
An efficient formulation of the Bayesian occupation filter for target tracking in dynamic environments

M.K. Tay*

INRIA Rhône-Alpes, Team e-Motion,
Montbonnot,
Saint Ismier Cedex, France
E-mail: Christopher.Tay_meng_keat@inrialpes.fr
*Corresponding author

K. Mekhnacha

ProBayes, Inovallée,
Saint Ismier cedex, France
E-mail: kamel.mekhnacha@probayes.com

C. Chen

Advanced Electronic Department (DEA),
API : TCR RUC T 62,
Technocentre Renault,
1 Avenue du Golf, Guyancourt cedex 78288, France
E-mail: cheng.chen@renault.com

M. Yguel and C. Laugier

INRIA Rhône-Alpes, Team e-Motion,
Montbonnot,
Saint Ismier Cedex, France
E-mail: manuel.yguel@inrialpes.fr
E-mail: christian.laugier@inrialpes.fr

Abstract: The Bayesian Occupation Filter (BOF) has proven successful for target tracking in the context of automotive applications. This paper describes an improved BOF for target tracking with lower computational costs while retaining the key advantages of the original BOF formulation. The BOF takes the form of a grid-based decomposition of the environment. In contrast to the original BOF, each cell of the newly proposed BOF contains an additional distribution over the velocity of the propagating cell occupancy. This is estimated using Bayesian filtering. We propose how to deal with the grid discretisation problem. Object-based representations do not exist in the BOF grids. However, there are often applications which require the definition and tracking at the object level. A simple clustering and target tracking methodology is used to illustrate how to obtain this object level representation. Experiments based on tracking humans in indoor environment were conducted.

Keywords: Bayesian filtering; occupancy grids; tracking.

Reference to this paper should be made as follows: Tay, M.K., Mekhnacha, K., Chen, C., Yguel, M. and Laugier, C. (2008) 'An efficient formulation of the Bayesian occupation filter for target tracking in dynamic environments', *Int. J. Vehicle Autonomous Systems*, Vol. 6, Nos. 1/2, pp.155–171.

Biographical notes: Meng Keat Tay (Christopher) received a Bachelor Honours in Computer Engineering from the Nanyang Technological University in 2003 and a MSc speciality Image Vision and Robotics (2004) from INPG (Institut National Polytechnique de Grenoble), France. He is currently a PhD candidate within the robotics team, e-Motion, situated in INRIA Rhône-Alpes. His research interests include machine learning, statistics and its applications to autonomous robots.

Kamel Mekhnacha received the Engineering degrees (1994) in Computer Science from Constantine University, Algeria, and the MS degree (1995) and the PhD (1999) both in Computer Science from INPG (Institut National Polytechnique de Grenoble), France. He worked as a Research Engineer at INRIA (The French National Institute for Research in Computer Science and Control) for four years where he worked on various projects involving the most recent approaches in Bayesian modelling, inference and learning. He participated especially to the design and the development of the 'ProBT' Bayesian programming tool. He is a co-founder of Probayes Company and is currently its Technological Director. His research interests include Bayesian reasoning and learning, and Monte Carlo methods. His application fields include robotics, geometric modelling, car driving assistance and bank risk estimation.

Cheng Chen received his BEng in 2001 from Huazhong University of Science and Technology, China. He did his PhD in Nanyang Technological University, Singapore. Currently, he is now a Research Engineer in Renault Technocentre, France. Before joining Renault, he has been affiliated with Singapore Technologies – Kinetics and INRIA Rhone-Alpes. His research interests include sensor fusion, tracking and pattern recognition.

Manuel Yguel is a PhD candidate at the Institut National Polytechnique de Grenoble (INPG) and Member of the robotic team, e-Motion, of INRIA. His subjects of interest are dense environment modelling for autonomous system, Bayesian inference and statistical modelling of sensors.

Christian Laugier is Research Director at INRIA and Scientific Leader of the e-Motion project-team common to INRIA Rhône-Alpes and to LIG Laboratory. Since January 2007, he is also Deputy Director of the LIG Laboratory. He received the PhD and 'State Doctor' degrees in Computer Science from the Institut National Polytechnique de Grenoble in 1976 and 1987, respectively. His current research interests mainly lie in the areas of motion autonomy, intelligent vehicles, decisional architectures, Bayesian reasoning and dynamic simulation. In 1997, he was awarded the Nakamura Prize for his contributions on 'Intelligent Robots'.

1 Introduction

Perception and reasoning with dynamic environments is pertinent for mobile robotics and still constitutes one of the major challenges. To work in these environments, the mobile robot needs to perceive the environment using its sensors, where measurements are uncertain and normally treated within the estimation framework. Systems for tracking the evolution of the environment has been traditionally a major component in robotics. The major requirement for such a system is a robust target tracking system. Most of the existing target tracking algorithms use an object-based representation of the environment. However, these existing techniques have to take into account explicitly data association and occlusion. In view of these problems, a grid-based framework, the Bayesian Occupancy Filter (BOF) (Coué et al., 2002, 2003), has been proposed.

1.1 Motivation

In classical tracking methodology (Bar-Shalov and Fortman, 1988), data association and state estimation are major problems to be addressed. The two problems are highly coupled together and an error in either portion leads to erroneous outputs. The BOF makes it possible to decompose this highly coupled relationship by avoiding the data association problem, in the sense that data association is to be handled at a higher level of abstraction.

In the BOF model, concepts such as *objects* or *tracks* do not exist; they are replaced by more useful properties such as occupancy or risk, which are directly estimated for each cell of the grid using both sensor observations and some prior knowledge.

It might seem strange to have no object representations when objects do exist in real life environments. However, an object-based representation is not always required for all applications. In cases where object-based representations are not pertinent, we argue that it is more useful to work with a more descriptive and richer sensory representation rather than constructing object-based representations along with its relevant complications in data association. For example, to calculate the risk of collision for a mobile robot, the only properties required are the probability distribution on occupancy and velocities for each cell in the grid. Variables such as the number of objects are inconsequential in this respect.

This model is especially useful when there is a need to fuse information from several sensors. In standard methods for sensor fusion in tracking applications, the problem of track-to-track association arises where each sensor contains its own local information. Under the standard tracking framework with multiple sensors, the problem of data association will be further complicated. On top of data association between two consecutive time instances from the same sensor, the association of tracks (or targets) amongst the different sensors will have to be taken into account as well. In contrast, the gridbased BOF will not encounter such a problem. A grid-based representation provides a conducive framework for performing sensor fusion (Moravec, 1988). Different sensor models can be specified to tailor to the different characteristics of the different sensors and facilitates efficient fusion onto the grids. The absence of an object-based representation permits the ease of fusing low level descriptive sensory information onto the grids without implicating data association.

Uncertainties characteristic of the different sensors are specified in the sensor models. This uncertainty is explicitly represented in the BOF grids in the form of occupancy probability. Thanks to the probabilistic reasoning paradigm, which is becoming a key paradigm in robotics, various approaches based on this paradigm have already been

successfully used to address several robotic problems, such as CAD modelling (Mekhnacha et al., 2001) or Simultaneous Map Building and Localisation (SLAM) (Arras et al., 2001; Kaelbling et al., 1998; Thrun, 1998).

In modelling the environment with BOF grids, the object model problem is non-existent because there are only cells representing the state of the environment at a certain position and time, and each sensor measurement changes the state of each cell. The fact that different kinds of object produce different kinds of measures is handled naturally by the cell space discretisation.

Another advantage of BOF grids is their rich representation of dynamic environments. This information includes the description of occupied and hidden areas (i.e. areas of the environment that are temporarily hidden to the sensors by an obstacle). The dynamics of the environment and its robustness relative to object occlusions are addressed using a novel two-step mechanism which permits taking into account the history of the sensor observations and the temporal consistency of the scene. This mechanism estimates, at each time step, the state of the occupancy grid by combining a prediction step (history) and an estimation step (incorporating new measurements). This approach is derived from the Bayes filters approach (Jazwinsky, 1970); which explains why the filter is called the BOF.

For real time applications, the BOF has been designed in order to be highly parallelised. A hardware implementation on a dedicated chip is possible, which will lead to an efficient real time representation of the environment of a mobile robot.

1.2 Contributions

Previous experiments based on the BOF techniques (Coué et al., 2006) relied on the assumption of a given constant velocity, and the problem of velocity estimation in this context has not been addressed. In particular the assumption that there could only be one object with one velocity in each cell was not part of the previous model. In this paper, a representation that has one probability distribution over velocities for each occupancy cell is presented. This model is similar in concept to optical flow, but with occupancy considerations rather than intensity. The general principle for the estimation of occupancy grids will be to include the velocity estimation in the prediction estimation loop of the classical BOF approach. For each grid in the BOF, the set of velocities that brings a set of corresponding cells in the previous time step to the current grid will be considered. The resulting distribution on the velocity of the current grid is updated by conditioning on the incoming velocities with respect to the current grid and on the observations.

An improved approach to estimate both the occupancy states of the grids and the distribution of grid velocity is presented. To avoid confusion, all ideas presented apply to the currently proposed formulation of BOF unless explicitly stated.

When an object level representation is pertinent, we demonstrate that a general clustering tracking approach makes it possible to recover objects and perform tracking at the object level despite having no object representation in the BOF.

This paper is organised as follows: in Section 2, related work on multiple target tracking systems and occupancy grids is presented. The fundamental concept of Bayesian filtering and the new filtering equations in the grids are presented in Section 3. We define the solutions and problems of discretisation from the spatial and velocity point of view in Section 4. The need for objects arises often and the generic clustering tracking is presented in Section 5. Results are presented in Section 6. Experiments were conducted

in an indoor environment based on data from the European Project CAVIAR. This paper is concluded in Section 7.

2 Related work

2.1 Multitarget tracking

The aim of multitarget tracking is to estimate, at each time step, the dynamics of each moving object observed by the sensors. Such an estimation is required due to uncertainty in observations that is sensor data. The estimation of the dynamics is performed, in a manner that is as robust as possible, after observations are obtained from the sensors. The main difficulty of multitarget tracking is known as the *Data Association* problem. It includes observation-to-track association and track management problems. The goal of observation-to-track association is to decide whether a new sensor observation corresponds to an existing track. Track management includes deciding whether existing tracks should be maintained, deleted or if new tracks should be created. Numerous methods exist to perform data association (Bar-Shalom and Li, 1995; Gauvrit et al., 1997; Streit and Luginbuhl, 1995). The reader is referred to Blackman and Popoli (2000) for a complete review of the existing tracking methods with one or more sensors.

Urban traffic scenarios are still a challenge in multitarget tracking area: the traditional data association problem is intractable in situations involving numerous appearances, disappearances and occlusions of a large number of rapidly manoeuvring targets.

In Wang et al. (2003), a classical Multiple Hypothesis Tracking technique is used to track moving objects while stationary objects are used for SLAM. Unfortunately, the authors did not explicitly address the problem of the interaction between tracked and stationary objects, for example, when a pedestrian is temporary hidden by a parked car. One of the aims of the BOF is to overcome such a problem.

2.2 Grid representation of the environment

The occupancy grids framework (Elfes, 1989; Moravec, 1988) is a classical way to describe the environment of a mobile robot. It has been extensively used for static indoor mapping using a two-dimensional grid (Thrun, 2002). The goal is to compute from the sensor observations the probability of each cell being occupied or empty. To avoid a combinatorial explosion of grid configuration, the cell states are estimated as independent random variables.

More recently, occupancy grids have been adapted to track multiple moving objects (Prassler et al., 2000). In this approach, spatio-temporal clustering applied to temporal maps is used to perform motion detection and tracking. A major drawback of this work is that a moving object may be lost due to occlusion effects.

3 The Bayesian occupancy filter

The BOF is represented as a two-dimensional planar grid-based decomposition of the environment. Each cell of the grid contains two probability distributions. A probability

distribution on the occupancy of the cell, and the probability distribution on the velocity of the cell occupancy.

This model of the dynamic grid is different from the approach adopted in the original BOF formulation by Coué et al. (2006). The grid model in Coué et al. (2006) is in four-dimensional space whereas this paper models the grid in two-dimensional space. The essential difference albeit subtle is that the original model allows the representation for overlapping objects but the model proposed in this paper does not. However, a common characteristic of the BOF presented in this paper and that of the original is estimation of the probability distributions of the cell occupancy and velocities by a Bayesian filter given the set of sensor data readings.

3.1 Grid-based Bayesian filtering

The Bayes filter (Jazwinsky, 1970) addresses the general problem of recursively estimating the probability distribution, $P(X^k | Z^k)$, of the state of a system conditioned on its observation. This expression is also known as the posterior distribution. The posterior distribution is obtained in two stages: prediction and estimation.

The prediction stage computes a priori prediction of the target's current state known as the prior distribution. The estimation stage then computes the posterior distribution by using the prediction with the current measurement of the sensor.

Exact solutions to this recursive propagation of the posterior density do exist for a restricted set of cases. In particular, the Kalman filter (Kalman, 1960; Welch and Bishop, 2006) is an optimal solution when the measurement and state transition model are linear with additive gaussian noise. But in general, solutions cannot be determined analytically, and hence an approximate solution has to be computed.

In the case of the BOF, the state of the system is given by the occupancy state and velocity of each cell of the grid, and the required conditions for being able to apply an exact solution such as the Kalman filter are not always verified. In addition, the particular structure of the model (grids) and the real-time constraint coming from most practical robotic applications leads to the development of the BOF.

The application of the Bayes filter to estimate the distribution of grid occupancy and velocities makes it possible to take past sensor observations into account. This is required to make robust estimations in changing environments (i.e. in order to be able to process temporary objects occlusions and detection problems).

It is also possible to focus the computation on the most probable velocities of grids instead of updating the grid occupancy values for every possible velocity in the standard BOF approach. Such an approach is not only more efficient computationally, but provides a more theoretically sound and systematic way of estimating grid velocities.

3.2 Bayesian model

The Bayesian model of the BOF is specified by first defining the joint distribution of all the relevant variables. The joint distribution is then decomposed into several subexpressions by assumption of independence between the variables. In the case of the BOF, most of the components of the decomposition are realised in the form of histograms. The BOF can be formulated mathematically as follows.

3.2.1 Probabilistic variable definitions

All the probabilistic variables below are defined within the context of a single cell c of the grid. This subscript is now omitted everywhere to maintain simplicity, except where ambiguity is possible.

- C is an index that identify each 2D cell of the grid.
- A is an index that identify each possible antecedent of the cell c over all the cells in the 2D grid.
- $Z_t \in \mathcal{Z}$ where Z_t is the random variable of the sensor measurement relative to the cell c .
- $V \in \mathcal{V} = \{v_1, \dots, v_n\}$ where V is the random variable of the velocities for the cell c and its possible values are discretised in n cases.
- $O, O^{-1} \in \mathcal{O} \equiv \{occ, emp\}$ where O represents the random variable of the state of c being either ‘occupied’ or ‘empty’. O^{-1} represents the random variable of the state of an antecedent cell of c through the possible motion through c . For a given velocity $v_k = (v_x, v_y)$ and a given time step δt , it is possible to define an antecedent for $c = (x, y)$ as $c^{-k} = (x - v_x \delta t, y - v_y \delta t)$.

3.2.2 Joint distributions

The following expression gives the decomposition of joint distribution of all the relevant variables according to Bayes’ rule and dependency assumptions.

$$\begin{aligned} P(C, A, Z, O, O^{-1}, V) \\ = P(A)P(V|A)P(C|V, A)P(O^{-1}|A)P(O|O^{-1})P(Z|O, V, C) \end{aligned} \quad (1)$$

Each subexpression of the joint distribution decomposition can be mapped to a semantic. The semantics of each distribution in the decomposition are interpreted as follows:

- $P(A)$ is the distribution over all the possible antecedent of the cell c . It is chosen to be uniform because the cell is considered reachable from all the antecedents with equal probability.
- $P(V|A)$ is the distribution over all the possible velocities of a certain antecedent of the cell c , its parametric form is in the form of a histogram.
- $P(C|V, A)$ is a distribution that explains if c is reachable from $[A = a]$ with the velocity $[V = v]$. In discrete spaces, this distribution is a dirac with value equal to one if and only if $c_x = a_x + v_x \delta t$ and $c_y = a_y + v_y \delta t$ which follows a dynamic model of constant velocity.
- $P(O^{-1}|A)$ is the conditional distribution over the occupancy of the antecedents. It gives the probability of the possible previous step of the current cell.
- $P(O|O^{-1})$ is the conditional distribution over the occupancy of the current cell, which depends on the occupancy state of the previous cell. It is defined as a transition matrix: $T = \begin{bmatrix} 1 - \epsilon & \epsilon \\ \epsilon & 1 - \epsilon \end{bmatrix}$, which allows the system to take in to account the fact that the null acceleration hypothesis is an approximation; in this matrix, ϵ is a

parameter representing the probability that the object in c does not follow the null acceleration model.

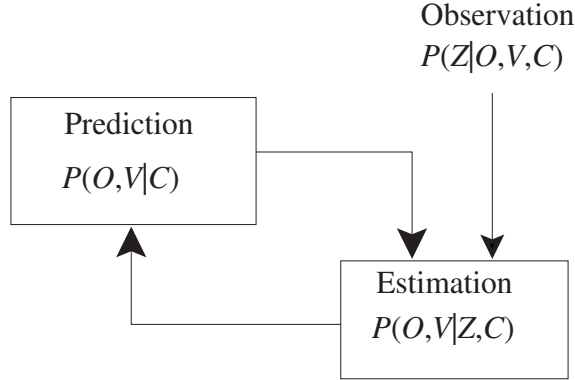
- $P(Z|O, V, C)$ is the conditional distribution over the sensor measurement values. It depends of the state of the cell, the velocity of the cell and obviously the position of the cell.

3.2.3 Filtering computation and representation

The aim of filtering in the BOF grid is to estimate the occupancy and grid velocity distributions for each cell of the grid, $P(O, V|Z, C)$.

Figure 1 shows how Bayesian filtering is performed in the BOF grids. Bayesian filtering consists of two stages, prediction and estimation, which are performed for each iteration. In the context of the BOF, prediction propagates cell occupation probabilities for each velocity and cell in the BOF grid ($P(O, V|C)$). During estimation, $P(O, V|C)$ is updated by taking into account its observation $P(Z|O, V, C)$ to obtain its final Bayesian filter estimation $P(O, V|Z, C)$. The result from the Bayesian filter estimation will then be used for prediction in the next iteration.

Figure 1 Bayesian filtering in the estimation of occupancy and velocity distribution in the BOF grids



From the implementation point of view, the set of possible velocities are discretised. One way of implementing the computation of the probability distribution is in the form of histograms. The following equations displayed are based on the discrete case. Therefore, the global filtering equation can be obtained by:

$$P(V, O|Z, C) = \frac{\sum_{A, O^{-1}} P(C, A, Z, O, O^{-1}, V)}{\sum_{A, O, O^{-1}, V} P(C, A, Z, O, O^{-1}, V)} \quad (2)$$

This can be equivalently represented as:

$$P(V, O, Z, C) = P(Z|O, V, C) \left[\sum_{A, O^{-1}} P(A)P(V|A)P(C|V, A)P(O^{-1}|A)P(O|O^{-1}) \right]$$

The summation in the above expression represents the prediction and its multiplication with the first term, $P(Z|O, V, C)$, gives the Bayesian filter estimation.

The global filtering equation (Equation 2) can actually be separated into three stages. The first stage being the prediction of the probability measure for each occupancy and velocity:

$$\begin{aligned}\alpha(\text{occ}, v_k) &= \sum_{A, O^{-1}} P(A)P(v_k|A)P(C|V, A)P(O^{-1}|A)P(\text{occ}|O^{-1}) \\ \alpha(\text{emp}, v_k) &= \sum_{A, O^{-1}} P(A)P(v_k|A)P(C|V, A)P(O^{-1}|A)P(\text{emp}|O^{-1})\end{aligned}\quad (3)$$

Equation 3 is performed for each cell in the grid, and for each velocity. Prediction for each cell is calculated by taking into account the velocity probability and occupation probability of the set of antecedent cells. The set of antecedent cells are cells with a velocity that will propagate itself in a certain time step to the current cell in question.

With the prediction of the grid occupancy and its velocities, the second stage consists of multiplying by its observation sensor model which gives the Bayesian filter estimation on occupation and velocity distribution which are not normalised:

$$\begin{aligned}\beta(\text{occ}, v_k) &= P(Z|\text{occ}, v_k)\alpha(\text{occ}, v_k) \\ \beta(\text{emp}, v_k) &= P(Z|\text{emp}, v_k)\alpha(\text{emp}, v_k)\end{aligned}$$

Similar to the prediction stage, Equation 4 is performed for each cell occupancy and each velocity. The marginalisation over the occupancy values gives the likelihood of a certain velocity:

$$l(v_k) = \beta(\text{occ}, v_k) + \beta(\text{emp}, v_k)$$

Finally the normalised Bayesian filter estimation on probability of occupancy for a cell C with a velocity v_k is obtained by:

$$P(\text{occ}, v_k|Z, C) = \frac{\beta(\text{occ}, v_k)}{\sum_{v_k} l(v_k)} \quad (4)$$

The occupancy distribution in a cell can be obtained by the marginalisation over the velocities and the velocity distribution by the marginalisation over the occupancy values:

$$P(O|Z, C) = \sum_V P(V, O|Z, C) \quad (5)$$

$$P(V|Z, C) = \sum_O P(V, O|Z, C) \quad (6)$$

4 Velocity discretisation

Discretisation is a recurring problem in different domains. In the case of a grid-based decomposition of the environment, the occupancy of a cell is considered to be the same for the entire cell, and for some arbitrary displacement of an occupied cell, it is highly likely that the final position does not fall exactly into the geometric confines of another cell¹.

The consequence of permitting such a displacement introduces errors in the prediction step. One way to overcome this error is to assign more believe in the observations. However, this leads to a filter sensitive to false detections and decreases the quality of the filter. Another way would be to assign occupancy probabilities proportionally to the area of overlap with the grid cells. The disadvantage of doing so will be to incur extra geometric computation.

An alternative prediction scheme that we propose for the steps in Equations (3) and (4) is to take the velocity of a cell into account. The key idea is to choose the velocity of a cell, during the prediction, such that it corresponds to a displacement of an exact integer number of cells in the grid during the discretisation of velocities. It strongly depends on the time step dt of filtering updates. We consider dt as a constant in this paper. The consequence is that the grid is regular and we choose a cartesian grid with a constant step size for each dimension: dx and dy in 2D.

Thus the set of possible velocities is:

$$\mathcal{V} = \left\{ \left(\frac{p}{n} \frac{dx}{dt}; \frac{q}{n} \frac{dy}{dt} \right) \mid (p, q, n) \in \mathbb{Z}^2 \times \mathbb{N}^* \right\}$$

where $1/n$ allows the consideration of movements of a cell that requires more than one time step to reach another cell totally. For example, a translation of dx which requires two time steps corresponds to the velocity vector: $(dx/2dt, 0)$ and the corresponding (p, q, n) is $(1, 0, 2)$.

4.1 Consequences for observation

The velocity plane with (p, q, n) velocity is thus observed with a frequency of $1/n$. The consequence is that all velocity planes are not checked at each time step, in particular slow velocity planes are observed less frequently. Indeed, slow moving objects required less attention than fast moving objects because their location in space change less between two time steps. This asynchronous updating scheme reflects an intuitive principle of attention allocation. This form of attention allocation leads to better allocation of computing resources.

The updating process gathers all velocity planes that share the same frequency. Then for each time step, only groups of velocity planes with a frequency that corresponds to the current time step are updated.

5 Object representations

There are often times where object level representations are required. The philosophy of the BOF is to delay the problem of data association and as a result does not contain information with respect to objects. A natural approach to obtain an object level representation from the BOF grids is to introduce grid-based clustering to extract object hypothesis and an object level tracker to handle the hypothesis extracted.

This section shows how this object level representation can be obtained based on this generalised clustering tracking methodology by illustrating this methodology with a simple clustering and tracking procedure.

5.1 Obtaining object hypotheses

Object hypotheses can be obtained by performing clustering on the BOF. A simple and naive clustering method would be to connect cells within its four or eight neighbourhood that has a sufficiently high occupancy probability. Such clustering methods, although simple, is fast and gives satisfactory results when coupled with an appropriate sensor model. However, it is not difficult to see that more sophisticated clustering algorithms can be constructed by taking occupancy values and velocity distributions into account.

5.2 Managing object hypotheses

Recall that one of the principal advantages of the BOF is to perform sensor fusion in the grids and inferring velocities for each cell in the BOF grid with relative ease. The data association problem is thus conveniently avoided in this level, which can be complicated further during sensor fusion.

However, with the set of object hypotheses $O_t = \{O_{1,t}, \dots, O_{N,t}\}$ at time t , it is now imperative to deal with the problem of data association since the notion of an object is now involved. This illustrates the point of delaying the data association to a stage as late as possible. Data association serves to associate object hypotheses O_{t+1} with object hypotheses at the previous time step O_t . The data association problem can be resolved using classical algorithms such as the Joint Probabilistic Data Association (JPDA) algorithm.

JPDA is a well-established data association algorithm that has gained popularity in various tracking applications, including human tracking (Schulz et al., 2003), aircraft tracking (Hwang et al., 2004), and visual tracking (Rasmussen and Hager, 2001). Its strength lies in its capability to jointly calculate the likelihood of association hypotheses, which precludes potentially erroneous assignment frequently obtained by greedy association algorithms like the Nearest Neighbourhood (NN).

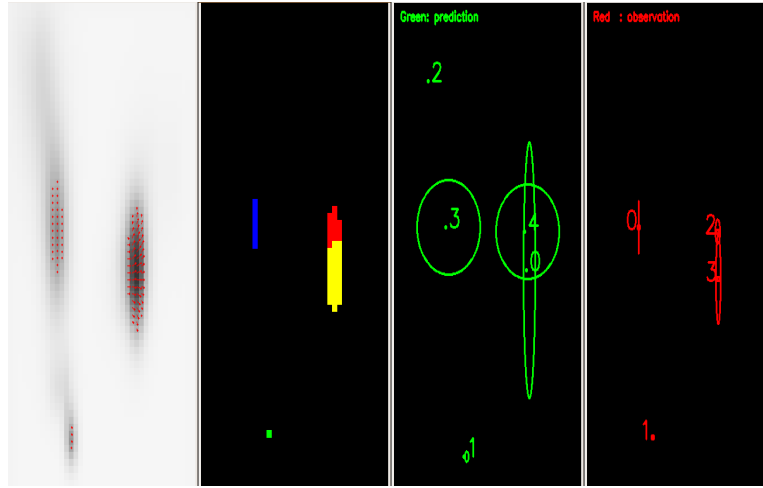
A screen shot from the running programme can be found in Figure 2, which depicts a typical data processing pipeline. In the first step, sensor inputs (image bounding boxes) are projected into the occupancy gridmap and fused together. The BOF then estimates the occupancy state of the surveillance region and the corresponding velocity. Thereafter, the BOF grid is processed by a simple but efficient clustering algorithm. Disconnected occupied regions are extracted from the BOF grid and deemed as observed target reports. These reports are further passed to the JPDA, where they are associated with existing tracks by a calculated likelihood. With such probabilistic association, the object-level representation of targets can be recursively updated over time (Table 1).

5.3 Track management

The generic JPDA algorithm assumes that the number of targets (tracks) are given, which is not the case in most real-life tracking scenarios. Therefore, mechanisms for track initialisation, merging and deleting must be further developed to support the core JPDA algorithm. These mechanisms are referred as ‘track management’ in this context. In the work of Schulz et al. (2003), the track number is calculated using a Bayesian filter, based on the number in previous iterations, and probabilistic models learned through offline training. This Bayesian filter tries to infer the number of target in current time instance. Although it has demonstrated some promising results, we noticed that its performance degrades considerably in the scenario when target number drastically changes within short period, which is often the case in car tracking. Therefore, in this work, we do not assume the dependency among the target number in different time instances. At each iteration, new tracks are introduced and some old tracks are merged or deleted. The finalised tracks are assumed to be true and used for JPDA in the next iteration.

5.3.1 Track initialisation

In JPDA, each report is associated with different tracks by a specific likelihood. Here these likelihoods are summed up and denoted as a , and a given threshold is used to examine

Figure 2 An illustration of a tracking instance (for colours see online version)

Note: From left to right, the BOF output, clustering, prediction and lastly observation. Predictions and observations covariances are shown in ellipses and are indexed. The association likelihoods for each track can be found in Table 1.

Table 1 The likelihood calculated by JPDA for report-to-track association corresponding to the situation in Figure 2

	<i>Report 0</i>	<i>Report 1</i>	<i>Report 2</i>	<i>Report 3</i>	<i>No-assoc.</i>
Track 0	0	0	0.790	0.153	0.056
Track 1	0	0	0	0	1
Track 2	0	0	0	0	1
Track 3	0.954	0	0	0	0.045
Track 4	0	0	0.145	0.796	0.058

whether a is sufficiently large. If not, the corresponding report is deemed as ‘unassociated’ and then a new track is initialised based on this report.

5.3.2 Track termination

In the tracking applications, it is common to set a ‘candidature flag’ for each new track, so that those not-so-confirmed tracks will not confuse the upper level modules that use the output of data association. However, as has been mentioned, in the vehicle tracking scenario, new targets come into and then leave the surveillance region within a short period. Consequently, too long a candidature period and often surplus the existence time of a target. For this reason, in this work, a very short confirmation period is employed, that is, two iterations. For each established track, that is, older than two iterations, its log likelihood (Bar-Shalov and Fortman, 1988) is calculated using a ‘memory window’ at the size of three iterations. Those tracks whose likelihoods are lower than a predefined threshold will be deemed as obsolete and eliminated from the track list.

5.3.3 Track merging

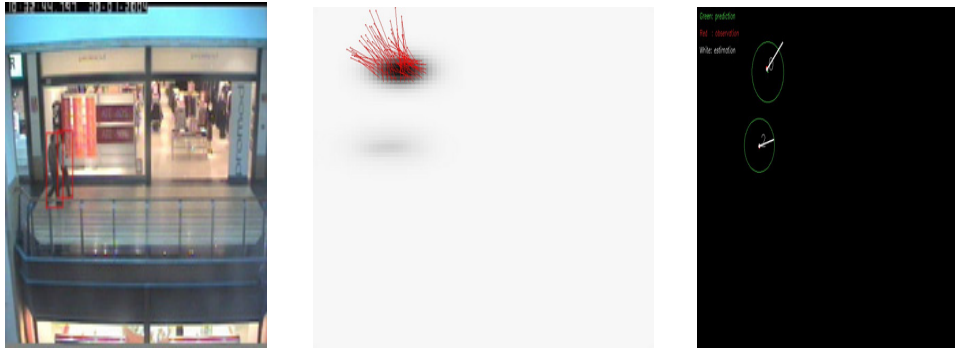
A very common problem of object detection in computer vision is that one single object in the image can be detected for more than one time in by the search algorithm. This is because a single object and several objects compactly placed often have very similar appearance. In the tracking context, this problem leads to the phenomenon that one single target is represented by several tracks that are compactly located and show very similar manoeuvring.

A track merging strategy is employed to handle such situation. At each time instance, for each track, its states within a certain time window (e.g. three iterations) are vectorised, and the χ^2 test is therefore a natural way to compare the similarity between two tracks. If the χ^2 test reports that the two tracks are too ‘dependent’, it is reasonable to assert that they actually represent the same object and should be merged into one track. For the merging, we used a very simple strategy which is to keep the ‘oldest’ one among the similar tracks. Although a more sophisticated fusion scheme may be more robust, we observed that this method is already sufficiently good.

6 Experiments

Experiments were conducted based on video sequence data from the European project CAVIAR. The selected video sequence presented in this paper is taken from the interior of a shopping centre in Portugal. An example is shown in the first column of the series of Figures (3–7). The data sequence from CAVIAR, which is freely available from the web² gives annotated ground truths for the detection of the pedestrians. In general, two different data sets can be found, one set taken from the entry hall of INRIA Rhône Alpes and the other from a shopping centre in Portugal.

Figure 3 CAVIAR tracking sequence involving two persons moving from left to right



Based on this given data, the uncertainties, false positives and occlusions have been simulated. This simulated data is then used as observations for the BOF. The BOF is a representation of the planar ground of the shopping centre within the field of view of the camera. With the noise and occlusion simulated bounding boxes which represents human detections, a gaussian sensor model is used. The gaussian sensor model gives a gaussian occupation uncertainty (in the BOF grids) of the lower edge of the image bounding box after being projected onto the ground plane.

Figure 4 Simulations of misdetections and occlusions

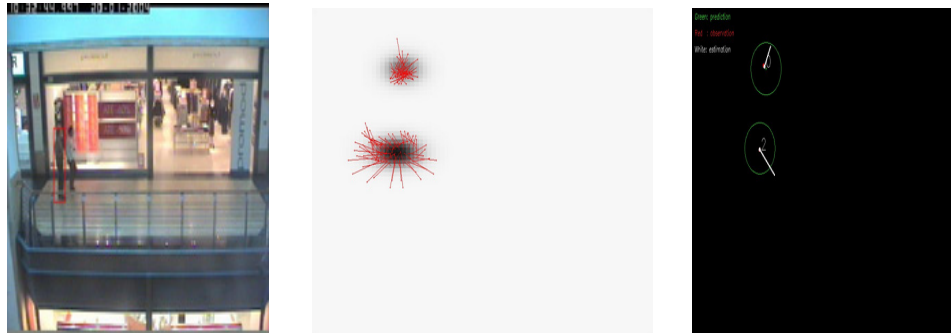


Figure 5 Another person enters the scene from the left. The person in the middle remains in position while the person on the right continues moving towards the right



Figure 6 An occlusion takes place when a person blocks the view of another

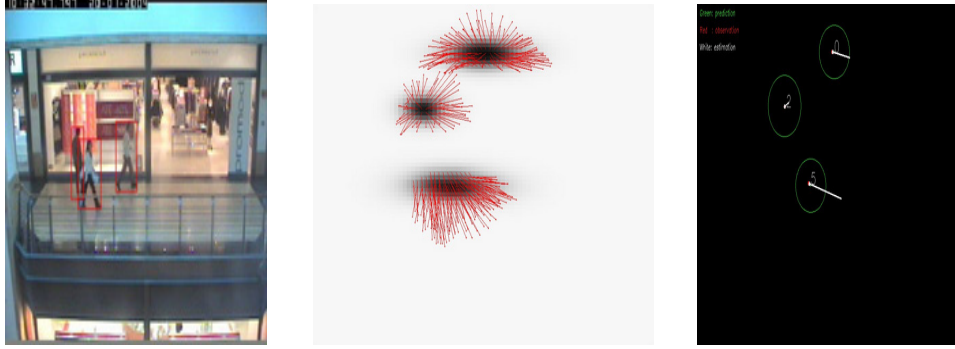


Recalling that there is no notion of objects in the BOF, object hypotheses are obtained from clustering and these object hypotheses are used as observations on a standard tracking module based on the JPDA.

The implementation of the tracker is in the C++ programming language without optimisations. Experiments were performed on a laptop with an Intel Centrino processor

with a clock speed of 1.6 GHz. It currently tracks with an average frame rate of 9.27 frames/sec. The computation time required for BOF, with a grid resolution of 80 cells 80 cells takes an average of 0.05 secs. The BOF represents the ground plane of the image sequence taking from a stationary camera, and represents a dimension of 30 metres by 20 metres.

Figure 7 The person previously occluded in Figure 6 is no longer occluded



From the equations in Section 3.2, BOF-based tracking on a moving camera or stationary camera does not take into account sensor movement and hence is the same. In the case of the Coué et al. (2006), the tracking of the BOF was on a moving laser sensor, and the tracking is done relative to the sensor frame. The same algorithm applies in the case of stationary or moving sensor, as tracking is performed in the sensor frame. We consider tracking with respect to the world frame, which involves localisation or SLAM, a different problem and is beyond the scope of the paper. The implementation issue involves taking into account the range of possible velocities on a moving platform being potentially more varied than that of a stationary platform.

The results shown in Figures 3–7 is shown in its time sequence. The first column of the figures shows the input image with the bounding boxes, each indicating the detection of a human, after the simulation of uncertainties and occlusions. The second column shows the corresponding visualisation of the BOF. The coloured intensity of the cells represents the occupation probability of the cell proportionally. The little arrows in a cell give the average velocity calculated from the velocity distribution of the cell. The third column gives the tracker output given by the JPDA tracker with the numbers indicating its track number. The sensor model used is a 2D planar gaussian model projected onto the ground. Its coordinates are given by the lower edge of the bounding box.

The characteristics of the BOF can be inferred from the Figures 3–7. A diminished occupancy of the person further away from the camera is seen from the data in Figures 3 and 4. This is caused by the simulated occasional instability in human detection. The occupancy in the BOF grids for the missed detection diminishes gradually with time rather than disappearing immediately for classical occupation grids. This mechanism provides a form of temporal smoothing to handle unstable detections.

A more challenging occlusion sequence is shown in Figures 5 to 7. Due to a relatively longer period of occlusion, the occupancy probability of the occluded person becomes weak. However, with an appropriately designed tracker, such problems can be handled at the object tracker level. The tracker manages to track the occlusion at the object tracker level as shown in the last column of Figure 7.

7 Conclusion

This paper presented a grid-based Bayesian formulation for perception and modelling of dynamic environments known as the BOF. Without the notion of objects in the BOF grids, the grid-based framework facilitates ease of sensor fusion. Another resulting advantage is the avoidance of the data association problem at the inter scan data association stage and during the sensor fusion stage.

There are cases where applications require an object level abstraction. This paper showed that it is possible to do so, using only information from the BOF grids, by using the general clustering tracking approach. A simple clustering and JPDA tracking technique was applied to experimental data from the European project CAVIAR and results were presented.

Further work involves more sophisticated and reliable clustering algorithms. Especially for cases where targets are very close to one another and form connected regions in the BOF grid. The challenge will be to cluster the connected regions in a coherent manner such that the clusters will represent the individual targets correctly. One way of doing so is to involve the tracking module as a form of feedback to the clustering module, which helps the clustering module to intelligently identify the number and positions of potential clusters.

We are also working on adding richer sensory data to complement the BOF. When working especially with camera data, imagery information is a valuable resource for data association, not only at the grid level, but on the tracker level as well. Doing so leads to more robust data association.

References

- Arras, K.O., Tomatis, N. and Siegwart, R. (2001) 'Multisensor on-the-fly localization : precision and reliability for applications', *Robotics and Autonomous Systems*, Vol. 44, pp.131–143.
- Bar-Shalom, Y. and Fortman, T.E. (1988) *Tracking and Data Association*, Academic Press.
- Bar-Shalom, Y. and Li, X. (1995) *Multitarget Multisensor Tracking : Principles and Techniques*, YBS Publishing.
- Blackman, S. and Popoli, R. (2000) *Design and Analysis of Modern Tracking Systems*, Artech House.
- Coué, C., Fraichard, T. Bessié, P. and Mazer, E. (2002) 'Multi-sensor data fusion using bayesian programming: an automotive application', *Proceedings of the IEEE-RSJ International Conferences on Intelligent Robots and Systems*, Lausanne, (CH), Octobre.
- Coué C., Fraichard, T., Bessié, P. and Mazer, E. (2003) 'Using bayesian programming for multi-sensor multi-target tracking in automotive applications', *Proceedings of IEEE International Conference on Robotics and Automation*, Taipei (TW), Septembre.
- Coué, C., Pradalier, C., Laugier, C., Fraichard, T. and Bessié, P. (2006) 'Bayesian occupancy filtering for multitarget tracking: an automotive application', *International Journal of Robotics Research*, Vol. 25, No. 1, pp.19–30.
- Elfes, A. 1989 'Using occupancy grids for mobile robot perception and navigation', *IEEE Computer, Special Issue on Autonomous Intelligent Machines*, Juin.
- Gauvrit, H., Le Cadre, J.P. and Jauffret, C. (1997) 'A formulation of multitarget tracking as an incomplete data problem', *IEEE Transactions on Aerospace and Electronic Systems*, Vol. 33, No. 4.
- Hwang, I., Balakrishanan, H., Roy, K. and Tomlin, C. (2004) 'Multiple-target tracking and identity management in clutter, with application to aircraft tracking', *Proceedings the IEEE American Control Conference*, pp.3422–3428, 30 June–2 July.
- Jazwinsky, A.H. (1970) *Stochastic Processes and Filtering Theory*, New York : Academic Press.

- Kaelbling, L.P., Littman, M.L. and Cassandra, A.R. (1998) 'Planning and acting in partially observable stochastic domains', *Artificial Intelligence*, p.101.
- Kalman, R.E. (1960) 'A new approach to linear filtering and prediction problems', *Journal of basic Engineering*, p.35 March.
- Mekhnacha, K., Mazer, E. and Bessiere, P. (2001) 'The design and implementation of a bayesian CAD modeler for robotic applications', *Advanced Robotics*, Vol. 15, No. 1, pp.45–70.
- Moravec, H.P. (1988) 'Sensor fusion in certainty grids for mobile robots', *AI Magazine*, Vol. 9, No. 2.
- Prassler, E. Scholz, J. and Elfes, A. (2000) 'Tracking multiple moving objects for real-time robot navigation', *Autonomous Robots*, Vol. 8, No. 2.
- Rasmussen, C. and Hager, G.D. (2001) 'Probabilistic data association methods for tracking complex visual objects', *IEEE Transactions Pattern analysis and machine intelligence*, Vol. 23, No. 6, pp.560–576.
- Schulz, D., Burgard, W., Fox, D. and Cremers, A.B. (2003) 'People tracking with a mobile robot using sample-based Joint Probabilistic Data Association Filters', *International Journal of Robotics Research*, Vol. 22, No. 2, pp.99–116.
- Streit, R.L. and Luginbuhl, T.E. (1995) 'Probabilistic multi-hypothesis tracking', *Technical Report 10,428, Naval Undersea Warfare Center Division Newport*.
- Thrun, S. (1998) 'Learning metric-topological maps for indoor mobile robot navigation', *Artificial Intelligence*, Vol. 99, No. 1.
- Thrun, S. (2002) 'Robotic mapping: a survey', *Exploring Artificial Intelligence in the New Millenium*, Morgan Kaufmann.
- Wang, C-C., Thorpes, C. and Thrun, S. (2003) 'Online simultaneous localization and mapping with detection and tracking of moving objects: theory and results from a ground vehicle in crowded urban areas', *Proceedings of the IEEE International Conferences on Robotics and Automation*, pp.842–849, Taipei (taiwan), September.
- Welchand, G. and Bishop, G. (2006) 'An introduction to the kalman filter', *Technical Report TR 95-041*, Department of Computer Science University of North Carolina at Chapel Hill.

Notes

¹This problem is also known as the aliasing problem.

²Available at: <http://groups.inf.ed.ac.uk/vision/CAVIAR/CAVIARDATA1/>.