

# SIMULTANEOUS LOCALIZATION AND MAPPING IN UNMODIFIED ENVIRONMENTS USING STEREO VISION

A. Gil, O. Reinoso, C. Fernández, M. A. Vicente  
*Universidad Miguel Hernández de Elche*  
*Avda. de la Universidad s/n, 03202 Elche (Alicante)*

A. Rottmann, O. Martínez Mozos  
*University of Freiburg*  
*Department of Computer Science*  
*D-79110 Freiburg*

Keywords: SLAM, Stereo Vision, Visual landmarks, Data Association.

Abstract: In this paper we describe an approach that builds three dimensional maps using visual landmarks extracted from images of an unmodified environment. We propose a solution to the Simultaneous Localization and Mapping (SLAM) problem for autonomous mobile robots using visual landmarks. Our map is represented by a set of three dimensional landmarks referred to a global reference frame, each landmark contains a visual descriptor that partially differentiates it from others. Significant points extracted from stereo images are used as natural landmarks, in particular we employ SIFT features found in the environment. We estimate both the map and the path of the robot using a Rao-Blackwellized particle filter, thus the problem is decomposed into two parts: one estimation over robot paths using a particle filter, and  $N$  independent estimations over landmark positions, each one conditioned on the path estimate. We actively track visual landmarks at a local neighbourhood and select only those that are more stable. When a visual feature has been observed from a significant number of frames it is then integrated in the filter. By this procedure, the total number of landmarks in the map is reduced, compared to prior approaches. Due to the tracking of each landmark, we obtain different examples that represent the same natural landmark. We use this fact to improve data association. Finally, efficient resampling techniques have been applied, which reduces the number of particles needed and avoids the particle depletion problem.

## 1 INTRODUCTION

Building an accurate map of a given environment is one of the hardest tasks for a mobile robot. It is inherently difficult, since noise in the estimation of robot's pose leads to errors in the estimation of the map and vice-versa. Recently, many authors have considered the problem of Simultaneous Localization and Mapping (SLAM). The aim of SLAM is to build a map of an unknown environment and simultaneously localise the robot with respect to this map. Here we consider this problem using a Rao-Blackwellized Particle Filter (RBPF).

Most work on SLAM so far has focussed on building 2D maps of environments using range sensors such as SONAR and laser (Wijk and Christensen, 2000), (Thrun, 2001). Recently, Rao-Blackwellized particle filters have been used as an effective mean of solving the SLAM problem using occupancy grid maps (Stachniss et al., 2004). In this approach, each particle constructs its own map based on the observations and the trajectory for that particle. Parti-

cles which best explain current observations are given higher weights and re-sampled.

While mapping a particular environment, one must be aware of two key problems:

- Odometry is subject to cumulative drift.
- Landmarks can be ambiguous.

Recently, some authors have been concentrating on building three dimensional maps using visual information extracted from cameras. In this scenario, the map is represented by a set of three dimensional landmarks related to a global reference frame. Stereo systems are typically less expensive than laser sensors and are able to provide directly 3D information from the scene. In (Little et al., 2001) and (Little et al., 2002) stereo vision is used to track 3D visual landmarks extracted from the environment. In this work, SIFT features are used as visual landmarks. SIFT (Scale Invariant Feature Transform) where initially invented by D. Lowe and used in object recognition tasks (Lowe, 1999). During exploration, the robot extracts SIFT features from stereo images and calcu-

lates relative measurements to them. Landmarks are then integrated in the map with an Extended Kalman Filter associated to it. However, this approach does not manage correctly 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 (Miró et al., 2005) a Kalman filter is used to estimate an augmented state constituted by the robot pose and  $N$  landmark positions (Disanayake et al., 2001). SIFT features are used too to manage the data association among visual landmarks. However, since only one hypothesis is maintained over the robot pose, the method would fail in the presence of incorrect data associations. In addition, in the presence of a significant number of landmarks the method would be computationally expensive.

The work presented in (Sim et al., 2005) uses SIFT features as significant points in space and tracks them over time. It uses a Rao-Blackwellized particle filter to estimate both the map and the path of the robot. The robot movement is here estimated from stereo ego-motion (Little et al., 2001), providing a corrected odometry that simplifies the SLAM problem, since no large odometric errors are introduced.

The most relevant contribution of this paper is twofold. First, we present a new mechanism to deal with the data association problem in the case of different landmarks with similar appearance. This fact may occur in most environments. Second, our approach actively tracks landmarks prior to its integration in the map. As a result only those landmarks that are more stable are incorporated in the map. By using this approach, our map typically consists of a reduced number of landmarks compared to those of (Little et al., 2002) and (Sim et al., 2005), for comparable map sizes. In addition, we have applied effective resampling techniques, as exposed in (Stachniss et al., 2005). This fact reduces the number of particles needed to construct the map, thus reducing computational time.

The remainder of the paper is structured as follows. Section 2 deals with visual landmarks and their utility in SLAM. Section 3 explains the basics of the Rao-Blackwellized particle filter. Next, section 4 presents our solution to the data association problem in the context of visual landmarks. In section 5 we present our experimental results. Finally, section 6 sums up the most important conclusions and proposes future extensions.

## 2 FEATURE-BASED METHODS: VISUAL FEATURES

Previous works in map building has revolved around two topics: occupancy or certainty grids, and feature-based methods. Feature based methods work by locating features in the environment, estimating their position, and then using them as known landmarks. Different kinds of features have been used to create a map of the environment. In our work, we use visual landmarks as features to build the map. In particular, we use SIFT features (Scale Invariant Feature Transform) which were developed for image feature generation, and used initially in object recognition applications (see (Lowe, 2004) and (Lowe, 1999) for some examples). Key locations are selected at maxima and minima of a difference of Gaussian function applied in scale space. The features are invariant to image translation, scaling, rotation, and partially invariant to illumination changes and affine or 3D projection. They are computed by building an image pyramid with resampling between each level. SIFT locations extracted by this procedure may be understood as significant points in space that are highly distinctive, thus can be found from a set of robot poses. In addition, each SIFT location is given a descriptor that describes this landmark. Thus, this process enables the same points in the space to be recognized from different viewpoints, which may occur while the robot moves around its workplace, thus providing information for the localization process. SIFT features have been used in robotic applications, showing its suitability for localization and SLAM tasks (Little et al., 2001), (Little et al., 2002), (Sim et al., 2005). Figure 1 shows a visual features extracted from an image. These SIFT locations are used as landmarks in the map. The size of the arrow on each feature is proportional to SIFT's scale.

## 3 RAO-BLACKWELLIZED SLAM

We estimate the map and the path of the robot using a Rao-Blackwellized particle filter. Using the nomenclature in this filter, we denote as  $s_t$  the robot pose at time  $t$ . On the other hand, the robot path until time  $t$  will be denoted  $s^t = \{s_1, s_2, \dots, s_t\}$ , the set of observations made by the robot until time  $t$  will be denoted  $z^t = \{z_1, z_2, \dots, z_t\}$  and the set of actions  $u^t = \{u_1, u_2, \dots, u_t\}$ . Therefore, the SLAM problem can be formulated as that of determining the location of all landmarks in the map  $\Theta$  and robot poses  $s^t$  from a set of measurements  $z^t$  and robot actions  $u^t$ . The map is composed as a set of different landmarks  $\Theta = \{\theta_1, \theta_2, \dots, \theta_i, \dots, \theta_N\}$ . We call  $c^t$  to the correspondence of the landmarks extracted from

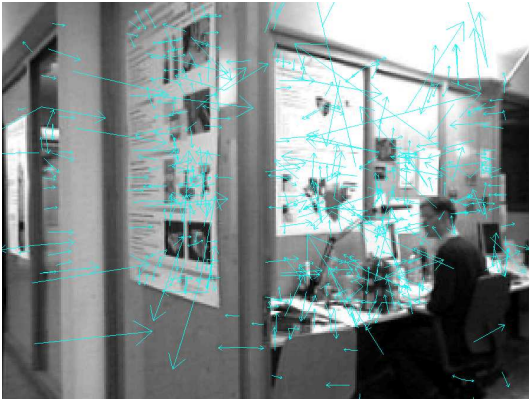


Figure 1: Image shows SIFT locations extracted from a typical image. The size of the arrow is proportional to SIFT's scale.

the data association problem. In consequence, the SLAM problem can be formulated as the estimation of the following probability density function over robot poses:

$$p(s^t, \Theta | z^t, u^t, c^t) \quad (1)$$

The map  $\Theta$  is represented by a collection of  $N$  landmarks. Each landmark is described as:  $\theta_k = \{\mu_k, \Sigma_k, d_k\}$ , where  $\mu_k = (X_k^g, Y_k^g, Z_k^g)$  is a vector describing the position of the landmark referred to a global reference frame  $O_g$  with associated covariance matrix  $\Sigma_k$ . In addition, each landmark  $\theta_k$  is associated with a SIFT 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 (Gil et al., 2005). We can see in figure 2 an example of the map represented by  $N$  landmarks in a section of the environment.

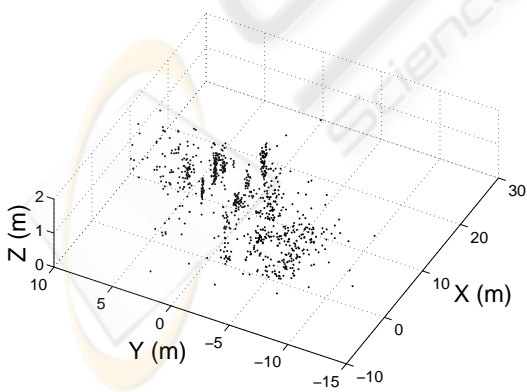


Figure 2: Example of the map created by a set of  $N$  3D landmarks in a section of the environment in which the robot is present.

While exploring a particular environment, the robot



Figure 3: Stereo correspondences using SIFT features. The epipolar constrain is used to find correspondences across images.

needs to determine whether a particular observation  $z_{t,k}$  corresponds to a previously mapped landmark or to a new one. For the moment, we consider this correspondence as known. Given that, at a time  $t$  the map is formed by  $N$  landmarks, the correspondence is represented by  $c_t = \{c_{t,1}, c_{t,2}, \dots, c_{t,B}\}$ , where  $c_{t,i} \in [1 \dots N]$ . In consequence at a time  $t$  the observation  $z_{t,k}$  corresponds to the landmark  $c_{t,k}$  in the map. When no correspondence is found we denote it as  $c_{t,i} = N + 1$ , indicating that it is a new landmark.

### 3.1 Stereo SIFT

Given two images  $I_t^L$  and  $I_t^R$ , captured with a stereo head at a time  $t$ , we extract natural landmarks which correspond to points in the 3-dimensional space. Each point is accompanied by its SIFT descriptor and then matched across images. The matching procedure is constrained by the well-known epipolar geometry of the stereo rig. In addition, a comparison between SIFT descriptors is used to avoid false correspondences. As a result, at a time  $t$  we obtain a set of  $B$  observations denoted by  $z_t = \{z_{t,1}, z_{t,2}, \dots, z_{t,B}\} = \{v_{t,1}, d_{t,1}, v_{t,2}, d_{t,2}, \dots, v_{t,B}, d_{t,B}\}$ . A particular observation is constituted by  $z_{t,k} = (v_{t,k}, d_{t,k})$ , where  $v_{t,k} = (X^r, Y^r, Z^r)$  is a three dimensional vector represented in the left camera reference frame and  $d_{t,k}$  is the SIFT descriptor associated to that point. Figure 3 shows two stereo images of the environment. Correspondences found across both images are shown. After stereo correspondence, a 3D reconstruction of the points is obtained, obtaining  $B$  measurements  $v_{t,k} = (X^r, Y^r, Z^r)$  relative to the left camera reference frame.



### 3.2 Particle Filter Estimation

The conditional independence property of the SLAM problem implies that the posterior (1) can be factored as (Montemerlo et al., 2002):

$$p(s^t, \Theta | z^t, u^t, c^t) = p(s^t | z^t, u^t, c^t) \prod_{k=1}^N p(\theta_k | s^t, z^t, u^t, c^t) \quad (2)$$

This equation states that the full SLAM posterior is decomposed into two parts: one estimator over robot paths, and  $N$  independent estimators over landmark positions, each conditioned on the path estimate. This factorization was first presented by Murphy in 1999 (Murphy, 1999). We approximate  $p(s^t | z^t, u^t, c^t)$  using a set of  $M$  particles, each particle having  $N$  independent landmark estimators (implemented as EKF), one for each landmark in the map. Each particle is thus defined as:

$$S_t^{[m]} = \{s_t^{t,[m]}, \mu_{t,1}^{[m]}, \Sigma_{t,1}^{[m]}, \dots, \mu_{t,N}^{[m]}, \Sigma_{t,N}^{[m]}\} \quad (3)$$

Where  $\mu_{t,i}^{[m]}$  is the best estimation at time  $t$  for the position of landmark  $\theta_i$  based on the path of the particle  $m$  and  $\Sigma_{t,i}^{[m]}$  its associated covariance matrix. The particle set  $S_t = \{S_t^{[1]}, S_t^{[2]}, \dots, S_t^{[M]}\}$  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  $s_t^{[m]} \sim p(s_t | s_{t-1}, u_t)$ . Next, and following the approach of (Montemerlo et al., 2002) each particle is then assigned a weight according to:

$$\omega_{t,i}^{[m]} = \frac{1}{\sqrt{|2\pi Z_{c_{t,i}}|}} \exp\left\{-\frac{1}{2}a\right\} \quad (4)$$

where

$$a = (v_{t,i} - \hat{v}_{t,c_{t,i}})^T [Z_{c_{t,i}}]^{-1} (v_{t,i} - \hat{v}_{t,c_{t,i}})$$

Where  $v_{t,i}$  is the current measurement and  $\hat{v}_{t,c_{t,i}}$  is the predicted measurement for the landmark  $c_{t,i}$  based on the pose  $s_t^{[i]}$ . The matrix  $Z_{c_{t,i}}$  is the covariance matrix associated with the innovation  $(v_{t,i} - \hat{v}_{t,c_{t,i}})$ . Note that we implicitly assume that each measurement  $v_{t,i}$  has been assigned to the landmark  $c_{t,i}$  of the map. This problem is, in general, hard to solve, since similar-looking landmarks may exist. In section 4 we describe our approach to this problem. In the case that  $B$  observations from different landmarks exist at a time  $t$ , we calculate the total weight assigned to the particle as:

$$\omega_t^{[m]} = \prod_{i=1}^B w_{t,i}^{[m]} \quad (5)$$

### 3.3 Efficient Resampling

In order to assess for the difference between the proposal and the target distribution, each particle is drawn with replacement with probability proportional to this importance weight. During resampling, particles with a low weight are normally replaced by others with a higher weight. It is a well known problem that the resampling step may delete good particles from the set and cause particle depletion. In order to avoid this problem we follow an approach similar to (Stachniss et al., 2004). Thus we calculate the number of efficient particles  $N_{eff}$  as:

$$N_{eff} = \frac{1}{\sum_{i=1}^M \omega_t^{[i]}} \quad (6)$$

We resample each time  $N_{eff}$  drops below a predefined threshold (set to  $M/2$  in our application). By using this approach we have verified that the number of particles needed to achieve good results is reduced.

## 4 DATA ASSOCIATION

While the robot explores the environment it must decide whether the observation  $z_{t,k} = (v_{t,k}, d_{t,k})$  corresponds to a previously mapped landmark or to a different landmark. We use each associated SIFT descriptor  $d_{t,k}$  to improve the data association process. Each SIFT descriptor is a 128-long vector computed from the image gradient at a local neighbourhood of the interest point. Experimental results in object recognition applications have showed that this description is robust against changes in scale, viewpoint and illumination (Lowe, 2004). In the approaches of (Little et al., 2001), (Little et al., 2002) and (Sim et al., 2005), data association is based on the squared Euclidean distance between descriptors. In consequence, given two SIFT descriptors  $d_i$  and  $d_j$  the following distance function is computed:

$$E = (d_i - d_j)(d_i - d_j)^T \quad (7)$$

Then, the landmark of the map that minimizes the distance  $E$  is chosen. Whenever the distance  $E$  is below a certain threshold, the two landmarks are considered to be the same. On the other hand, a new landmark is created whenever the distance  $E$  exceeds a pre-defined threshold. When the same point is viewed from slightly different viewpoints and distances, the values in its SIFT descriptor remain almost unchanged. However, when the same point is viewed from significantly different viewpoints (e.g. 30 degrees apart) the difference in the descriptor is remarkable. In the presence of similar looking landmarks,

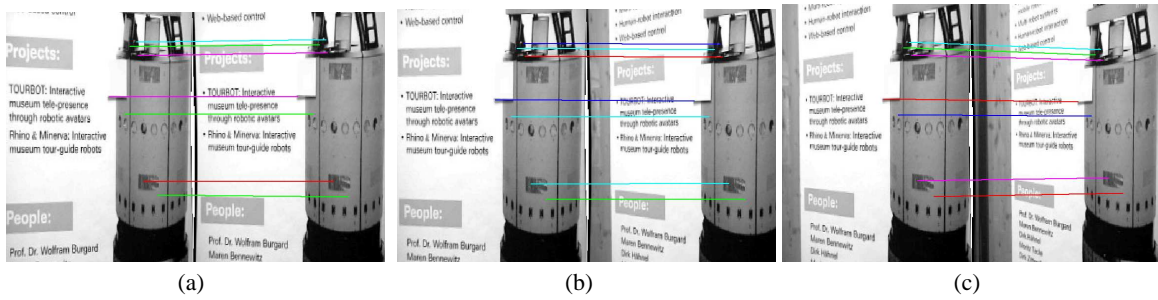


Figure 4: Figures (a)-(c) show different SIFT points tracked along different images with variations in scale and orientation.

Table 1: Comparison of correct and incorrect matches using the Euclidean distance and the Mahalanobis distance.

	Correct matches	Incorrect matches
Euclidean distance	83.85	16.15
Mahalanobis distance	94.04	5.96

this fact causes a significant number of false correspondences, as can be seen in the first column of table 1.

We propose a different method to deal with the data association in the context of SIFT features. We address the problem from a pattern classification point of view. We consider the problem of assigning a pattern  $d_j$  to a class  $C_i$ . Each class  $C_i$  models a landmark. We consider different views of the same visual landmark as different patterns belonging to  $C_i$ . Whenever a landmark is found, it is tracked along  $p$  frames and its descriptors  $d_1, d_2, \dots, d_p$  are stored. Then, for each landmark  $C_i$  we compute a mean value  $\bar{d}_i$  and estimate a covariance matrix  $S_i$ , assuming the elements in the SIFT descriptor independent. Based on this data we compute the Mahalanobis distance:

$$L = (\bar{d}_i - \bar{d}_j) S_i^{-1} (\bar{d}_i - \bar{d}_j)^T \quad (8)$$

Where  $S_i$  is a diagonal covariance matrix associated with the class  $C_i$  with mean value  $\bar{d}_i$ . We compute the distance  $L$  for all the landmarks in the map of each particle and assign the correspondence to the landmark that minimizes  $L$ . If none of the values exceeds a predefined threshold then we consider it a new landmark.

In order to test this distance function we have recorded a set of images with little variations of viewpoint and distance (see figure 4). SIFT landmarks are easily tracked across consecutive frames, since the variance in the descriptor is low. In addition, we visually judged the correspondence across images. Based on these data we compute the matrix  $S_i$  for each SIFT

point tracked for more than 5 frames. Following, we compute the distance to the same class using equation (7) and (8). For each experiment, we select the class that minimises its distance function and as we already know the correspondences, we can compute the number of mistakes and correct matches. Table 1 shows the results based on our experiments. A total of 3000 examples were used. As can be clearly seen, a raw comparison of two SIFT descriptors using the Euclidean distance does not provide total separation between landmarks, since the descriptor can vary significantly from different viewpoints. As can be seen, the number of false correspondences is reduced by using the Mahalanobis distance.

By viewing different examples of the same landmark we are able to build a more complete model of it and this permits us to better separate each landmark from others. We consider that this approach reduces the number of false correspondences and, consequently produces better results in the estimation of the map and the path of the robot, as will be shown in section 5.

## 5 EXPERIMENTAL RESULTS

During the experiments we used a B21r robot equipped with a stereo head and a LMS laser range finder. We manually steered the robot and moved it through the rooms of the building 79 of the University of Freiburg. A total of 507 stereo images at a resolution of 320x240 were collected. The total traversed distance of the robot is approximately 80m. For each pair of stereo images a number of correspondences were established and observations  $z_t = \{z_{t,1}, z_{t,2}, \dots, z_{t,B}\} = \{v_{t,1}, d_{t,1}, v_{t,2}, d_{t,2}, \dots, v_{t,B}, d_{t,B}\}$  were obtained. After stereo correspondence, each point is tracked for a number of frames. By this procedure we can assure that the SIFT point is stable and can be viewed from a significant number of robot poses. Practically, when a landmark has been tracked for more than 5 frames it is

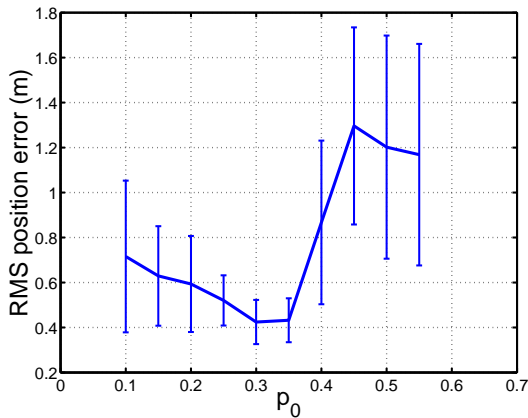


Figure 5: Results in RMS localization error when varying  $p_0$ .

considered a new observation and is integrated in the filter. As mentioned in section 4, each descriptor is now represented by  $d_{t,i} = \{\bar{d}_{t,i}, S_i\}$  where  $\bar{d}_{t,i}$  is the SIFT vector computed as the mean of the  $p$  tracked landmarks and  $S_i$  is the corresponding diagonal covariance matrix.

## 5.1 Threshold Selection

In a practical way, given a SIFT descriptor and a predicted landmark position, the robot needs to decide whether a particular observation belongs to a landmark  $c_{t,k}$  in the map or it is a new one. In our approach we used equation 8 and set up a particular threshold  $p_0$ . Based on a particular observation  $z_{t,k} = \{v_{t,k}, d_{t,k}\}$ , we find the landmark that minimizes  $L$  for the landmarks that lay in a neighbourhood. When the distance  $L$  exceeds  $p_0$  for all the landmarks in the neighbourhood, we create a new landmark. If  $L < p_0$  for a particular landmark in the map, we update its Kalman filter. In consequence if  $p_0$  is set too low, landmarks will be frequently confused. On the contrary, if we set  $p_0$  too high, new landmarks will be frequently created, which affects the quality of the estimated path and the resulting map.

To test the influence of  $p_0$  we repeatedly ran the algorithm changing its value. Figure 5 shows the RMS error in the position of the robot when the value of  $p_0$  is changed. In order to test the quality of our results, we have compared the estimated pose of our method with the estimated pose using laser data recorded during exploration. To calculate the pose based on laser measurements, the method exposed in (Stachniss et al., 2004) was used. The SLAM algorithm was run several times for each value of  $p_0$ . Finally, to build the map we choose the value of  $p_0$  that produces better results. During SLAM, the true pose of the robot is not known, in consequence, the value of

$p_0$  that produces better results cannot be calculated by this way. However, we believe that the chosen value of  $p_0$  will produce good results when mapping different environments.

Figure 6 shows the map constructed with 1, 10, and 100 particles. A total number of 1500 landmarks were estimated. It can be seen that, with only 10 particles, the map is topologically correct. As can be seen in the figures, some areas of the map do not possess any landmark. This is due to the existence of featureless areas in the environment (i.e. texture-less walls), where no SIFT features can be found.

Figure 7(a) shows the error in localization for each movement of the robot during exploration using 200 particles. Again, we compare the estimated position of the robot using our approach to the estimation using laser data.

In addition, we have compared both approaches to data association as described in section 4. To do this, we have made a number of simulations varying the number of particles used in each simulation. The process was repeated using both data association methods. As can be seen in figure 7(b) for the same number of particles, better localization results are obtained when the Mahalanobis distance is used, thus improving the quality of the estimated map.

Compared to preceding approaches our method uses less particles to achieve good results. For example, in (Sim et al., 2005), a total of 400 particles are needed to compute a topologically correct map, while correct maps have been built using 50 particles with our method. In addition, our maps typically consists of about 1500 landmarks, a much more compact representation than the presented in (Sim et al., 2005), where the map contains typically around 100.000 landmarks.

## 6 CONCLUSION AND FUTURE WORK

In this paper a solution to SLAM based on a Rao-Blackwellized particle filter has been presented. This filter uses visual information extracted from cameras. We have used natural landmarks as features for the construction of the map. The method is able to build 3D maps of a particular environment using relative measurements extracted from a stereo pair of cameras.

We have also proposed an alternative method to deal with the data association problem in the context of visual landmarks, addressing the problem from a pattern classification point of view. When different examples of a particular SIFT descriptor exist (belonging to the same landmark) we obtain a probabilistic model for it. Also we have compared the re-

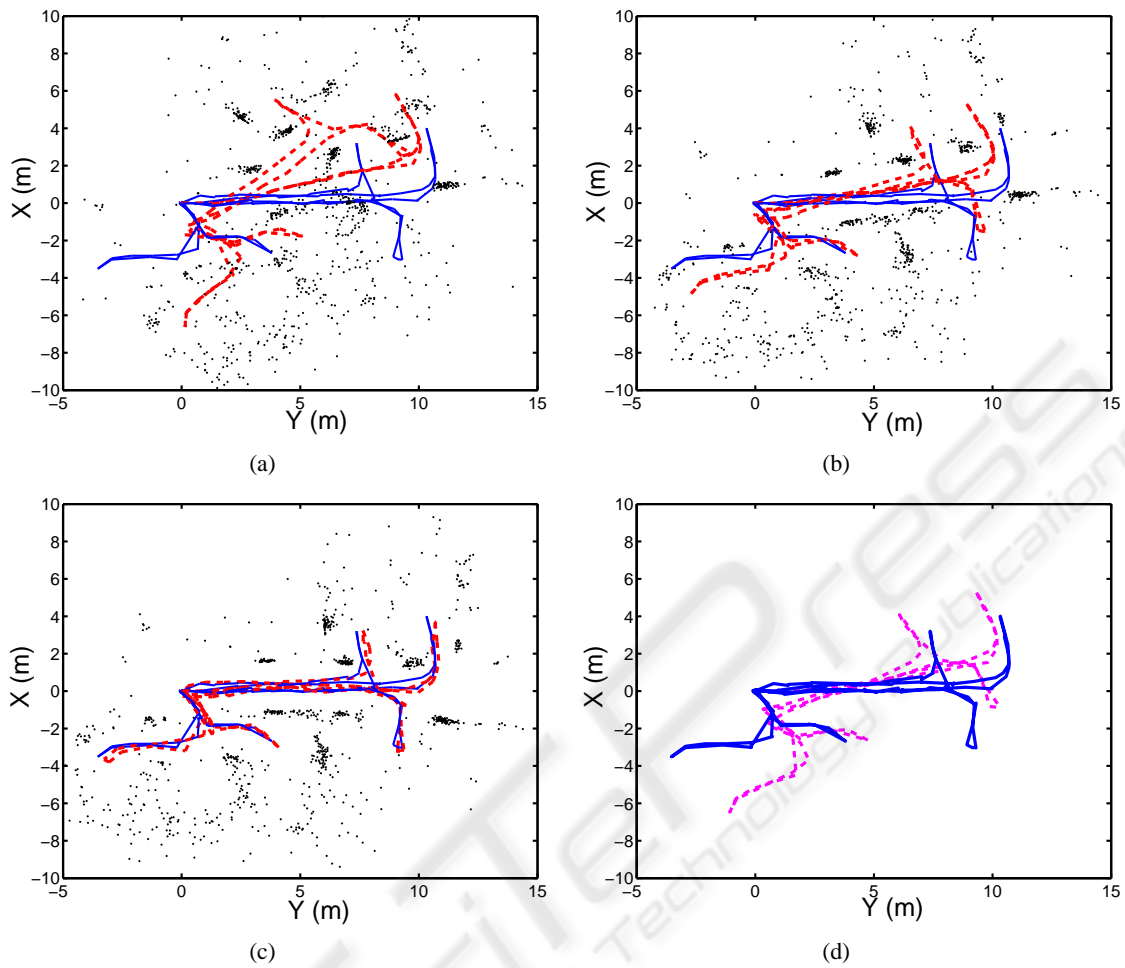


Figure 6: Figure 6(a) shows a map created using 1 particle. Figure 6(b) shows a map created using 10 particles and image 6(c) used 100 particles. We also show superposed the real path (continuous) and the estimated path using our approach (dashed). Figure 6(d) shows the real path (continuous) and the odometry of the robot (dashed).

sults obtained using the Mahalanobis distance and the Euclidean distance. By using a Mahalanobis distance the data association is improved, and, consequently better results are obtained since most of the false correspondences are avoided.

Opposite to maps created by means of occupancy or certainty grids, the visual map generated by the approach presented in this paper does not represent directly the occupied or free areas of the environment. In consequence, some areas totally lack of landmarks, but are not necessary free areas where the robot may navigate through. For example, featureless areas such as blank walls provide no information to the robot. In consequence, the map may be used to effectively localize the robot, but cannot be directly used for navigation. We believe, that this fact is originated from the nature of the sensors and it is not a failure of the proposed approach. Other low-cost sensors such as

SONAR would definitely help the robot in its navigation tasks.

As a future work we think that it is of particular interest to further research in exploration techniques when this representation of the world is used. We would also like to extend the method to the case where several robots explore an unmodified environment and construct a visual map of it.

## ACKNOWLEDGEMENTS

This research has been supported by a Marie Curie Host fellowship with contract number HPMT-CT-2001-00251 and the spanish government (Ministerio de Educación y Ciencia. Projects Ref: DPI2004-07433-C02-01, and PCT-G54016977-2005). The au-



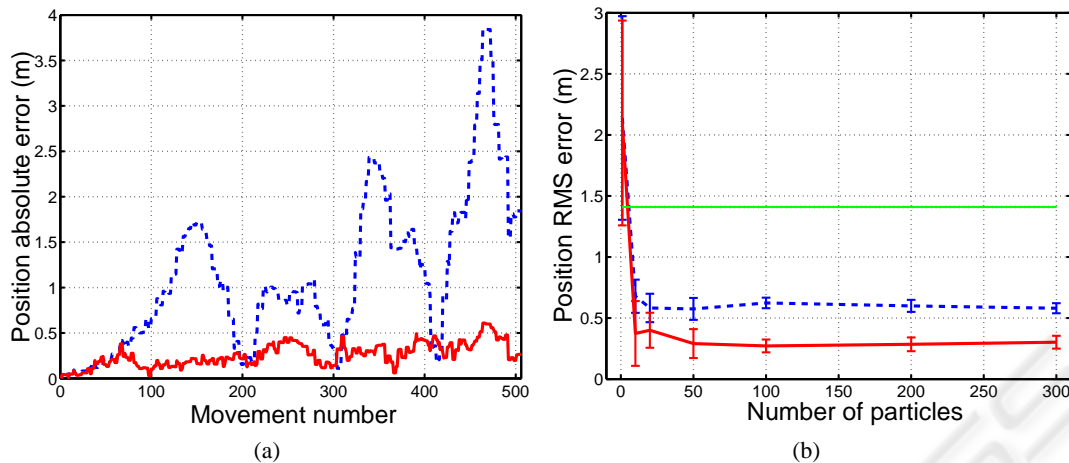


Figure 7: Figure 7(a) shows the absolute position error during the SLAM process. The final error in position is 0.34m. Figure 7(b) shows results in localization error when varying the number  $M$  of particles. The RMS error in odometry is shown as a dotted line. Results using equation (7) are shown as a dashed line and results using equation (8) are shown as a continuous line.

thors would also like to thank, Rudolph Triebel and Manuel Mucientes for their help and stimulating discussions.

## REFERENCES

- Dissanayake, G., Newman, P., Clark, S., Durrant-Whyte, H., and Csorba, M. (2001). A solution to the simultaneous localization and map building (slam) problem. *IEEE Trans. on Robotics and Automation*, 17:229–241.
- Gil, A., Reinoso, O., Vicente, A., Fernández, C., and Payá, L. (2005). Monte carlo localization using sift features. *Lecture Notes in Computer Science (LNCS)*, 1(3523):623–630.
- Little, J., Se, S., and Lowe, D. (2001). Vision-based mobile robot localization and mapping using scale-invariant features. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, pages 2051–2058.
- Little, J., Se, S., and Lowe, D. (2002). Global localization using distinctive visual features. In *Proceedings of the 2002 IEEE/RSJ Intl. Conference on Intelligent Robots and Systems*.
- Lowe, D. (1999). Object recognition from local scale-invariant features. In *International Conference on Computer Vision*, pages 1150–1157.
- Lowe, D. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 2(60):91–110.
- Miró, J. V., Dissanayake, G., and Zhou, W. (2005). Vision-based slam using natural features in indoor environments. In *Proceedings of the 2005 IEEE International Conference on Intelligent Networks, Sensor Networks and Information Processing*, pages 151–156.
- Montemerlo, M., Thrun, S., Koller, D., and Wegbreit, B. (2002). Fastslam: A factored solution to the simultaneous localization and mapping problem. In *AAAI*.
- Murphy, K. (1999). Bayesian map learning in dynamic environments. In *In Neural Information Processing Systems (NIPS)*.
- Sim, R., Elinas, P., Griffin, M., and Little, J. J. (2005). Vision-based slam using the rao-blackwellised particle filter. In *IJCAI Workshop on Reasoning with Uncertainty in Robotics*.
- Stachniss, C., Haehnel, D., and Burgard, W. (2004). Exploration with active loop-closing for FastSLAM. In *IEEE/RSJ Int. Conference on Intelligent Robots and Systems*.
- Stachniss, C., Haehnel, D., and Burgard, W. (2005). Improving grid-based slam with rao-blackwellized particle filters by adaptive proposals and selective resampling. In *IEEE Int. Conference on Robotics and Automation (ICRA)*.
- Thrun, S. (2001). A probabilistic online mapping algorithm for teams of mobile robots. *International Journal of Robotics Research*, 20(5):335–363.
- Wijk, O. and Christensen, H. I. (2000). Localization and navigation of a mobile robot using natural point landmark extracted from sonar data. *Robotics and Autonomous Systems*, 1(31):31–42.