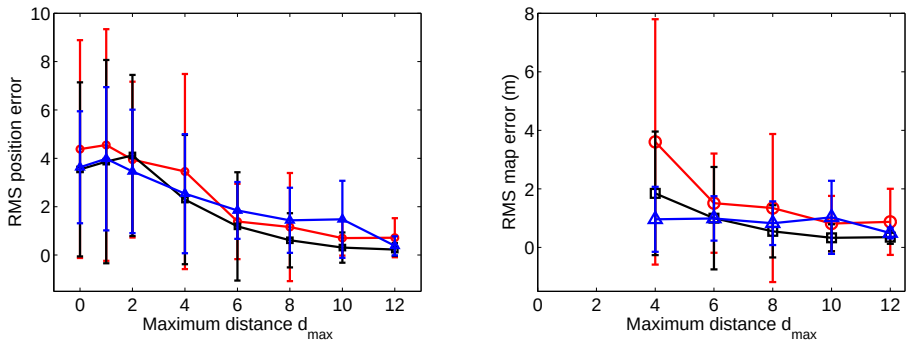
(a) (b)

Fig. 5. Fig. (a) shows the root mean square (RMS) error in the estimated path of the robots when varying the number of particles $M = \{1\ 10\ 50\ 100\ 200\ 300\ 400\ 500\ 1000\ 1500\ 2000\}$. Fig. (b) shows the RMS error in the estimated map when varying the number of particles M. The error bars are defined as two times the standard deviation. We show simultaneously results with one robot (◯), two robots (□) and three robots (△).

particles, the map can be estimated with comparable precision using 1, 2 or 3 robots. The time needed to explore the environment is obviously reduced when more robots are used.

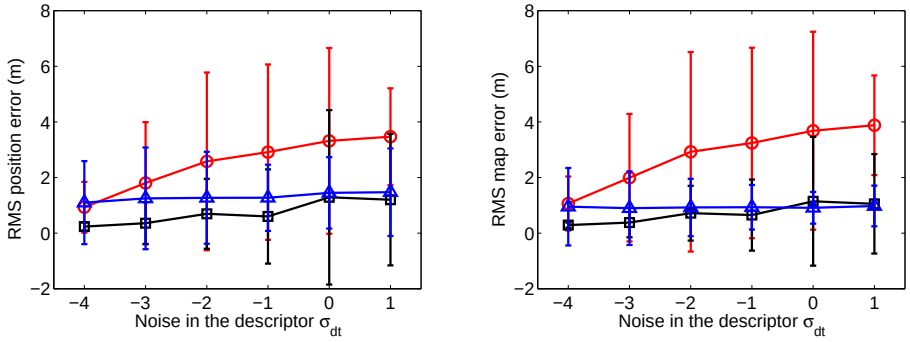
During the simulations, we assumed that the robot could detect a landmark when it was placed at a distance below d_{max} and inside the field of view of the cameras. We performed different simulations changing the maximum detection distance d_{max} while maintaining constant the rest of parameters shown in Table 2. The results are presented in Figure 6(a) and 6(b). It can be observed that the localization error decreases when the maximum distance increases. This result can be explained in the following way: when the robot observes landmarks only at a close distance, it continuously explores new space, and the uncertainty over its pose increases rapidly. On the contrary, when the observation distance is large, the robot maintains its localization inside its local map, and its uncertainty remains low. When using more robots, a bigger observation distance d_{max} implies that with more probability a robot may be able to observe landmarks detected by other robots, thus improving the localization and the map, as previously justified.

We have also studied the effects on the map contruction that are produced by adding noise to the visual descriptor. As explained in Section 6 each landmark in the map is associated with a SIFT descriptor. Once the robot observes a landmark, it obtains an observation $z_{t,\langle i\rangle} = (v_{t,\langle i\rangle}, d_{t,\langle i\rangle})$. The relative measurement is corrupted with a gaussian noise defined by $N(0, R_t)$ using Equations (15), whereas a descriptor $d_{t,\langle i\rangle}$ is associated considering $d_{t,\langle i\rangle} = d_j + N(0, \sigma^2_{d_t})$, where d_j is the SIFT descriptor of the landmark θ_j in the map and $N(0, \sigma^2_{d_t})$ represents a 128-dimensional vector of uncorrelated



(a) (b)

Fig. 6. Figure (a) shows the root mean square error (RMS) computed over the estimated path computed for different values of d_{max}. Figure (b) shows the RMS error computed between the real and the estimated map when the distance d_{max} is changed. The error bars are defined as two times the standard deviation.

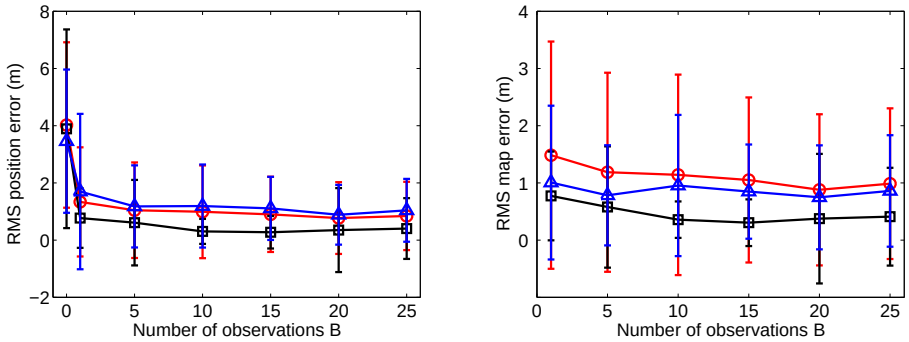


(a) (b)

Fig. 7. Figure (a) shows the RMS error computed over the estimated path when noise is added to the visual descriptor σ_{desc}. Figure (b) shows the RMS error computed between the real and the estimated map. We show simultaneously results with one robot (◯), two robots (□) and three robots (△).

gaussian noise with variance $\sigma_{d_t}^2$. As explained in Section 5 the data association makes use of the descriptor associated to each landmark. In consequence, adding more noise to the observed descriptor implies that more false data associations will be produced, and this affects negatively the quality of the results. We performed a series of simulations increasing the noise added to the descriptor while maintaining the rest of parameters constant (with the values shown in Table 2). The results are presented in Figure 7. It can be noticed that, as the noise in the descriptor increases, false data associations are likely to occur, and the results get worse. When using more robots, the number of observations integrated is greater, and this fact makes the process more robust to noise.

Next, we present the influence in the results when a different number of observations are integrated at each iteration of the algorithm. At each simulation



(a) (b)

Fig. 8. Fig. (a) shows the root mean square (RMS) error in the estimated path of the robots when varying the number of observations integrated at each iteration of the algorithm $B = \{0\ 1\ 5\ 10\ 15\ 20\ 25\}$. Fig. (b) shows the RMS error in the estimated map when varying the number of particles M. The error bars are defined as two times the standard deviation. We show simultaneously results with one robot (◯), two robots (□) and three robots (△).

we used the parameters shown in Table 2 and obtained different results when changing B. The results are presented in Figure 8. Figure 8(a) present the RMS error in position computed for the paths and Figure 8(b) shows the RMS error in the map. It can be noticed that the results in the estimation of the robot paths and the map improve when more observations are integrated at each time step. In addition, the results exhibit the same trend as observed in the experiments previously presented.

Finally, we compare the processing time needed at each iteration of the algorithm when using a different number of robots. To do this, we performed the exploration using different number of robots and measured the time needed at each iteration. The time needed at each iteration of the algorithm depends on a series of parameters, basically the number of particles M used, the number of observations B integrated at each time step. However, this time is not constant and increases with the number of landmarks that exist in the map: with more landmarks in the map, finding the correct data association becomes computationally expensive. Figure 9(a) presents the mean time needed to process the observations at each iteration of the algorithm, using a different number of particles used in the filter and considering that each of the robots integrates B measurements. It can be seen that, as the number of particles M increases, the time needed to process the observations grows exponentially, and is significantly greater when using 2 or 3 robots.

In a real exploration situation, when the robots navigate through the environment, the observations must be processed within time constraints. First, the observations must be integrated in the map. Second, based on this map, the exploration algorithm plans the next movement for each robot. If the process-

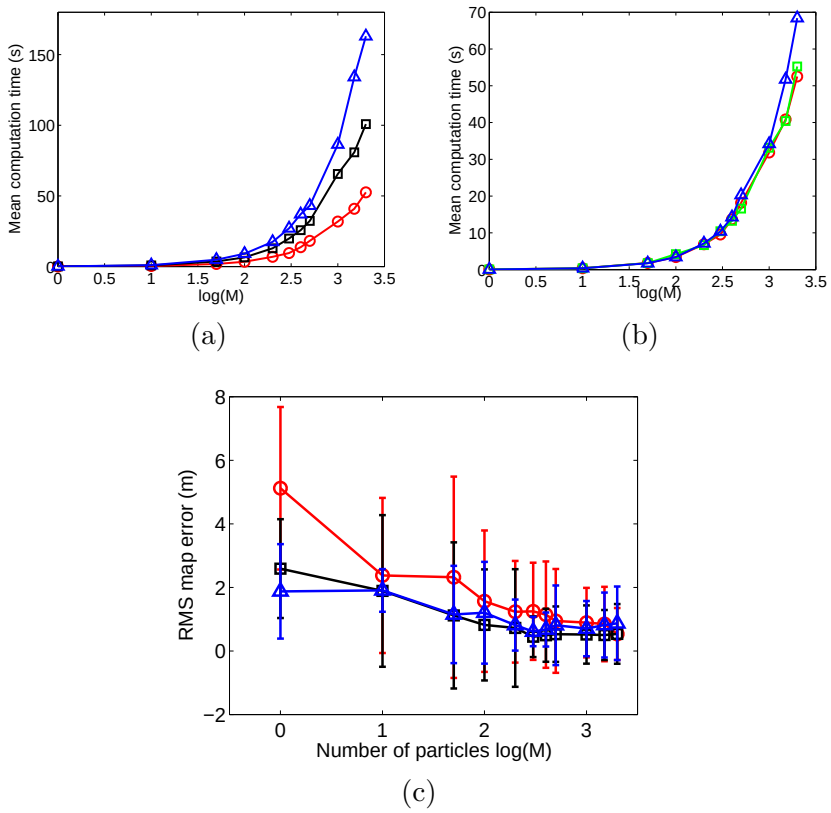


Fig. 9. Figure (a) and (b) shows the mean computation time at each iteration of the algorithm for different number of particles $M = \{1\ 10\ 50\ 100\ 200\ 300\ 400\ 500\ 1000\ 1500\ 2000\}$. Figure (c)shows the RMS error in the map when B/K observations are used.

ing time is too high, then the robots would need to stop their movements and wait until the new observations have been integrated in the filter. In order to decrease the computational cost, we can reduce the number of observations that each robot introduces at each iteration of the algorithm. We consider that each robot obtains B/K observations. That is, when using one robot, we integrate $B = 15$ observations and when using three robots, we integrate $B = 5$ observations. In Figure 9(b) we present the mean processing time if the robots perform B/K observations at each iteration of the algorithm. As can be seen in the figure, the processing time in the case of 1, 2 or 3 robots is similar, for any number of particles. In this case, the total number of observations integrated in the filter is the same for the case of 1, 2 or 3 robots. With this restriction, we evaluated the error in the map, which is shown in Figure 9(c). It can be observed that the results are similar to the presented in Figure 5(b), since the estimation of the map and the paths improves when more robots are used. As a consequence, we can reduce the computational cost of the algorithm by integrating less observations by each robot and still obtain comparable good results.

8 Conclusion

We have presented an approach to visual SLAM that builds a 3D map of the environment using a team of cooperative robots. We consider that the robots are equipped with a stereo heads that allow to perform observations over visual landmarks, each one consisting of a relative measurement and a visual descriptor. The approach builds a unique map of the environment that is shared by all the robots in the team, using a Rao-Blackwellized particle filter to estimate both the map and the trajectories of the robots. The data association problem is solved using the relative measurements obtained by the robots and comparing the visual descriptor. With this strategy we have found that the number of false correspondences is low, thus allowing to obtain good estimations of the map and the path.

We have tested the validness of the approach by means of simulations. For this purpose, an indoor environment has been simulated. We used a noise model to represent the relative measurements obtained by the stereo cameras and used typical values for the noise in the odometry model, as estimated in real robots in our laboratory. In the experiments we assumed that the robots were able to explore the environment efficiently, thus we compare the case in which the total distance traveled by one robot to explore the environment is divided uniformly along the team members. A great number of simulations have been performed, that precisely resemble real exploring situation. The simulations allowed us to test the approach under different conditions, by changing the parameters of the Rao-Blackwellized particle filter. We could also evaluate the algorithm when the descriptor associated to a visual landmarks was corrupted with noise, and evaluated its effects on the results. With this procedure we could emulate the effects of observing a landmark from different viewpoints. The results presented show that the algorithm is robust to false data associations, and is able to produce a good solution with a wide range of parameter settings.

In the presented approach, the dimension of the state to estimate grows with the number of robots. In principle, the number of particles needed to obtain a good estimation grows exponentially with the dimension of the state to estimate. Under typical conditions we show that using the same number of particles M, the precision of the map estimated using a single robot is similar to the precision obtained using 2 or 3 robots. In addition, the results show that, when more robots are used, a precise map can be obtained without increasing the computational cost of the method. As a consequence, since more robots exist in the environment, the time needed to explore the environment will be reduced significantly, whereas the computational cost is not increased.

As a conclusion, under the assumptions considered, we have demonstrated that

the proposed SLAM algorithm is suitable for small groups of robots exploring a given environment and obtaining observations over visual landmarks. We show that precise results can be obtained, without increasing the computational cost considerably.

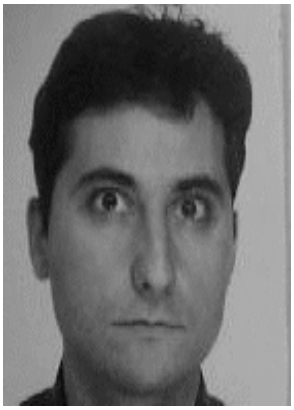
As a future work, we plan to validate the simulations by using experimental data obtained by real robots and natural landmarks occurring in the environment. We further plan to use the mutual observations of the team, that is, when a robot observes the relative position of another member in the team. We expect that this information will lead to further improve the results.

9 Acknowledgment

This work has been supported by the spanish government by means of project: 'Sistemas de Percepción Visual Móvil y Cooperativo como Soporte para la Realización de Tareas con Redes de Robots'. Project reference: CICYT DPI2007-61107, Ministerio de Educación y Ciencia.

10 Biographies

Arturo Gil Aparicio received the M. Eng. degree in Industrial Engineering from the Miguel Hernández University (UMH), Elche, Spain, in 2002, receiving also the best student academic award in Industrial Engineering by the UMH. Since 2003, he works as a lecturer and researcher at the UMH, teaching subjects related to Control and Computer Vision. His research interests are focussed on mobile robotics, visual SLAM and cooperative robotics. He is currently working on techniques to build visual maps using teams of mobile robots.



Óscar Reinoso García received the M. Eng. degree from the Polytechnical University of Madrid (UPM), Madrid, Spain in 1991. Later, he obtained de Ph.D. degree in 1996. He worked at Protos Desarrollo S.A. company in the development and research of artificial vision systems from 1994 to 1997. Since 1997, he works as a professor at the Miguel Hernández University (UMH), teaching subjects related to Control, Robotics and Computer Vision. His research interests are mobile robotics, climbing robots and visual inspection systems. He is member of CEA-IFAC and IEEE.

Mónica Ballesta Galdeano received the M. Eng. degree in Industrial Engineering from the Miguel Hernández University (UMH), Elche, Spain in 2005, receiving also the academic award in Industrial Engineering by the UMH. Since 2007, she has a research position as scholarship holder in the area of Systems Engineering and Automation of the UMH, receiving a pre-doctoral grant (FPI) by the Valencian Government (Generalitat Valenciana). Her research interests are focussed on mobile robots, visual feature extraction and visual SLAM.

Miguel Juliá Cristóbal Miguel Juliá received a M. Eng. degree in Telecommunications Engineering at the Miguel Hernández University (UMH), Elche, Spain. During 2006, he worked as a researcher at the Computer Science and Communications department, at Murcia University. Since 2007 he is involved in research projects at the Systems Engineering department at the UMH. Currently, his research interests focus on mobile robotics, exploration and computer vision.

References

[1] M. Ballesta, O. Martínez-Mozos, A. Gil and O. Reinoso. A Comparison of Local Descriptors for Visual SLAM. In Proceedings of the Workshop on Robotics and Mathematics (RoboMat 2007), 2007.

[2] A.J. Davison. Real-time simultaneous localisation and mapping with a single camera. *Proc. of the Int. Conf. on Computer Vision (ICCV)*, 2003.

[3] M. W. M. Gamini Dissanayake, P. Newman, S. Clark, H. Durrant-Whyte, and M. Csorba. A solution to the simultaneous localization and map building (SLAM) problem. *IEEE Transactions on Robotics and Automation (ICRA)*, 17, 2001.

[4] J.W. Fenwick, P.M. Newman, and J.J. Leonard. Cooperative concurrent mapping and localization. *IEEE Int. Conf. on Robotics and Automation (ICRA)*, pages 1810–1817, 2002.

[5] A. Gil, O. Reinoso, O. Martínez-Mozos, C. Stachniss and W. Burgard. Improving data association in vision-based SLAM. In *Proc. of the IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*, China, 2006.

[6] A. Gil, O. Reinoso, M. A. Vicente, C. Fernández and L. Payá. Monte carlo localization using SIFT features. *Lecture Notes in Computer Science*, I(3523):623–630, 2005.

[7] P. Jensfelt, D. Kragic, J. Folkesson and M. Björkman A Framework for Vision Based Bearing Only 3D SLAM *IEEE Proc. of the Int. Conf. on Robotics & Automation*, 2006.

[8] D. Lowe. Object recognition from local scale-invariant features. In *Proc. of the Int. Conf. on Computer Vision (ICCV)*, pages 1150–1157, 1999.

[9] D. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 2(60):91–110, 2004.

[10] O. Martínez-Mozos, A. Gil, M. Ballesta and O. Reinoso. Interest Point Detectors for Visual SLAM. Lectures Notes in Artificial Intelligence (LNAI), vol. 4788, 2007.

[11] M. Montemerlo and S. Thrun. Simultaneous Localization and Mapping with Unknown Data Association using FastSLAM. *IEEE Int. Conf. on Robotics and Automation (ICRA)*, 2003.

[12] A.C. Murillo, J.J. Guerrero and C. Sagüés. SURF features for efficient robot localization with omnidirectional images. *IEEE Int. Conf. on Robotics and Automation*, 2007.

[13] K. Murphy. Bayesian map learning in dynamic environments. *Neural Information Processing Systems (NIPS)*, 1999.

[14] J. Neira and J. D. Tardós. Data association in stochastic mapping using the joint compatibility test. *IEEE Transactions on Robotics and Automation*, 17(6):890–897, 2001.

[15] L. Parker. Current state of the art in distributed autonomous mobile robotics. *Distributed Autonomous Robotic Systems*, 14(6):3–12, 2000.

[16] S. Se, D. Lowe and J. Little. Vision-based mobile robot localization and mapping using scale-invariant features. *IEEE Int. Conf. on Robotics and Automation (ICRA)*, pages 2051–2058, 2001.

[17] R. Sim, P. Elinas, M. Griffin and J. J. Little. Vision-based slam using the rao-blackwellised particle filter. In *IJCAI Workshop on Reasoning with Uncertainty in Robotics*, 2005.

[18] C. Stachniss, G. Grisetti, D. Hähnel and W. Burgard. Improved rao-blackwellized mapping by adaptive sampling and active loop-closure. In *Proc. of the Workshop on Self-Organization of AdaptiVE behavior (SOAVE)*, pages 1–15, 2004.

[19] B. Stewart, J. Ko, D. Fox, and K. Konolige. A hierarchical bayesian approach to mobile robot map structure estimation. In *In Proceedings of the Conference on Uncertainty in AI (UAI)*, 2003.

[20] S. Thrun. An online mapping algorithm for teams of mobile robots. *Int. Journal of Robotics Research*, 20(5):335–363, 2001.

[21] S. Thrun, W. Burgard and D. Fox. *Probabilistic Robotics*. The MIT Press, 2005. ISBN: 0-262-20162-3.

[22] R. Triebel and W. Burgard. Improving Simultaneous Mapping and Localization in 3D Using Global Constraints. *Proc. of the National Conference on Artificial Intelligence (AAAI)*, 2005.