Simultaneous camera orientation estimation and road target tracking

Per Skoglar and David Törnqvist

Original Publication:

Per Skoglar and David Törnqvist, Simultaneous camera orientation estimation and road target tracking, 2012, Proceedings of 15th International Conference on Information Fusion, 802-807.

Simultaneous Camera Orientation Estimation and Road Target Tracking

Per Skoglar*† and David Törnqvist*
* Division of Automatic Control
Department of Electrical Engineering (ISY)
Linköpings universitet SE-581 83 Linköping, Sweden
Email: {skoglar, tornqvist}@isy.liu.se
† Department of Sensor and EW Systems
Swedish Defence Research Agency (FOI)
Box 1165, SE-581 11 Linköping, Sweden

***Abstract*—Airborne surveillance systems equipped with a vision/infrared camera require good knowledge about the position and orientation of the camera for successful tracking of ground targets. In particular, this is essential when incorporating prior information, like road maps, that is expressed relative a global reference system. Usually, it is possible to obtain good positioning with inertial/satellite navigation systems, but estimating the orientation is generally more difficult. It might be possible to use SLAM (Simultaneous Localization and Mapping) or image registration approaches to support the navigation system, but not always since such approaches require stable features in the images. In this paper the problem of simultaneous orientation error estimation and road target tracking is considered by assuming that the target is constrained to a known road network. A particle filter approach is proposed and it is shown that the result of this filter is close to the performance of the ideal case where the orientation error is perfectly known. However, the performance depends on how informative the road path is and in rare cases the orientation error is unobservable.**

I. Introduction

Utilizing road network information in navigation and target tracking will reduce the location error significantly, compared to the case when the map information is ignored [1]. In surveillance applications, road maps have been used for road targets tracking based on detections from a radar sensor (typically ground moving target indicator, GMTI) [2], [3] or a vision sensor [4].

In target tracking, the navigation error of the observing sensor is typically neglected and this is reasonable as long as the tracks are expressed relative the sensor itself. However, when external information, such as road maps, represented relative a global reference system is included, un-modeled navigation error biases can have severe effects on the tracking performance. In particular this is a problem when the accuracy of the sensor is similar or better, in some sense, than the accuracy of the navigation estimate. For instance, even though an INS/GPS navigation system is used, there might be robustness issues for vision/infrared camera based systems due to the high angular resolution of cameras. Just a small camera orientation error can lead to a large error for a distant object when an observation is transformed to a global reference frame.

There is a number of approaches to handle these robustness issues. For estimation methods relying on a representation with bad support in low probability areas (such as the particle filter), increasing the measurement noise might be tempting to increase the robustness of the filter, but this is a questionable solution since the tracking performance will suffer. An alternative is to use a multiple-model filter with both an on-road mode and an off-road mode [4] where the off-road mode serves as a fall-back solution. A third alternative is to use track landmarks in a vision SLAM (Simultaneous Localization and Mapping) framework [5], [6], [7] to estimate the orientation error, but this requires that a number of suitable landmarks are available over time and this might not always be the case. A related approach is to use image registration to estimate the orientation error by aligning the camera image frame to a known scene model [8], [9]. However, this approach also relies on stable features in the image frames and parallax effects are complicated to handle.

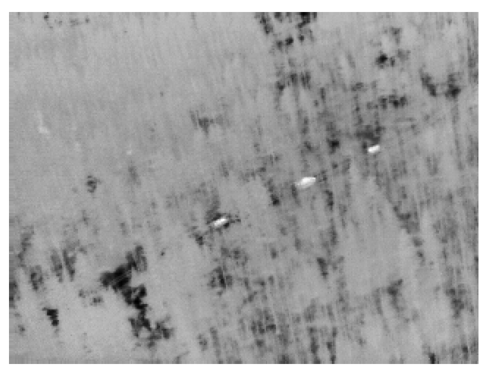

Fig. 1. An infrared image from an airborne surveillance system. Three cars on a road are visible. The image is dominated by trees and it is unlikely that any visual landmarks can be used to support the navigation. This paper considers an estimation problem where the orientation error and the targets are tracked simultaneously.

Figure 1 shows an image frame from an airborne surveillance system with an infrared sensor. Three cars are traveling on a road through a wood, but there are barely any useful static landmarks for a SLAM framework to use. Since the targets are visible and they are on a known road network, an interesting question is if the orientation error can be estimated and the target can be tracked simultaneously. The purpose of this paper is to investigate this problem. The reason for just considering the orientation error is that this error has much larger impact on the overall result compared to the positioning error of the camera that is assumed to be known with high accuracy by a satellite navigation system. This estimation problem is related to SLAM, but a significant difference is that constrained dynamic targets are used here, unlike SLAM where static landmarks are considered.

The paper is organized as follows. Section II presents the problem and models in detail. In Section III a particle filter is proposed as the estimator of the problem. Simulation results are presented and discussed in Section IV and finally the work is summarized and some conclusions are drawn in Section V.

II. Problem Description

In this section the models of the simultaneous orientation error estimation and road target tracking problem are given. The orientation error is represented by a quaternion and the targets are assumed to be constrained to a known road network. In this work a simplified version of the road target tracking approach in [4] is used; just one single road is here considered and the association problem is ignored. The observation model is based on a pinhole camera model.

A. Road Target Model

In this paper it is assumed that the target is on the same road all the time. Road intersections and target transitions to other roads are not considered in this work. The lateral and vertical locations of the target relative the road are also assumed to be zero. However, it should be quite straightforward to implement a more general road target representation by following the approach in e.g. [4].

A curve-linear coordinate system is defined for the road. Let x and v be the longitudinal position and velocity, respectively, along the road relative the road start. The on-road state vector is defined as $x^r \triangleq (\mathsf{x}\ \mathsf{v})^\mathsf{T}$ and the dynamic target model can, as long as the target stays on the same road, be expressed as the linear discrete-time model

$$x^r_{t+1} = f^r(x^r_t, v^r_t) = \underbrace{\begin{pmatrix} 1 & T \\ 0 & 1 \end{pmatrix}}_{\triangleq F} x^r_t + \underbrace{\begin{pmatrix} T^2/2 \\ T \end{pmatrix}}_{\triangleq G} w^r_t \tag{1}$$

where the process noise is i.i.d. as $w^r_t \sim \mathcal{N}(0, Q^r)$ and T is the sampling time.

Let $g^{gr}(\cdot)$ represent the transformation that is transforming a coordinate x^r in the local road aligned system to a coordinate x^g in a global Cartesian reference system, i.e., $x^g = g^{gr}(x^r)$.

B. Orientation Bias Model

The orientation bias is represented by a quaternion $x^a \triangleq (q_0\ q_1\ q_2\ q_3)^\mathsf{T}$ and the bias is modeled [10] as

$$x^a_{t+1} = f^a(x^a_t, v^a_t) = x^a_t + \frac{T}{2}\tilde{S}(x^a_t)v^a_t \tag{2}$$

where the process noise is i.i.d. as $w^a_t \sim \mathcal{N}(0, Q^a)$ and

$$\tilde{S}(x^a) = \begin{pmatrix} -q_1 & -q_2 & -q_3 \\ q_0 & -q_3 & q_2 \\ q_3 & q_0 & -q_1 \\ -q_2 & q_1 & q_0 \end{pmatrix}. \tag{3}$$

C. Observation Model

The camera is a staring-array vision sensor with limited field-of-view and it is assumed that the pinhole camera model can be used. Thus, an observation y_t at time t is a detection of the target in the image plane corresponding to the azimuth and elevation angles from the sensor location to the target location under the influence of the orientation bias. Let

$$y_t = h(x_t) + e_t \tag{4}$$

where the measurement noise is i.i.d. as $e_t \sim \mathcal{N}(0, R)$. According to the pinhole camera model a point $x^c_t = (\mathsf{x}^c_t\ \mathsf{y}^c_t\ \mathsf{z}^c_t)^\mathsf{T}$ expressed in Cartesian coordinates relative a camera fixed reference system, is projected on a virtual image plane onto the image point $(\mathsf{u}\ \ \mathsf{v})^\mathsf{T}$ according to the ideal perspective projection formula

$$h(x_t) = \begin{pmatrix} \mathsf{u}_t \\ \mathsf{v}_t \end{pmatrix} = \frac{1}{\mathsf{z}^c_t}\begin{pmatrix} \mathsf{x}^c_t \\ \mathsf{y}^c_t \end{pmatrix}. \tag{5}$$

The point x^c is computed as the difference between the camera location x^s_t and target location, relative the global reference system, followed by a transformation from the global system to the camera fixed system, i.e.,

$$x^c_t = R(x^a_t)R^{cg}_t(g^{gr}(x^r_t) - x^s_t) \tag{6}$$

where R^{cg}_t is a rotational matrix representing the (unbiased) orientation of the camera. Note that both R^{cg} and x^s_t is assumed to be known in this work. The rotation matrix $R(x^a_t)$ is the standard rotation matrix based on the quaternion and can for instance be found in [11, p. 158].

D. Augmented Model

The model where both the orientation error and the target tracking of n targets can be defined as follows. The state and noise input vectors are defined as

$$x \triangleq \begin{pmatrix} x^a \\ x^{r_1} \\ \vdots \\ x^{r_n} \end{pmatrix},\ w \triangleq \begin{pmatrix} w^a \\ w^{r_1} \\ \vdots \\ w^{r_n} \end{pmatrix}, \tag{7}$$

respectively, and the augmented dynamic model is defined as

$$\begin{aligned} x_{t+1} &= f(x_t, v_t) \\ &= \begin{pmatrix} I & 0 & \dots & 0 \\ 0 & F & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & F \end{pmatrix} x_t + \begin{pmatrix} \frac{T}{2}\tilde{S}(x^a_t) & 0 & \dots & 0 \\ 0 & G & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & G \end{pmatrix} v_t. \end{aligned} \tag{8}$$

An observation of target i is modeled as

$$y^{r_i}_t = h(x^a_t, x^{r_i}_t) + e^{r_i}_t. \tag{9}$$

III. Estimation Approach

If the uncertainty is relatively low, the estimation problem will be close to linear and could be handled with an Extended Kalman Filter (EKF). However, with a larger uncertainty and for an arbitrary road network the estimated state distribution is likely to be multimodal and thus an EKF would no longer work. Here, we have instead used the particle filter for estimation. The particle filter is a very flexible estimation approach where it is relatively straightforward to use non-linear model like road maps. However, one disadvantage of the particle filter is its bad support in low probability areas which sometimes makes it sensitive to un-modeled errors. Another problem is when the number of tracked targets increases. Then the state dimension increases and the computational load rapidly becomes infeasible to handle. To alleviate this, it is proposed to do the estimation with a Rao-Blackwellized particle filter which reduces the state dimension handled by the particle filter.

A. Bootstrap Particle Filter

As the non-linear estimation problem treated in this article does not have an analytical solution, an approximate solution with the particle filter is used. The particle filter was introduced in [12] and a good tutorial can be found in [13]. To use the particle filter, the dynamic model is rewritten as conditional distributions. The dynamic model in (1) and (2) can with knowledge about the process noise distribution be reformulated into the conditional distribution $p(x_t|x_{t-1}) = p(x_t^r|x_{t-1}^r)p(x_t^a|x_{t-1}^a)$, where

$$p(x_t^r|x_{t-1}^r) = \mathcal{N}(Fx_{t-1}^r,\, GQ^rG^T) \tag{10}$$

and

$$p(x_t^a|x_{t-1}^a) = \mathcal{N}\big(x_{t-1}^a,\, \frac{T^2}{4}\tilde{S}(x_{t-1}^a)Q^a\tilde{S}(x_{t-1}^a)^T\big). \tag{11}$$

In the same way, the measurement model (5) can be rewritten as the density

$$p(y_t|x_t) = \mathcal{N}\big(h(x_{t-1}),\, R\big). \tag{12}$$

With knowledge about these distributions, the particle filter in Algorithm 1 can be applied.

B. Rao-Blackwellized Particle Filter

To reduce the dimension handled by the particle filter, the state space is partitioned into two parts $x_t = (x_t^{pT}\ x_t^{kT})^T$. The posterior distribution can then also be partitioned into two parts as

$$p(x_t^p, x_t^k|y_{1:t}) = p(x_t^p|y_{1:t})p(x_t^k|x_t^p, y_{1:t}). \tag{16}$$

Now, if the system is linear given the states x_t^p (i.e., conditionally linear) then the second density in (16) can be computed by a Kalman filter. The particle filter is then only needed for computing the density over x_t^p, hence the dimension reduction. This partitioning is referenced to as the Rao-Blackwellized Particle Filter and more informatiom can be found in [15], [16], [17].

Algorithm 1 Particle Filter

1) Initialize the particles according to $x_0^{(i)} \sim p(x_0)$ and set appropriate weights $q_{0|0}^{(i)}$ for all $i = 1, \ldots, N$.
2) Time update: Generate new particles according to the proposal distribution

$$x_t^{(i)} \sim \pi(x_t|x_{t-1}^{(i)}, y_t), \quad i = 1, \ldots, N. \tag{13}$$

Update the weights according to

$$q_{t|t-1}^{(i)} = \frac{p(x_t^{(i)}|x_{t-1}^{(i)})}{\pi(x_t^{(i)}|x_{t-1}^{(i)}, y_t)} q_{t-1|t-1}^{(i)}, \quad i = 1, \ldots, N, \tag{14}$$

which in case $\pi(x_t|x_{t-1}^{(i)}, y_t) = p(x_t|x_{t-1}^{(i)})$ simplifies to $q_{t|t-1}^{(i)} = q_{t-1|t-1}^{(i)}$.
3) Measurement update: Calculate importance weights according to

$$q_{t|t}^{(i)} = \frac{q_{t|t-1}^{(i)} p(y_t|x_t^{(i)})}{\sum_{j=1}^N q_{t|t-1}^{(j)} p(y_t|x_t^{(j)})}, \quad i = 1, \ldots, N. \tag{15}$$

4) Resampling: Use a method of choice, in this work systematic resampling is used [14].
5) Set $t := t + 1$ and repeat from step 2.

In this application, the states are divided as $x^p = \mathsf{x}$ and $x^k = (\mathsf{v}\, x^{aT})^T$. The dynamics for x^k will be close to linear since small orientation errors are expected and targets typically move far away from the sensor. Hence, this part can be linearized. Multimodality is expected for x^p depending on the shape of the road network, hence this part is estimated with the particle filter.

In case of multiple targets, the state space is divided as $x^p = (\mathsf{x}^{r_1} \cdots \mathsf{x}^{r_n})^T$ and $x^k = (\mathsf{v}^{r_1} \cdots \mathsf{v}^{r_n}\, x^{aT})^T$.

IV. Results

In this section a number of simulation examples are presented. The road paths that are used are defined by

$$\begin{aligned} \mathsf{x}^g &= r(\mathsf{x})\cos(\mathsf{x}) \\ \mathsf{y}^g &= r(\mathsf{x})\sin(\mathsf{x}) \\ \mathsf{z}^g &= -100 + 10\sin(c\mathsf{x}) \\ r(\mathsf{x}) &= 100 + 20\sin(2c\mathsf{x}) \end{aligned} \tag{17}$$

where c is a constant. Three different cases will be used, $c = 0, 2, 4$, and the these paths are illustrated in Figure 2.

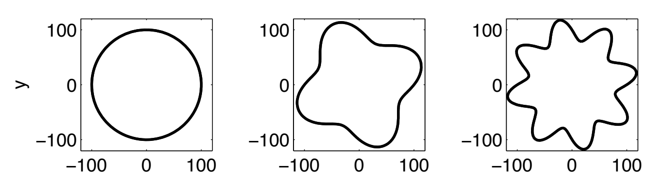

Fig. 2. Three different road paths (17) used in the simulations. *Left:* $c = 0$. *Middle:* $c = 2$. *Right:* $c = 4$.

Three different filters are applied to the single road target tracking problem. The RMSE (root mean square error) result of Monte-Carlo (MC) simulations with 100 runs are presented for evaluation of the estimation results. The filter algorithm, for all cases, is the bootstrap particle filter presented in Section III, but the models and initial conditions are different.

- The dynamic model of filter (IT) (as in Ideal Tracking) only contains the target tracking part in (1), and the orientation error is assumed to be perfectively known. This filter is providing an *ideal* estimation result of the target tracking and gives a lower bound of the target tracking performance.
- Filter (TO) (as in Tracking Only) is also just considering the target tracking part (1), but unlike the ideal filter (IT), this filter does not know the orientation error and acts as if this error is zero. Thus, filter (TO) will show the impact of un-modeled orientation error.
- Filter (AT) (as in Augmented Tracking) is using the model in (8) which includes both the target tracking and the orientation error parts.

The camera is assumed to be located in the origin and it is aiming downwards. The initial orientation error is a rotation α around the pointing direction vector. The orientation error then evolves according to (2). For each MC run α is sampled uniformly from the interval $[-0.1, 0.1]$ (rad). The initial position and velocity errors of the target are sampled uniformly from $[-10, 10]$ (m) and $[-1, 1]$ (m/s), respectively. The sampling time is 0.2 s and the number of particles is 1000 in each filter. The true state trajectory is generated by using the following covariance matrices: $Q^r = 0.5^2$, $Q^a = 0.01^2 I$, $R = 0.02^2 I$, but in the filters twice as big Q^r and R are used.

Figures 3, 4 and 5 show the filter results for the road cases $c = 4$, $c = 2$ and $c = 0$, respectively. The more curved the path is, the easier it is to estimate the orientation error and, hence, the target position. This can be seen by noting that the position error of the target $\tilde{x}$ of filter (AT) (black line) is close to the position error of filter (IT) (dashed gray line) quicker in Figure 3 compared to Figure 4 (position error is shown in the second plot from the bottom in each figure).

The case $c = 0$ in Figure 5 is a special case where the orientation error is not observable in filter (AT) since the circle path is invariant w.r.t. a rotation around a vertical axis. For this particular case, the positioning error of filter (AT) is similar to the error of filter (TO) where the orientation error is ignored. However, even for less informative paths filter (AT) usually performs better than filter (TO). This is illustrated by another example of the case $c = 0$ in Figure 6 where the orientation error is around a vector in the xy-plane instead. Although the position error of the target in filter (AT) does not approach the ideal case that much, the result is better than for filter (TO) where the orientation error is ignored. Note that the impact of the orientation error on the position error of the target in filter (TO) is dependent on the target location on the circle path.

Just a few examples are presented here, but the overall behavior of the filters is similar for other conditions, e.g. other initial orientation errors. Filter (AT) performs good and, if the path is informative, the position error of the target is close to the ideal case. The impact of un-modeled orientation errors in filter (TO) are also dependent on the path and also the direction of the orientation error, but in all circumstances the errors have a severe effect on the target tracking performance.

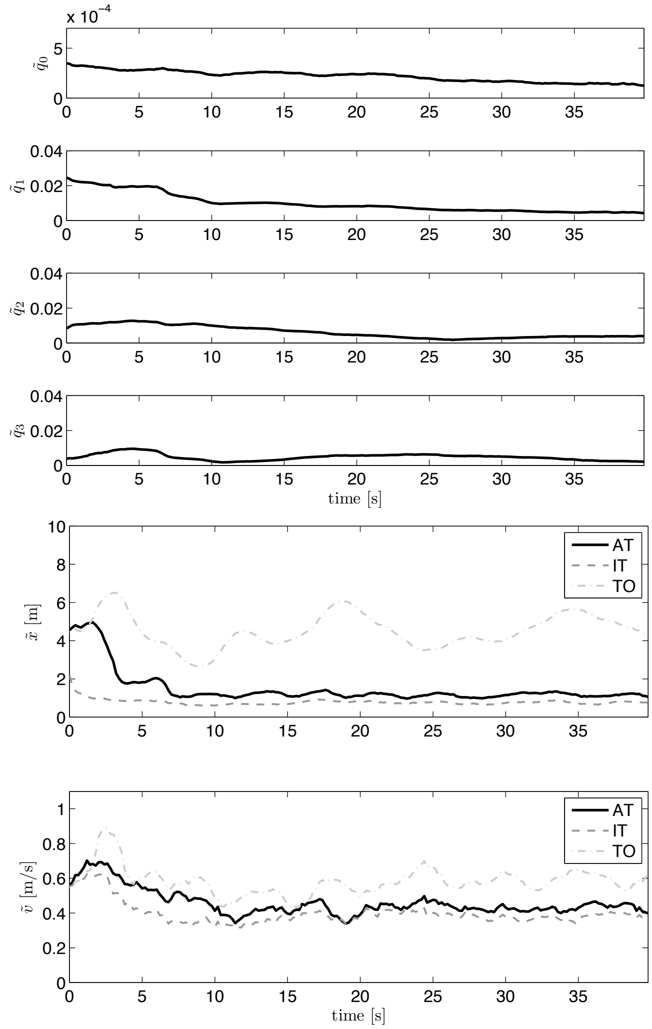

Fig. 3. RMSE for path $c = 4$, the right case in Figure 2. Black line shows the RMSE errors (denoted by the tilde symbol on each state variable) of filter (AT) with simultaneous orientation error and target tracking. Dashed dark gray line shows the RMSE errors of filter (IT) with the target tracking part only and perfect orientation error knowledge. Dashed-dotted light gray line shows the RMSE errors of filter (TO) with the target tracking part only and ignored orientation error. The filter (AT) is successfully estimating the orientation error and the positioning error $\tilde{x}$ is close to the ideal case (second plot from the bottom).

V. Discussion and Conclusions

This paper treats the problem of simultaneous orientation error estimation and road target tracking. The application in mind is an airborne surveillance system with a vision/infrared camera for road target tracking and an INS/GPS system for navigation. Although the navigation system provides the position and orientation of the camera there might be errors

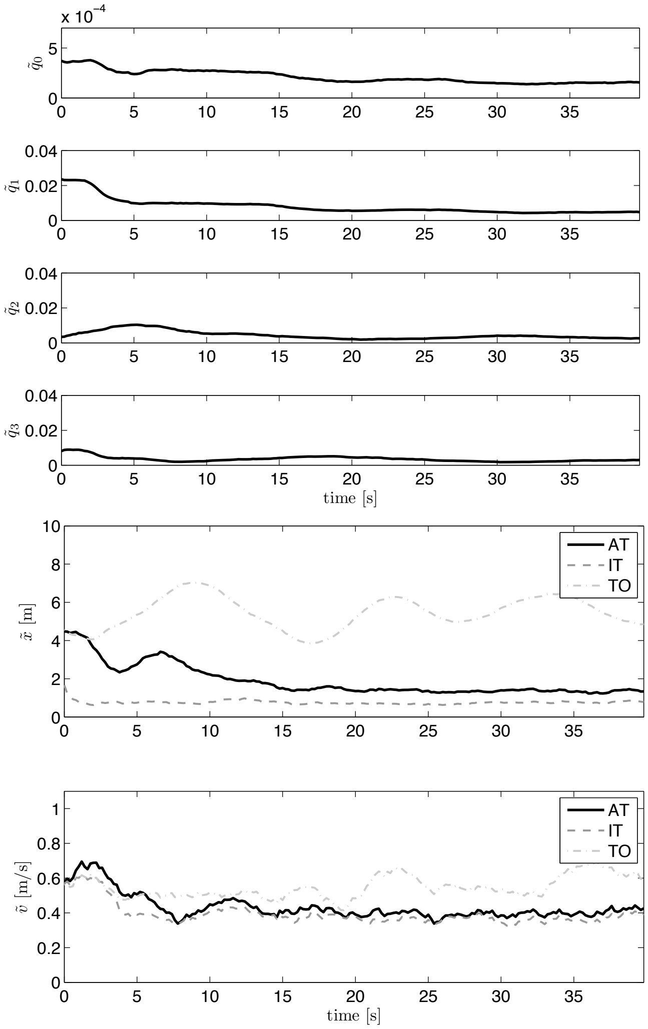

Fig. 4. RMSE for path $c = 2$, the middle case in Figure 2. Here the path is less "informative" and the result of filter (AT) close to the ideal case slower, compared to the case in Figure 3.

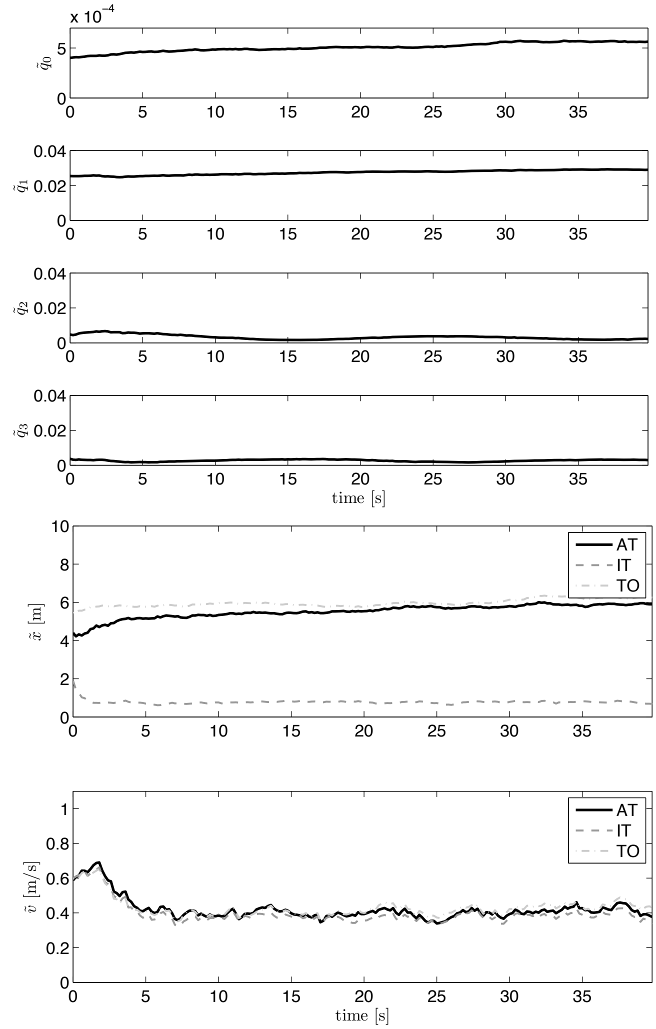

Fig. 5. RMSE for path $c = 0$, the left case in Figure 2. This path is just a circle and the orientation error is not observable in filter (AT) for this special case when the orientation error is an angle around a vertical vector from the sensor in the origin to the circle center. This causes the positioning error of filter (AT) to be similar to the result of filter (TO) where the orientation error is ignored.

that can have severe effects on the tracking performance since observations are incorrectly aligned with prior information expressed in a global reference frame. In this work it assumed that SLAM and image registration techniques could not be used due to bad/unstable image features. Instead a filter is proposed that estimates the orientation error and the road target state simultaneously by exploiting the knowledge about the road map. The position error is neglected since it is assumed to be very small compared with the range to the target. The filter is based on the well-known bootstrap particle filter, but if several targets are tracked the Rao-Blackwellized version is a better choice to handle the curse-of-dimensionality. Particle filters are very flexible and relatively straightforward to implement and adapt to non-linear systems, like road maps in this work. However, it is important to remember that it has bad support in low probability areas.

In the simulation examples it is shown that the tracking performance of a road target is close to the ideal case where the orientation error is known. However, observability is always an issue in vision based target tracking. In fact, the target tracking problem for a stationary angle-only sensor, e.g. a camera, is not observable and some external information is needed to support the estimation process. In this work the knowledge about the road path is used, but as seen in one example, there exist cases where the orientation error is unobservable. Such cases are quite rare, but still different paths can be more or less informative from an estimation point of view. Basically, curvy paths are better than straight or very smooth paths, and if the sensor platform is moving the conditions for target tracking is usually better.

References

[1] F. Gustafsson, U. Orguner, T. B. Schön, P. Skoglar, and R. Karlsson, *Handbook of Intelligent Vehicles*. Springer-Verlag London Limited,

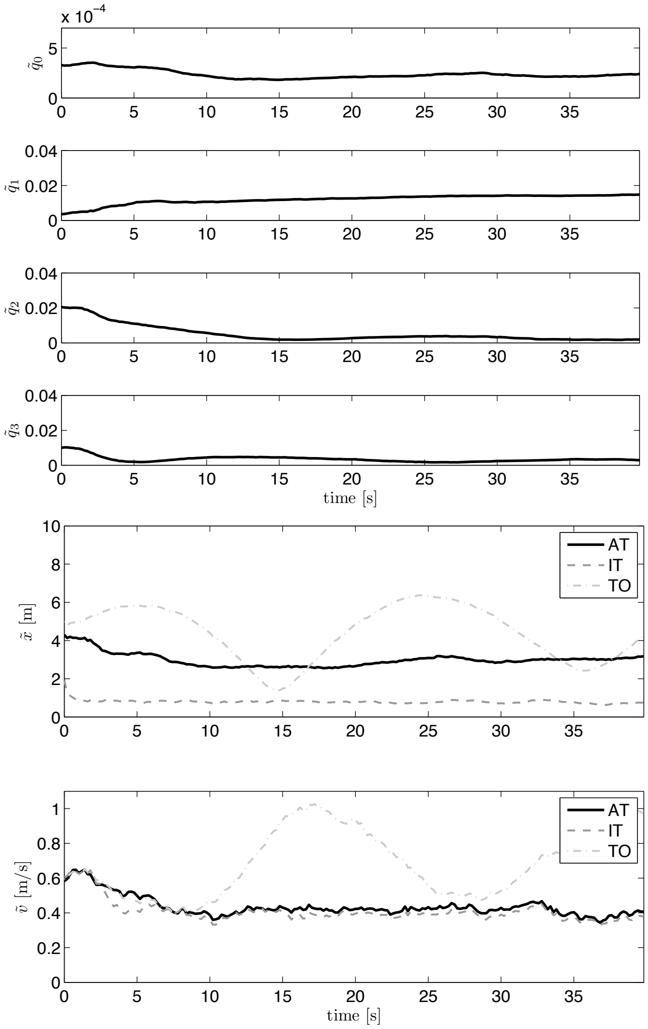

Fig. 6. RMSE for path $c = 0$. The same case as in Figure 5, but the orientation error is instead around a vector in the xy-plane. Although the information richness of the path is very low, the filter (AT) performs significantly better than the filter (TO).

2012, ch. Navigation and Tracking of Road-Bound Vehicles Using Map Support, to appear.

[2] M. Ulmke and W. Koch, "Road-Map assisted ground target tracking," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 42, no. 3, pp. 1264–1274, Oct. 2006.

[3] B. Ristic, S. Arulampalam, and N. Gordon, *Beyond the Kalman Filter - Particle filters for tracking applications*. Artech House, 2004.

[4] P. Skoglar, U. Orguner, D. Törnqvist, and F. Gustafsson, "Pedestrian tracking with an infrared sensor using road network information," *EURASIP J Adv. Signal Process.*, vol. 2012, no. 1, p. 26, 2012. [Online]. Available: http://asp.eurasipjournals.com/content/2012/1/26

[5] A. J. Davison, I. Reid, N. Molton, and O. Strasse, "MonoSLAM: Real-time single camera SLAM," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 29, no. 6, pp. 1052–1067, Jun. 2007.

[6] H. Durrant-Whyte and T. Bailey, "Simultaneous localization and mapping (SLAM): Part I," *IEEE Robot. Automat. Mag.*, vol. 13, no. 2, pp. 99–110, Jun. 2006.

[7] T. Bailey and H. Durrant-Whyte, "Simultaneous localization and mapping (SLAM): Part II," *IEEE Robot. Automat. Mag.*, vol. 13, no. 3, pp. 108–117, Sep. 2006.

[8] D.-G. Sim, R.-H. Park, R.-C. Kim, S. U. Lee, and I.-C. Kim, "Integrated position estimation using aerial image sequences," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 24, no. 1, pp. 1 –18, jan 2002.

[9] B. Zitova and J. Flusser, "Image registration methods: a survey," *Image Vis. Comput.*, vol. 21, no. 11, pp. 977 – 1000, 2003.

[10] D. Törnqvist, "Estimation and detection with applications to navigation," Dissertations No 1216, Dept. Electr. Eng, Linköpings universitet, Sweden, SE-581 83 Linköping, Sweden, Nov. 2008.

[11] J. B. Kuipers, *Quaternions and Rotation Sequences*. Princeton Univerity Press, 1999.

[12] N. J. Gordon, D. J. Salmond, and A. F. M. Smith, "Novel approach to nonlinear/non-Gaussian Bayesian state estimation," *IEE Proc.-F*, vol. 140, no. 2, pp. 107–113, Apr. 1993.

[13] S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for on-line non-linear/non-Gaussian Bayesian tracking," *IEEE Trans. Signal Processing*, vol. 50, no. 2, pp. 174–188, Feb. 2002.

[14] G. Kitagawa, "Monte Carlo filter and smoother for non-gaussian nonlinear state space models," *J. Comput. Graph. Stat.*, vol. 5, no. 1, pp. pp. 1–25, 1996. [Online]. Available: http://www.jstor.org/stable/1390750

[15] R. Chen and J. S. Liu, "Mixture Kalman filters," *J. Royal Statist. Soc. B*, vol. 62, no. 3, pp. 493–508, 2000.

[16] C. Andrieu and A. Doucet, "Particle filtering for partially observed Gaussian state space models," *J. Royal Statist. Soc. B*, vol. 64, no. 4, pp. 827–836, 2002.

[17] T. Schön, F. Gustafsson, and P.-J. Nordlund, "Marginalized particle filters for mixed linear/nonlinear state-space models," *IEEE Trans. Signal Processing*, vol. 53, no. 7, pp. 2279–2289, Jul. 2005.