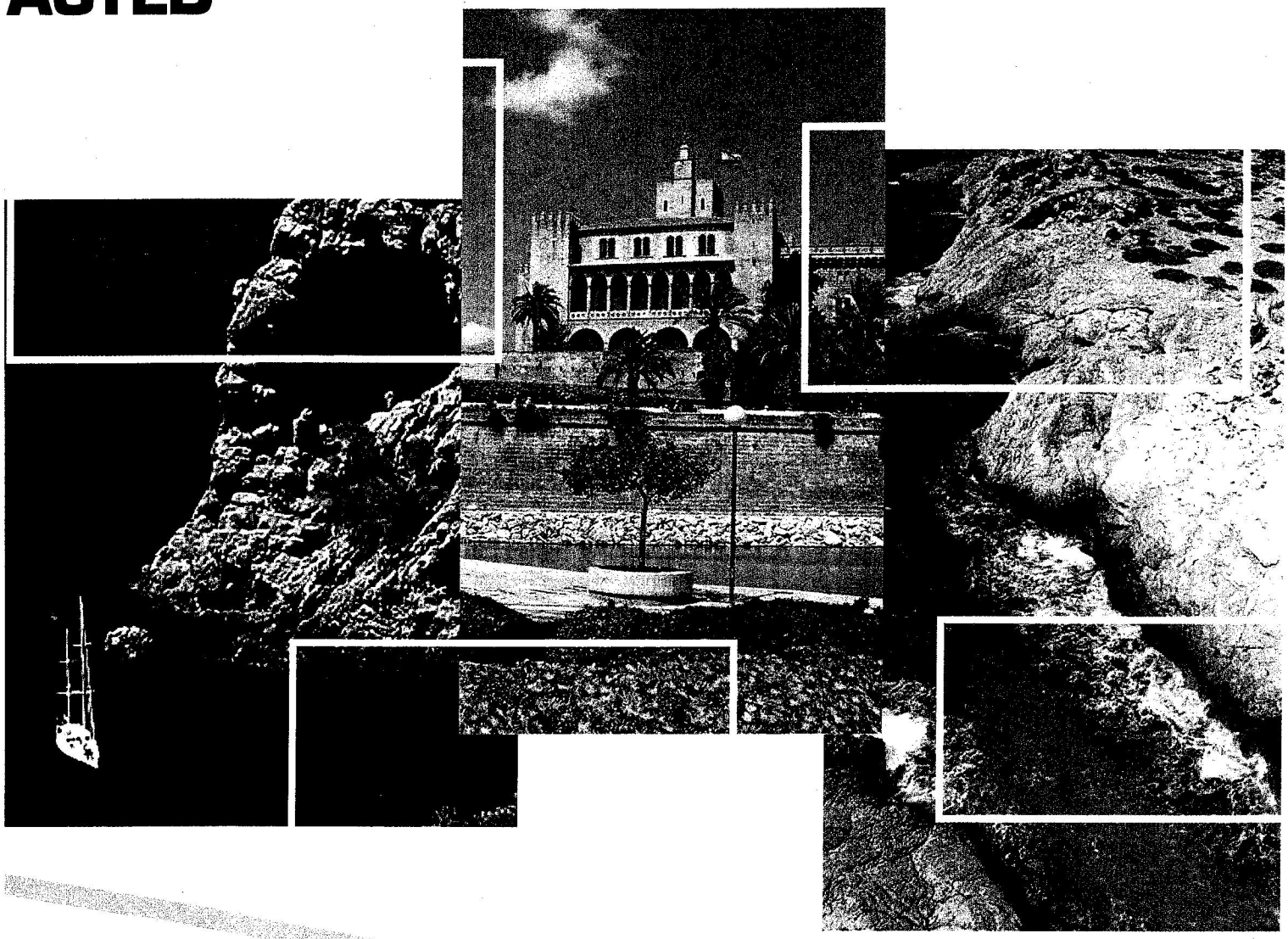
ASTED

A Publication of the International Association of Science and Technology for Development



oceedings of the Sixth IASTED International Conference on

VISUALIZATION, IMAGING, AND IMAGE PROCESSING

August 28 – 30, 2006
Palma de Mallorca, Spain

ditor: J.J. Villanueva

BN: 0-88986-598-1
SN: 1482-7921

ACTA Press Anaheim | Calgary | Zurich

Proceedings of the Sixth IASTED International Conference on **Visualization, Imaging, and Image Processing**, held August 28 – 30, 2006, Palma de Mallorca, Spain.

SPONSORS

The International Association of Science and Technology for Development (IASTED)

- Technical Committee on Image Processing
- Technical Committee on Computers
- Technical Committee on Visualization

World Modelling and Simulation Forum (WMSF)

EDITOR

J.J. Villanueva – Computer Vision Centre and Autonomous University of Barcelona, Spain

PROGRAM CO-CHAIRS

A. Ebert – University of Kaiserslautern, Germany
A. Sappa – Computer Vision Centre, Spain

KEYNOTE SPEAKER

N. Petkov – University of Groningen, The Netherlands

INTERNATIONAL PROGRAM COMMITTEE

M.J. Abásolo Guerrero – University of the Balearic Islands, Spain
M. Alam – University of South Alabama, USA
J. Angulo – Paris School of Mines, France
M. Bicego – University of Sassari, Italy
N. Bonnet – University of Reims, France
Y. Cai – Nanyang Technological University, Singapore
F.A. Candelas Herías – University of Alicante, Spain
J.R. Casas – Technical University of Catalonia, Spain
A. Castro Martinez – University of Coruña, Spain
C. Charrier – University of Caen Basse-Normandie, France
L.-H. Chen – National Chiao Tung University, Taiwan
Y.-T. Ching – National Chiao Tung University, Taiwan
R.S. Choras – University of Technology and Agriculture, Poland
M.F. Costabile – University of Bari, Italy
P. Dannenmann – German Research Center for Artificial Intelligence, Germany
A. Ebert – University of Kaiserslautern, Germany
E. Edirisinghe – Loughborough University, UK
Z. Fan – Volume Interactions Pte. Ltd., Singapore
A. Fernández – University of Vigo, Spain
E. Ferreira – University of Minho, Portugal
R.B. Fisher – University of Edinburgh, UK
I. Fujishiro – Tohoku University, Japan
J. Gao – University of Texas at Arlington, USA
G. Georgiev – University of Wisconsin Oshkosh, USA
L. Giubbolini – WaveBand Corporation, USA
E.J. Gómez – Polytechnic University of Madrid, Spain
J. Gonzàlez – Autonomous University of Barcelona, Spain
P. Gouton – University of Burgundy, France
N. Gueorguieva – City University of NY / College of Staten Island, USA
H. Hagen – University of Kaiserslautern, Germany
D. Harvey – Liverpool John Moores University, UK
I. Ideses – Tel Aviv University, Israel
J.M. Iñesta – University of Alicante, Spain
S. Jaeger – University of Maryland, USA
K.M. Jambi – King Abdul Aziz University, Saudi Arabia
X. Jiang – University of Münster, Germany
Z. Jin – Nanjing University of Science and Technology, PRC
L. Jofre – Politechnic University of Catalunya, Spain
I.A. Kakadiaris – University of Houston, USA
K. Koyamada – Kyoto University, Japan
H. Laggoune – University of Quebec at Montreal, Canada
J. Lladós – Autonomous University of Barcelona, Spain
S. Loncaric – University of Zagreb, Croatia
S. Loskovska – University of St. Cyril and Methodius, FYROM
H. Ma – Beijing University of Posts & Telecommunication, PRC
S. Madenda – Gunadarma University, Indonesia
Y. Manabe – Nara Institute of Science and Technology, Japan
F. Marques – Polytechnic University of Catalunya, Spain
P.L. Mazzeo – National Research Council Italy, Italy
D.B. Megherbi – University of Massachusetts Lowell, USA
J. Meyer – University of California, Irvine, USA
M. Mirmehdi – University of Bristol, UK
V. Murino – University of Verona, Italy
A. Naftel – University of Manchester, UK
T. Noma – Kyushu Institute of Technology, Japan
J.-M. Ogier – University of the Rochelle, France
O. Okun – University of Oulu, Finland
M.M. Oliveira – Federal University of Rio Grande do Sul, Brazil
V. Ostromoukhov – University of Montreal, Canada
E. Paquette – School of Superior Technology, Canada
M. Pardàs – Polytechnic University of Catalunya, Spain
V. Pascucci – Lawrence Livermore National Laboratory, USA
A.A. Pasko – Hosei University, Japan
F.J. Perales – University of the Balearic Islands, Spain
M.J. Perez Malumbres – Miguel Hernández University, Spain
A. Pinho – University of Aveiro, Portugal
X. Pueyo – University of Girona, Spain
H. Qin – State University of New York at Stony Brook, USA
T. Rabie – University of Toronto, Canada
A.L. Reznik – Siberian Branch of the Russian Academy of Sciences, Russia
R. Rizo Aldeguer – University of Alicante, Spain
S. Robila – Montclair State University, USA
W. Roque – Federal University of Rio Grande do Sul, Brazil
R. Sablatnig – Vienna University of Technology, Austria
K. Saeed – Technical University of Bialystok, Poland
C. Sansone – University of Napoli Federico II, Italy
A. Sappa – Computer Vision Centre, Spain
M. Sarfraz – King Fahd University of Petroleum and Minerals, Saudi Arabia
A. Savva – Cyprus Intercollege, Cyprus
G. Scheuermann – University of Leipzig, Germany
J. Serra-Sagrista – Autonomous University of Barcelona, Spain
J. Serrat – Autonomous University of Barcelona, Spain
J.H. Sossa Azuela – National Polytechnic Institute, Mexico
J. Stewart – Queen's University, Canada
S.R. Subramanya – LGE Mobile Research, USA
A.Z. Talib – University of Science Malaysia, Malaysia
L.J. Tardón García – University of Málaga, Spain
J.C. Teixeira – University of Coimbra, Portugal
G.Y. Tian – University of Huddersfield, UK
A. Tonazzini – National Research Council Italy, Italy
D.-Y. Tsai – Niigata University, Japan
T. Tsuta – Hiroshima International University, Japan
R. van Liere – Center for Mathematics and Information, The Netherlands
L. Velho – Institute of Pure Applied Mathematics, Brazil
R.A. Vivó Hernando – Technical University of Valencia, Spain
D. Weiskopf – Simon Fraser University, Canada
Y. Xiao – University of Akron, USA
J. Zara – Czech Technical University in Prague, Czech Republic
J.F. Zelasco – University of Buenos Aires, Argentina

ADDITIONAL REVIEWERS

P.D. Cristea – Romania
C. Grecos – UK
M. Jeon – Korea
A. Liverani – Italy
J.P. Oakley – UK
S. Qin – UK
J. Shen – USA
J. Tavares – Portugal
B. Zagar – Austria

For each IASTED conference, the following review process is used to ensure the highest level of academic content. Each full manuscript submission is peer reviewed by a minimum of two separate reviewers on the International Program Committee/Additional Reviewers list. The review results are then compiled. If there are conflicting reviews, the paper is sent to a third reviewer.

ACTA Press
P.O. Box 5124
Anaheim, CA 92814–5124
USA

ACTA Press
B6, Suite #101, 2509 Dieppe Ave SW
Calgary, Alberta T3E 7J9
Canada

ACTA Press
P.O. Box 354
CH–8053 Zurich
Switzerland

Publication Code: 541

TABLE OF CONTENTS
VIIP 2006

HEAD AND FACE MODELING AND ANALYSIS

MEDICAL IMAGING

BIOMETRICS

IMAGE CODING AND COMPRESSION

3D MODELLING

PATTERN ANALYSIS AND RECOGNITION

IMAGE PROCESSING

APPLICATION

IMAGE ANALYSIS

DOCUMENT AND VIDEO ANALYSIS

ADDITIONAL PAPER FROM SIGNAL AND IMAGE PROCESSING CONFERENCE

Proceedings of the Sixth IASTED International Conference
Visualization, Imaging, and Image Processing
August 28–30, 2006, Palma de Mallorca, Spain
ISBN Hardcopy: 0-88986-598-1 / CD: 0-88986-600-7

SIMULTANEOUS LOCALIZATION AND MAPPING IN INDOOR ENVIRONMENTS USING SIFT FEATURES

A. Gil, L. Payá, O. Reinoso, C. Fernández, R. Puerto
System Engineering Department
Universidad Miguel Hernández
Avda. de la Universidad s/n
03202 Elche (Alicante)
email: arturo.gil@umh.es

ABSTRACT
We consider the problem of building a map of an unmodified environment using only visual information extracted from cameras. In order to build a map, we must estimate both robot's location and a map of its surrounding environment. In general, this problem is known as Simultaneous Localization and Mapping (SLAM). It is an inherently hard problem because noise in the estimate of the robot's pose leads to noise in the estimate of the map and vice versa. Past work on this area has centered on building maps using distance sensors (i.e. laser and SONAR sensors). However, in our case, we propose a method to build a map based only on visual information . While the robot moves along the environment, it extracts a number of interesting points from images (i.e. corners) and calculates a relative measurement vector $V_r = (X_r, Y_r, Z_r)$ to each one of them using stereo vision. We are using SIFT features as the relevant points extracted from images. SIFT features, are said to be invariant to image translation, scaling and rotation and partially invariant to illumination changes and affine projection. In consequence those points are suitable for localizing the robot in a particular environment. Our map consist of a number of L three dimensional landmarks referred to a global frame S_g. In addition, each of the 3D landmarks is assigned a SIFT descriptor that enables us to partially differentiate that particular landmark from the rest. Our approach to SLAM is based on a Rao-Blackwellised Particle Filter. This permits us to separate the estimation process in two parts: On the one hand, we estimate the path of the robot using a particle filter and estimate the map conditioned to each path of the robot. We present experimental results that validate our approach to vision-based SLAM in large unmodified environments.

KEY WORDS
SLAM, Stereo Vision, Visual landmarks, Data Association

1 Introduction

Mobile robots rely normally on a map of the environment in order to complete their navigation tasks. However, frequently an accurate map of the environment is not available, thus the robot faces the problem of building it up from readings obtained from its sensors. 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. This problem is generally known as Simultaneous Localization and Mapping (SLAM). The aim of SLAM is to build an accurate map of an unknown environment and simultaneously localise the robot with respect to this map. Our approach solves this problem using a Rao-Blackwellized Particle Filter (RBPF) and integrating measurements extracted from a stereo head.

Most work on SLAM so far has focussed on building 2D maps of environments using range sensors such as SONAR and laser [1], [2]. SONAR sensors are typically inaccurate, thus providing little information for the localization process. Laser sensors, provide generally highly accurate measurements, but need typically expensive and heavy devices. In addition, the information acquired lies normally on a plane, in consequence the robot is not able to build a complete representation of the world. Cameras are typically low-weight devices that provide a vast quantity of information. Stereo systems are typically less expensive than laser sensors and are able to provide directly 3D information from the scene. 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. In [3] and [4] stereo vision is used to track 3D visual landmarks extracted from the environment. In this work, SIFT features are used as visual landmarks. During exploration, the robot extracts SIFT features from stereo images and calculates relative measurements to them. Landmarks are then integrated in the map with and an EKF is 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 [5] a Kalman filter is used to estimate an augmented state constituted by the robot pose and N landmark positions [6]. SIFT features are used too to manage data association among visual landmarks. 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 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 which look quite similar. This fact may occur in most environments. Second, our approach actively tracks landmarks prior to integrating them 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 [4] and [7], for comparable map sizes. In addition, we have applied effective resampling techniques. As exposed in [8], this fact reduces the number of particles needed to build 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 SIFT features

Our representation of the world is formed by set of three dimensional points referred to a common reference system. In this sense, each landmark is constituted by a significant point in space, which can be perceived by the robot from different viewpoints. In this paper we use visual landmarks as features to make the map. SIFT features (Scale Invariant Feature Transform) were developed for image feature generation, and used initially in object recognition applications (see [9] and [10] for some examples). These significant points (key point locations in SIFT nomenclature) 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. The 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 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 [3], [4], [7]. Figure 1 shows a set of visual features extracted from an image.

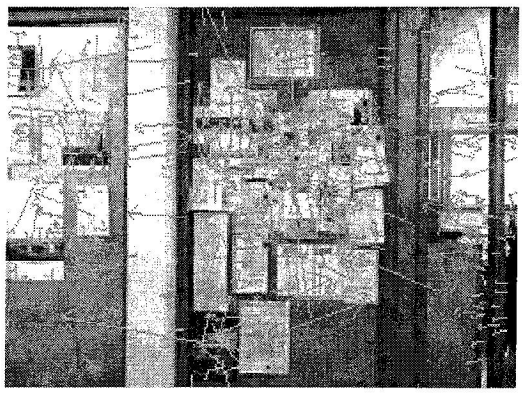


Figure 1. Image shows SIFT locations extracted from a typical image. The size of the arrow is proportional to SIFT's scale.

3 Rao-Blackwellized SLAM

We estimate the map and the path of the robot using a Rao-Blackwellized particle filter. Using the typical nomenclature in this context, we denote as s_t the robot pose at time t. On the other hand, the robot path until time t will be denoted $s^t = \{s_1, s_2, \ldots, s_t\}$, the set of observations made by the robot until time t will be denoted $z^t = \{z_1, z_2, \ldots, z_t\}$ and the set of actions $u^t = \{u_1, u_2, \ldots, u_t\}$. Therefore, the SLAM problem can be formulated as that of determining the location 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 map is composed as a set of differente landmarks $\Theta = \{\theta_1, \theta_2, \ldots \theta_i, \ldots, \theta_N\}$. In consequence, the SLAM can be formulated as the problem of estimating the following:

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

where c^t are landmark correspondences of the landmarks extracted from the association.

The map Θ 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 matriz Σ_k. In addition, each landmark θ_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 [11].

While exploring a particular environment, the robot 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 $c_t = \{c_{t,1}, c_{t,2}, \ldots, c_{t,B}\}$, where $c_{t,i} \in [1 \ldots N]$. 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

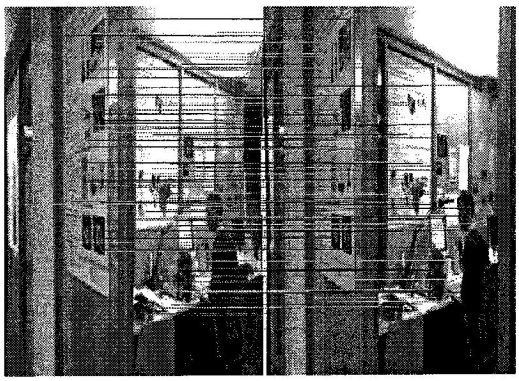


Figure 2. Stereo correspondences using SIFT features. The epipolar constrain is used to find correspondences across images.

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 following constrains are applied during stereo correspondence:

- Epipolarity restriction: The feature location in the right image must be placed in the same row as the in the left image. In practice, this condition was relaxed, permitting a maximum ± 2 pixel displacement.

- SIFT restriction: The euclidean distance between two SIFT descriptors must not surpass a certain threshold (determined experimentally).

Each time a SIFT feature is matched correctly in both images, a 3D reconstruction of the point is obtained relative to the left camera reference frame. As a result, at a time t we obtain a set of B observations denoted by $z_t = \{z_{t,1}, z_{t,2}, \ldots, z_{t,B}\} = \{v_{t,1}, d_{t,1}, v_{t,2}, d_{t,2}, \ldots, v_{t,B}, d_{t,B}\}$. 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 2 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 [12]:

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

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 [13]. 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 EKFs), one for each landmark in the map. Each particle is thus defined as:

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

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 $\Sigma_{t,i}^{[m]}$ 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. 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 [12] each particle is then assigned a weight according to:

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

where

$$a = (v_{t,i} - \hat{v}_{t,c_{t,i}})^T [Z_{c_{t,i}}]^{-1} (v_{t,i} - \hat{v}_{t,c_{t,i}}) \tag{5}$$

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:

$$\omega_t^{[m]} = \prod_{i=1}^{B} w_{t,i}^{[m]} \tag{6}$$

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

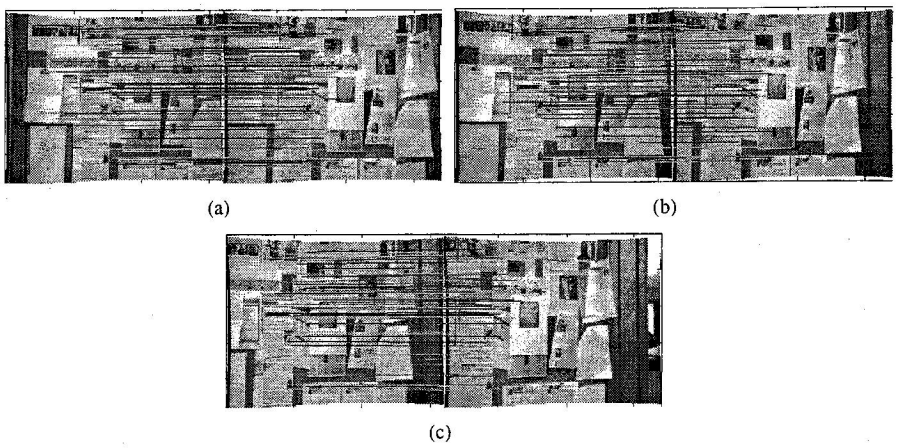


Figure 3. Figures (a)-(c) show different SIFT points tracked along different images with variations in scale and orientation.

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 [14]. Thus we calculate the number of efficient particles N_{eff} as:

$$N_{eff} = \frac{1}{\sum_{i=1}^{M} \omega_t^{[i]}} \tag{7}$$

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

When navigating through the world, the robot finds different landmarks, then, it must decide whether the observation $z_{t,k} = (v_{t,k}, d_{t,k})$ corresponds to a previously mapped landmark or to a new one. Data association is based in the SIFT descriptor and in the spatial position of the landmark. 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 [9]. In the approaches of [3], [4] and [7], 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:

$$E = (d_i - d_j)(d_i - d_j)^T \tag{8}$$

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

A significant number of false data associations may produce a divergence of the filter, thus failing to produce a correct map. 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. While mapping the environment, the robot tracks each landmark for a significant number of frames, thus, we have different patterns for the same class. When a landmark has been tracked for p frames, its descriptors $d_1, d_2, \ldots, d_p$ 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:

$$L = (\bar{d}_i - \bar{d}_j)S_i^{-1}(\bar{d}_i - \bar{d}_j)^T \tag{9}$$

Where S_i is a diagonal covariance matrix associated with the class C_i with mean value $\bar{d}_i$. When a landmark is detected, 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 it is considered 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 3). SIFT landmarks are easily tracked across consecutive frames, since the variance in the descriptor is low. In table 1 we compare results using equations 9 and 8. A total of 3000 SIFT descriptors were used.

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

By using a Mahalanobis we obtain the data association is more robust, thus improving the map building process.

5 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 $z_t = \{z_{t,1}, z_{t,2}, \ldots, z_{t,B}\} = \{v_{t,1}, d_{t,1}, v_{t,2}, d_{t,2}, \ldots, v_{t,B}, d_{t,B}\}$ 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 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.

Figure 4 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. These 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 similar results. For example, in [7], 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 [7], where the map contains typically around 100.000 landmarks.

6 Conclusion

In this paper a solution to SLAM based on a Rao-Blackwellized particle filter has been presented. The proposed solution is based on information extracted from a stereo head. We calculate both a representation of each landmark (SIFT descriptor) and spatial information from the stereo reconstruction of the points. Both data are then integrated in the map, estimating both the map and the path of the robot.

Differing from prior approaches, we propose an alternative method to deal with the data association problem in the context of visual landmarks. In the case where a significant number of similar looking points exist, a solution based on the Mahalanobis distance helps us improve the results of our map and avoids false correspondences.

Opposite to maps created by means of occupancy or certainty grids, the visual map generated by the process presented in this paper does not represent directly the occupied or free areas of the environment. Also, in some areas there may not exist landmarks in the map (i.e. featureless areas such as blank walls of glass). The map can 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 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.

Acknowledgments

This research is supported by the spanish government (Ministerio de Educación y Ciencia. Project Reference: DPI2004- 07433-C02-01) and Fundacin Quorum: Parque Cientfico y empresarial de la Universidad Miguel Herndez, through project ref. PCT-G54016977-2005 (Ministerio de Educacin y Ciencia).

References

[1] 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.

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

[3] J. Little, S. Se, and D. Lowe. 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, 2001.

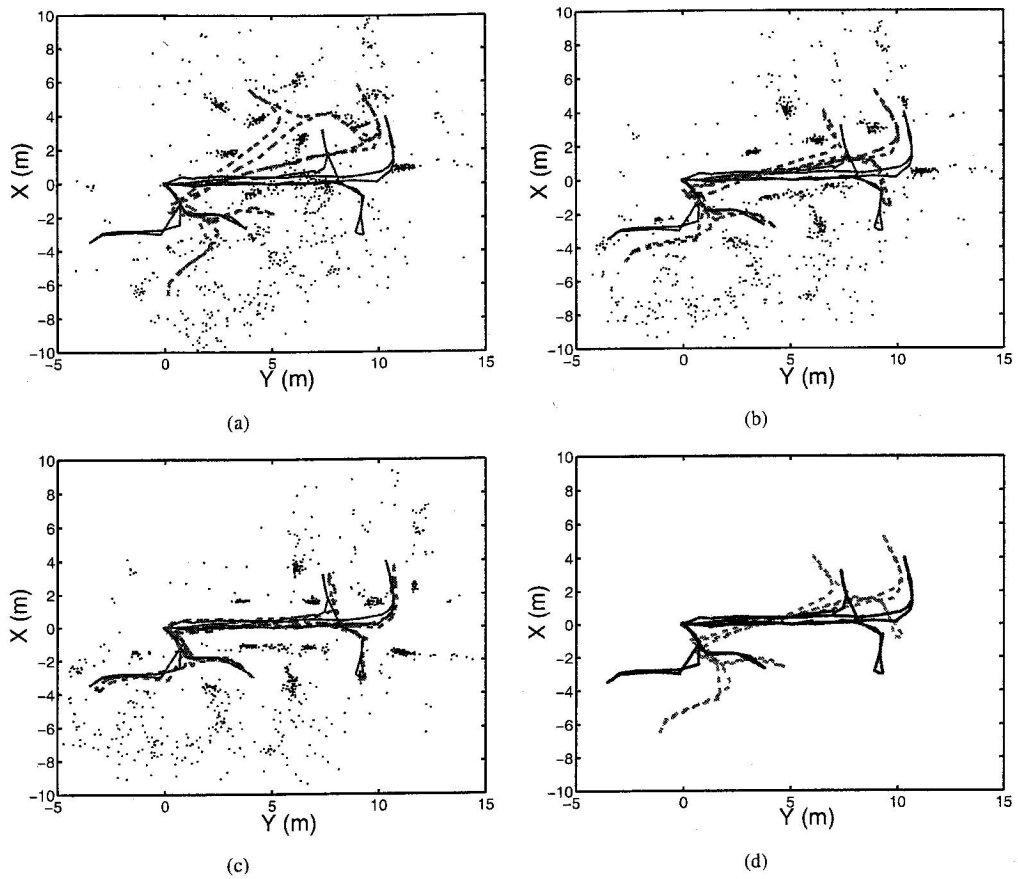


Figure 4. Figure 4(a) shows a map created using 1 particle. Figure 4(b) shows a map created using 10 particles and image 4(c) used 100 particles. We also show superposed the real path (continuous) and the estimated path using our approach (dashed). Figure 4(d) shows the real path (continuous) and the odometry of the robot (dashed).

[4] J. Little, S. Se, and D. Lowe. Global localization using distinctive visual features. In *Proceedings of the 2002 IEEE/RSJ Intl. Conference on Intelligent Robots and Systems*, 2002.

[5] 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.

[6] 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.

[7] 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.

[8] C. Stachniss, D. Haehnel, and W. Burgard. Improving grid-based slam with rao-blackwellized particle filters by adaptive proposals and selective resampling. In *IEEE Int. Conference on Robotics and Automation (ICRA)*, 2005.

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

[10] D. Lowe. Object recognition from local scale-invariant features. In *International Conference on Computer Vision*, pages 1150–1157, 1999.

[11] A. Gil, O. Reinoso, A. Vicente, C. Fernández, and L. Payá. Monte carlo localization using sift features. *Lecture Notes in Computer Science (LNCS)*, 1(3523):623–630, 2005.

[12] M. Montemerlo, S. Thrun, D. Koller, and B. Wegbreit. Fastslam: A factored solution to the simultaneous localization and mapping problem. In *AAAI*, 2002.

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

[14] C. Stachniss, D. Haehnel, and W. Burgard. Exploration with active loop-closing for FastSLAM. In *IEEE/RSJ Int. Conference on Intelligent Robots and Systems*, 2004.