

RESEARCH

Open Access

# Pedestrian tracking with an infrared sensor using road network information

Per Skoglar<sup>1,2\*</sup>, Umut Orguner<sup>1</sup>, David Törnqvist<sup>1</sup> and Fredrik Gustafsson<sup>1</sup>

## Abstract

This article presents a pedestrian tracking methodology using an infrared sensor for surveillance applications. A distinctive feature of this study compared to the existing pedestrian tracking approaches is that the road network information is utilized for performance enhancement. A multiple model particle filter, which uses two different motion models, is designed for enabling the tracking of both road-constrained (on-road) and unconstrained (off-road) targets. The lateral position of the pedestrians on the walkways are taken into account by a specific on-road target model. The overall framework seamlessly integrates the negative information of occlusion events into the algorithm for which the required modifications are discussed. The resulting algorithm is illustrated on real data from a field trial for different scenarios.

**Keywords:** pedestrian tracking, infrared sensor, road network, particle filter, multiple model, occlusion

## 1 Introduction

Accurate pedestrian tracking and anomaly detection are important hot topics in surveillance applications in the security area (see the surveys [1,2] and the special issue [3]), where currently the demands on the operator are very high. Further, the tracking algorithms integrated in the sensors have the potential to solve some of the integrity problems currently associated with video surveillance. In order to obtain efficient solutions, in terms of both performance and cost, there is a need for automatic processing and analysis of imagery. Multiple pedestrian tracking is a very challenging task due to clutter, occlusion, etc. The exploitation of contextual information, such as maps and terrain information, is therefore highly desirable not only for the enhancement of the tracking performance, but also for behavior analysis and anomaly detection.

This article presents a sensor system with an infrared camera and sophisticated algorithms for pedestrian detection and tracking. The focus is here on the tracking part rather than the detector which is a classifier that is trained using a variant of boosting. The multiple pedestrian tracker is proposed to be a multiple-model

particle filter that uses prior information about walkways to enhance the estimation performance. The state of the art multiple model particle filters are used with two different models, namely, an on-road (road-constrained) model and an off-road (unconstrained) model to perform tracking in 3D global coordinates. The proposed algorithms are applied to real-world imagery data where a number of pedestrians are walking around in a park-like environment.

The related literature is vast and spans the areas of research related to several academic communities. For this reason, we defer a more comprehensive survey until Section 2.4 and summarize below just the main contributions of this study compared to the existing literature.

1. The use of the road network for pedestrian tracking, enabling multiple model approaches is novel. This has, to the best of the authors' knowledge, not been presented in literature before.

2. The road network information use in target tracking have indeed been proposed earlier for road vehicles observed by a radar sensor, typically GMTI (ground moving target indicator). Compared to the state of the art GMTI based approaches, the following distinct properties of the pedestrian tracking make our study a significant contribution to road-constrained tracking literature (see [4] and the references therein):

\* Correspondence: skoglar@isy.liu.se

<sup>1</sup>Division of Automatic Control, Department of Electrical Engineering, Linköping University, SE-581 83 Linköping, Sweden  
Full list of author information is available at the end of the article

- Better angular resolution of the sensors (compared to radar) enables tracking the lateral position on the road.
- Pedestrians move much more freely and independently than cars, so the algorithm cannot rely on the motion model to the same extent.
- Switches between on-road and off-road modes occur more frequently, increasing the need for robust mode tracking.

### 3. The multiple model framework with on-road and off-road modes

- gives better tracking performance, independently of which state of the art algorithm that is used (MMPF or IMMPPF);
- provides improved predictions during occlusion by using the concept of negative information;
- serves well for planning the pan/tilt/zoom of the camera via improved predictions;
- includes statistical tools that can be used to calculate the switching times, frequency, corresponding positions, and correlation for such events between different pedestrians which makes it possible to learn what is normal behavior. This is in fact a technical enabler for future anomaly detection algorithms.

4. Although the road network information has been used in GMTI based target tracking before, the number of examples in which real world experiments were performed is very few. Our algorithm presented in this study is applied to a real world data set and the resulting estimates are compared to GPS data which answers some fundamental questions as to what the achievable accuracy in this type of application would be.

We finalize this section with a brief outline of the remaining parts of the article as follows. Section 2 introduces the elements of the surveillance problem considered in this article such as surveillance environment, prior knowledge, and sensor system. In particular, a global overview of the multiple pedestrian motion models is given, and the pedestrian image detector is described. The section ends with a literature survey of the related research. Section 3 gives a brief introduction to estimation theory and multiple target tracking from a particle filter perspective. In Section 4 the specific models of on/off-road pedestrian motion and the infrared sensor are described in detail and the proposed multiple model pedestrian tracking particle filter is presented. The filter is applied to a real-world data set and the results are illustrated in Section 5. Finally, in Section 6 some conclusions are drawn along with the discussion of the results.

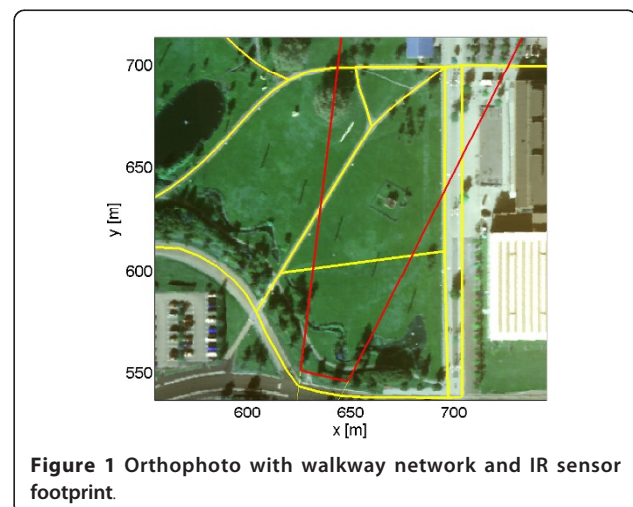
## 2 Problem description

We consider a surveillance scenario where a sensor system with an infrared camera is monitoring a certain area with a number of known walkways. Detected pedestrians must be tracked simultaneously. The detector and tracking modules would be an essential part of (semi-)autonomous surveillance systems corresponding to the autonomous unmanned aerial vehicle (UAV) framework presented in [5,6] where also sensor management is an important part. The sensor management controls the movement of the sensor platform and the pointing direction of the pan/tilt infrared camera such that the performance of the tracking and monitoring is as good as possible.

One major tool for providing a “situation awareness” of the scene is to estimate interesting states of the environment. These states can have very different properties, depending on the mission and the user requirements, but in this study the position, velocity, etc., of the pedestrians are important. In order to improve the tracking performance prior knowledge about the walkway network will facilitate the estimation process.

### 2.1 Multiple pedestrian motion models

The walkway network is available for a park like environment, see the orthophoto with the network overlaid in Figure 1. An infrared sensor is located south of the area pointing upwards, the approximate sensor footprint on the ground is also shown. One image frame is shown in Figure 2 with the walkway network projected onto the image. We will use the symbolic notation  $\mathcal{I}_{RN}$  to denote the road network information. (The terms *road* and *walkway* are used interchangeably in this article. The terms *pedestrian* and *target* are also used interchangeably.)





**Figure 2** An IR image frame with the projected walkway network

Suppose we would like to track pedestrians which can move both on-road and off-road. We consider two different state space representations corresponding to on-road and off-road target modes,

$$x_{t+1}^r = f^r(x_t^r, \mathcal{I}_{RN}, \eta_t^r, v_t^r), \quad (1)$$

$$x_{t+1}^s = f^s(x_t^s, \eta_t^s), \quad (2)$$

where the vectors  $x_t^r \in \mathbb{R}^{n_x}$  and  $x_t^s \in \mathbb{R}^{n_x}$  represent the state vectors of the target in on-road and off-road (global) coordinates, respectively. The functions  $f(\cdot)$  and  $f^s(\cdot)$  are in general nonlinear functions. The process noise terms  $\eta_t^r \in \mathbb{R}^{n_x}$  and  $\eta_t^s \in \mathbb{R}^{n_x}$  are assumed to be white. The process noise  $v_{t+1}^r \in \{1, 2, \dots, N_r(x_t^r)\}$  determines which road segment the target will follow in the next sampling interval in case more than one alternative exists. We assume the availability of prior probability density functions (or probability mass functions in the discrete case)  $p_{\eta_t^r}(\cdot)$ ,  $p_{\eta_t^s}(\cdot)$ , and  $p_{v_t^r}(\cdot)$  for the random variables  $\eta_t^r$ ,  $\eta_t^s$ , and  $v_t^r$ , respectively.

In order to be able to use both models at the same time, one always needs the appropriate functions to convert the state vectors given in one of the representations into the other representation. For this purpose we assume the availability of two transformation functions named  $T^{sr}(\cdot)$  (transformation from road coordinates to global coordinates) and  $T^{rg}(\cdot)$  (transformation from global coordinates to road coordinates).

The measurements associated with the target are modeled according to relations

$$y_t = h^s(x_t^s) + e_t^s, \quad (3)$$

where  $h^s(\cdot)$  is in general a nonlinear function of the global state of the target and  $e_t^s$  is white measurement noise. We assume that the probability density function  $p_{e_t^s}(\cdot)$  is available. Note that with this notation, the

measurements related to on-road coordinates of the target can be written to satisfy

$$y_t = h^s(T^{sr}(x_{t-1}^r, \mathcal{I}_{RN})) + e_t^s. \quad (4)$$

The hypothesis (event) that the target is moving on-road or off-road is modeled by a discrete variable  $q_t \in \{1, 2\}$  where the events  $\{q_t = 1\}$  and  $\{q_t = 2\}$  correspond to the hypotheses that the target is on-road and off-road, respectively. According to the value of the variable  $q_t$  the corresponding dynamics of the target given in (1) and (2) must be used. It is assumed that  $q_t$  is a homogeneous possibly state dependent Markov chain with transition probability matrix denoted as  $\Pi = [\pi_{ij}]$  where

$$\pi_{ij} \triangleq P(q_t = j | q_{t-1} = i, x_t^s, \mathcal{I}_{RN}). \quad (5)$$

This modeling framework where the underlying dynamics of the target evolves based on a Markov chain belongs to the class of so called jump Markov nonlinear systems in the literature (see [7] and the references therein).

## 2.2 Infrared sensor system

The experimental sensor system consists of a gyro-stabilized gimbal with IR and CCD video sensors, and an integrated high-performance navigation system. The navigation system combines GPS with data from an inertial measurement unit (IMU) mounted with reference to the optical sensors. However, in the experiments presented in this article external landmarks with known location have also been used to estimate the orientation of the camera relative the world frame by using standard camera calibration techniques [8].

The IR sensor in the gimbal is a FLIR systems ThermaCAM SC3000, which is a long-wave infrared (LWIR) sensor with a quantum well infrared photodetector (QWIP) focal plane array. It has a low noise equivalent temperature difference (NETD) of 30 mK. The detector array is composed of  $320 \times 240$  pixels with a comparatively narrow spectral sensitivity of 8.0 - 9.2  $\mu\text{m}$ , which corresponds to the wavelength peak of an equivalent black body radiator at 25°C. The digital output has a resolution of 14 bits/pixel and a frame rate of 50 Hz. The mounted optics has a field-of-view of  $20^\circ \times 15^\circ$  which gives a spatial angular resolution of 1.1 *mrad* per pixel.

## 2.3 Target detector

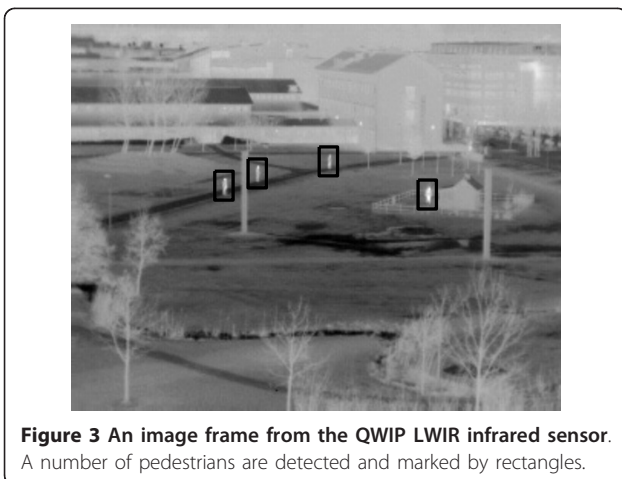
The detection problem is to find targets in cluttered backgrounds and the output from the detector is a set of image coordinates for all detections in each video frame. In this study a sliding window approach is used to detect pedestrians in cluttered backgrounds [9]. At each image position, the content of a local image region

is fed into a classifier that decides whether or not the region contains a target.

The classifier is trained using a variant of boosting [10]. Boosting iteratively builds a highly discriminative classifier by combining the outputs of many component functions often referred to as “weak learners”. Applying the resulting classifier to an image window  $x$ , the output can be written as  $F(x) = \sum_i f_i(x)$  and the window is classified as containing a target if the confidence sum  $F(x)$  is greater than a threshold that is set to achieve an acceptable false alarm rate. Viola and Jones [11] proposed a highly efficient cascade-structured detector architecture where each stage is a boosting classifier that is trained to reject a moderate fraction of the remaining background examples, while retaining a large fraction of the target examples. This leads to an exponential decay in the probability that a retained window belongs to the background class. Another important contribution by [11] is the design of weak learners that can be computed very efficiently.

In the Viola-Jones detection framework each weak learner bases its decision on the response of a single Haar-like image feature, which can be computed very efficiently using a so-called integral image representation. In addition to Haar-like features, our implementation also uses more discriminative (but computationally more expensive) gradient histogram features, similarly to Laptev [12]. We adopt the soft cascade detector architecture [13] which allows for efficient trade-off between accuracy and speed.

Figure 3 shows an infrared image frame with a number of pedestrian detections. The false alarm rate is very low, and persistent false alarms can easily be handled by the tracking filter, or ignored if the detection location is in unreasonable areas according to prior information of the buildings and environment. Non-persistent clutter is



handled by a suitable initiator logic that prevents the false alarms to give rise to new tracks.

## 2.4 Related research

Visual surveillance and crowd analysis in dynamic scenes with humans are very active research topics in computer vision [2,14]. The possible applications are numerous, and so are the number of publications in the area.

This article’s focus is on the object tracker part of the surveillance system, see [15] and the references therein for an overview. The study [16] is an early publication where a particle filter is used for visual contour tracking. In [17] a mixture particle filter and an Ada-boost detector is used to track multiple objects (hockey players) in a video stream. Visual tracking is often performed in the image plane with the benefit of keeping the state dimension low and avoiding the calibration of extrinsic camera parameters, i.e., the location and orientation of the camera relative a world reference frame. In this study, tracking is performed in global coordinates which simplifies the motion model of the target and also makes it easier to combine with other tracking systems and contextual knowledge about the environment. Tracking in global coordinates with a vision sensor is essentially equivalent to tracking with a bearings-only sensor which has been traditionally treated in the target tracking community, see [18, Chapter 6] and the references therein.

Association is a hard problem, especially with a single camera in crowded environments with occlusions. A hierarchical association approach is proposed in [19] to form the trajectories of the pedestrians. The method also contains an automatic scene structure estimator. The study [20] estimates the probabilities of the occupancy bins in the ground plane represented as a grid. The Viterbi algorithm is then used to estimate target trajectories in a sequence of frames. One common approach for handling occlusion is to use multiple views in order to be able to utilize the depth information. In [21] a planar homography constraint is used to locate the targets on the ground plane. Only the types of occlusion which are due to stationary and known objects like buildings and trees are considered in this study.

In a classic surveillance setup the vision sensors are stationary, but in recent years a number of pedestrian detection and tracking systems have been proposed for moving cameras in automotive applications, see e.g., [22]. The study [23] uses structure-from-motion to estimate the ground plane that supports the target tracking.

Target tracking with road network information requires methodologies which can keep the inherent multi-modality of the underlying probability densities.

The first attempts [24-26] used the jump-Markov (non) linear systems in combination with the interacting multiple model (IMM) algorithm [27,28] with extended Kalman filters (EKFs) as sub-blocks. Since the different road segments correspond to different modes in these IMM algorithms, there are too many of them to be considered at a single step of the multiple model filter. Hence, these algorithms applied the so-called variable structure interacting multiple model (VS-IMM) algorithm [29] which adds/removes modes into/from the filter when necessary.

Important alternatives to IMM based methods appear in [30],[18, Chapter 10] which propose variable structure multiple model particle filters (VS-MMPF) as an extension of the VS-IMM approaches. Since the particle filters can handle nonlinear and non-Gaussian models, the user has much more freedom than in VS-IMM modeling. The road constraints are handled using the concept of directional process noise. In [31] the roads are 3D curves represented by linear segments and the road network is represented as a graph with roads and intersections as the edges and nodes, respectively. The position and velocity along a single road are modeled by a standard linear Gauss-Markov model. The target can be masked both by the clutter notch of the sensor and by terrain obstacles. The results for a Gaussian sum filter (see also [32]) and a standard bootstrap particle filter approach are presented.

A considerable amount of research effort has been made in the literature for improving particle filter based methods in terms of both performance and computational efficiency. The so-called optimal proposals and Rao-Blackwellization have been utilized to produce more efficient particle filters. In this respect [33] proposes an unscented particle filter (UPF) in a GMTI context and it is shown that fewer particles are needed compared to VS-MMPF. Optimal proposal densities are also used in [34]. However, the use of them unfortunately requires the combinatorial enumeration of all the possible models and the road segments the target can use in the next sampling period which might, at the same time, be a computational bottleneck. The proposed filter is applied to a GMTI target tracking example and it also utilizes Rao-Blackwellization of the full kinematic state in order to minimize the number of particles, i.e., given the road segment the target is on, the whole kinematic target state is represented by a Gaussian density. A more recent example of the Rao-Blackwellized particle filter is given in [35] to solve the road target tracking problem with a bearings-only observation model. Compared to other Rao-Blackwellized and filter bank approaches [33,34], this study treats not only the road identity, but also the position along the road as a nonlinear state. This means that the probability densities

with multiple modes along a single road can be handled, and this is often the case in tracking applications with a vision sensor when buildings and vegetation are possibly occluding the road.

In the standard bootstrap version of the particle filter, the number of particles in each mode is determined by the posterior probability of that mode. In the case of some unexpected events, like a sudden on-road to off-road transition, particle degeneracy happens if the new mode has too few particles. There are already some alternatives in the literature proposed for establishing robustness against this phenomenon with road networks. An example using the VS-MMPF methodology is presented in [36] where user selected number of particles can be used in each mode of the filter by making use of the so-called “variable-mass” idea. Another important alternative is the interacting multiple model particle filter (IMM-PF) of [7] which is applied to the road target tracking case in [37] with an on-road and off-road mode.

Recent advances in multiple target tracking [38,39] have resulted in random set theoretic methods [40] and in [41], an instance of such methods, namely a cardinalized probability hypothesis density (CPHD) filter [42] was presented for multiple ground target tracking. An example, with two groups of targets with four single targets in each group, is given. Track extraction is shown to be faster if the road information is used with the same road network model and observation model (GMTI) as in [31].

### 3 Multi-target tracking

Classical multi-target tracking consists of three sub-problems; detection, association, and estimation [39,28]. The multi-target tracker used in this study follows this structure, i.e., the detections are treated by an association step where each observation is associated with a known target track. The state of each target is estimated and predicted by a single target filter, and the observations are used to improve the result. If an observation cannot be associated with a known target, a new tentative filter is initialized.

In this section the target tracking problem is described by presenting first the general estimation solution and then the particle filter that is used to compute the posterior estimates. The association problem is briefly described and, in particular, a classical association technique is tailored to the particle representation.

#### 3.1 The general estimation solution

The aim of this section is to introduce the recursive state estimation theory. Let  $x_t$  denote the state of the target at time  $t$  and let  $y_t$  be an observation of the target at time  $t$ . Assume that the target state evolution can be

represented as a hidden Markov model composed of the transition model  $p(x_{t+1}|x_t)$  and the observation likelihood function  $p(y_t|x_t)$ . Let  $y_{1:t} = \{y_1, y_2, \dots, y_t\}$  denote the set of all observations up to and including time  $t$ . A recursive state estimator is given by the Bayes rule and can be expressed as the well-known measurement update formula

$$p(x_t|y_{1:t}) = \alpha_t^{-1} p(y_t|x_t) p(x_t|y_{1:t-1}) \quad (6)$$

and the one step ahead prediction

$$p(x_t|y_{1:t-1}) = \int p(x_t|x_{t-1}) p(x_{t-1}|y_{1:t-1}) dx_{t-1}. \quad (7)$$

The normalizing factor  $\alpha_t$  can be calculated as

$$\alpha_t = p(y_t|y_{1:t-1}) = \int p(y_t|x_t) p(x_t|y_{1:t-1}) dx_t. \quad (8)$$

The above equations represent the so called Bayesian filter and there are only few cases when it is possible to derive the analytical solutions for them. One case is the linear Gaussian case, leading to the well known Kalman filter (KF). In the general case, numerical approximations are necessary. One common technique is to approximate the target density  $p(x_t|y_{1:t})$  by a particle mixture as in the particle filter (PF).

### 3.2 Particle filter

In a PF the target density  $p(x_t|y_{1:t})$  is approximated by a particle mixture, containing  $N$  particles  $\{x_t^{(i)}\}_{i=1}^N$  and their corresponding importance weights  $\{w_t^{(i)}\}_{i=1}^N$ . Thus, the approximation is expressed as

$$p(x_t|y_{1:t}) \approx \sum_{i=1}^N w_t^{(i)} \delta(x_t - x_t^{(i)}) \quad (9)$$

where

$$\sum_{i=1}^N w_t^{(i)} = 1, \quad w_t^{(i)} \geq 0, \quad \forall i \quad (10)$$

and  $\delta(\cdot)$  is the Dirac delta distribution. This approximation is very suitable for calculating the integral in (7) and it can be shown that this approximation converges to the true solution as the number of particles goes to infinity, see [43] and [44] for the details on particle filtering. The importance weights  $\{w_t^{(i)}\}_{i=1}^N$  are computed using importance sampling where samples  $\{x_t^{(i)}\}_{i=1}^N$  are drawn from a proposal density  $q(x_t|x_{t-1}, y_t)$ . The filter recursion (6) and (7) can be expressed as

$$\begin{aligned} x_t^{(i)} &\sim q(x_t^{(i)}|x_{t-1}^{(i)}, y_t) \\ w_t^{(i)} &\propto \frac{p(y_t|x_t^{(i)}) p(x_t^{(i)}|x_{t-1}^{(i)})}{q(x_t^{(i)}|x_{t-1}^{(i)}, y_t)} w_{t-1}^{(i)} \end{aligned} \quad (11)$$

where the weights are normalized such that  $\sum_{j=1}^N w_t^{(j)} = 1$ . If the proposal density is selected as the state transition model, the filter recursion is simplified to

$$\begin{aligned} x_t^{(i)} &\sim p(x_t^{(i)}|x_{t-1}^{(i)}) \\ w_t^{(i)} &\propto p(y_t|x_t^{(i)}) w_{t-1}^{(i)}. \end{aligned} \quad (12)$$

This is perhaps the simplest particle filter and is called Bootstrap particle filter (BSPF) [44].

A resampling step is needed to prevent degeneration, see [45] for details. The so called systematic resampling algorithm was used in this study.

### 3.3 Association

The detector provides image coordinates of the measurements in each video frame, but it does not provide any information about the correspondence between the measurements at different times. An association method is used to handle this problem. Association is the process of assigning measurements to existing tracks, or existing tracks to measurements.

The association method used in this study is based on the global nearest neighbor (GNN) algorithm [39], but in contrast to the classical GNN where the target densities are assumed to be Gaussians, a more general approach is here used with the particle mixture approximation. Basically, the method computes the likelihood of each possible measurement to track correspondence and chooses the most likely global association hypothesis which gives the origins of all the measurements in current measurement set. The most likely association of measurements and tracks (or false alarms) is determined using the auction algorithm [39]. Letting  $P_D$  be the probability of detection, the log likelihood that the measurement  $j$  belongs to target  $k$  is defined as

$$l_{jk} \triangleq \log \left( \frac{P_D p^k(y_t^j|y_{1:t-1})}{1 - P_D} \right) \quad (13)$$

A suitable approximation, in the particle filter context, of the predictive likelihood  $p^k(y_t^j|y_{1:t-1})$  is

$$p^k(y_t^j|y_{1:t-1}) = \int p(y_t^j|x_t) p^k(x_t|y_{1:t-1}) dx_t \quad (14)$$

$$\approx \int p(y_t|x_t) \sum_{i=1}^N w_{t|t-1}^{k(i)} \delta(x_t - x_{t|t-1}^{k(i)}) dx_t \quad (15)$$

$$= \sum_{i=1}^N w_{t|t-1}^{k(i)} p(y_t^i|x_{t|t-1}^{k(i)}) \quad (16)$$

where the particles  $x_{t|t-1}^{k(i)}$  are sampled from a proposal density  $q(x_t^{k(i)}|x_{t-1}^{k(i)}, y_t^j)$  and the predictive weights are

$$w_{t|t-1}^{k(i)} = \frac{p(x_t^{k(i)}|x_{t-1}^{k(i)})}{q(x_t^{k(i)}|x_{t-1}^{k(i)}, y_t^j)}. \quad (17)$$

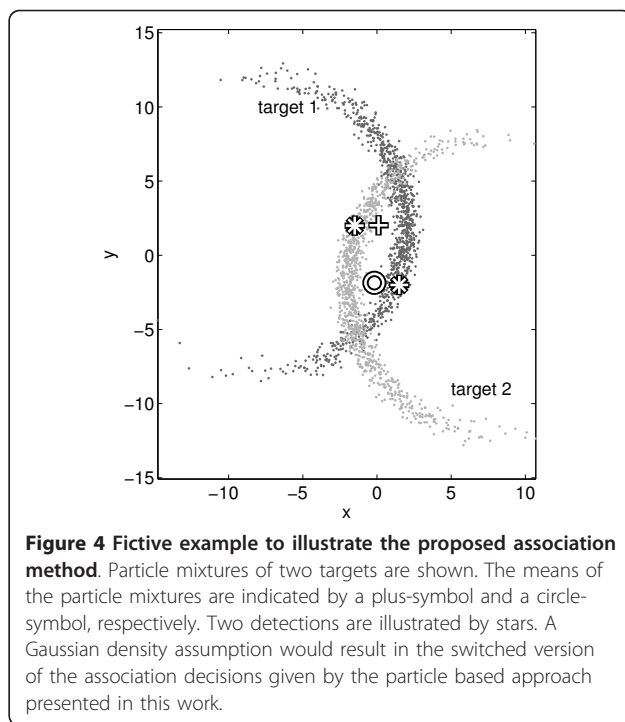
A similar calculation was used in [46] in a joint probabilistic data association framework. If observation model is represented as  $y_t^j = h(x_t) + e_t$ ,  $e_t \sim \mathcal{N}(e_t; 0, R)$ , then  $p(y_t^j|x_{t|t-1}^{k(i)}) = \mathcal{N}(y_t^j; h(x_{t|t-1}^{k(i)}), R)$ . If the bootstrap particle filter is used, the weights are  $w_{t|t-1}^{k(i)} = w_{t-1}^{k(i)}$  and the particles  $x_{t|t-1}^{k(i)}$  are obtained by simulating the particles  $x_{t-1}^{k(i)}$  according to the motion model.

It is assumed that the non-persistent false alarms are uniformly distributed in the image plane and their number is Poisson distributed with rate  $\beta_{FA}$ . The log likelihood that measurement  $j$  belongs to a non-persistent false alarm, is then given as  $l_{j,FA} \triangleq \log(\beta_{FA})$ .

Measurements that are not associated to any confirmed or tentative tracks are used to create new tentative tracks. A basic  $M/N$ -logic [28] is used for determining when a tentative track will be considered as confirmed. If a tentative track is updated with measurements for  $M$  out of  $N$  consecutive frames, it is considered as a confirmed track. Furthermore, a target is considered as lost and the track is deleted if no measurements are associated to the track for a number of consecutive frames, or the state covariance is too large.

*Remark 1* Classical target tracking also uses a gating step to exclude very unlikely measurement to the track assignments. The main purpose is to reduce the overall computational load since the gating is much cheaper to evaluate compared to association. In this study the gating step is removed since a reasonable gating criterion needs a similar amount of computational power as the log likelihood measures above. Furthermore, the number of detections and targets are quite low in our application, hence, the number of possible assignments are reasonably low.

*Example 1 (Association: Particle mixtures vs. Gaussianity assumption)* Note that classical association methods often assume Gaussian target densities. The association method presented here does not have such assumptions and will handle the possible multi-modal



**Figure 4** Fictive example to illustrate the proposed association method. Particle mixtures of two targets are shown. The means of the particle mixtures are indicated by a plus-symbol and a circle-symbol, respectively. Two detections are illustrated by stars. A Gaussian density assumption would result in the switched version of the association decisions given by the particle based approach presented in this work.

and/or non-Gaussian target densities in a reasonable way. See the example in Figure 4 where the particle mixtures of two targets are shown. The means of the particle mixtures are indicated by a plus-symbol and a circle-symbol, respectively. Now assume that two detections, which are denoted by stars, have been received. The association methods proposed here will associate the lower right detection with target 1 (if the measurement noise is reasonably small, and the  $P_{FA}$  is low). This is despite the fact that the mean of target 2 is very close to that detection. A Gaussian density assumption would in fact switch the association decisions yielding an unreasonable matching.

#### 4 Road constrained pedestrian tracking with MMPF

In this section the on-road and off-road motion models and the observation model are described in more detail compared to the introduction in Section 2.1. After the specific models are presented, the multi-model particle filter algorithm is described and some implementation issues will also be considered.

##### 4.1 On-road motion model

In a geographic information system (GIS) different forms of geographically referenced information can be analyzed and displayed. There are two classical methods to store GIS data: raster data (images) and vector data. Different geometrical types can be described by vector

data and basically there are three broad type categories; zero-dimensional points are used to represent points-of-interest, lines are used to represent linear features such as roads and topological lines, and polygons are used to represent particular areas such as lakes. There exist many approaches to store geo-spatial vector data and one common representation is the environmental systems research institute (ESRI) shapefile [47].

For target tracking purposes it is sometimes convenient to have a slightly different representation with redundant information to facilitate and speed up the data processing. In such a case, one data structure represents the roads and this structure contains the road stretch and the corresponding attributes. This structure is more or less the raw shape data plus an ID number for each road and an intersection ID for the each road end. An additional structure is used for the intersections and it contains the location and all connected roads (IDs) of each intersection. The exact structure of the data depends on what type of additional information is included, such as travel direction and prior probabilities for roads at an intersection.

In this study the road network information  $I_{RN}$  contains the two data structures mentioned above. The road information structures contain the following fields

- $ID$  - unique road ID
- $N$  - number of road segments
- $X$  -  $(3 \times N)$  vector with 3D coordinates
- $d$  -  $(1 \times N)$  vector with the cumulative distances of all road segments
- $w$  - width of the road
- $i_1$  -  $(1 \times N_1)$  vector containing the intersection ID of  $N_1$  roads connected to the start intersection
- $i_2$  -  $(1 \times N_2)$  vector containing the intersection ID of  $N_2$  roads connected to the end intersection
- $p_1$  -  $(1 \times N_1)$  vector containing prior probabilities of each connected road to the start intersection
- $p_2$  -  $(1 \times N_2)$  vector containing prior probabilities of each connected road to the end intersection and the intersection structure contains
- $ID^i$  - unique intersection ID
- $X^i$  -  $(3 \times 1)$  3D location of the intersection.
- $N^r$  - number of connecting roads
- $ID^r$  -  $(1 \times N^r)$  vector with IDs of the connecting roads

Assume that a road network description as defined above is available. The target is assumed to be on one of the roads all the time. A curve-linear coordinate system is defined for each road. Which road a target currently travels on is described by a mode parameter  $m$ . Let  $x^r \in [0, [d]_N]$  be the longitudinal position along the road relative the road start ( $[d]_N$  is the last element in the cumulative distance vector  $d$ , or in other words, the

total length of the road).  $v^r$  is the longitudinal speed and  $y^r$  and  $z^r$  are the lateral and the vertical deviation relative the road, respectively.

The on-road state vector is defined as  $x^r \triangleq (x^r \ y^r \ z^r \ v^r)^T$  and the dynamic target model  $f^r(x_t^r, \mathcal{I}_{RN}, \eta_t^r, v_t^r)$  in (1) can, as long as the target stays on the same road, be expressed as the linear discrete-time model

$$x_{t+1} = f^r(x_t^r, \mathcal{I}_{RN}, \eta_t^r, v_t^r) = \begin{pmatrix} 1 & 0 & 0 & T \\ 0 & \beta_{y^r} & 0 & 0 \\ 0 & 0 & \beta_{z^r} & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} x_t + \eta_t^r \quad (18)$$

where the process noise is  $\eta_t^r \sim \mathcal{N}(0, Q^r)$  and  $\beta_i \in \{\beta \mid 0 < \beta \leq 1\}$ ,  $i = y^r, z^r$ , are constants.

Thus, the target state is updated according to the linear model in (18), but a feasibility check is needed after every update. If the target has passed an intersection and is outside the current road, a nonlinear state update is also needed. A new road connected to that intersection is selected randomly among the roads  $i_{1/2}$  according to some discrete random variable  $v^r$  given the road probabilities  $p_{1/2}$ . In such a case, the mode parameter  $m_{t+1}$  is set to the new road and the longitudinal distance outside the old road is used to update  $x_{t+1}^r$ . Note that the direction of the old and new roads affects the update of  $x_{t+1}^r$ . Furthermore, the sign of the longitudinal velocity  $v_{t+1}^r$  needs to be changed if the travel directions on the roads are opposite.

The standard choice for the constants  $\beta_{y^r}$  and  $\beta_{z^r}$  is 1, but  $\beta_i < 1$  can be used to constrain the standard deviation of the state  $i$ . In practice, if  $0 < \beta_i < 1$  and no observations of the target is received, the state  $i$  will approach zero. This is in general a reasonable behavior since we do not want the prediction to deviate too much from the actual road network.

#### 4.2 Off-road motion model

The off-road motion model  $f^g(x_t^g, \eta_t^g)$  in (2) is selected to be the following constant velocity model with the state vector  $x^g = (x^g \ y^g \ z^g \ v^g \ \psi)^T$ , where  $x^g, y^g, z^g$  is the 3D location in a global Cartesian reference system,  $v^g$  is the translational speed in the  $x^g y^g$ -plane, and  $\psi$  is the course. The model is expressed as

$$x_{t+1}^g = f^g(x_t^g, \eta_t^g) = \begin{pmatrix} x_t^g + v_t^g T \cos(\psi_t) \\ y_t^g + v_t^g T \sin(\psi_t) \\ \beta_{z^g} z_t^g \\ v_t^g \\ \psi_t \end{pmatrix} + \eta_t^g \quad (19)$$

where  $\beta_{z^g} \in \{\beta \mid 0 < \beta \leq 1\}$  is a constant design parameter. The process noise is distributed as  $\eta_t^g \sim \mathcal{N}(0, Q^g)$  and ideally  $Q^g$  is state dependent, but in this study only



constant covariance matrices are considered for simplicity.

*Remark 2 (Incorporating the ground model)* The default value of the constant  $\beta_{zs}$  is 1, but in the case of a stationary bearings-only sensor the constant needs to be less than 1 to make the estimation problem observable. An implicit incorporation of a known ground model into the problem is possible here by defining the state  $z^g$  as the deviation from the ground model.

### 4.3 On/off-road transformations

As mentioned in Section 2.1 we need appropriate functions to convert the state vectors given in one of the representations into the other representation.

The function  $T^{gr}(\cdot)$  converts a state vector given in on-road coordinates to off-road (global) coordinates. This is generally an easy task and the global 3D position is found by interpolation. The underlying function is given by the array  $X$  in  $\mathcal{I}_{RN}$  in the points  $d$ .

The function  $T^{rg}(\cdot)$ , on the other hand, has to find the closest on-road coordinate state corresponding to a state vector in global coordinates. This is more involved in that one generally has to search in the road database for the closest point on the road network to the position component of the global state vector and has to project the velocity and other quantities onto their equivalents in the road network. It might also be useful to have a feasibility test by just checking if the lateral deviation state  $|y^r|$  is smaller than the road width (denoted as  $w$  in  $\mathcal{I}_{RN}$ ).

A globalization function  $T^g(\cdot, \cdot)$  is defined for later use as

$$T^g(x_t^{(i)}, g_t^{(i)}) \triangleq \begin{cases} T^{gr}(x_t^{(i)}), & q_t^{(i)} = 1 \\ x_t^{(i)}, & q_t^{(i)} = 2. \end{cases} \quad (20)$$

### 4.4 Observation model

A detection consists of the image coordinate and the height and width of the detection window. In the tracking filter the location of the feet of the pedestrian is used, so a foot detector is also needed. The position of the feet are transformed to azimuth and inclination angles given the perspective projection formula and knowledge of the sensor orientation and the intrinsic camera parameters. Thus, the observation model is a bearings-only model where the azimuth and inclination describe the direction to the target relative the sensor platform.

Let  $x^s = (x^s \ y^s \ z^s)^T$  be the position of the sensor relative to a global Cartesian reference system. An observation at time  $t$  is the relative angles between the sensor in  $x^s$  and the target in  $x^g$ , i.e.,

$$y_t = h(x_t^s; x_t^g) + e_t \\ = \left( \begin{array}{c} \arctan_2(y_t^g - y_t^s, x_t^g - x_t^s) \\ \arctan_2(z_t^g - z_t^s, \sqrt{(x_t^g - x_t^s)^2 + (y_t^g - y_t^s)^2}) \end{array} \right) + e_t \quad (21)$$

where  $e_t$  is the measurement noise modeled according to the Student's T-distribution

$$e_t \sim p_e(x_t) = \mathcal{S}_t(x_t; 0, \sigma_e^2 I_{2 \times 2}, \nu) \quad (22)$$

where  $\nu$  is the degree-of-freedom. Note that the Gaussian probability distribution  $\mathcal{N}(x; \mu, \Sigma)$  is a special case of Student's T-distribution  $\mathcal{S}_t(x; \mu, \Sigma, \nu)$  when the degree-of-freedom  $\nu$  is  $\infty$ . For  $1 \leq \nu < \infty$  the distribution resembles a Gaussian function but with heavier tails. The reason for selecting the Student's T-distribution is that it has been seen in early empirical trials to make the PF more robust to outliers.

*Remark 3 (Observability)* It is a well known fact that the observability in bearings-only tracking is highly dependent on the sensor trajectory, see [6] and references therein. In particular, for a stationary camera some additional information is required, e.g., a road network or a ground elevation model, see Remark 2.

### 4.5 Multiple-model PF

In a multi-model particle filter (MMPF) one keeps the particles  $\{x_t^{(i)}, q_t^{(i)}\}_{i=1}^{N_p}$  and their weights  $\{w_t^{(i)}\}_{i=1}^{N_p}$ , where  $x_t^{(i)}$  is the state of the particle with respect to either road coordinates  $(x_t^{r,(i)})$  or global coordinates  $(x_t^{g,(i)})$  according to the value of the on-road/off-road hypothesis variable  $q_t^{(i)}$  i.e.,

$$x_t^{(i)} = \begin{cases} x_t^{r,(i)}, & q_t^{(i)} = 1 \\ x_t^{g,(i)}, & q_t^{(i)} = 2. \end{cases} \quad (23)$$

Having these particles one can always calculate the density of the state of the target in global coordinates as

$$p(x_t | \gamma_{0:t}) = \sum_{i=1}^N w_t^{(i)} \delta(T^g(x_t^{(i)}, q_t^{(i)}) - x_t). \quad (24)$$

Using the density function (24), the minimum mean square error estimate of the target state in global coordinates is given by

$$\hat{x}_{t|t} = \sum_{i=1}^{N_p} w_t^{(i)} T^g(x_t^{(i)}, q_t^{(i)}) \quad (25)$$

with a covariance

$$P_{t|t} = \sum_{i=1}^{N_p} w_t^{(i)} \left( T^g(x_t^{(i)}, q_t^{(i)}) - \hat{x}_{t|t} \right) \left( T^g(x_t^{(i)}, q_t^{(i)}) - \hat{x}_{t|t} \right)^T \quad (26)$$

The particle filter calculates with each measurement the updated particles  $\{x_t^{(i)}, q_t^{(i)}\}_{i=1}^{N_p}$  and their weights  $\{w_t^{(i)}\}_{i=1}^{N_p}$  from the corresponding previous particles  $\{x_{t-1}^{(i)}, q_{t-1}^{(i)}\}_{i=1}^{N_p}$  and weights  $\{w_{t-1}^{(i)}\}_{i=1}^{N_p}$ .

A single step of the bootstrap version of the MMPF is summarized below.

**Algorithm 1 (MMPF)** Suppose we have the previous particles  $\{x_{t-1}^{(i)}, q_{t-1}^{(i)}\}_{i=1}^{N_p}$  and weights  $\{w_{t-1}^{(i)}\}_{i=1}^{N_p}$  available and we have received a new measurement  $y_t$ .

1. *Resampling:* Sample  $\{\tilde{x}_{t-1}^{(i)}, \tilde{q}_{t-1}^{(i)}\}_{i=1}^{N_p}$  from  $\{x_{t-1}^{(i)}, q_{t-1}^{(i)}\}_{i=1}^{N_p}$  according to weights  $\{w_{t-1}^{(i)}\}_{i=1}^{N_p}$  such that

$$P(\tilde{x}_{t-1}^{(i)} = x_{t-1}^{(j)}, \tilde{q}_{t-1}^{(i)} = q_{t-1}^{(j)}) = w_{t-1}^{(j)} \quad (27)$$

for each  $i = 1, \dots, N_p$ .

2. *Prediction Step:*

(a) Sample  $q_t^{(i)}$  from  $\tilde{q}_{t-1}^{(i)}$  such that

$$P(q_t^{(i)} | \tilde{q}_{t-1}^{(i)}) = \pi_{q_t^{(i)}, \tilde{q}_{t-1}^{(i)}} \quad (28)$$

for each  $i = 1, \dots, N_p$ .

(b) For each  $i = 1, \dots, N_p$ , generate  $x_t^{(i)}$  from  $\tilde{x}_{t-1}^{(i)}$  and  $\tilde{q}_{t-1}^{(i)}$  by using samples from the process noise sequences  $\eta_t^{r,(i)} \sim p_{\eta_t^r}(\cdot)$ ,  $\eta_t^{s,(i)} \sim p_{\eta_t^s}(\cdot)$  and  $v_t^{r,(i)} \sim p_{v_t^r}(\cdot)$  according to:

- If  $\tilde{q}_{t-1}^{(i)} = 1$ ,  $q_t^{(i)} = 1$  then

$$x_t^{(i)} = f^r(\tilde{x}_{t-1}^{(i)}, \mathcal{I}_{RN}, \eta_t^{r,(i)}, v_t^{r,(i)}) \quad (29)$$

- If  $\tilde{q}_{t-1}^{(i)} = 1$ ,  $q_t^{(i)} = 2$  then

$$x_t^{(i)} = f^s(T^{gr}(\tilde{x}_{t-1}^{(i)}, \mathcal{I}_{RN}), \eta_t^{s,(i)}), \quad (30)$$

- If  $\tilde{q}_{t-1}^{(i)} = 2$ ,  $q_t^{(i)} = 1$  then

$$x_t^{(i)} = f^r(T^{gr}(\tilde{x}_{t-1}^{(i)}, \mathcal{I}_{RN}), \mathcal{I}_{RN}, \eta_t^{r,(i)}, v_t^{r,(i)}), \quad (31)$$

- If  $\tilde{q}_{t-1}^{(i)} = 2$ ,  $q_t^{(i)} = 2$  then

$$x_t^{(i)} = f^s(\tilde{x}_{t-1}^{(i)}, \eta_t^{s,(i)}). \quad (32)$$

3. *Update Step:* Set  $w_t^{(i)}$  as

$$w_t^{(i)} \propto p_e(y_t - h(T^g(x_t^{(i)}, q_t^{(i)}))) \quad (33)$$

such that  $\sum_{i=1}^{N_p} w_t^{(i)} = 1$ .

*Remark 4 (Feasibility Check)* When a particle is selected to be transformed from the off-road mode to the on-road mode, a feasibility check of the new on-road state is done according to Section 4.3 (basically, check if the particle close to a road or not). If the state is not feasible, that particle will not be transformed and will therefore continue being in the off-road mode. Since an on-road state can always be transformed to an off-road state, a similar feasibility test is not needed in the opposite case. This will formally mean that the transition probability matrix (5)

$$\Pi = \begin{pmatrix} \pi_{11} & \pi_{12} \\ \pi_{21} & \pi_{22} \end{pmatrix} \quad (34)$$

is state dependent where

$$(\pi_{21} \ \pi_{22}) = \begin{cases} (0 \ 1) & \text{if } T^g(x_t^{(i)}, 2) \text{ is infeasible,} \\ (\bar{\pi}_{21} \ \bar{\pi}_{22}) & \text{otherwise,} \end{cases} \quad (35)$$

and  $\pi_{11}$ ,  $\pi_{12}$ ,  $\bar{\pi}_{21}$  and  $\bar{\pi}_{22}$  are constants.

*Remark 5 (Initialization)* Measurements that are not associated to any confirmed or tentative tracks are used to create new tentative tracks. When a new filter is created,  $N$  particles are generated for both models using different Gaussian prior distributions, one for each model. The initial position is computed by projecting the observation onto the ground plane. The feasibility check in Remark 4 is here also used for all the on-road particles, so particles outside the roads are discarded. The prior should be quite flat since the initial measurement is directly used in a measurement update step plus a resampling step to set the total number of particles to  $N$  in the MMPF.

*Remark 6 (Other Multiple Model Particle Filters)* There are other instances of multiple model particle filters in the literature [7,36]. The particular selection of MMPF in our study was made only because of the fact that it is the most well-known and the earliest of its kind. In general, all of the different multiple model particle filters are expected to give similar performance results for our application, which is also confirmed by the comparison between MMPF and IMM-PF of [7] we present in Section 5.3. Nevertheless, it must still be noted that there might be pathological examples (see e. g., [37]) for which these algorithms would yield significantly different performances especially during mode transitions.

#### 4.6 Occlusion and information from non-detections

The standard approach in target tracking is to update the filter statistics if an observation is received, otherwise, if no observation is received the target state remains intact in the update step. However, a measurement indicating no target in the field of view can also be considered as an observation and this is sometimes called “negative information” [48,49]. Negative information represents conclusions that are drawn from expected but actually missed detections. Despite that no observation data is available, these conclusions can be used to improve the current target estimate. Let  $\gamma_t = \emptyset$  denote that no detection was obtained at time  $t$ . The density  $p(x_t^r | \gamma_t = \emptyset, \gamma_{1:t-1})$  is not just the prediction  $p(x_t^r | \gamma_{1:t-1})$ , it also has to incorporate the (negative) information of a non-detection as

$$p(x_t^r | \gamma_t = \emptyset, \gamma_{1:t-1}) \propto p(\gamma_t = \emptyset | x_t^r) p(x_t^r | \gamma_{1:t-1}) \quad (36)$$

where  $p(\gamma_t = \emptyset | x_t^r) = 1 - P_D(x_t^r)$ . In the particle filter this means that the weight  $i$  is updated according to

$$w_{it}^{(i)} \propto \left(1 - \alpha P_D(x_{it-1}^{(i)})\right) w_{it-1}^{(i)} \quad (37)$$

where  $\alpha = 1$ . When the possibly occluded regions in the scene are known, this information can be used as a form of negative information in the particle filter at time instants with no detection. In such a case, the (negative information) update (37) tends to increase the weights of the particles in the occluded regions and reduce the weights of particles in the non-occluded regions.

Note that this requires that the model of the probability of detection be correct, otherwise, the risk of degeneracy increases in a particle filter with a limited number of particles. In practice a more conservative approach with  $0 \leq \alpha < 1$  is recommended when  $P_d(\cdot)$  may have significant modeling errors.

## 5 Results

In this section some results of the proposed pedestrian tracker are presented. First, in Section 5.1 the multiple-pedestrian tracker is applied to a real-world data set of an infrared sensor placed on top of a roof and pointing at a parklike environment with some trees, buildings and walkways. In Section 5.2 the comparison between MMPF and a standard off-road tracker is made on a similar data set with a GPS trajectory as the ground truth. A Monte-Carlo (MC) study based on synthetic data is presented in Section 5.3 where the IMM-PF [37,7] is also evaluated in order to come to a judgement about the expected differences between different multiple model particle filters. Finally, in Section 5.4 an

example illustrating the use and the performance gain of negative information is shown.

### 5.1 Pedestrian tracking field trial

The task presented in this section is to track a number of pedestrians in an infrared image sequence acquired by the infrared sensor described in Section 2.2. The pedestrians were walking both on and off the walkways and trees/buildings were occluding the pedestrians in some areas. The detector in Section 2.3 is used and the resulting detections are fed into the multi-target tracker based on the MMPF of Section 4.

The infrared sensor is located south of the surveillance area on a roof and the sampling frequency is 50 Hz, but just every 5th frame is used, i.e., the sampling frequency of the filter is 10 Hz. This makes it possible to use a time interleaved approach for increased robustness, where the same algorithm runs in parallel, each one time interleaved and operating on different data.

The number of particles in the MMPF is 1000 and the transition probabilities in (5) and (35) are

$$\bar{\Pi} = (0.95 \ 0.05 \ 0.1 \ 0.9). \quad (38)$$

The measurement noise is assumed to be distributed as

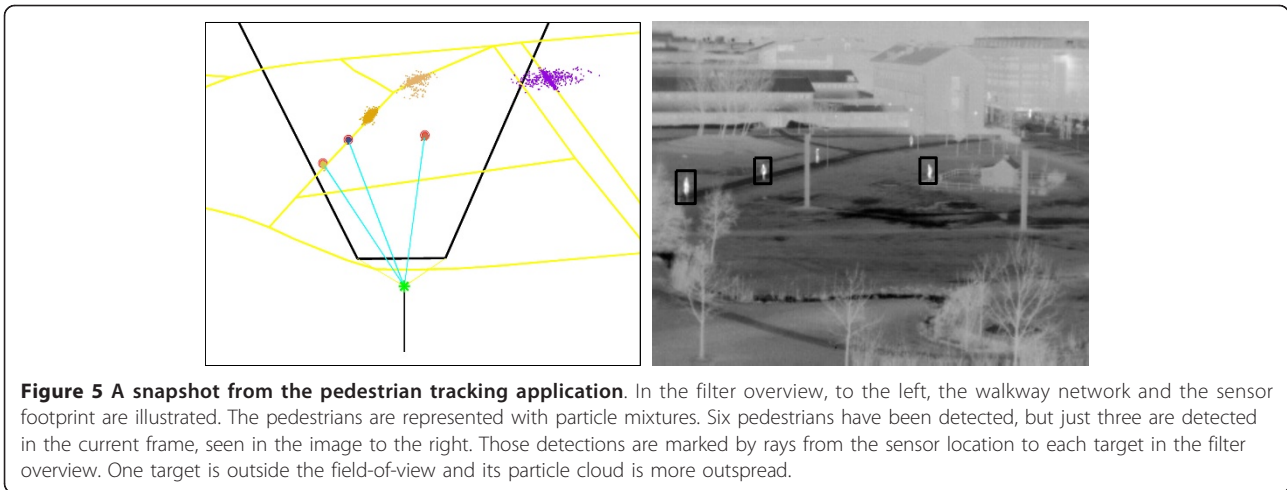
$$e_t \sim St(0, 0.004^2 I, 10). \quad (39)$$

When the standard deviation of the angle noise is set to be  $\sigma_e = 0.004$  as above, the projected uncertainty on the ground plane (with 68% confidence) corresponds to 4 m and 9 m for Gaussian distributions when a target is 130 m and 200 m away from the sensor, respectively. These uncertainty values become slightly larger for Student's T distributions ( $\nu = 10$ ) due to its heavier tails. The covariance matrices of the process noise are

$$\begin{aligned} Q^r &= \text{diag}(6.25 \cdot 10^{-3}, 6.25 \cdot 10^{-3}, 2.5 \cdot 10^{-4}, 6.25 \cdot 10^{-3}) \\ Q^s &= \text{diag}(6.25 \cdot 10^{-3}, 6.25 \cdot 6.25 \cdot 10^{-3}, 2.5 \cdot 10^{-4}, 6.25 \cdot 10^{-3}, 3 \cdot 10^{-4}) \end{aligned} \quad (40)$$

for the on-road and the off-road models, respectively. The  $\beta$  parameters are set to  $\beta_{y^r} = 0.96$  and  $\beta_{z^r} = \beta_{z^s} = 0.99$ . The initial state distribution is selected as Gaussian. It has been observed that the tracking results are quite insensitive to the initial state covariance. A target must be detected for three consecutive frames after its first appearance, and then for two out of three consecutive frames (after the first three consecutive frames) in order to be confirmed. A target is deleted if no detection is received for 5 s.

The results of the experiment are illustrated in a number of figures below. (Movies are available, see [50].) A snapshot where the particle mixtures can be seen is



shown in Figure 5. In Figures 6, 7, and 8 the focus is only on three selected pedestrians for the sake of clarity.

The estimated paths, based on the point estimates (25), for these three pedestrians are shown in Figure 6. One target is starting off-road, but ending on-road and vice-versa for another pedestrian. This mode transition can easily be seen in Figure 7 where the on-road mode probabilities are shown. Note that when a pedestrian is off-road, the on-road mode probability is very close to zero, but when the pedestrian is on-road the mode probability is just about 0.7-0.8. The reason for this is that the off-road model is valid when the pedestrian is on-road as well, but the opposite is not true if the target is too far from the road. The improvement of using a road network model can be seen in Figure 8 where

the uncertainty is shown. The uncertainty is here defined as

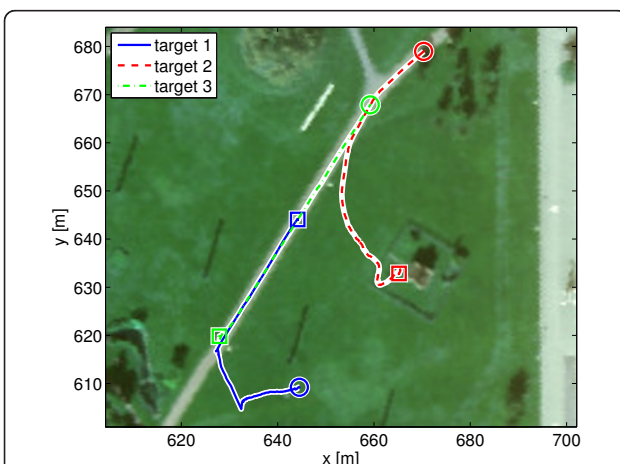
$$\sqrt{\text{tr } P_t^{pos}} \quad (41)$$

where  $P_t^{pos}$  is the position part of the state covariance matrix (26).

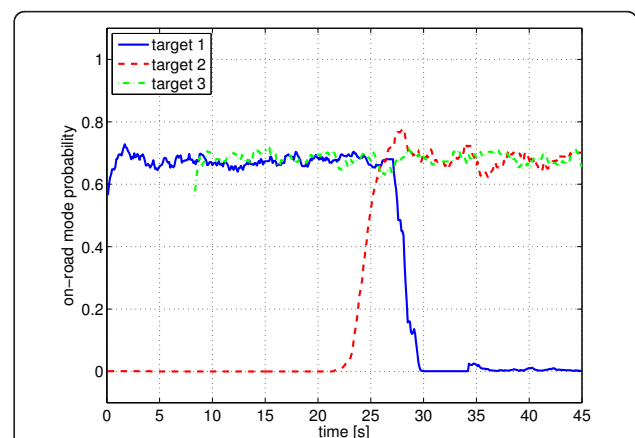
### 5.2 Performance evaluation with GPS ground truth

In this section a similar real data set to the one described above is used to evaluate the tracking performance for a single pedestrian by using the GPS trajectory of the pedestrian as ground truth. The MMPF pedestrian tracker with both on-road and an off-road models is compared with a standard off-road PF with no road network knowledge.

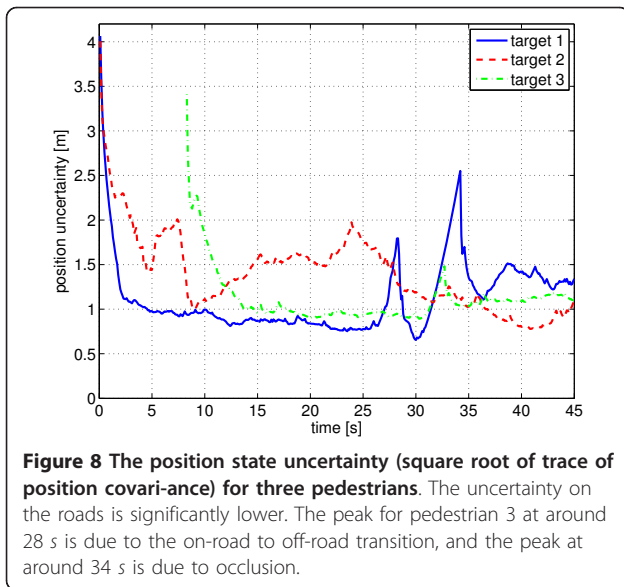
The scenario and the filter parameters of the MMPF and the PF are similar to those in Section 5.1, but the frame rate of the filters here is 12.5 Hz. The number of



**Figure 6 Estimated paths for three pedestrians.** Start locations (where the pedestrians were first observed) are marked with a square. Pedestrian 1 is walking toward south on a walkway, but is suddenly going off-road. Pedestrian 2 starts off-road, but eventually approaches a walkway moving north. Pedestrian 3 is on the walkway for the whole sequence.



**Figure 7 The on-road mode probability.** The on/off-road transitions can be clearly seen around 25-30 s.



particles in both filters is 1000 and the transition probabilities in MMPF are

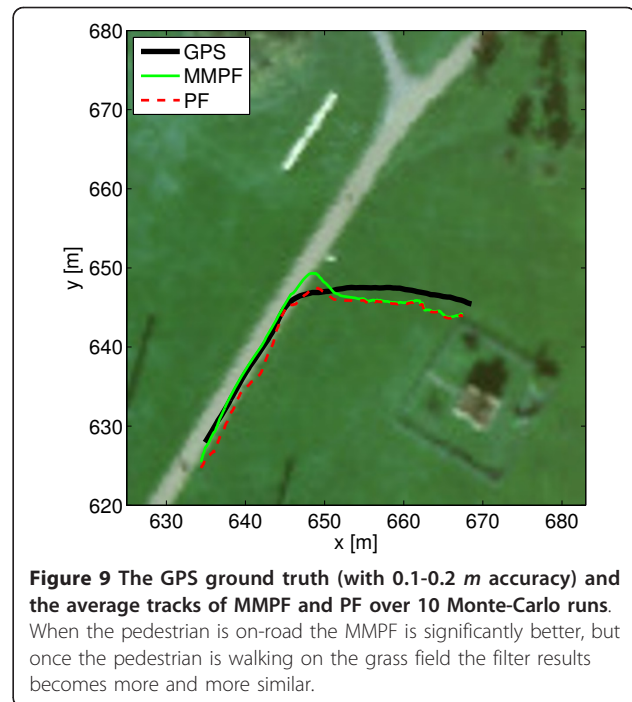
$$\bar{\Pi} = (0.95 \ 0.05 \ 0.05 \ 0.95). \quad (42)$$

The covariance matrices of the process noise are

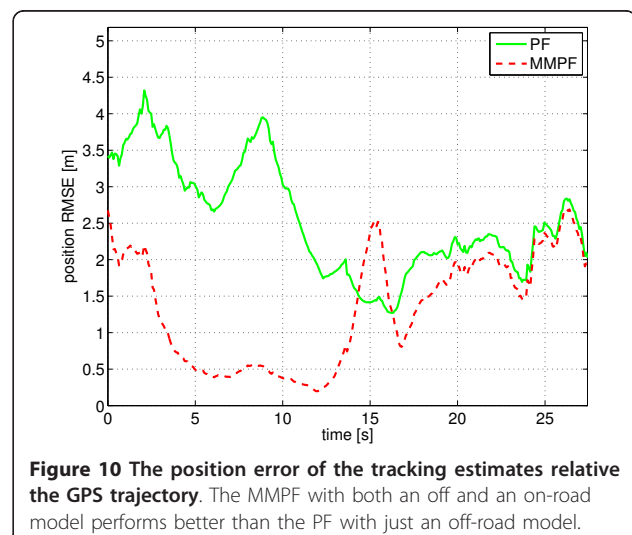
$$\begin{aligned} Q^r &= \text{diag}(8 \cdot 10^{-4}, 4 \cdot 10^{-4}, 1.6 \cdot 10^{-5}, 8 \cdot 10^{-4}) \\ Q^g &= \text{diag}(8 \cdot 10^{-4}, 8 \cdot 10^{-4}, 3.2 \cdot 10^{-5}, 8 \cdot 10^{-4}, 3.8 \cdot 10^{-2}) \end{aligned} \quad (43)$$

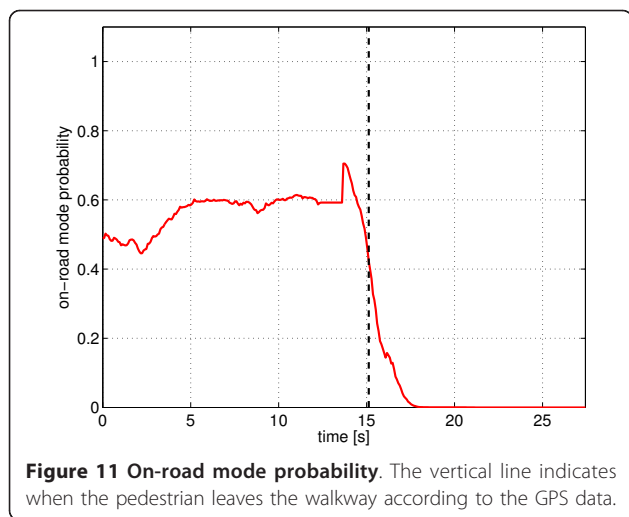
for the on-road and the off-road models, respectively. The  $\beta$  parameters are set to  $\beta_{y^r} = 0.95$  and  $\beta_{z^r} = \beta_{z^g} = 0.99$ . The altitudes of the roads are given by GPS measurements. Since no ground model is available, in order to get observability for the off-road model, the ground is simply assumed to be a plane. For each Monte-Carlo run, the fixed ground plane elevation is selected randomly by sampling uniformly from an interval of length 0.3 m which is determined by the altitude range for the closest road segment.

Since there is only a single set of measurements in the experiment (as opposed to the Monte-Carlo runs where a different realization of the measurement process is generated for each run) and since the results of the particle filters hardly differ for different runs, only 10 Monte-Carlo runs were found to be trustable. The true (GPS) path of the pedestrian with the expected accuracy around 0.1 - 0.2 m is shown in Figure 9. In addition, Figure 9 illustrates the average path estimate of each filter over the Monte-Carlo runs. The RMS position errors corresponding to both filters are presented in Figure 10. Figure 11 shows the average on-road mode probabilities provided by MMPF. As expected, the tracking result is significantly better for the MMPF when the target is on-road. When the target switches to off-road motion, the



accuracy difference between the filters starts to get smaller. The peak in the MMPF error occurs at the on-road to off-road switching of the target during which the on-road model of MMPF pulls the overall estimate towards the road segment. As soon as the mode probabilities of MMPF converge, the MMPF estimate becomes slightly better than that of PF. The PF estimates are more erroneous than those of MMPF during the off-road motion of the pedestrian. The reason is that the initial error of PF (just after the switching occurs) takes some time before decaying to the steady state level where both filters are expected to reach the



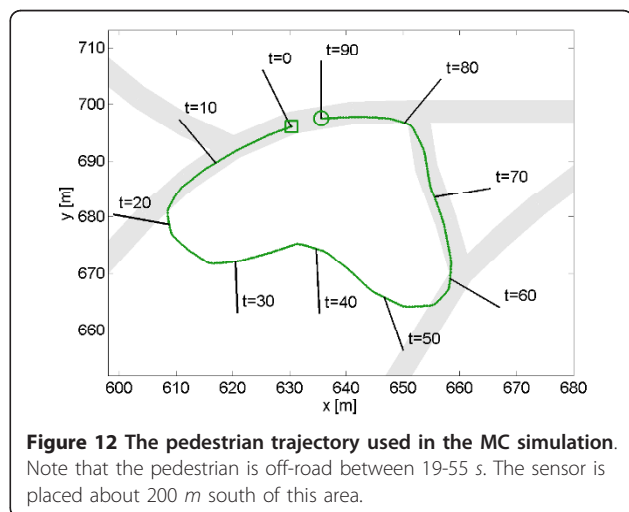


**Figure 11 On-road mode probability.** The vertical line indicates when the pedestrian leaves the walkway according to the GPS data.

same performance. The short period around 15 s where the PF error curve makes a dip below the MMPF error curve is a scenario specific phenomenon which is confirmed by the average path of PF intersecting the true GPS path in Figure 9.

### 5.3 Monte-Carlo simulation study

In order to compare the performance of different multiple model particle filters, and different mode transition probabilities in a controlled manner, a Monte-Carlo simulation based on synthetic data is presented in this section. The task is to track a single pedestrian that is walking both on and off the walkways according to Figure 12. The main objective in this section is to show the advantage of using a multiple model particle filter with road network knowledge over a standard PF. In addition to MMPF another multiple model particle filter, the IMMPF [7,37], is also evaluated. IMMPF is similar to



**Figure 12 The pedestrian trajectory used in the MC simulation.** Note that the pedestrian is off-road between 19-55 s. The sensor is placed about 200 m south of this area.

MMPF, but the number of particles is constant and pre-defined for each mode, unlike the MMPF where the number of particles in each mode is varying according to the posterior mode probabilities. We here emphasize that the IMMPF-MMPF comparison is included here only to show whether the particular selection of MMPF as the tracking algorithm in this study is critical or not. In fact, IMMPF, being a well-known method, was not used in pedestrian tracking before either and might as well have been selected as the tracking algorithm in this study.

In the MC-simulation the total number of particles is 1000 for all filters: PF, MMPF, and IMMPF. For the IMMPF, the total number of particles is divided equally between the modes, i.e., each model has 500 particles. The vision sensor is running at 10 *fps* and is located about 200 m south and 17 m above the surveillance area in Figure 12. To achieve a better triangulation behavior the sensor is moving slowly with 1 m/s to the east. The measurement noise of the vision sensor is distributed as

$$e_t \sim \mathcal{N}(0, 0.004^2 I). \quad (44)$$

The projected uncertainty (with 68% confidence) on the ground corresponds to 9 m.

We run two instances of each multiple model particle filter with different transition probabilities in order to see the algorithms sensitivity. In the literature, the convention for selecting the transition probabilities for multiple model methods is to use almost always diagonally dominant transition probability matrices (TPM). We here follow the same tradition and select the different transition probabilities for MMPF and IMMPF as

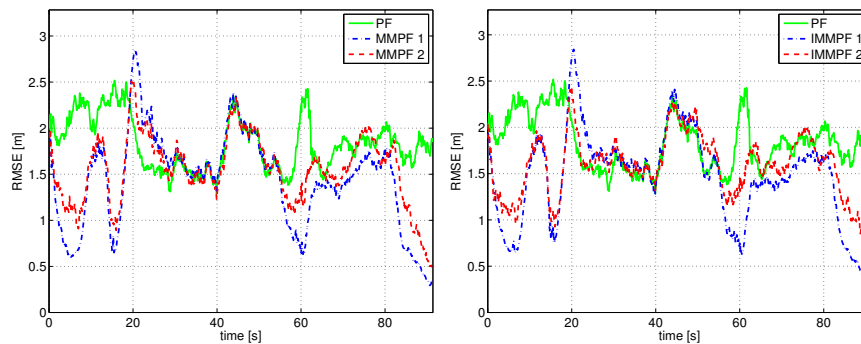
$$\begin{aligned} \bar{\Pi}^1 &= (0.95 \ 0.05 \ 0.05 \ 0.95) \\ \bar{\Pi}^2 &= (0.99 \ 0.01 \ 0.01 \ 0.99) \end{aligned} \quad (45)$$

where  $\bar{\Pi} \triangleq (\pi_{11} \ \pi_{12} \ \bar{\pi}_{21} \ \bar{\pi}_{22})$  whose elements are defined in (5) and (35).

In this MC simulation, the covariance matrices of the process noise are set as

$$\begin{aligned} Q^r &= \text{diag}(1 \cdot 10^{-3}, 1 \cdot 10^{-3}, 2.5 \cdot 10^{-4}, 2.25 \cdot 10^{-3}) \\ Q^g &= \text{diag}(1 \cdot 10^{-3}, 1 \cdot 10^{-3}, 2.5 \cdot 10^{-4}, 2.25 \cdot 10^{-3}, 8.3 \cdot 10^{-3}) \end{aligned} \quad (46)$$

for the on-road and the off-road models, respectively. Suitable value of the model parameter  $\beta_{yr}$  depends on the target behavior, but also the width of the road/walkway. Low value of  $\beta_{yr}$  will decrease the state uncertainty and force the state estimate towards the center line of the road, but at the cost of possible state bias and decreased ability of on-road to off-road change detection. A value between 0.95 and 0.99 is reasonable in



**Figure 13** The RMSE position errors of the MC simulation. Five different filters are considered: one PF with a single off-road model, two MMPFs and two IMMPFs with transition probabilities  $\bar{\Pi}^1$  and  $\bar{\Pi}^2$ , respectively. The MMPF results are compared to the off-road PF filter on the left and the IMMPF filters results are compared to the off-road PF filter on the right. The results of the MMPF and IMMPF are very similar.

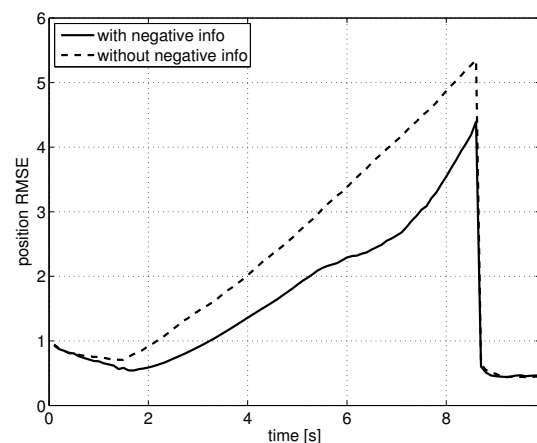
most cases. In this MC simulation, the  $\beta$  parameters are set to  $\beta_{y^r} = 0.95$  and  $\beta_{z^r} = \beta_{z^s} = 0.99$ .

The position RMSE values for the MC simulation with 100 runs are shown in Figure 13. Five different filters are considered: one PF with a single off-road mode, two MMPFs and two IMMPFs with transition probabilities  $\bar{\Pi}^1$  and  $\bar{\Pi}^2$ , respectively. When the pedestrian is off-road the performances of all filters are basically identical once the effects of the mode transition have died out. The only part where the single mode filter is best is at the on-road to off-road transition which is due to the fact that PF has no particles locked to the on-road model pulling the estimates towards the road. The differences between the MMPF and the IMMPF are quite small, even during the mode transitions. When the target is on-road the MMPF on-road mode probability is about 0.5 and, hence, the number of particles in each mode is then similar to IMMPF and the behavior during the on-road to off-road transition becomes similar. In the off-road to on-road transition the IMMPF cannot benefit from the reserved on-road particles since those are infeasible, therefore the behavior of the two filters will be similar for this case too. The MMPF is slightly better when the target is off-road since it can use twice as many particles, but the difference is too small to be clearly visible in the figure. The direction of the roads affects significantly how much the multiple model filters would gain from the on-road model. The more perpendicular the road stretch is to the line of sight of the sensor, the more useful the road information is. For example, compare the errors of the MMPF (or IMMPF) during the time intervals 70-80 s and 80-90 s. Although the effects of the transition probabilities on the performance of the two multiple model particle filters are quite visible, the changes due to different transition probabilities seem to be rather small compared to the gain in using the road network information (i.e., the on-road model).

#### 5.4 Use of negative information in pedestrian tracking

In Section 4.6 the concept of negative information was introduced, i.e., how one can draw conclusions from non-detections. This section will provide a simple example to illustrate the gain in using the negative information. Note, that in this study we are only considering occlusions caused by stationary objects, like trees and buildings, with known locations. Occlusions caused by other pedestrians are not handled.

Two particle filters using the on-road motion model are applied to a scenario where a fictitious building is placed in the area in front of the path of one pedestrian. The detections are removed manually when the pedestrian is occluded by the hypothetical building. The only difference between the filters is that one filter is using the so-called negative information. The position RMSE for the two filters are compared in Figure 14. A non-occluded filter result is used as ground truth.



**Figure 14** The position RMSE for two almost identical filters, the only difference is that one filter is using the so-called negative information.

The filter that is using negative information performs better since the effect of the particles that are visible from the point-of-view of the sensor is suppressed. An intuitive explanation for reducing the effect of the visible particles is given as follows. If the particles that are visible represented the true state, then the pedestrian would have been detected, but he/she is not, and therefore such particles should be less probable.

## 6 Conclusions

The pedestrian tracker proposed in this study is a multiple-model particle filter that uses prior information about the walkways to enhance the estimation performance. The tracking is performed in 3D global coordinates by utilizing the road network information. The states of the pedestrians are estimated by separate filters. Thus, the correlation between pedestrians are neglected, but experiments show that this is a reasonable approximation. For example, cars on a road are in general much more correlated than pedestrians.

The sampling based GNN association method works very well since the detector performs well with few false detections and the measurement noise is quite small for vision/infrared sensors compared to, for instance, radar. Using the Student's T-distribution for the measurement noise makes the filter more robust against minor outliers caused by the detector.

There are a number of advantages of using a road model. The tracking performance is significantly better if the road network information is used. On the other hand, filters based only on an off-road model perform quite well too as long as the detections are received on a regular basis and a reliable ground model is available. The gains in incorporating an on-road model into the estimation are significant not only for pedestrian motion prediction (e.g., due to occlusion or not in the field-of-view), but also for enhanced sensor management, track analysis, and anomaly detection.

On the other hand, there can also be some unpredicted disadvantages of using a road model. Using contextual information that is described relative a global reference system requires that the knowledge of the location and the orientation of the sensor be very accurate, otherwise unmodeled navigation error biases can have severe effects on the tracking performance. For a sensor system, in a known environment with known landmarks, the location and the orientation are usually straightforward to estimate with good accuracy. If this is not the case, algorithms that rely much on prior information should always be used with a fail-safe algorithm that can take over when the prior information is wrong or erroneous. In our case the off-road model provides the filter with both an off-road tracking capability and

increased robustness against model and navigation errors in on-road target tracking.

Observability is always an issue in vision based target tracking. Since the infrared sensor was stationary in the field trial, the off-road filter also needs a ground elevation model. This external information can be included explicitly by computing a range measurement or implicitly in the motion model. Regardless of the method used, erroneous orientation estimate and/or ground model will cause problems as in the erroneous road model case. However, note that a road network model is in general much easier to acquire and verify than a complete ground model. If no shape data exists for the roads of interest, it is quite straightforward to use GPS or orthophotos to create the road network and then to verify the result by projecting the network onto the camera image. If the sensor platform is moving the observability improves and the robustness against road and ground plane model errors increases.

In tracking applications, the performance depends always on a number of tuning parameters which are usually scenario dependent. As usual, there is a compromise between low uncertainty and robustness against unexpected events. In the end, it is the user, with certain experiences and preferences, that decides which models and parameter values to use. Our conclusion here is that if a reliable road network model is available, it is very beneficial to use it even in a pedestrian tracking application where the apparent gains, at first sight, might be shadowed by the properties of the accurate sensor.

According to our simulation results, incorporating both on-road and off-road models into the tracking seems to be much more important than the specific multiple model particle filter (MMPF or IMMPPF) that is used for implementing the incorporation. Similarly, the sensitivity to the transition probabilities used in multiple model filters proves to be less important compared to the gain obtained by using an additional on-road model.

In this article it has also been shown how a probability of detection model, e.g., 3D models of buildings etc., can be used to draw conclusions from non-detections. In practice the gain in using negative information depends on several factors, e.g., the environment (many or few buildings and trees), the target motion characteristics (highly predictable or not) etc., and the decision to use negative information must be made after taking such factors into consideration.

## Acknowledgements

This study had been supported by the Swedish Research Council under the Linnaeus Center CADICS and the frame project grant Extended Target Tracking (621-2010-4301). The study was a part of the graduate school Forum Securitatis in Security Link. The data acquisition was done in the FOI project "Signalbehandling for styrbara sensorsystem" founded by FM



(Swedish Armed Forces) and the detector described in this work has been developed by FOI. The authors would like to thank Fredrik Näsström, Fredrik Hemström, Gustav Haapalahti, Philip Engström, Karl-Göran Stenborg, Jörgen Karlholm, Joakim Rydell, Staffan Cronström and Morgan Ulvklo at FOI.

#### Author details

<sup>1</sup>Division of Automatic Control, Department of Electrical Engineering, Linköping University, SE-581 83 Linköping, Sweden <sup>2</sup>Department of Information Systems, Swedish Defence Research Agency, Box 1165, SE-581 11 Linköping, Sweden

#### Competing interests

The authors declare that they have no competing interests.

Received: 14 May 2011 Accepted: 14 February 2012

Published: 14 February 2012

#### References

1. T Rätty, Survey on contemporary remote surveillance systems for public safety. *IEEE Trans Syst Man Cybern C*. **40**(5):493–515 (2010)
2. W Hu, T Tan, L Wang, S Maybank, A survey on visual surveillance of object motion and behaviors. *IEEE Trans Syst Man Cybern C*. **34**(3):334–352 (2004). doi:10.1109/TSMCC.2004.829274
3. I Ahmad, Z He, M Liao, F Pereira, MT Sun, Special issue on video surveillance. *IEEE Trans Circuits Syst Video Technol*. **18**(8):1001–1005 (2008)
4. F Gustafsson, U Orguner, TB Schön, P Skoglar, R Karlsson, Navigation and tracking of road-bound vehicles. In *Handbook of Intelligent Vehicles*, ed. by Eskandarian, A Springer, (2011)
5. J Rydell, G Haapalahti, J Karlholm, F Näsström, P Skoglar, KG Stenborg, M Ulvklo, Autonomous functions for UAV surveillance. In: *International Conference on Intelligent Unmanned Systems (ICIUS)* (2010)
6. P Skoglar, Planning methods for aerial exploration and ground target tracking. Licentiate thesis no. 1420, Department of Electrical Engineering, Linköping University, SE-581 83 Linköping. Sweden, (2009)
7. Y Boers, J Driessen, Interacting multiple model particle filter. *IEE P-Radar Son Nav*. **150**(5):344–349 (2003). doi:10.1049/ip-rsn:20030741
8. Y Ma, S Soatto, J Kosecka, SS Sastry, An Invitation to 3-D Vision: From Images to Geometric Models. Springer Verlag, (2003)
9. J Karlholm, Design and evaluation of a hierarchy of boosted classifiers for detection of ground targets in aerial surveillance imagery. *Automatic Target Recognition XIV, Proc SPIE*. **5426** (2004)
10. RE Schapire, Y Singer, Improved boosting algorithms using confidence-rated predictions. *Mach Learn*. **37**, 297–336 (1999). doi:10.1023/A:1007614523901
11. P Viola, MJ Jones, Robust real-time face detection. *Internat J Comput Vision*. **57**, 137–154 (2004)
12. I Laptev, Improving object detection with boosted histograms. *Image Vision Comput*. **27**(5):535–544 (2009). doi:10.1016/j.imavis.2008.08.010
13. L Bourdev, J Brandt, Robust object detection via soft cascade. *Computer Vision and Pattern Recognition. IEEE Computer Society Conference on*, **2**, 236–243 (2005)
14. B Zhan, D Monekoso, P Remagnino, S Velastin, LQ Xu, Crowd analysis: a survey. *Mach Vision Appl*. **19**, 345–357 (2008). doi:10.1007/s00138-008-0132-4
15. A Yilmaz, O Javed, M Shah, Object tracking: A survey. *ACM J Com-put Surv*. **38**(4):1–45 (2006)
16. M Isard, A Blake, Condensation - conditional density propagation for visual tracking. *Internat J Comput Vision*. **29**, 5–28 (1998). doi:10.1023/A:1008078328650
17. K Okuma, A Taleghani, Nd Freitas, JJ Little, DG Lowe, A boosted particle filter: Multitarget detection and tracking. In *Computer Vision - ECCV 2004, Lecture Notes in Computer Science*, ed. by Pajdla, T, Matas, J Springer Berlin/Heidelberg, **3021**, 28–39 (2004). doi:10.1007/978-3-540-24670-1\_3
18. B Ristic, S Arulampalam, N Gordon, Beyond the Kalman Filter: Particle Filters for Tracking Applications (Artech House Radar Library). Artech House, Norwood, MA, (2004)
19. C Huang, B Wu, R Nevatia, Robust object tracking by hierarchical association of detection responses. In: *Proceedings of the 10th European Conference on Computer Vision: Part II, ECCV '08*, 788–801 (2008)
20. F Fleuret, J Berclaz, R Lengagne, P Fua, Multicamera people tracking with a probabilistic occupancy map. *IEEE Trans Pattern Anal Mach Intell*. **30**, 267–282 (2008)
21. S Khan, M Shah, A multiview approach to tracking people in crowded scenes using a planar homography constraint. In *Computer Vision - ECCV 2006, Lecture Notes in Computer Science*, ed. by Leonardis, A, Bischof, H, Pinz, A Springer Berlin/Heidelberg, **3954**, 133–146 (2006)
22. F Xu, X Liu, K Fujimura, Pedestrian detection and tracking with night vision. *IEEE Trans Intell Transp Syst*. **6**(1):63–71 (2005). doi:10.1109/TITS.2004.838222
23. B Leibe, K Schindler, N Cornelis, L Van Gool, Coupled object detection and tracking from static cameras and moving vehicles. *IEEE Trans Pattern Anal Mach Intell*. **30**, 1683–1698 (2008)
24. T Kirubarajan, Y Bar-Shalom, KR Pattipati, I Kadar, Ground target tracking with variable structure IMM estimator. *IEEE Trans Aerosp Electron Syst*. **36**(1):26–46 (2000). doi:10.1109/7.826310
25. PJ Shea, T Zadra, D Klamer, E Frangione, R Brouillard, Improved state estimation through use of roads in ground tracking. In: *Proceedings of Signal and Data Processing of Small Targets, SPIE*. **4048**, 312–332 (2000)
26. PJ Shea, T Zadra, D Klamer, E Frangione, R Brouillard, Precision tracking of ground targets. In: *Proceedings of IEEE Aerospace Conference, IEEE*. **3**, 473–482 (2000)
27. H Blom, Y Bar-Shalom, The interacting multiple model algorithm for systems with Markov switching coefficients. *IEEE Trans Automat Contr*. **33**(8):780–783 (1988). doi:10.1109/9.1299
28. Y Bar-Shalom, XR Li, Estimation and Tracking: Principles, Techniques and Software. Artech House, Inc., Storrs, CT, Norwood, MA, (1993)
29. XR Li, Y Bar-Shalom, Multiple-model estimation with variable structure. *IEEE Trans Automat Contr*. **41**(4):478–493 (1996). doi:10.1109/9.489270
30. MS Arulampalam, N Gordon, M Orton, B Ristic, A variable structure multiple model particle filter for GMTI tracking. In: *Proceedings of International Conference on Information Fusion*. **2**, 927–934 (2002)
31. M Ulmke, W Koch, Road-map assisted ground moving target tracking. *IEEE Trans Aerosp Electron Syst*. **42**(4):1264–1274 (2006)
32. J Koller, M Ulmke, Road-map assisted ground target tracking. *Aerosp Sci Technol*. **11**(4):261–270 (2007). doi:10.1016/j.jast.2006.10.010
33. O Payne, A Marrs, An unscented particle filter for GMTI tracking. In: *Proceedings of IEEE Aerospace Conference*. **3**, 1869–1875 (2004)
34. Y Cheng, T Singh, Efficient particle filtering for road-constrained target tracking. *IEEE Trans Aerosp Electron Syst*. **43**(4):1454–1469 (2007)
35. P Skoglar, U Orguner, D Törnqvist, F Gustafsson, Road target tracking with an approximative Rao-Blackwellized Particle filter. In: *Proceedings of International Conference on Information Fusion* (2009)
36. G Kravaritis, B Mulgrew, Variable-mass particle filter for road-constrained vehicle tracking. *EURASIP J Adv Signal Process*. **2008** (2008)
37. U Orguner, TB Schön, F Gustafsson, Improved target tracking with road network information. In: *Proceedings of IEEE Aerospace Conference, Big Sky, Montana, USA* (2009)
38. Y Bar-Shalom, XR Li, Multitarget-Multisensor Tracking: Principles, Techniques. YBS Publishing, Storrs, CT, (1995)
39. S Blackman, R Popoli, Design and Analysis of Modern Tracking Systems. Artech House, Inc., Norwood, MA, (1999)
40. R Mahler, Statistical Multisource-Multitarget Information Fusion. Artech House, Norwood, MA, USA, (2007)
41. M Ulmke, O Erdin, c, P Willett, Gaussian mixture cardinalized PHD filter for ground moving target tracking. In: *Proceedings of International Conference on Information Fusion, Quebec, Que*, 1–8 (2007)
42. R Mahler, PHD filters of higher order in target number. *IEEE Trans Aerosp Electron Syst*. **43**(4):1523–1543 (2007)
43. F Gustafsson, F Gunnarsson, N Bergman, U Forssell, J Jansson, R Karls-son, PJ Nordlund, Particle filters for positioning, navigation, and tracking. *IEEE Trans Signal Process*. **50**(2):425–437 (2002). doi:10.1109/78.978396
44. NJ Gordon, DJ Salmond, AFM Smith, Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEE Proc Radar Signal Process*. **140**(2):107–113 (1993). doi:10.1049/ip-f-2.1993.0015
45. JD Hol, TB Schön, F Gustafsson, On resampling algorithms for particle filters. In: *Nonlinear Statistical Signal Processing Workshop* (2006)
46. J Vermaak, S Godsill, P Perez, Monte-Carlo filtering for multi target tracking and data association. *IEEE Trans Aerosp Electron Syst*. **41**(1):309–332 (2005). doi:10.1109/TAES.2005.1413764
47. ESRI - Environmental Systems Research Institute: ESRI shapefile technical description - an ESRI white paper. URL. <http://www.esri.com/library/whitepapers/pdfs/shapefile.pdf> (1998). accessed 21 August 2011
48. LD Stone, CA Barlow, TL Corwin, Bayesian Multiple Target Tracking. Artech House Publishers, Norwood, MA, (1999)

49. W Koch, On exploiting "negative" sensor evidence for target tracking and sensor data fusion. *Inf Fusion*. **8**(1):28–39 (2007). doi:10.1016/j.inffus.2005.09.002
50. P Skoglar, Pedestrian tracking movies. URL. <http://www.control.isy.liu.se/~skoglar/asp2011/> (2011)

doi:10.1186/1687-6180-2012-26

**Cite this article as:** Skoglar et al.: Pedestrian tracking with an infrared sensor using road network information. *EURASIP Journal on Advances in Signal Processing* 2012 **2012**:26.

**Submit your manuscript to a SpringerOpen<sup>®</sup> journal and benefit from:**

- ▶ Convenient online submission
- ▶ Rigorous peer review
- ▶ Immediate publication on acceptance
- ▶ Open access: articles freely available online
- ▶ High visibility within the field
- ▶ Retaining the copyright to your article

---

Submit your next manuscript at ▶ [springeropen.com](http://springeropen.com)

---