Available online at www.sciencedirect.com

SCIENCE DIRECT®

Robotics and Autonomous Systems 54 (2006) 159–164

Robotics and Autonomous Systems

www.elsevier.com/locate/robot

Appearance-based concurrent map building and localization

J.M. Porta*, B.J.A. Kröse

IAS Group, Informatics Institute, University of Amsterdam, Kruislaan 403, 1098SJ, Amsterdam, The Netherlands

Received 2 October 2004; accepted 13 September 2005
Available online 9 December 2005

Abstract

In appearance-based robot localization the environment map does not represent geometrical features but consists of an appearance map, which is a collection of robot poses and corresponding sensor observations. In this paper, we describe a concurrent map-building and localization (CML) system based on a multi-hypotheses tracker that is able to build and refine autonomously the appearance map required for localization as the robot moves in the environment. The results included in this paper validate our approach.

Keywords: Appearance-based localization; Concurrent map building and localization

1. Introduction

Autonomous robot localization requires some kind of *representation* or *map* of the environment. If we pay attention to the type of features included in the map, robot localization methods can be divided into two families: geometric methods [8,15,17], and methods based on appearance modeling of the environment [6,12,14,20].

Geometric-based localization methods rely on the assumption that geometric information (i.e., position of landmarks, etc.) can be accurately extracted from the sensor readings. However, this transformation is, in general, complex and prone to errors. As a counterpart, in appearance-based methods, the environment is not modeled geometrically, but as an appearance map that is a collection of sensor readings obtained at known positions. The advantage of the appearance-based approach is that the sensor readings obtained at a given moment can be matched directly with the observations stored in the appearance-based map and, thus, we can obtain pose information without requiring any intermediate processes to obtain geometric information (see [20] for more details).

A comparison between two localization methods—one geometric-based and one appearance-based using vision as sensory input—can be found in [23], showing that the appearance-based method is more robust to noise, certain types of occlusions, and changes in illumination (when an edge detector is used to pre-process the images) than the geometric-based method. The main drawback of appearance-based methods is that the construction of an appearance map is usually a supervised process that can be quite time-consuming and that is only valid as long as no important modifications of the environment occur. While much work has been done on concurrent mapping and localization (CML) using landmarks [7,9,16,25], this is not the case within the appearance-based approach. Recent work in this line [22] does not exploit all the potential of the appearance-based framework such as, for instance, the ability to perform global localization (i.e., localization without any prior information on the robot's position).

In this paper, we replace the supervised map of the environment used in appearance-based localization by an approximation to it obtained autonomously by the robot. The basic idea we exploit is that, if the robot re-visits an already explored area, it can use the information previously stored to reduce the uncertainty in its position. Additionally, the improvements in the robot's position can be back-propagated to map points stored in previous time slices using trajectory reconstruction techniques. The result is a correction of both the robot's position and the map and, thus, we achieve the objective of concurrently localizing and building a map of the environment. Similar ideas are exploited in map building based on cyclic trajectories [2,10].

* Corresponding author. Present address: Institut de Robòtica i Informàtica Industrial (UPC-CSIC), Llorens i Artigas 4–6, 08028, Barcelona, Spain. Tel.: +34 93 4015751; fax: +34 93 4015750.

E-mail addresses: porta@iri.upc.edu (J.M. Porta), krose@science.uva.nl (B.J.A. Kröse).

0921-8890/$ - see front matter © 2005 Published by Elsevier B.V.
doi:10.1016/j.robot.2005.09.025

However, these works aim at building a geometric map of the environment and not an appearance-based one.

In the following sections, we first describe how to estimate the position of the robot (assuming that we have a map). After that, we describe how to extract features from the input images and how to approximate on-line the feature-based map that is necessary for localization. Next, we show the preliminary results obtained with the new CML system and, finally, we summarize our work and extract some conclusions from it.

2. Robot position estimation

The probabilistic localization methods aim at improving the estimation of the pose (position and orientation) of the robot at time t, i.e. x_t, taking into account the movements of the robot $\{u_1, \ldots, u_t\}$ and the observations of the environment taken by the robot $\{y_1, \ldots, y_t\}$ up to that time. In our notation, the Markov process goes through the following sequence $x_0 \xrightarrow{u_1} (x_1, y_1) \xrightarrow{u_2} \cdots \xrightarrow{u_t} (x_t, y_t)$. The Markov assumption states that the robot's pose can be updated from the previous state probability $p(x_{t-1})$, the last executed action u_t, and the current observation y_t. Applying Bayes, $p(x_t|u_t, y_t)$ gives

$$p(x_t|u_t, y_t) \propto p(y_t|x_t)p(x_t|u_t), \tag{1}$$

where the probability $p(x_t|u_t)$ can be computed propagating from $p(x_{t-1}|u_{t-1}, y_{t-1})$ using the action model

$$p(x_t|u_t) = \int p(x_t|u_t, x_{t-1})p(x_{t-1}|u_{t-1}, y_{t-1})\mathrm{d}x_{t-1}. \tag{2}$$

Eqs. (1) and (2) define a recursive system for estimating the position of the robot.

To compute the integral in Eq. (2) we have to make some assumption on the representation of $p(x_{t-1}|u_{t-1}, y_{t-1})$. Sometimes this probability is represented as a Gaussian[15], but this is a rather restrictive assumption on the shape of $p(x_{t-1}|u_{t-1}, y_{t-1})$. When a probabilistic occupancy grid [3,26] or a particle filter [19,28] is used, we can represent probability distributions with any shape. However, occupancy grids and particle filters are computationally expensive in memory and execution time.

In our work, the probability on the state $p(x_{t-1}|u_{t-1}, y_{t-1})$ is represented using a Gaussian Mixture (GM) X_{t-1} with N components and parameters $X_{t-1} = \{(x^i_{t-1}, \Sigma^i_{t-1}, w^i_{t-1}) \mid i \in [1, N]\}$. As noted by several authors, GMs provide a good trade-off between flexibility and efficiency [1,5,11]. Thus, we have that

$$p(x_{t-1}|u_{t-1}, y_{t-1}) \propto \sum_{i=1}^{N} w^i_{t-1}\phi(x_{t-1}|x^i_{t-1}, \Sigma^i_{t-1}),$$

where $\phi(x_{t-1}|x^i_{t-1}, \Sigma^i_{t-1})$ is a Gaussian centered at x^i_{t-1} with covariance matrix Σ^i_{t-1}. The weight $w^i_{t-1} (0 < w^i_{t-1} \leq 1)$ provides information on the certainty of the hypothesis represented by the corresponding Gaussian.

The motion of the robot is modeled as $x_t = f(x_{t-1}, u_t, v_t)$, where v_t is a Gaussian noise with zero mean and covariance Q. Thus, using a linear approximation, we can express $p(x_t|u_t)$ in Eq. (2) as the GM resulting from applying f to the elements in X_{t-1}, i.e. a GM with the same number of components as X_{t-1} and with parameters $X_{u_t} = \{(x^i_{u_t}, \Sigma^i_{u_t}, w^i_t) \mid i \in [1, N]\}$ with

$$\begin{aligned} x^i_{u_t} &= f(x^i_{t-1}, u_t), \\ \Sigma^i_{u_t} &= F\Sigma^i_{t-1}F^\top + GQG^\top, \end{aligned} \tag{3}$$

where F is the Jacobian of f with respect to x^i_{t-1}, and G is the Jacobian of f with respect to v_{t-1}.

After we have an approximation of $p(x_t|u_t)$, we need to integrate the information provided by the sensor readings $p(y_t|x_t)$ to obtain an estimation of the new robot's pose $p(x_t|u_t, y_t)$ (see Eq. (1)). As $p(x_{t-1}|u_{t-1}, y_{t-1})$ is represented as a GM, so will $p(x_t|u_t, y_t)$. Therefore, the problem is to determine a new GM, X_t, from the one representing X_{t-1} and the additional information provided by the sensors. At this point, we assume we have an appearance map of the environment from which we can define $p(y_t|x_t)$. A classical method for approximating the sensor model $p(y|x)$ from a supervised training set is to use kernel smoothing. This method scales linearly with the size of the map and, thus, it is inefficient for practical applications. For these reason, as proposed by Vlassis et al. in [28], we use a GM to approximate the sensor model. The parameters of the GM $X_{y_t} = \{(x^j_{y_t}, \Sigma^j_{y_t}, w^j_{y_t}) \mid j \in [1, N']\}$ are obtained from the map point with an observation closest to the current one y_t. x_{y_t} denotes the poses from which an observation similar to y_t has been observed, and Σ_{y_t} and w_{y_t} are the associated covariance and weight factors. In Section 4, we describe how to create and update the map from which X_{y_t} is defined. If X_{y_t} has no components ($N' = 0$), the estimation on the robot's pose obtained by applying Eq. (3) can not be improved and we have $X_t = X_{u_t}$. If $N' > 0$, we have to fuse the Gaussian functions in X_{u_t} with those in X_{y_t}. The direct application of Eq. (1) amounts to multiplying each one of the elements in X_{u_t} (the predicted state after applying the action model) with those in X_{y_t} (the sensor model). This would produce a quadratic ($N \times N'$) number of hypotheses. To keep the number of hypotheses below a reasonable limit, we will only associate elements of X_{u_t} and X_{y_t} that are alternative approximations of the same positioning hypothesis. This raises the problem of *data association*: to determine which elements of X_{y_t} and X_{u_t} are to be combined. We perform the data association using an innovation-based criterion. For each couple (i, j) with $(x^i_{u_t}, \Sigma^i_{u_t}, w^i_t) \in X_{u_t}$ and $(x^j_{y_t}, \Sigma^j_{y_t}, w^j_{y_t}) \in X_{y_t}$, we compute the innovation as

$$\begin{aligned} \upsilon_{i,j} &= x^i_{u_t} - x^j_{y_t} \\ S_{i,j} &= \Sigma^i_{u_t} + \Sigma^j_{y_t}, \end{aligned}$$

and we assume that that hypothesis on the robot position i and sensor reading j match if the following condition holds

$$\upsilon_{i,j}S^{-1}_{i,j}\upsilon^\top_{i,j} \leq \gamma, \tag{4}$$

where γ is a user-defined threshold. A small innovation $\upsilon_{i,j}S^{-1}_{i,j}\upsilon^\top_{i,j}$ indicates two similar Gaussian functions and, thus, two Gaussian functions that are very likely to refer to the same pose hypothesis.