

# Vehicle Motion Estimation Using an Infrared Camera

Emil Nilsson\* Christian Lundquist\*\* Thomas B. Schön\*\*  
David Forslund\* Jacob Roll\*

\* *Autoliv Electronics AB, SE-583 30, Linköping, Sweden*  
(e-mail: *firstname.lastname@autoliv.com*)

\*\* *Division of Automatic Control, Linköping University,*  
*SE-581 83, Linköping, Sweden (e-mail: {lundquist, schon}@isy.liu.se)*

---

**Abstract:** This paper is concerned with vehicle motion estimation. The problem is formulated as a sensor fusion problem, where the vehicle motion is estimated based on the information from a far infrared camera, inertial sensors and the vehicle speed. This information is already present in premium cars. This work is concerned with the off-line situation and the approach taken is to formulate the problem as a nonlinear least squares problem. In order to illustrate the performance of the proposed method experiments on rural roads in Sweden during night time driving have been performed. The results clearly indicates the efficacy of the approach.

*Keywords:* Motion estimation, Smoothing, Automotive industry, Vision, Far infrared camera.

---

## 1. INTRODUCTION

This work is concerned with the problem of estimating the vehicle motion using measurements from a far infrared (FIR) camera, along with proprioceptive sensors measuring acceleration, speed and yaw rate. FIR cameras are currently used for pedestrian detection in several different vehicle models already in series production. This implies that the rich sensor information available from the FIR camera is already available in the vehicle for free. Fig. 1 illustrates the difference between a visual image and a FIR image (from Autoliv's Night Vision system, which explains the pedestrian warning symbol) taken during night time driving. From this figure it is clear that there is a lot of information about the surroundings available (also during night time driving). The goal of this work is to show how this information can be used in order to compute smoothed estimates of the vehicle motion.



Fig. 1. This figure shows a visual image and a corresponding FIR image.

Initial studies along this line have already been performed in Schön and Roll (2009). That work indicates that the information from the FIR camera is indeed valuable for motion estimation in a filtering setting, i.e. when the on-line motion estimation problem is considered. More specifically, an extended Kalman filter (EKF) was used in order to compute the estimate. The present work targets the off-line problem, i.e. the problem of estimating the motion of the vehicle after the measurements have been collected. As always, it is expected that the smoothed estimate is better than the filtered estimate, since more information is available.

The way in which the camera data is utilised is highly related to the, by now, fairly well studied Simultaneous Localization and Mapping (SLAM) problem, see e.g. (Thrun et al., 2005; Davison et al., 2007; Durrant-Whyte and Bailey, 2006; Bailey and Durrant-Whyte, 2006). However, the current work only aims at estimating the motion of the vehicle and not the map, which implies that this work has even stronger ties to the visual odometry problem (Cheng et al., 2006; Nistér et al., 2006). Furthermore, a sensor fusion problem is considered here, where measurements from several different sensors, not just the camera, are used. The problem is formulated as a nonlinear least squares problem, inspired by the work of Dellaert and Kaess (2006).

An important component of the proposed solution is the motion model for the vehicle. Some effort has therefore been spent in deriving and evaluating an appropriate vehicle motion model. The perhaps slightly non-standard component introduced in this model is the vehicle pitch dynamics. More specifically, a constant offset and the influence from acceleration, are suggested by Dickmanns (2007).

## 2. MODELLING

The system is modeled by a state space model. At time  $t$  the vehicle state is denoted  $x_t^v$  and the input  $u_t$ , resulting in the vehicle motion model

$$x_t^v = f(x_{t-1}^v, u_t) + w_t, \quad (1a)$$

where  $w_t$  is Gaussian process noise. The position at time  $t$  of the  $j$ th landmark is parametrized by its state  $x_{j,t}^l$  and the landmark model is

$$x_{j,t}^l = x_{j,t-1}^l, \quad (1b)$$

since it is assumed that all landmarks are stationary. At time  $t$ , the vehicle measurements  $y_t^v$  are given by

$$y_t^v = h^v(x_t^v) + e_t^v, \quad (1c)$$

where  $e_t^v$  is Gaussian measurement noise, and the landmark measurement  $y_{j,t}^l$  of the  $j$ th landmark is given by

$$y_{j,t}^l = h^l(x_t^v, x_{j,t}^l) + e_{j,t}^l, \quad (1d)$$

where  $e_{j,t}^l$  is Gaussian measurement noise.

### 2.1 Coordinate Frames

There are three relevant coordinate frames for the combined vehicle and camera system:

- **World ( $w$ ):** This is considered an inertial frame and is fixed to the surroundings of the vehicle.
- **Vehicle body ( $b$ ):** This frame is fixed to the vehicle, with its origin located in the middle of the rear axis. Coordinate frame  $b$  coincides with  $w$  at the start of a scenario.
- **Camera ( $c$ ):** This frame is fixed relative to  $b$  and is positioned in the optical center of the camera.

The rotation matrix  $R(\alpha, \beta, \gamma)$  transforms coordinates from coordinate frame  $B$  to coordinate frame  $A$ , where the orientation of  $B$  relative to  $A$  is  $\alpha$  (yaw),  $\beta$  (pitch) and  $\gamma$  (roll).

### 2.2 Vehicle Process Models

This section starts by describing the vehicle process model used in Schön and Roll (2009), from now on referred to as the basic model, followed by a few model extension proposals aiming at increasing the pose estimation accuracy.

With the states described in Table 1, the vehicle state vector at time  $t$  is given by

$$x_t^v = (p_t^T \ v_{x,t} \ \psi_t \ \delta_t \ \alpha_t \ \varphi_t)^T, \quad (2a)$$

$$p_t = (p_{x,t} \ p_{y,t} \ p_{z,t})^T. \quad (2b)$$

The input signal  $u_t$  is given by

$$u_t = \dot{v}_{x,t}, \quad (3)$$

where the vehicle acceleration  $\dot{v}_{x,t}$  is measured. By treating  $\dot{v}_{x,t}$  as an input signal instead of as a measurement it does not have to be incorporated in the vehicle state.

With  $T$  as the sampling time,  $L$  as the wheel base of the vehicle and  $C$  as a pitch damping parameter, the process model becomes

Table 1. Vehicle states.

| State     | Description  |
|-----------|--|
| $p$       | Vehicle position in world coordinates.                               |
| $v_x$     | Velocity of the vehicle in its longitudinal direction.               |
| $\psi$    | Yaw angle (z-axis rotation), relative to the world coordinate frame. |
| $\delta$  | Front wheel angle, relative to the vehicle's longitudinal direction. |
| $\alpha$  | Road pitch angle, relative to the world coordinate frame xy-plane.   |
| $\varphi$ | Pitch angle of the vehicle, relative to the road.                    |

$$f(x_t^v, u_{t+1}) = \begin{pmatrix} p_{x,t} + T v_{x,t} \cos \psi_t \cos \alpha_t \\ p_{y,t} + T v_{x,t} \sin \psi_t \cos \alpha_t \\ p_{z,t} - T v_{x,t} \sin \alpha_t \\ v_{x,t} + T u_{t+1} \\ \psi_t + \frac{T v_{x,t}}{L} \tan \delta_t \\ \delta_t \\ \alpha_t \\ C \varphi_t \end{pmatrix}. \quad (4)$$

The vehicle process noise  $w_t$  is independent and Gaussian, according to

$$w_t \sim \mathcal{N}(0, Q(x_{t-1}^v)), \quad (5a)$$

$$Q(x_t^v) = B(x_t^v) Q^\omega B(x_t^v)^T, \quad (5b)$$

$$B(x_t^v) = \begin{pmatrix} T \cos \psi_t \cos \alpha_t & 0 \\ T \sin \psi_t \cos \alpha_t & 0 \\ T \sin \alpha_t & 0 \\ 1 & 0 \\ 0 & I^{4 \times 4} \end{pmatrix}, \quad (5c)$$

$$Q^\omega = \text{diag}(q^{v_x} \ q^\psi \ q^\delta \ q^\alpha \ q^\varphi), \quad (5d)$$

where all the  $q$ -variables are process noise variance parameters.

The rest of this section describes the extensions to the vehicle process model tested during this work. Note that these model extensions can be used together in any combination. Some new notation has to be introduced: In order to describe the vehicle process model for one of the vehicle states, superscripts are used. For example: for the yaw angle ( $\psi$ ) process model the notation  $f^\psi$  is used.

*Constant Offset in Car Pitch Angle* Since the stationary pitch angle of the camera, relative to the road, might be non-zero, due to the current load in the vehicle or misaligned camera mounting, the state vector is augmented with a state for the camera pitch offset  $\varphi^0$ . This offset state is interpreted as the stationary car pitch angle, around which the car pitch angle oscillates, according to

$$f^\varphi(x_t^v, u_{t+1}) = C(\varphi_t - \varphi_t^0) + \varphi_t^0, \quad (6a)$$

$$f^{\varphi^0}(x_t^v, u_{t+1}) = \varphi_t^0, \quad (6b)$$

where  $C$  is the pitch damping parameter from (4).

The camera pitch offset models a constant offset angle, so the process noise variance for  $\varphi_t^0$  is zero. The process noise variance for  $\varphi$  is independent of whether  $\varphi^0$  is included in the vehicle state and process model or not.

*Acceleration Offset in Car Pitch Angle* The vehicle acceleration significantly influences the pitch angle of the car. The model of this effect is that the stationary car pitch angle, when the acceleration  $u$  is constant, becomes

$Ku$ , where  $K$  is a parameter that depends on the vehicle geometry. This leads to

$$f^\varphi(x_t^v, u_{t+1}) = C(\varphi_t - Ku_{t+1}) + Ku_{t+1}. \quad (7)$$

The process noise variance for  $\varphi$  should not be changed when adding the car pitch acceleration offset to the process model.

*Roll Angle* For several reasons, such as that curves of country roads are banked, and that the car may roll when driving on uneven roads, letting the roll angle be constant at zero, might be an inadequate approximation. The process model for the combined roll angle  $\gamma$  of the car and road is given by

$$f^\gamma(x_t^v, u_{t+1}) = \gamma_t. \quad (8)$$

Since the roll and pitch angles of automobiles have similar behaviour in terms of amplitude and natural frequency, the process noise variance for the roll angle is set to be approximately the same as the basic model car pitch process noise variance.

### 2.3 Landmark Parametrization

In order to extract measurements from the FIR images a Harris corner detector (Harris and Stephens, 1988) is used to find features, and normalized cross-correlation (NCC), see e.g. Ma et al. (2004), is used to associate previously extracted features with new images. The landmarks are parameterized using the so-called inverse depth parameterization introduced by Montiel et al. (2006). More specifically, the landmark states are described in Table 2 and the landmark state vector is given by

$$x_t^l = \begin{pmatrix} x_{j_t(1),t}^l \\ x_{j_t(2),t}^l \\ \vdots \\ x_{j_t(M_t),t}^l \end{pmatrix}, \quad (9a)$$

$$x_{j,t}^l = ((k_{j,t}^w)^T \theta_{j,t}^w \phi_{j,t}^w \rho_{j,t})^T, \quad (9b)$$

$$k_{j,t}^w = (k_{j,t,x}^w \ k_{j,t,y}^w \ k_{j,t,z}^w)^T. \quad (9c)$$

The landmark state  $x_{j,t}^l$  is a parametrization of the position  $l_{j,t}^w$  of landmark  $j$  at time  $t$ . The relationship between position and state, with landmark position given in world coordinates, is given by

$$l_{j,t}^w = k_{j,t}^w + \frac{1}{\rho_{j,t}} \underbrace{\begin{pmatrix} \cos \phi_{j,t}^w \cos \theta_{j,t}^w \\ \cos \phi_{j,t}^w \sin \theta_{j,t}^w \\ \sin \phi_{j,t}^w \end{pmatrix}}_{m_{j,t}^w}. \quad (10)$$

Table 2. Landmark states.

| State      | Description   |
|------------|---|
| $k^w$      | The position in world coordinates of the camera at the time when the landmark was first seen.   |
| $\theta^w$ | The azimuth angle of the landmark as seen from $k^w$ , relative to world coordinate frame directions.   |
| $\phi^w$   | The elevation angle of the landmark as seen from $k^w$ , relative to world coordinate frame directions, with positive angles towards the positive z-axis. |
| $\rho$     | The inverse depth (which is the inverse of the distance) from $k^w$ to the landmark.  |

At time  $t$  there are  $M_t$  visible landmarks. Visible means that the landmark has been measured; a landmark may very well be non-visible although it is present in the FIR image, but it cannot be visible if it is not in the image. The landmark index of the visible landmark number  $i \in \{1, 2, \dots, M_t\}$  at time  $t$  is denoted  $j_t(i)$ .

### 2.4 Measurement Model

The measurement model (1c) relating to the proprioceptive sensors of the vehicle is given by

$$h^v(x_t^v) = \begin{pmatrix} v_{x,t} \\ \psi_t \end{pmatrix} = \begin{pmatrix} v_{x,t} \\ \frac{v_{x,t}}{L} \tan \delta_t \end{pmatrix}, \quad (11)$$

where  $L$  is the wheel base of the vehicle. Furthermore, the measurement model (1d) for landmark  $j$  at time  $t$  is given by

$$h^l(x_t^v, x_{j,t}^l) = \mathcal{P}_n(p_{j,t}^c) = \frac{1}{p_{j,t,x}^c} \begin{pmatrix} p_{j,t,y}^c \\ p_{j,t,z}^c \end{pmatrix}, \quad (12a)$$

$$\begin{aligned} p_{j,t}^c &= \begin{pmatrix} p_{j,t,x}^c \\ p_{j,t,y}^c \\ p_{j,t,z}^c \end{pmatrix} = p^c(x_t^v, x_{j,t}^l) = \\ &= \frac{1}{\rho_{j,t}} R^{cb} (R^{bw} (\rho_{j,t} (k_{j,t}^w - p_t) + m_{j,t}^w) - \rho_{j,t} c^b), \end{aligned} \quad (12b)$$

$$R^{cb} = R(\alpha^c, \beta^c, \gamma^c)^T, \quad (12c)$$

$$R^{bw} = R(\psi_t, \alpha_t + \varphi_t, \gamma_t)^T, \quad (12d)$$

where  $c^b$  is the position of the camera in the vehicle body coordinate frame, and  $\mathcal{P}_n(p^c)$  is the so-called normalized pinhole projection of a point  $p^c$ , which is given in camera coordinates. Furthermore,  $\mathcal{P}_n$  generates normalized camera coordinates, and  $\alpha^c$ ,  $\beta^c$  and  $\gamma^c$  are the yaw, pitch and roll angles of the camera, relative to the vehicle body coordinate frame.

Both of the two measurement noises  $e_t^v$  and  $e_{j,t}^l$  in (1c) and (1d) are independent and Gaussian, according to

$$e_t^v \sim \mathcal{N}(0, S^v), \quad S^v = \text{diag}(s^{v_x} \ s^{v_\psi}), \quad (13a)$$

$$e_{j,t}^l \sim \mathcal{N}(0, S^l), \quad S^l = \text{diag}(s^c \ s^c), \quad (13b)$$

where all the  $s$ -variables are measurement noise variance parameters.

The translation between pixel coordinates  $(\tilde{y} \ \tilde{z})^T$  and normalized camera coordinates  $(y \ z)^T$ , in which the landmark measurements  $y_{j,t}^l$  are given, is

$$\begin{pmatrix} y \\ z \end{pmatrix} = \begin{pmatrix} \frac{\tilde{y} - \tilde{y}^{ic}}{f_y} \\ \frac{\tilde{z} - \tilde{z}^{ic}}{f_z} \end{pmatrix}, \quad (14)$$

where  $(\tilde{y}^{ic} \ \tilde{z}^{ic})^T$  denotes the image center, and  $f_y$  and  $f_z$  are the focal lengths (given in pixels) in the  $y$ -direction and the  $z$ -direction, respectively.

## 3. NONLINEAR LEAST SQUARES FORMULATION AND SOLUTION

In this section the nonlinear least squares problem that is solved in order to find the smoothed estimates of the vehicle motion is formulated. In other words, the vehicle

states and the landmark states are estimated simultaneously, based on the information in all the measurements. Notation that will be needed is first introduced.

The complete state vector  $x$  is given by

$$x = \left( (x^v)^T \ (x^l)^T \right)^T, \quad (15a)$$

where  $x^v$  denotes the vehicle states (2) for all time steps  $t = 1, 2, \dots, N$  and  $x^l$  denotes the stationary landmark states. To be specific,

$$x^v = \left( (x_1^v)^T \ (x_2^v)^T \ \dots \ (x_N^v)^T \right)^T, \quad (15b)$$

$$x^l = \left( (x_{j(1)}^l)^T \ (x_{j(2)}^l)^T \ \dots \ (x_{j(M)}^l)^T \right)^T, \quad (15c)$$

$$x_j^l = (\theta_j^v \ \phi_j^w \ \rho_j)^T. \quad (15d)$$

Furthermore, let  $\bar{x}$  denote the current estimate of  $x$ . Note that instead of including the camera positions in the landmark states  $x_j^l$ , the vehicle state  $x_{t_c(j)}^v$ , from the time  $t_c(j)$  when landmark  $j$  was first seen, may be used. This is possible since the camera is rigidly attached to the car. In other words there is a known transformation  $g$ ,

$$g(x_{t_c(j)}^v, x_j^l) = \begin{pmatrix} p_{t_c(j)} + R^{wb} c^b \\ x_j^l \end{pmatrix}, \quad (16a)$$

$$R^{wb} = R(\psi_{t_c(j)}, \alpha_{t_c(j)} + \varphi_{t_c(j)}, \gamma_{t_c(j)}), \quad (16b)$$

which returns the complete landmark state (i.e. camera position, azimuth angle, elevation angle and inverse depth).

Due to the fact that the number of landmark measurements  $y_{j,t}^l$  are not the same for every time step,  $k$  is used to enumerate the complete series of landmark measurements. The notation  $j_k$  for the index of the landmark associated with measurement  $k$  is introduced, and similarly  $t_k$  for the time when measurement  $k$  was acquired. Using the notation just introduced,  $y_{j_k, t_k}^l$  is the landmark measurement number  $k$ .

The nonlinear least squares problem is solved by using a locally linear approximation, obtained using a linearization around  $x = \bar{x}$ . Linear approximations of the process model and the measurement model are straightforwardly obtained according to

$$\bar{x}_t^v + \delta x_t^v = f(\bar{x}_{t-1}^v, u_t) + F_t \delta x_{t-1}^v + w_t, \quad (17a)$$

$$y_t^v = h^v(\bar{x}_t^v) + H_t^v \delta x_t^v + e_t^v, \quad (17b)$$

$$y_{j_k, t_k}^l = h^l(\bar{x}_{t_k}^v, g(\bar{x}_{t_c(j_k)}^v, \bar{x}_{j_k}^l)) + H_k^l \delta x_{t_k}^v + H_k^{l,c} \delta x_{t_c(j_k)}^v + J_k \delta x_{j_k}^l + e_{j_k, t_k}^l, \quad (17c)$$

where

$$F_t = \left. \frac{\partial f(x_{t-1}^v, u_t)}{\partial x_{t-1}^v} \right|_{x_{t-1}^v = \bar{x}_{t-1}^v}, \quad (18a)$$

$$H_t^v = \left. \frac{\partial h^v(x_t^v)}{\partial x_t^v} \right|_{x_t^v = \bar{x}_t^v}, \quad (18b)$$

$$H_k^l = \left. \frac{\partial h^l(x_t^v, g(\bar{x}_{t_c(j_k)}^v, \bar{x}_{j_k}^l))}{\partial x_t^v} \right|_{x_t^v = \bar{x}_{t_k}^v}, \quad (18c)$$

$$H_k^{l,c} = \left. \frac{\partial h^l(\bar{x}_{t_k}^v, g(x_{t_c}^v, \bar{x}_{j_k}^l))}{\partial x_{t_c}^v} \right|_{x_{t_c}^v = \bar{x}_{t_c(j_k)}^v}, \quad (18d)$$

$$J_k = \left. \frac{\partial h^l(\bar{x}_{t_k}^v, g(\bar{x}_{t_c(j_k)}^v, x_{j_k}^l))}{\partial x_{j_k}^l} \right|_{x_{j_k}^l = \bar{x}_{j_k}^l}, \quad (18e)$$

are the Jacobians of the process model and measurement model. Define the residuals according to

$$a_t = \bar{x}_t^v - f(\bar{x}_{t-1}^v, u_t), \quad (19a)$$

$$c_t^v = y_t^v - h^v(\bar{x}_t^v), \quad (19b)$$

$$c_k^l = y_{j_k, t_k}^l - h^l(\bar{x}_{t_k}^v, g(\bar{x}_{t_c(j_k)}^v, \bar{x}_{j_k}^l)), \quad (19c)$$

which allows the following least squares problem to be formulated,

$$\begin{aligned} \delta x^* &= \arg \min_{\delta x} \left( \sum_t \left( \|w_t\|_{Q_t^{-1}}^2 + \|e_t^v\|_{(S^v)^{-1}}^2 \right) + \right. \\ &\quad \left. + \sum_k \|e_{j_k, t_k}^l\|_{(S^l)^{-1}}^2 \right) = \\ &= \arg \min_{\delta x} \left( \sum_t \left( \|F_t \delta x_{t-1}^v - \delta x_t^v - a_t\|_{Q_t^{-1}}^2 + \right. \right. \\ &\quad \left. \left. + \|H_t^v \delta x_t^v - c_t^v\|_{(S^v)^{-1}}^2 \right) + \right. \\ &\quad \left. + \sum_k \|H_k^l \delta x_{t_k}^v + H_k^{l,c} \delta x_{t_c(j_k)}^v + J_k x_{j_k}^l - c_k^l\|_{(S^l)^{-1}}^2 \right), \end{aligned} \quad (20)$$

where  $Q_t = Q(\bar{x}_{t-1}^v)$  (see (5)), and the vector norm  $\|\cdot\|_{P^{-1}}$  is the Mahalanobis distance, i.e.,

$$\begin{aligned} \|e\|_{P^{-1}}^2 &= e^T P^{-1} e = (P^{-T/2} e)^T (P^{-T/2} e) = \\ &= \|P^{-T/2} e\|_2^2. \end{aligned} \quad (21)$$

Note that  $\bar{x}_0^v$  is at time  $t = 1$  required, but not known. By letting  $F_1 = 0$ ,  $a_1 = 0$  and  $Q_1 = P_{1|0}^v$ , which is the initial vehicle state covariance for the extended Kalman filter,  $\bar{x}_0^v$  is no longer required, and the resulting term  $\|\delta x_1^v\|_{Q_1^{-1}}^2$  makes sure that the smoothed state estimate  $\bar{x}^s$  stays reasonably close to  $\bar{x}$ .

By collecting all the weighted Jacobians ( $Q_t^{-T/2} F_t$ ,  $Q_t^{-T/2} (S^v)^{-T/2} H_t^v$ , ...) in a matrix  $A$ , and stacking all the weighted residuals ( $Q_t^{-T/2} a_t$ ,  $(S^v)^{-T/2} c_k^l$  and  $(S^l)^{-T/2} c_t^v$ ) in a vector  $b$ , the least squares problem (20) can be rewritten in the form of a standard linear least squares problem, according to

$$\delta x^* = \arg \min_{\delta x} \|A \delta x - b\|_2^2. \quad (22)$$

The resulting smoothed state  $\bar{x}^s$  is now obtained according to

$$\bar{x}^s = \bar{x} + \delta x^*. \quad (23)$$

In order to get good accuracy for the state estimate, the procedure described above is iterated, using  $\hat{x}^s$  as a starting point for the next iteration, until  $\delta x^*$  becomes smaller than some predefined threshold. The initial guess is provided by the extended Kalman filter, as derived in Schön and Roll (2009) and further elaborated in Nilsson (2010).

#### 4. EXPERIMENTAL RESULTS

In order to illustrate the performance of the smoothed estimates 12 measurement sequences recorded during nighttime driving on rural roads in Sweden have been used. There is no ground truth available. However, the results still indicates that the FIR camera<sup>1</sup> is very useful in order to solve the vehicle motion estimation problem under study. This will be shown in two ways. First, the smoothed estimates of the vehicle position are reprojected onto an image, i.e., the plot shows the estimated position of the vehicle expressed in the world coordinate frame  $\hat{p}_t^w$ . This is a direct, but non-quantified validation and also to some extent subjective. However, it is a clear indication that the smoothed estimates provide a good estimate of the vehicle motion. An example of this trajectory visualization is shown in Fig. 2, and as expected the estimates from the smoothing approach (solid) appears to be better than the estimates based on the filtering (EKF) approach (dash-dotted). The second performance measure is the root



Fig. 2. The vehicle motion estimates (solid: from smoothing, dash-dotted: from filtering) are here reprojected into an image.

mean squared landmark measurement noise (i.e. the mean landmark measurement residuals, given in pixels). This is an indirect, but quantified measure of the estimation accuracy and the result is presented in Fig. 3, which compares the mean landmark measurement residuals for the smoothed and the filtered estimates, for all combinations of sequences and model extensions. The lines in the figure illustrates the ratios of the measurement residuals for the smoothed estimate compared to the filtered estimate. The

<sup>1</sup> The FIR camera used in this work registers radiation in the far infrared region at 30 Hz, with a resolution of  $320 \times 240$  pixels, and a spectral range of 8–14  $\mu\text{m}$ .

results presented in the figure indicates that the smoothed estimate is better than the filtered estimate, as expected.

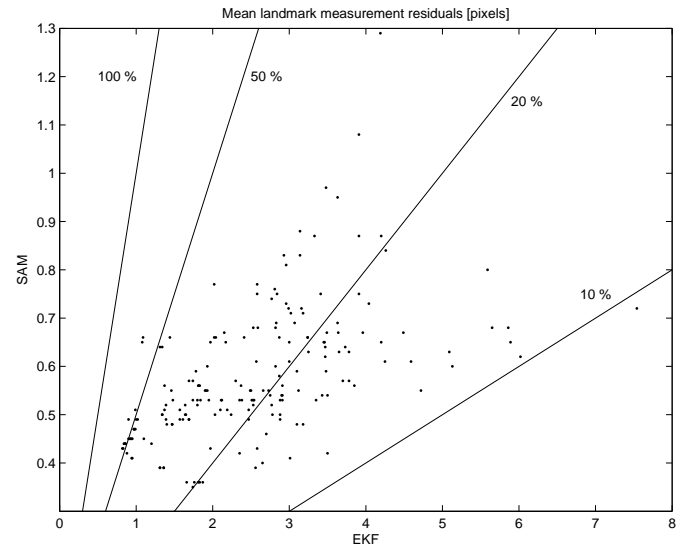


Fig. 3. This figure compares the mean landmark measurement residuals for the smoothed and the filtered estimates.

The change in performance due to the model extensions, in terms of mean landmark measurement residuals, is presented in Fig. 4, where the lines illustrates equal performance when using and not using the particular model extension.

It is important to note that not all sequences contain behaviour that the model extensions are meant to handle. If, for example, the true car pitch offset is zero, the offset model extension cannot be expected to improve the pose estimates for that particular sequence. The same goes for the acceleration offset in the car pitch angle for any sequence in which the vehicle moves at constant speed.

Fig. 4a shows the landmark measurement residuals, indicating that the model extension can give some improvements in pose estimation accuracy. The magnitude of the improvement is of course heavily dependent on how big the car pitch angle offset really is, which is why this model extension cannot provide improvements in accuracy for all sequences. It can be seen in Fig. 4b that the model extension for offset in pitch angle due to acceleration is able to reduce the value of the landmark measurement residual measure, indicating improved estimation accuracy. The roll model extension differs from the others in that it introduces a new dimension in the model of the vehicle motion. The results in Fig. 4c reflect this by showing that it is the model extension which provides the largest performance improvement, at least in terms of landmark measurement residuals.

#### 5. CONCLUSIONS AND FUTURE WORK

This work has showed how information from a FIR camera can be used to compute a smoothed estimate of the vehicle motion, by solving an appropriate nonlinear least squares problem. The results were evaluated using measurements from test drives on rural roads during night time.

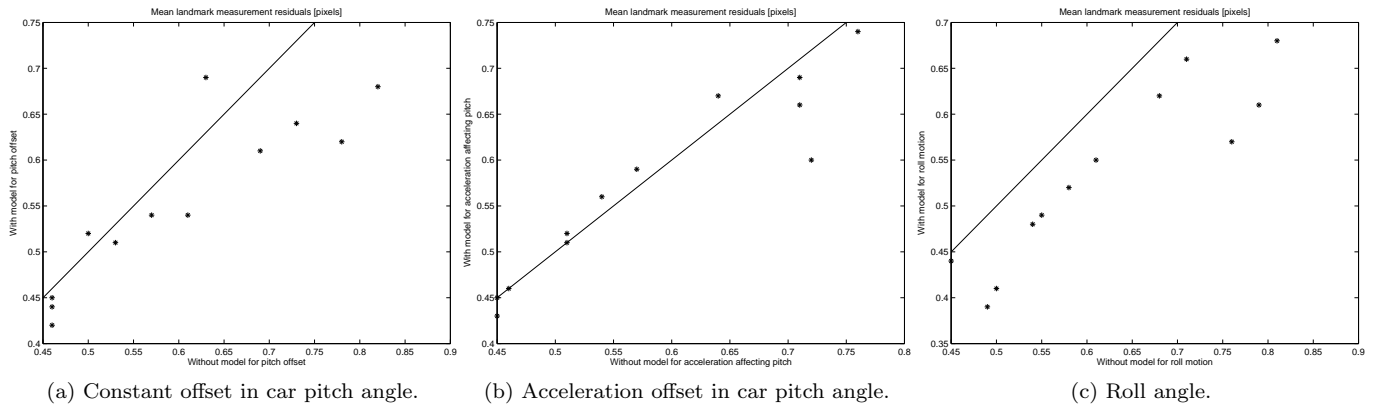


Fig. 4. These figures compares the mean landmark measurement residuals for using or not using the different model extensions.

Furthermore, it was also showed that including the roll motion in the vehicle motion model clearly improved the estimation accuracy. Furthermore, modelling the offset in the car pitch angle, both constant offset and offset due to acceleration, can result in slightly improved estimation accuracy.

During this work it has become more and more clear that the problem with erroneous association of features is an important issue to be dealt with, since the feature measurements are the foundation on which the resulting estimate relies. The natural way to improve the landmark association quality is to make use of a better outlier rejection scheme. Another way to improve the measurements derived from the camera images is to extract image measurements of yaw, pitch and roll change rates that are not based on landmarks, but instead uses e.g. phase correlation.

Regarding computational demand, the current implementation of the smoothing algorithm is, when applied to a 15 seconds long data sequence and using a standard desktop computer, approximately a factor 50 slower than required for real-time applications. It is possible that this obstacle can be overcome by doing the following:

- Implement the iSAM algorithm (Kaess et al., 2008) which performs the smoothing and mapping incrementally, and uses e.g. matrix factorization to reduce computational demand.
- Derive and use analytic expressions for the Jacobians of the process and measurement models.

#### ACKNOWLEDGEMENTS

The second and third authors were supported by CADICS, a Linneaus Center funded by the Swedish Research Council.

#### REFERENCES

Bailey, T. and Durrant-Whyte, H. (2006). Simultaneous localization and mapping (SLAM): Part II. *IEEE Robotics & Automation Magazine*, 13(3), 108–117.

Cheng, Y., Maimone, M.W., and Matthies, L. (2006). Visual odometry on the Mars exploration rovers. *IEEE Robotics & Automation Magazine*, 13(2), 54–62.

Davison, A.J., Reid, I., Molton, N., and Strasse, O. (2007). MonoSLAM: Real-time single camera SLAM. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(6), 1052–1067.

Dellaert, F. and Kaess, M. (2006). Square root SAM: Simultaneous localization and mapping via square root information smoothing. *International Journal of Robotics Research*, 25(12), 1181–1203.

Dickmanns, E. (2007). *Dynamic Vision for Perception and Control of Motion*. Springer, Secaucus, NJ, USA.

Durrant-Whyte, H. and Bailey, T. (2006). Simultaneous localization and mapping (SLAM): Part I. *IEEE Robotics & Automation Magazine*, 13(2), 99–110.

Harris, C. and Stephens, M. (1988). A combined corner and edge detector. In *Proceedings of the 4th Alvey Vision Conference*, 147–151. Manchester, UK.

Kaess, M., Ranganathan, A., and Dellaert, F. (2008). iSAM: Incremental Smoothing and Mapping. *IEEE Transactions on Robotics*, 24(6), 1365–1378.

Ma, Y., Soatto, S., Kosecka, J., and Sastry, S. (2004). *An Invitation to 3-D Vision: From Images to Geometric Models*. Springer.

Montiel, J., Civera, J., and Davison, A. (2006). Unified inverse depth parametrization for monocular SLAM. In *Proceedings of Robotics: Science and Systems (RSS)*. Philadelphia, USA.

Nilsson, E. (2010). *An Optimization Based Approach to Visual Odometry Using Infrared Images*. Master’s thesis, Linköping University.

Nistér, D., Neroditsky, O., and Bergen, J. (2006). Visual odometry for ground vehicle applications. *Journal of Field Robotics*, 23(1), 3–20.

Schön, T.B. and Roll, J. (2009). Ego-motion and indirect road geometry estimation using night vision. In *Proceedings of the IEEE Intelligent Vehicles Symposium (IV)*, 30–35. Xi’an, Shaanxi, China.

Thrun, S., Burgard, W., and Fox, D. (2005). *Probabilistic Robotics*. Intelligent Robotics and Autonomous Agents. The MIT Press, Cambridge, MA, USA.