Proceedings of the 2006 IEEE/RSJ
International Conference on Intelligent Robots and Systems
October 9 - 15, 2006, Beijing, China

Improving Data Association in Vision-based SLAM

Arturo Gil* Óscar Reinoso* Óscar Martínez Mozos† Cyrill Stachniss‡ Wolfram Burgard†

* *Miguel Hernández University, Department of Systems Engineering, 03202 Elche (Alicante), Spain*
† *University of Freiburg, Department of Computer Science, 79110 Freiburg, Germany*
‡ *Eidgenössische Technische Hochschule (ETH), Inst. of Robotics and Intelligent Systems, 8092 Zurich, Switzerland*
Email: {arturo.gil|o.reinoso}@umh.es, {omartine|burgard}@informatik.uni-freiburg.de, {scyrill}@ethz.ch

***Abstract*— This paper presents an approach to vision-based simultaneous localization and mapping (SLAM). Our approach uses the scale invariant feature transform (SIFT) as features and applies a rejection technique to concentrate on a reduced set of distinguishable, stable features. We track detected SIFT features over consecutive frames obtained by a stereo camera and select only those features that appear to be stable from different views. Whenever a feature is selected, we compute a representative feature given the previous observations. This approach is applied within a Rao-Blackwellized particle filter to make the data association easier and furthermore to reduce the number of landmarks that need to be maintained in the map. Our system has been implemented and tested on data gathered with a mobile robot in a typical office environment. Experiments presented in this paper demonstrate that our method improves the data association and in this way leads to more accurate maps.**

I. Introduction

Learning maps is a fundamental problem of mobile robots, since maps are required for a series of high-level robotic applications. As a result, several researchers have focused on the problem of simultaneous localization and mapping (SLAM). 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. A large number of papers on SLAM focused on building maps of environments using range sensors like sonars and laser (see, for example, [6]–[8], [14], [17], [19] for two-dimensional maps or [1], [5], [16], [18] for three-dimensional maps).

This paper considers the feature-based SLAM problem using stereo camera images. In general, cameras are less expensive than laser range finders and are also able to provide 3D information from the scene using two cameras in a stereo setting. Our map is represented by a set of landmarks whose position is given by 3D coordinates (X, Y, Z) related to a global reference frame. In our approach, we use distinctive points extracted from these stereo images as natural landmarks. In particular, we use the SIFT feature extractor provided by Lowe [11]. We describe each landmark by two vectors. The first one represents the 3D position of the landmark in the map. The second vector is given by a SIFT descriptor, which contains the visual appearance of the landmark. SIFT descriptors are invariant to image translation, scaling, rotation and partially invariant to illumination changes and affine transformation.

Our system furthermore applies a Rao-Blackwellized particle filter, which has originally been introduced by Murphy, Doucet, and colleagues [3], [15] as an effective means to solve the SLAM problem. The key idea of this approach is to estimate the joint posterior about the trajectory of the robot and the map m of the environment given the observations and odometry measurements.

One of the key problems, especially in feature-based SLAM, is the problem of finding the correct data association. The robot has to determine whether a detected landmark corresponds to a previously seen landmark or to a new one. Often, vision-based SLAM approaches use the Euclidean distance of the SIFT descriptors as a similarity measurement. If the squared Euclidean distance between both descriptors is below a certain threshold, the features are considered to be the same. This technique provides good correspondences in case the feature has been observed from similar viewing angles. Since SIFT descriptors are only partially invariant to affine projection, the SIFT descriptor of the same feature may be significantly different when observing it from different viewpoints.

Figure 1 illustrates a feature recorded from different perspectives. In all the images, an identical point is marked with a circle. The Euclidian distance of the descriptor vectors is small between consecutively recorded images but around one order of magnitude larger between the first and the last image. This difference in the descriptor vectors can lead to serious problems in the context of SLAM. For example, a robot may be unable to make the correct data association when moving through the same corridor but from different directions.

The key idea of this work is to track visual landmarks during several consecutive frames and select only those features that stay comparably stable under different viewing angles. This reduces the number of landmarks in the resulting map representation. In the map, we represent landmarks by a representative descriptor given the individual observations. This descriptor is then used to solve the global data association problem. Instead of using the squared Euclidean distance we propose to use an alternative distance measure which is based on the Mahalanobis distance. This approach reduces the number of false correspondences and consequently produces better maps than an approach based on the Euclidian distance.

II. Related Work

In the past, different approaches have been proposed to solve the SLAM problem in 3D using visual information. Little *et al.* [9], [10] also use a stereo vision system to track

Fig. 1. This figure illustrates the dependency of the SIFT descriptor from the viewing angle. The images depict the same landmark (marked with a circle) viewed from different viewpoints. The squared Euclidean distance between consecutive images is between 0.03 and 0.06. In contrast to that, the squared Euclidean distance between non-consecutive images is between 0.3 and 0.4, which is around one order of magnitude larger.

3D visual landmarks extracted from the environment. In their approach, the landmarks are represented by SIFT features and an Euclidean distance function is used to find the SIFT in the database that is closest to each landmark. Miró *et al.* [13] used an extended Kalman filter (EKF) to estimate an augmented state constituted by the robot pose and N landmark positions using a method proposed by [2]. In this work, SIFT features are used to manage the data association among visual landmarks.

The work presented by Sim *et al.* [16] uses SIFT features as distinctive points in the environment. It also applies a Rao-Blackwellized particle filter to estimate the map of the environment as well as the path of the robot. The movements of the robot are estimated from stereo ego-motion. Compared to these approaches, we actively track the visual landmarks during several consecutive frames and select only those that appear to be more stable. Based on this procedure, we reduce the number of landmarks. We additionally apply the Mahalanobis distance in the data association.

Additionally, several authors have used a Rao-Blackwellized particle filter to solve the SLAM problem. Montemerlo *et al.* [14] applies a Rao-Blackwellized particle filter using 2D point landmarks extracted from laser range data. Their system was the first mapping system based on a Rao-Blackwellized particle filter that was able to deal with large numbers of landmarks. Hähnel *et al.* [7] applies the same filter but using occupancy grid maps. In their approach, an incremental scan-matching technique is applied to pre-correct the odometry. As a result, the error in the robot motion is significantly reduced so that a substantially smaller number of particles is needed to build an accurate map. Recently, Grisetti *et al.* [6] proposed a Rao-Blackwellized particle filter together with a selective resampling to compute an informed proposal distribute and to reduce the risk of the particle depletion problem. In contrast to these approaches, the algorithm described in this paper constructs 3D landmark maps using vision data extracted from camera images. The major contribution of this paper lies in the improvement of the data association. Due to the viewpoint-dependent SIFT feature descriptor, our approach concentrates on the features that appear to be stable under different viewing angles. This approach reduces the risk of making a wrong data association and consequently produces more accurate maps of the environment.

The remainder of the paper is organized as follows. Section III describes SIFT features and their utility in SLAM. Section IV explains the basic idea of the Rao-Blackwellized particle filter for SLAM. Then, Section V explains the tracking of SIFT features across consecutive frames. Section VI presents our solution to the data association problem in the context of SIFT features extracted from stereo images. Finally, in Section VII, we present our experimental results.

III. Visual Landmarks

The scale invariant feature transform (SIFT) features was originally developed for image feature extraction in the context of object recognition applications [11], [12]. SIFT features are represented by a 128-dimensional vector and are computed by building an image pyramid and by considering local image gradients. The image gradients are computed at a local neighborhood that provides invariance to image translation, scaling, rotation, and partial invariance to illumination changes and affine projection.

Given two images (I_t^L, I_t^R) from the left and right camera of a stereo head captured at a time t, we extract landmarks which correspond to points in the 3-dimensional space using SIFT. Each point is accompanied by its SIFT descriptor and then matched across the images. The matching procedure is constrained by the epipolar geometry of the stereo rig. Figure 2 shows an example of a matching between the features of two stereo images.

In our approach, we obtain at each point in time t a set of B observations denoted by $z_t = \{z_{t,1}, z_{t,2}, \ldots, z_{t,B}\}$, where each observation consists of $z_{t,k} = (v_{t,k}, d_{t,k})$, 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. After calculating the stereo correspondence, we calculate the 3D reconstruction of the points using epipolar geometry.

IV. Rao-Blackwellized Visual SLAM

The goal of this section is to give a brief description of how to use a Rao-Blackwellized particle filter to solve the SLAM problem. Additionally, we describe how the map is represented in our current system.

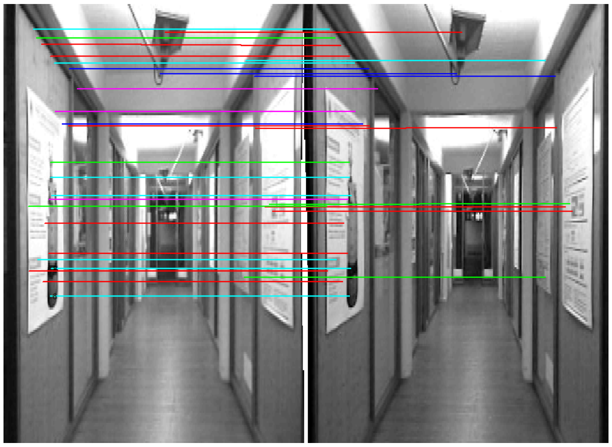


Fig. 2. Stereo correspondences using SIFT features. Epipolar geometry is used to find correspondences across images.

A. Map Representation

According to FastSLAM [14], the map Θ is represented by a collection of N landmarks $\Theta = \{\theta_1, \theta_2, \ldots, \theta_N\}$. Each landmark is described as: $\theta_k = \{\mu_k, \Sigma_k, d_k\}$, where $\mu_k = (X_k, Y_k, Z_k)$ is a vector describing the position of the landmark in the global reference frame and Σ_k a covariance matrix. In addition to that, a SIFT descriptor d_k is added to each landmark θ_k that partially differentiates it from other landmarks.

B. Particle Filter Estimation

While mapping an environment, the robot has to determine whether a particular observation $z_{t,k} = (v_{t,k}, d_{t,k})$ corresponds to a previously observed landmark or to a new one. For the moment, we consider this correspondence as known (we will drop this assumption in the following sections). Given that at time t the map is formed by N landmarks, the correspondence between the observations $z_t = \{z_{t,1}, z_{t,2}, \ldots, z_{t,B}\}$ and the landmarks in the map, is represented by an index vector $c_t = \{c_{t,1}, c_{t,2}, \ldots, c_{t,B}\}$, where $c_{t,i} \in [1 \ldots N]$. In other words, at time t the observation $z_{t,k} = (v_{t,k}, d_{t,k})$ corresponds to the landmark $c_{t,k}$ in the map. When there is no corresponding landmark, we denote it as $c_{t,i} = N + 1$, indicating that it is a new landmark.

Following the usual nomenclature of Montemerlo *et al.* [14], s_t is the robot pose at time t and $s^t = \{s_1, s_2, \ldots, s_t\}$ is the robot path until time t. The set of observations up to time t is denoted as $z^t = \{z_1, z_2, \ldots, z_t\}$ and the set of actions as $u^t = \{u_1, u_2, \ldots, u_t\}$. We formulate the SLAM problem as that of determining the locations of all landmarks in the map Θ and robot poses s^t from a set of measurements z^t and robot actions u^t.

The conditional independence property of the SLAM problem implies that the SLAM posterior can be factored as

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

This factorization was first presented by Murphy [15]. It states that the full SLAM posterior is decomposed into two parts: an estimator for robot paths and N independent estimators for landmark positions, each conditioned on the path estimate. In the Rao-Blackwellized particle filter, one estimates $p(s^t | z^t, u^t, c^t)$ by a set of M particles. Each particle maintains N independent landmark estimators (implemented as EKFs), one for each landmark in the map. Each particle is thus defined as

$$S_t^{[m]} = \{s^{t,[m]}, \mu_{t,1}^{[m]}, \Sigma_{t,1}^{[m]}, \ldots, \mu_{t,N}^{[m]}, \Sigma_{t,N}^{[m]}\}, \quad (2)$$

where $\mu_{t,i}^{[m]}$ is the best estimation at time t for the position of landmark θ_i based on the path of the particle m and $\Sigma_{t,i}^{[m]}$ is its associated covariance matrix. 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. Each particle is sampled from a proposal distribution $s_t^{[m]} \sim p(s_t | s_{t-1}, u_t)$. Furthermore, a weight is assigned to each sample according to

$$\omega_{t,i}^{[m]} = \frac{1}{\sqrt{|2\pi Z_{c_{t,i}}|}} \cdot \exp\left[-\frac{1}{2}(v_{t,i} - \hat{v}_{t,c_{t,i}})^T Z_{c_{t,i}}^{-1} (v_{t,i} - \hat{v}_{t,c_{t,i}})\right]. \quad (3)$$

In this equation, $v_{t,i}$ is the actual 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 the equations presented above assume that each measurement $v_{t,i}$ has been assigned to one landmark $c_{t,i}$ in the map. In Section VI, 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

$$\omega_t^{[m]} = \prod_{i=1}^{B} w_{t,i}^{[m]}. \quad (4)$$

In order to reduce the risk of particle depletion, we use the approach proposed by Doucet [4] to trigger the resampling. We compute the effective sample size and carry out a resampling operation only of the resulting number is smaller than a threshold (here chosen as $M/2$). In the context of mapping with Rao-Blackwellized particle filters, this approach has first been applied by Grisetti *et al.* [6].

V. Tracking of SIFT Features

To obtain multiple observations of the same feature from several view points, we track each SIFT feature along p consecutive frames. For this purpose the robot takes two stereo images and extracts a number of SIFT features from them. Next, SIFT features that comply with the epipolar constraint are matched across the stereo images. Two SIFT features are matched from the left to the right image if the Euclidean distance between the descriptors is below a predefined threshold. This procedure has shown to be very robust since both images are taken from very close points of view and consequently their SIFT descriptors are similar [12]. Figure 2 shows the matching performed across two stereo images. Once the matching is performed, a measurement $v_{t,k} = (X_r, Y_r, Z_r)$ relative to the robot reference frame is obtained for each SIFT point. After a short movement of the robot given by $(\Delta x, \Delta y, \Delta\theta)$, we estimate the new coordinates at time $t+1$ $v_{t+1,k} = (X'_r, Y'_r, Z'_r)$ given $v_{t,k} = (X_r, Y_r, Z_r)$ as:

$$\begin{pmatrix} X'_r \\ Y'_r \\ Z'_r \end{pmatrix} = \begin{pmatrix} (X - \Delta x)\cos(\Delta\theta) - (Z - \Delta y)\sin(\Delta\theta) \\ Y \\ (X - \Delta x)\sin(\Delta\theta) + (Z - \Delta y)\cos(\Delta\theta) \end{pmatrix} \quad (5)$$

In the frame obtained at the next time step the SIFT point is projected at image coordinates (r', c'):

$$\begin{pmatrix} r' \\ c' \end{pmatrix} = \begin{pmatrix} v_0 - f\frac{Y'_r}{Z'_r} \\ u_0 + f\frac{X'_r}{Z'_r} \end{pmatrix}, \quad (6)$$

where f and the central point of the left camera are calibrated values. We then look for the SIFT points in a local neighborhood of the predicted projection of the point (r', c'). Again, this matching is performed using the Euclidean distance, since the variation in the SIFT descriptor across two consecutive frames is low, assuming the robot has performed a short movement as explained before (see also the motivating example given in Figure 1).

VI. Data Association

While the robot moves through 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. In most existing approaches, data association is based on the squared Euclidean distance between SIFT descriptors

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

where d_i and d_j are the SIFT descriptors.

Then, the landmark of the map that minimizes the distance E is regarded as the correct data association. Whenever the distance E is below a certain threshold, the two landmarks are considered to be the same. Otherwise, a new landmark is created. As explained in Section V, when the same point is viewed from slightly different viewpoints and distances, the values in its SIFT descriptor remain quite similar. However, when the same point is viewed from significantly different viewpoints (e.g., 30 degrees apart) the difference in the descriptor is remarkable and the check using the Euclidian distance is likely to produce a wrong data association.

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, where each class C_i models a landmark. We take different views of the same visual landmark as different elements of class C_i. Whenever a landmark is found, it is tracked along p frames as shown in Section V, and its descriptors $d_1, d_2, \ldots, d_p$ are stored.

For the landmark represented by C_i we compute a mean value $\bar{d}_i$ and estimate a covariance matrix S_i, assuming the elements in the SIFT vector are independent of each other. Whenever a new landmark d_j is found, we compute the Mahalanobis distance to each stored landmark, represented by $\bar{d}_i$ and S_i as

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

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, we consider it as a new landmark. As we will show in the experiments, this technique allows us to make better data associations and as a result produce better maps of the environment.

VII. Experimental Results

To carry out the experiments, we used a B21r robot equipped with a stereo head and an LMS laser range finder. We manually steered the robot through the rooms of the building 79 at the University of Freiburg. For each pair of stereo images we calculated the correspondences and the feature tracks over 5 consecutive frames to improve the stability of the SIFT points. As mentioned in Section VI, 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 views of the same landmark and S_i is the corresponding diagonal covariance matrix.

In our first experiment, we apply the Rao-Blackwellized particle filter to create the map using our data association method. A total of 507 stereo images at a resolution of 320x240 were collected. The distance traveled by the robot is approximately $80m$. Figure 5 shows the results with 10, and 100 particles. A total number of 1500 landmarks were estimated. It can be seen that, with only 10 particles, the map is a good approximation. In the figures, some areas of the map do not possess any landmark, which correspond to feature-less areas (i.e., texture-less walls), where no SIFT features have been found. Compared to preceding approaches, our method uses less particles to achieve good results. For example, in [16], 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 substantially more compact representation than obtained with the algorithm presented in [16], in which the map contains typically around 10.000 landmarks.

We additionally compared the estimated pose of our method with the estimated pose using a Rao-Blackwellized filter with observations consisting in laser range data as described in work of Grisette *et al.* [6]. Figure 3 shows the error in position using our approach in comparison with the position using laser data. As can be seen, the absolute error is maintained always under $0.6m$.

In a second experiment, we compared both distance functions, namely Euclidean and Mahalanobis, for solving the data association problem. For both approaches, we tracked SIFT

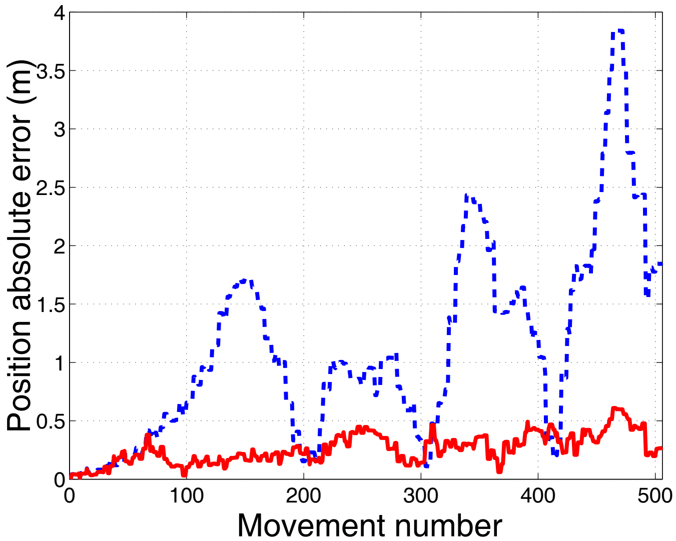


Fig. 3. Figure shows the position error using our approach in comparison with the position using laser data (continuous line). For comparison, we also plot the error in odometry (dashed).

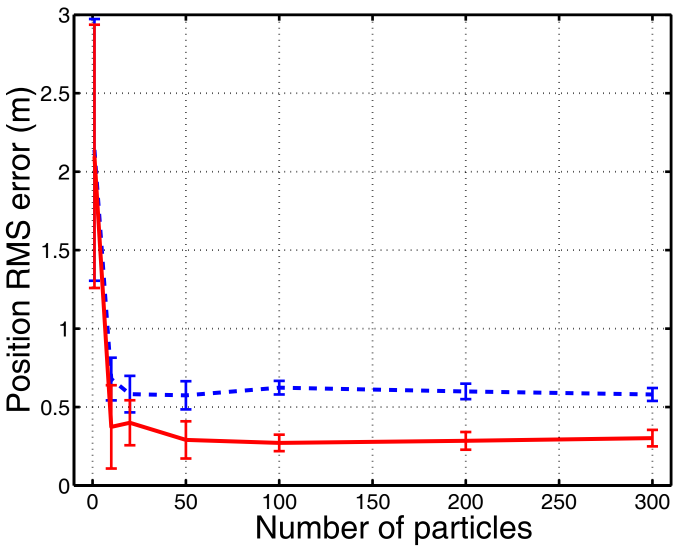
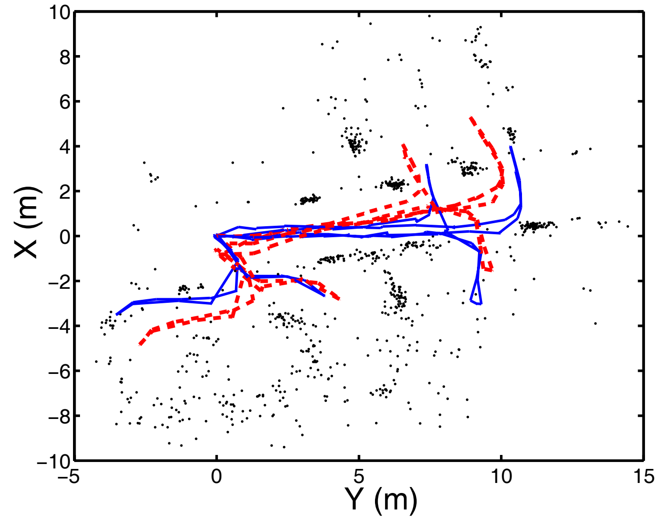


Fig. 4. The figure shows the RMS error in localization depending on the number M of particles. The results using Equation (7) are shown as a dashed line and results using Equation (8) are shown as a continuous line.

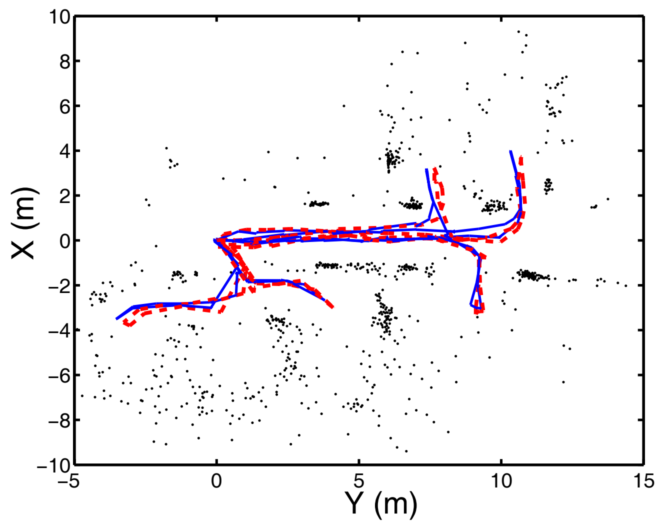
features and calculated the Root Mean Square (RMS) error of the position of the robot with respect to the position given by the localization using laser data. To do this, we made a number of simulations varying the number of particles used in each simulation. As can be seen in Figure 4, we obtained better localization results for the same number of particles when the Mahalanobis distance is used.

VIII. Conclusion and Future Work

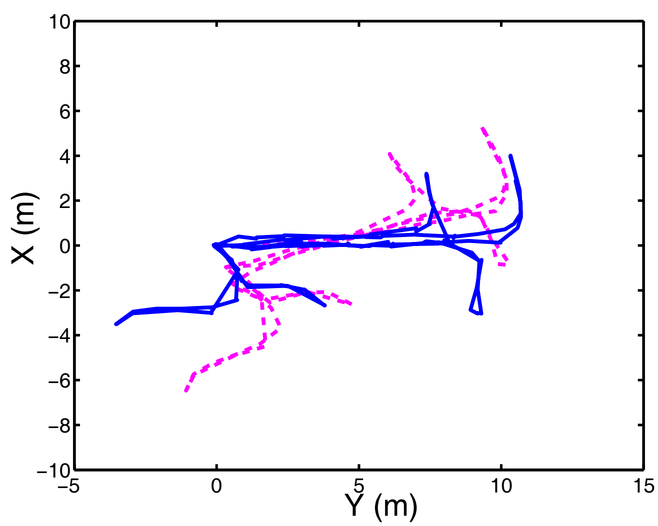
In this paper, we presented a solution to the SLAM problem based on a Rao-Blackwellized particle filter that uses visual information extracted from cameras. In particular, we track



(a)



(b)



(c)

Fig. 5. Whereas Figure (a) shows a map created using 10 particles, Figure (b) has been created with 100 particles. We also superposed the real path (continuous) and the estimated path using our approach (dashed). Figure (c) shows the real path (continuous) and the odometry of the robot (dashed).

SIFT features extracted from stereo images and use those that are stable as landmarks. To solve the data association problem when the robot closes a loop, our approach calculates SIFT prototypes and applies the Mahalanobis distance for calculating the similarity between landmarks. As a result, the data association is improved and we obtained better maps, since most wrong correspondences can be avoided more reliably. In practical experiments, we have shown that our approach is able to build 3D maps.

Despite these results, we are aware that there still are important issues that warrant future research. For example, the maps created by our algorithm do not correctly represent the occupied or free areas of the environment. For example, featureless areas such as blank walls provide no information to the robot. In such environments, the map cannot be used to localize the robot. We believe, that this fact is originated from the nature of the sensors and it is not a failure of the proposed approach. In such environments, alternative sensors like sonars might be needed for navigation.

ACKNOWLEDGMENTS

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. Ref: DPI2004- 07433-C02-01. Título: Herramientas de teleoperación colobarativa: aplicación al control cooperativo de robots). It has further been supported by the EC under contract number FP6-004250-CoSy, FP6-IST-027140-BACS, and FP6-2005-IST-5-muFly, and by the German Research Foundation (DFG) under contract number SFB/TR-8 (A3). The author would also like to thank Axel Rottmann, Rudolph Triebel, and Manuel Mucientes for their help and stimulating discussions.

REFERENCES

[1] P. Biber, H. Andreasson, T. Duckett, and A. Schilling. 3d modelling of indoor environments by a mobile robot with a laser scanner and panoramic camera. In *Proc. of the IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*, pages 1530–1536, Sendai, Japan, 2004.

[2] G. Dissanayake, P. Newman, S. Clark, H. Durrant-Whyte, and M. Csorba. A solution to the simultaneous localization and map building (slam) problem. *IEEE Trans. on Robotics and Automation*, 17:229–241, 2001.

[3] A. Doucet, J.F.G. de Freitas, K. Murphy, and S. Russel. Rao-Blackwellized partcile filtering for dynamic bayesian networks. In *Proc. of the Conf. on Uncertainty in Artificial Intelligence (UAI)*, pages 176–183, Stanford, CA, USA, 2000.

[4] A. Doucet, N. de Freitas, and N. Gordan, editors. *Sequential Monte-Carlo Methods in Practice*. Springer Verlag, 2001.

[5] R. Eustice, H. Singh, and J.J. Leonard. Exactly sparse delayed-state filters. In *Proc. of the IEEE Int. Conf. on Robotics & Automation (ICRA)*, pages 2428–2435, Barcelona, Spain, 2005.

[6] G. Grisetti, C. Stachniss, and W. Burgard. Improving grid-based slam with rao-blackwellized particle filters by adaptive proposals and selective resampling. In *Proc. of the IEEE Int. Conf. on Robotics & Automation (ICRA)*, pages 2443–2448, Barcelona, Spain, 2005.

[7] D. Hähnel, W. Burgard, D. Fox, and S. Thrun. An efficient FastSLAM algorithm for generating maps of large-scale cyclic environments from raw laser range measurements. In *Proc. of the IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*, pages 206–211, Las Vegas, NV, USA, 2003.

[8] J.J. Leonard and H.F. Durrant-Whyte. Mobile robot localization by tracking geometric beacons. *IEEE Transactions on Robotics and Automation*, 7(4):376–382, 1991.

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

[10] J. Little, S. Se, and D.G. Lowe. Global localization using distinctive visual features. In *Proc. of the IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*, 2002.

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

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

[13] J. Valls Miró, G. Dissanayake, and W. Zhou. 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, 2005.

[14] M. Montemerlo, S. Thrun, D. Koller, and B. Wegbreit. FastSLAM: A factored solution to simultaneous localization and mapping. In *Proc. of the National Conference on Artificial Intelligence (AAAI)*, pages 593–598, Edmonton, Canada, 2002.

[15] K. Murphy. Bayesian map learning in dynamic environments. In *Proc. of the Conf. on Neural Information Processing Systems (NIPS)*, pages 1015–1021, Denver, CO, USA, 1999.

[16] 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.

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

[18] R. Triebel and W. Burgard. Improving simultaneous mapping and localization in 3d using global constraints. In *Proc. of the National Conference on Artificial Intelligence (AAAI)*, 2005.

[19] O. Wijk and H. I. Christensen. Localization and navigation of a mobile robot using natural point landmarkd extracted from sonar data. *Robotics and Autonomous Systems*, 1(31):31–42, 2000.