intechweb.org

Home | Books | Journals | Shop

Search

Books

Browse by Title | Browse by Subject | **New Books** | Most Downloaded Books

Computer Vision
ISBN 978-953-7619-21-3
Hard cover, 538 pages
Edited by: Xiong Zhihui
Publisher: IN-TECH
Publication date: November 2008
Buy this Book
Price: 80 Euro incl. package & postage
Download the full text of this book
FREE

About the Book

This book presents research trends on computer vision, especially on application of robotics, and on advanced a omnidirectional vision). Among them, research on RFID technology integrating stereo vision to localize an indoor mobi this book includes many research on omnidirectional vision, and the combination of omnidirectional vision with robotics. the computer vision, and it puts more focus on robotics vision and omnidirectioal vision. The intended audience is anyor latest research work on computer vision, especially its applications on robots. The contents of this book allow the rea applications of computer vision. Researchers and instructors will benefit from this book.

Table of Content

01 Computer Vision
Xiong Zhihui
02 Behavior Fusion for Visually-Guided Service Robots
Mohamed Abdellatif
03 Dynamic Omnidirectional Vision Localization Using a Beacon Tracker Based on Particle Filter
Zuoliang Cao, Xianqiu Meng and Shiyu Liu
04 Paracatadioptric Geometry using Conformal Geometric Algebra
Carlos Lopez-Franco
05 Treating Image Loss by using the Vision/Motion Link:
David Folio and Viviane Cadenat
06 Nonlinear Stable Formation Control using Omnidirectional Images
Christiano Couto Gava, Raquel Frizera Vassallo, Flavio Roberti and Ricardo Carelli
07 Dealing with Data Association in Visual SLAM
Arturo Gil, Oscar Reinoso, Monica Ballesta and David Ubeda
08 Precise and Robust Large-Shape Formation using Uncalibrated Vision for a Virtual Mold
Biao Zhang, Emilio J. Gonzalez-Galvan, Jesse Batsche, Steven B. Skaar, Luis A. Raygoza and Ambrocio Loredo
09 Humanoid with Interaction Ability Using Vision and Speech Information
Junichi Ido, Ryuichi Nisimura, Yoshio Matsumoto and Tsukasa Ogasawara
10 Development of Localization Method of Mobile Robot with RFID Technology and Stereo Vision
Songmin Jia, Jinbuo Sheng and Kunikatsu Takase
11 An Implementation of Humanoid Vision - Analysis of Eye Movement and Implementation to Robot
Kunihito Kato, Masayuki Shamoto and Kazuhiko Yamamot
12 Methods for Postprocessing in Single-Step Diffuse Optical Tomography
Alexander B. Konovalov, Vitaly V. Vlasov, Dmitry V. Mogilenskikh, Olga V. Kravtsenyuk and Vladimir V. Lyubimov
13 Towards High-Speed Vision for Attention and Navigation of Autonomous City Explorer (ACE)
Tingting Xu, Tianguang Zhang, Kolja Kühnlenz and Martin Buss
14 New Hierarchical Approaches in Real-Time Robust Image Feature Detection and Matching
M. Langer and K.-D. Kuhnert

15 Image Acquisition Rate Control Based on Object State Information in Physical and Image Coordinates
Feng-Li Lian and Shih-Yuan Peng
16 Active Tracking System with Rapid Eye Movement Involving Simultaneous Top-down and Bottom-up Attention Contr
Masakazu Matsugu, Kan Torii and Yoshinori Ito
17 Parallel Processing System for Sensory Information Controlled by Mathematical Activation-Input-Modulation Model
Masahiko Mikawa, Takeshi Tsujimura and Kazuyo Tanaka
18 Development of Pilot Assistance System with Stereo Vision for Robot Manipulation
Takeshi Nishida, Shuichi Kurogi, Koichi Yamanaka, Wataru Kogushi and Yuichi Arimura
19 Camera Modelling and Calibration - with Applications
Anders Ryberg, Anna-Karin Christiansson, Bengt Lennartson and Kenneth Eriksson
20 Algorithms of Digital Processing and the Analysis of Underwater Sonar Images
S.V. Sai, A.G. Shoberg and L.A. Naumov
21 Indoor Mobile Robot Navigation by Center Following based on Monocular Vision
Takeshi Saitoh, Naoya Tada and Ryosuke Konishi
22 Temporal Coordination among Two Vision-Guided Vehicles: A Nonlinear Dynamical Systems Approach
Cristina P Santos and Manuel Joao Ferreira
23 Machine Vision: Approaches and Limitations
Moises Rivas Lopez, Oleg Sergiyenko and Vera Tyrsa
24 Image Processing for Next-Generation Robots
Gabor Sziebig, Bjørn Solvang and Peter Korondi
25 Projective Reconstruction and Its Application in Object Recognition for Robot Vision System
Ferenc Tél and Béla Lantos
26 Vision-based Augmented Reality Applications
Yuko Uematsu and Hideo Saito
27 Catadioptric Omni-directional Stereo Vision and Its Applications in Moving Objects Detection
Xiong Zhihui, Chen Wang and and Zhang Maojun
28 Person Following Robot with Vision-based and Sensor Fusion Tracking Algorithm
Takafumi Sonoura, Takashi Yoshimi, Manabu Nishiyama, Hideichi Nakamoto, Seiji Tokura and Nobuto Matsuhira

Computer Vision

Computer Vision

Edited by

Xiong Zhihui

I-Tech

Published by In-Teh

In-Teh is Croatian branch of I-Tech Education and Publishing KG, Vienna, Austria.

www.in-teh.org
Additional copies can be obtained from:
publication@ars-journal.com

First published November 2008
Printed in Croatia

A catalogue record for this book is available from the University Library Rijeka under no. 120110068
Computer Vision, Edited by Xiong Zhihui
p. cm.
ISBN 978-953-7619-21-3
1. Computer Vision, Xiong Zhihui

Preface

Computer vision uses digital computer techniques to extract, characterize, and interpret information in visual images of a three-dimensional world. The goal of computer vision is primarily to enable engineering systems to model and manipulate the environment by using visual sensing.

The field of computer vision can be characterized as immature and diverse. Even though earlier work exists, it was not until the late 1970s that a more focused study of the field started when computers could manage the processing of large data sets such as images.

There are numerous applications of computer vision, including robotic systems that sense their environment, people detection in surveillance systems, object inspection on an assembly line, image database organization and medical scans.

Application of computer vision on robotics attempt to identify objects represented in digitized images provided by video cameras, thus enabling robots to "see". Much work has been done on stereo vision as an aid to object identification and location within a three-dimensional field of view. Recognition of objects in real time, as would be needed for active robots in complex environments, usually requires computing power beyond the capabilities of present-day technology.

This book presents some research trends on computer vision, especially on application of robotics, and on advanced approachs for computer vision (such as omnidirectional vision). Among them, research on RFID technology integrating stereo vision to localize an indoor mobile robot is included in this book. Besides, this book includes many research on omnidirectional vision, and the combination of omnidirectional vision with robotics.

This book features representative work on the computer vision, and it puts more focus on robotics vision and omnidirectioal vision. The intended audience is anyone who wishes to become familiar with the latest research work on computer vision, especially its applications on robots. The contents of this book allow the reader to know more technical aspects and applications of computer vision. Researchers and instructors will benefit from this book.

Editor

Xiong Zhihui
College of Information System and Management,
National University of Defense Technology,
P.R. China

Contents

Preface V

1. Behavior Fusion for Visually-Guided Service Robots 001
Mohamed Abdellatif

2. Dynamic Omnidirectional Vision Localization Using a Beacon Tracker Based on Particle Filter 013
Zuoliang Cao, Xianqiu Meng and Shiyu Liu

3. Paracatadioptric Geometry using Conformal Geometric Algebra 029
Carlos López-Franco

4. Treating Image Loss by using the Vision/Motion Link: A Generic Framework 045
David Folio and Viviane Cadenat

5. Nonlinear Stable Formation Control using Omnidirectional Images 071
Christiano Couto Gava, Raquel Frizera Vassallo, Flavio Roberti and Ricardo Carelli

6. Dealing with Data Association in Visual SLAM 99
Arturo Gil, Óscar Reinoso, Mónica Ballesta and David Úbeda

7. Precise and Robust Large-Shape Formation using Uncalibrated Vision for a Virtual Mold 111
Biao Zhang, Emilio J. Gonzalez-Galvan, Jesse Batsche, Steven B. Skaar, Luis A. Raygoza and Ambrocio Loredo

8. Humanoid with Interaction Ability Using Vision and Speech Information 125
Junichi Ido, Ryuichi Nisimura, Yoshio Matsumoto and Tsukasa Ogasawara

9. Development of Localization Method of Mobile Robot with RFID Technology and Stereo Vision 139
Songmin Jia, Jinbuo Sheng and Kunikatsu Takase

10. An Implementation of Humanoid Vision - Analysis of Eye Movement and Implementation to Robot 159
Kunihito Kato, Masayuki Shamoto and Kazuhiko Yamamot

11. Methods for Postprocessing in Single-Step Diffuse Optical Tomography 169
Alexander B. Konovalov, Vitaly V. Vlasov, Dmitry V. Mogilenskikh, Olga V. Kravtsenyuk and Vladimir V. Lyubimov

12. Towards High-Speed Vision for Attention and Navigation of Autonomous City Explorer (ACE) 189
Tingting Xu, Tianguang Zhang, Kolja Kühnlenz and Martin Buss

13. New Hierarchical Approaches in Real-Time Robust Image Feature Detection and Matching 215
M. Langer and K.-D. Kuhnert

14. Image Acquisition Rate Control Based on Object State Information in Physical and Image Coordinates 231
Feng-Li Lian and Shih-Yuan Peng

15. Active Tracking System with Rapid Eye Movement Involving Simultaneous Top-down and Bottom-up Attention Control 255
Masakazu Matsugu, Kan Torii and Yoshinori Ito

16. Parallel Processing System for Sensory Information Controlled by Mathematical Activation-Input-Modulation Model 271
Masahiko Mikawa, Takeshi Tsujimura, and Kazuyo Tanaka

17. Development of Pilot Assistance System with Stereo Vision for Robot Manipulation 287
Takeshi Nishida, Shuichi Kurogi, Koichi Yamanaka, Wataru Kogushi and Yuichi Arimura

18. Camera Modelling and Calibration - with Applications 303
Anders Ryberg, Anna-Karin Christiansson, Bengt Lennartson and Kenneth Eriksson

19. Algorithms of Digital Processing and the Analysis of Underwater Sonar Images 333
S.V. Sai, A.G. Shoberg and L.A. Naumov

20. Indoor Mobile Robot Navigation by Center Following based on Monocular Vision 351
Takeshi Saitoh, Naoya Tada and Ryosuke Konishi

21. Temporal Coordination among Two Vision-Guided Vehicles: A Nonlinear Dynamical Systems Approach 367
Cristina P Santos and Manuel João Ferreira

22. Machine Vision: Approaches and Limitations 395
Moisés Rivas López, Oleg Sergiyenko and Vera Tyrsa

23. Image Processing for Next-Generation Robots 429
Gabor Sziebig, Bjørn Solvang and Peter Korondi

24. Projective Reconstruction and Its Application in Object Recognition for Robot Vision System 441
Ferenc Tél and Béla Lantos

25. Vision-based Augmented Reality Applications 471
Yuko Uematsu and Hideo Saito

26. Catadioptric Omni-directional Stereo Vision and Its Applications in Moving Objects Detection 493
Xiong Zhihui, Chen Wang and Zhang Maojun

27. Person Following Robot with Vision-based and Sensor Fusion Tracking Algorithm 519
Takafumi Sonoura, Takashi Yoshimi, Manabu Nishiyama, Hideichi Nakamoto, Seiji Tokura and Nobuto Matsuhira

6

Dealing with Data Association in Visual SLAM

Arturo Gil, Óscar Reinoso, Mónica Ballesta and David Úbeda
Miguel Hernández University
Systems Engineering Department
Elche (Alicante), Spain

1. Introduction

This chapter presents a stereo vision application to Mobile Robotics. In particular we deal with the problem of simultaneous localization and mapping (SLAM) (Dissanayake et al., 2001; Montemerlo et al., 2002) and propose a stereo vision-based technique to solve it (Gil et al., 2006). The problem of SLAM is of paramount importance in the mobile robotics community, since it copes with the problem of incrementally building a map of the environment while simultaneously localizing the robot within this map. Building a map of the environment is a fundamental task for autonomous mobile robots, since the maps are required for different higher level tasks, such as path planning or exploration. It is certainly an ability necessary to achieve a true autonomous operation of the robot. In consequence, this problem has received significant attention in the past two decades.

The SLAM problem is inherently a hard problem, because noise in the estimate of the robot pose leads to noise in the estimate of the map and viceversa. The approach presented here is feature based, since it concentrates on a number of points extracted from images in the environment which are used as visual landmarks. The map is formed by the 3D position of these landmarks, referred to a common reference frame. The visual landmarks are extracted by means of the Scale Invariant Feature Transform (SIFT) (Lowe, 2004). A rejection technique is applied in order to concentrate on a reduced set of highly distinguishable, stable features. The SIFT transform detects distinctive points in images by means of a difference of gaussian function (DoG) applied in scale space. Next, a descriptor is computed for each detected point, based on local image information at the characteristic scale (Lowe, 2004). We track detected SIFT features along 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 more representative feature model given the previous observations. This model allows to improve the Data Association within the landmarks in the map and, in addition, permits to reduce the number of landmarks that need to be maintained in the map. The visual SLAM approach is applied within a Rao-Blackwellized particle filter (Montemerlo et al., 2002; Grisetti et al., 2005).

In this chapter we propose two relevant contributions to the visual SLAM solution. First, we present a new mechanism to deal with the data association problem for the case of visual landmarks. Second, our approach actively tracks landmarks prior to its integration in the map. As a result, we concentrate on a small set of stable landmarks and incorporate them in the map. With this approach, our map typically consists of a reduced number of landmarks

Source: Computer Vision, Book edited by: Xiong Zhihui,
ISBN 978-953-7619-21-3, pp. 538, November 2008, I-Tech, Vienna, Austria

compared to those of (Little et al., 2002) and (Sim et al., 2006), for comparable map sizes. In addition, we have applied effective resampling techniques, as exposed in (Stachniss et al., 2004). This fact reduces the number of particles needed to construct the map, thus reducing computational burden.
Our system has been implemented and tested on data gathered with a mobile robot in a typical office environment. Experiments presented in this chapter demonstrate that our method improves the data association and in this way leads to more accurate maps.
The remainder of the chapter is structured as follows. Section 2 introduces related work in the context of visual SLAM. Next, Section 3 defines the concept of visual landmark and their utility in SLAM. Section 4 explains the basics of the Rao-Blackwellized particle filter employed in the solution. Next, Section 5 presents our solution to the data association problem in the context of visual landmarks. In Section 6 we present our experimental results. Finally, Section 7 sums up the most important conclusions and proposes future extensions.

2. Related work

Most work on SLAM so far has focussed on building 2D maps of environments using range sensors such as SONAR or laser (Wijk and Christensen, 2000; Thrun, 2001). Recently, Rao-Blackwellized particle filters have been used as an effective means to solve the SLAM problem using occupancy grid maps (Stachniss et al., 2004) or landmark-based maps (Montemerlo et al., 2002). Fig. 1 shows an example of both kind of maps. Recently, some authors have been concentrating on building three dimensional maps using visual information extracted from cameras. Typically, in this scenario, the map is represented by a set of three dimensional landmarks related to a global reference frame. The reasons that motivate the use of vision systems in the SLAM problem are: cameras are typically less expensive than laser sensors, have a lower power consumption and are able to provide 3D information from the scene.

a) b)

Fig. 1. Two typical maps. Fig. 1(a) occupancy-grid map. Fig. 1(b) landmark-based map: landmarks are indicated with (grey/yellow dots).

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. During exploration, the robot extracts SIFT features from stereo images and

computes 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 (Dissanayake 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 may 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., 2006) 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.

3. Visual landmarks

In our work, we use visual landmarks as features to build the map. Two main processes can be distinguished when observing a visual landmark:

- The detection phase: This involves extracting a point in the space by means of images captured from the environment. The detection algorithm should be stable to scale and viewpoint changes, i.e. should be able to extract the same points in space when the robot observes them from different angles and distances.
- The description phase: Which aims at describing the appearance of the point based on local image information. The visual descriptor computed in this phase should also be invariant to scale and viewpoint changes. 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. The descriptor is employed in the data association problem, described in Section 5.

Nowadays, a great variety of detection and description methods have been proposed in the context of visual SLAM. In particular, in the work presented here we use SIFT features (Scale Invariant Feature Transform) which were developed for image feature generation, and used initially in object recognition applications (Lowe, 2004; Lowe, 1999). The Scale-Invariant Feature Transform (SIFT) is an algorithm that detects distinctive keypoints from images and computes a descriptor for them. The interest points extracted are said to be invariant to image scale, rotation, and partially invariant to changes in viewpoint and illumination. SIFT features are located at maxima and minima of a difference of Gaussians (DoG) function applied in scale space. They can be computed by building an image pyramid with resampling between each level. Next, the descriptors are computed based on orientation histograms at a 4x4 subregion around the interest point, resulting in a 128 dimensional vector. The features are said to be invariant to image translation, scaling, rotation, and partially invariant to illumination changes and affine or 3D projection. 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., 2006).
Recently, a method called Speeded Up Robust Features (SURF) was presented (Bay et al., 2006). The detection process is based on the Hessian matrix. SURF descriptors are based on sums of 2D Haar wavelet responses, calculated in a 4x4 subregion around each interest point. For example, in (Murillo et al., 2007) a localization method based on SURF features is presented.

Finally, in (Davison & Murray, 2002) monocular SLAM is implemented using the Harris corner detector and the landmarks are described by means of a gray patch centered at the points.
To sum up, different detectors and descriptors have been used in visual SLAM approaches. In our opinion, there exists no consensus on this matter and this means that the question of which interest point detector and descriptor is more suitable for visual SLAM is still open. However, the evaluation presented by (Mikolajczyk & Schmid, 2005) proved the great invariability and discriminant power of the SIFT descriptors. On the other hand, the study presented in (Ballesta et al.,2007), demonstrated that the points obtained with the DoG detector where highly unstable. As a consequence, in the work presented here, a tracking of the points is performed in order to reject unstable points.

4. Rao-Blackwellized SLAM

We estimate the map and the path of the robot using a Rao-Blackwellized particle filter. Using the most usual nomenclature, we denote as s_t the robot pose at time t. On the other hand, the robot path until time t will be denoted $s_t = \{s_1, s_2, \cdots, s_t\}$. We assume that at time t the robot obtains an observation z_t from a landmark. The set of observations made by the robot until time t will be denoted $z^t = \{z_1, z_2, \cdots, z_t\}$ and the set of actions $u^t = \{u_1, u_2, \cdots, u_t\}$. The map is composed by a set of different landmarks $L = \{l_1, l_2, \cdots, l_N\}$. Therefore, the SLAM problem can be formulated as that of determining the location of all landmarks in the map L and robot poses s_t from a set of measurements z^t and robot actions u^t. Thus, the SLAM problem can be posed as the estimation of the probability:

$$p(L \mid s_t, z_t, u_t, c_t) \tag{1}$$

While exploring the environment, the robot has to determine whether a particular observation z_t corresponds to a previously seen landmark or to a new one. This problem is known as the Data Association problem and will be further explained in Section 5. Provided that, at a time t the map consists of N landmarks, the correspondence is represented by c_t, where $c_t \in [1 \ldots N]$. In consequence, at a time t the observation z_t corresponds to the landmark c_t in the map. When no correspondence is found we denote it as $c_t = N+1$, indicating that a new landmark should be initialized. The conditional independence property of the SLAM problem implies that the posterior (1) can be factored as (Montemerlo et al., 2002):

$$p\left(s^t, L \mid z^t, u^t, c^t\right) = p\left(s^t \mid z^t, u^t, c^t\right) \prod_{k=1}^{N} p\left(l_k \mid s^t, z^t, u^t, c^t\right) \tag{2}$$

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 (Murphy, 1999). We approximate $p(s_t \mid z_t, u_t, c_t)$ 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:

$$S_t^{[m]} = \left\{ s^{t,[m]}, \mu_{1,t}^{[m]}, \Sigma_{1,t}^{[m]}, \mu_{2,t}^{[m]}, \Sigma_{2,t}^{[m]}, \cdots, \mu_{N,t}^{[m]}, \Sigma_{N,t}^{[m]} \right\} \tag{3}$$

where $\mu_{k,t}^{[m]}$ is the best estimation at time t for the position of landmark based on the path of the particle m and $\Sigma_{k,t}^{[m]}$ its associated covariance matrix. Each landmark is thus described as: $l_k=\{\mu_k, \Sigma_k, d_k\}$, where d_k is the associated SIFT descriptor. The SIFT descriptor allows to differentiate between landmarks, based on their visual appearance. The set of M particles, each one with its associated map will be denoted $S_t=\{S_t^1, S_t^2, \cdots, S_t^M\}$. The particle set S_t is calculated incrementally from the set S_{t-1}, computed at time t-1 and the robot control u_t. Thus, each particle is sampled from a proposal distribution $p(s_t \mid s_{t-1}, u_t)$, which defines the movement model of the robot. Particles generated by the movement model are distributed following the probability distribution $p(s^t \mid z^{t-1}, u^t, c^{t-1})$, since the last observation of the robot z_t has not been considered. On the contrary, we would like to estimate the posterior: $p(s^t \mid z^t, u^t, c^t)$, in which all the information from the odometry u^t and observations z^t is included. This difference is corrected by means of a process denoted sample importance resampling (SIR). Essentially, a weigth is assigned to each particle in the set according to the quality by which the pose and map of the particle match the current observation z_t. Following the approach of (Montemerlo et al., 2002) we compute the weight assigned to each particle as:

$$\omega_t^{[m]} = \frac{1}{\sqrt{\left|2\pi Z_{c_t,t}\right|}} \exp\left\{-\frac{1}{2}\left(z_t - \hat{z}_{c_t,t}\right)^T \left[Z_{c_t,t}\right]^{-1}\left(z_t - \hat{z}_{c_t,t}\right)\right\} \tag{4}$$

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

$$\omega_t^{[m]} = \prod_{i=1}^{B} \omega_{t,i}^{[m]} \tag{5}$$

The weights are normalized so that $\sum_{i=1}^{M} \omega_t^{[i]} = 1$, so that they ressemble a probability function. 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:

$$N_{eff} = \frac{1}{\sum_{i=1}^{M}\left(\omega_t^{[i]}\right)^2} \tag{6}$$

We resample each time N_{eff} drops below a pre-defined 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.

5. Data association

While the robot explores the environment it must decide whether the observation z_t corresponds to a previously mapped landmark or to a different one. The observation z_t is a relative three-dimensional relative measurement obtained with a stereo camera. Associated to the observation is a visual SIFT descriptor d_t. To find the data association we find a set of landmark candidates using the current measurement z_t and the following Mahalanobis distance function:

$$d = \left(z_t - \hat{z}_{c_t,t}\right)^T \left[Z_{c_t,t}\right]^{-1} \left(z_t - \hat{z}_{c_t,t}\right) \tag{7}$$

The landmarks with d below a pre-defined threshold d_0 are considered as candidates. Next, we use the associated SIFT descriptor d_t to find the correct data association among the candidates. 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., 2006), data association is based on the squared Euclidean distance between descriptors. In consequence, given a current SIFT descriptor, associated to the observation z_t and the SIFT descriptor d_i, associated to the i landmark in the map, the following distance function is computed:

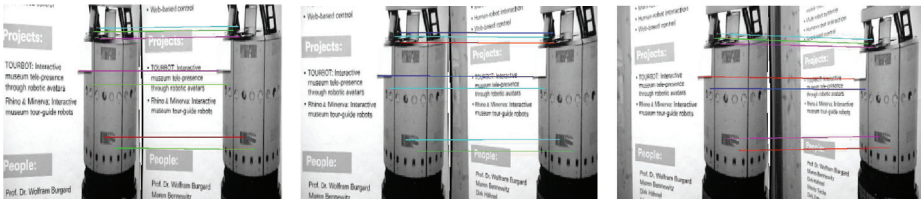
$$E = (d_t - d_i)(d_t - d_i) \tag{8}$$

Then, the landmark i of the map that minimizes the distance E is chosen. Whenever the distance E is below a certain threshold, the observation and the landmark are associated. 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 of 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 approach produces a remarkable number of incorrect data associations, normally causing an inconsistent map.

We propose a different method to deal with the data association in the context of visual SLAM. We address the problem from a pattern classification point of view. We consider the problem of assigning a pattern d_t 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 the same class C_i. Whenever a landmark is found in an image, it is tracked along p frames and its descriptors $\{d_1, d_2, \ldots, d_p\}$ are stored. Then, for each landmark C_i we compute a mean value 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:

$$D = (d_t - d_i)S_i^{-1}(d_t - d_i) \tag{9}$$

We compute the distance D for all the landmarks in the map of each particle and assign the correspondence to the landmark that minimizes D. 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 2). 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 (8) and (9). For each observation, we select the class that minimises its distance function and as we already know the correspondences, we can compute the number of incorrect and correct matches. Table 1 shows the results based on our experiments. A total of 3000 examples where 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 6.



a) b) c)

Fig. 2. Tracking of points viewed from different angles and distances.

6. 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 $z_t = \{z_{t,1}, z_{t,2}, \cdots, z_{t,B}\}$ were obtained, each observation accompanied by a SIFT descriptor $\{d_{t,1}, d_{t,2}, \cdots, d_{t,B}\}$. 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. In a practical way, when a landmark has been tracked for more than p=5 frames it is considered a new observation and is integrated in the filter. After the tracking, a mean value d_i is computed using the SIFT descriptors in the p views and a diagonal covariance matrix is also computed. In consequence, as mentioned in Section 5, each landmark is now represented by (d_i, S_i). Along with the images, we captured laser data using the SICK laser range finder. These data allowed us to estimate the path followed by the robot using the approach of (Stachniss, 2004). This path has shown to be very precise and is used as ground truth.

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

Table 1. Comparison of correct and incorrect matches using the Euclidean distance and the Mahalanobis distance in the data association.

Figure 3 shows the map constructed with 1, 10, and 100 particles. A total number of 1500 landmarks were estimated. With only 1 particle the method fails to compute a coherent map, since only one hypothesis is maintained over the robot path. It can be seen that, with only 10 particles, the map is topologically correct. Using only 100 particles the map is very precise. On every figure we show the path followed by the robot (blue continuous line), the odometry of the robot (magenta dotted line) and the path estimated using the visual SLAM approach presented here (red dashed line). 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 4 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 5. 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 5 for the same number of particles, better localization results are obtained when the Mahalanobis distance is used (red continuous line), compared to the results obtained using the Euclidean distance (blue dashed line). Better results in the path estimation imply an in the quality of the estimated map.
Compared to preceeding approaches our method uses less particles to achieve good results. For example, in (Sim et al., 2006), 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., 2006), where the map contains typically around 100.000 landmarks.

7. Conclusion

In this Chapter 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 results 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.

8. Acknowledgment

The work has been supported by the Spanish government under the project: SISTEMAS DE PERCEPCION VISUAL MOVIL Y COOPERATIVO COMO SOPORTE PARA LA REALIZACIÓN DE TAREAS CON REDES DE ROBOTS. Ref. DPI2007-61197, (Ministerio de Educación y Ciencia).

9. References

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

Bay, H.; Tuytelaars, T. & Van Gool, L. (2006). Object recognition from local scale-invariant features. In: *European Conference on Computer Vision*.

Davison, A.J. & Murray, D.W. (2002). Simultaneous localisation and map-building using active vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence.*

Dissanayake, G.; Newman, P.; Clark, S.; Durrant-Whyte, H. & 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.; Burgard, W.; Stachniss, C. & Martínez Mozos, O. (2006). Improving data association in rao-blackwellized visual SLAM. *In: IEEE/RSJ Int. Conf. on Intelligent Robots & Systems*. Beijing, China.

Grisetti, G.; Stachniss, C. & Burgard, W. (2007). Improved techniques for grid mapping with Rao-blackwellized particle filters. *IEEE Transactions on Robotics* 23(1).

Little, J.; Se, S. & Lowe, D. (2002). Global localization using distinctive visual features. *In: IEEE/RSJ Int. Conf. on Intelligent Robots & Systems (ICRA)*.

Little, J.; Se, S. & Lowe, D. (2001). Vision-based mobile robot localization and mapping using scale-invariant features. *In: IEEE Int. Conf. on Robotics & Automation*.

Lowe, D. (2004). Distinctive image features from scale-invariant keypoints. *Int. Journal of computer Vision,* 2(60).

Lowe, D. (1999). Object recognition from local scale invariant features. In *International Conference on Computer Vision*, pages 1150–1157.

Mikolajczyk, K. & Schmid, C. (2005). A performance evaluation of local descriptors. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 27(10).

Miró, J. V.; Dissanayake, G. & 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. & Wegbreit, B. (2002). Fastslam: A factored solution to the simultaneous localization and mapping problem. *In Proc.~of the National Conference on Artificial Intelligence* (AAAI). Edmonton, Canada.

Murillo, A. C.; Guerrero, J.J. & Sagüés, C. (2007). SURF features for efficient robot localization with omnidirectional images. In: *IEEE Int. Conf. on Robotics & Automation*.

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

Sim, R.; & Little, J. J. (2006). Autonomous vision-based exploration and mapping using hybrid maps and Rao-Blackwellised particle filters. *In: IEEE/RSJ Int. Conf. on Intelligent Robots & Systems*. Beijing, China.

Stachniss, C.; Hähnel, D. & Burgard, W. (2004). Exploration with active loop-closing for FastSLAM. In *IEEE/RSJ Int. Conference on Intelligent Robots and Systems*.

Stachniss, C.; Hähnel, D. & 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. & Christensen, H. I. (2000). Localization and navigation of a mobile robot using natural point landmarks extracted from sonar data. *Robotics and Autonomous Systems*, 1(31):31–42.

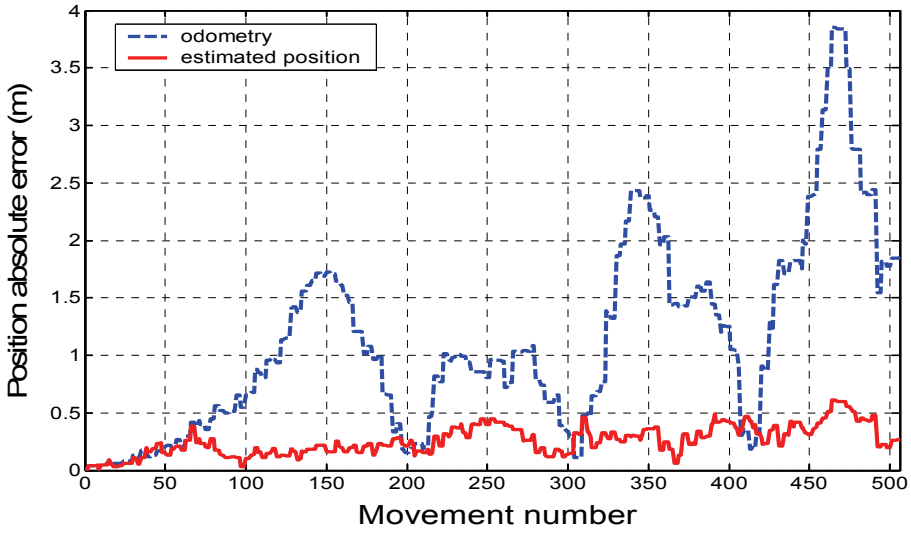


Fig. 4. Absolute position error in odometry and visual estimation.

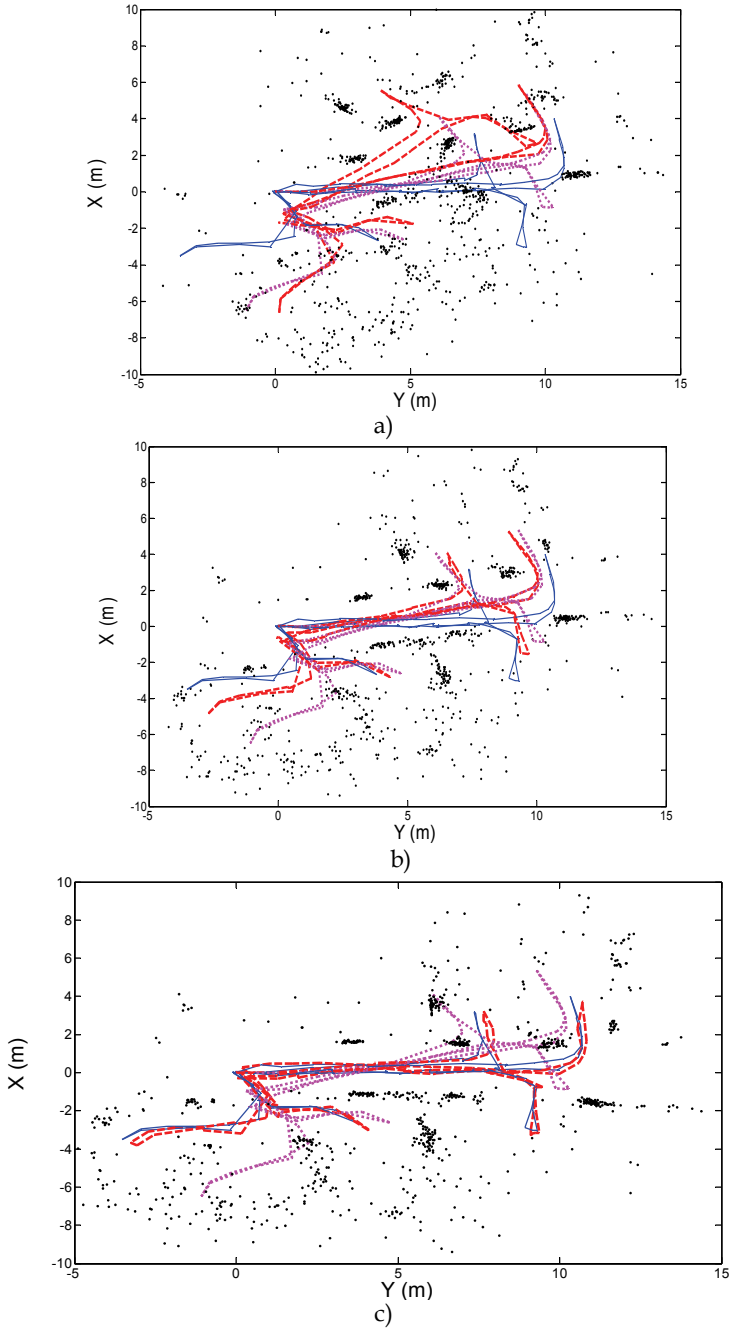


Fig. 3. Maps built using 1, 10 and 100 particles. A 2d view is showed where landmarks are indicated with black dots.

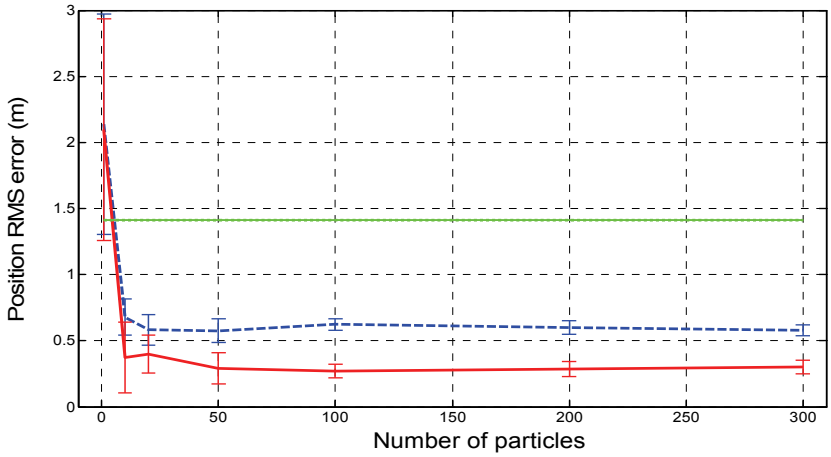


Fig. 5. RMS error in position for different number of particles.