

Motion Interpretation for In-Car Vision Systems

Marinus B. van Leeuwen, Frans C.A. Groen

Faculty of Science, University of Amsterdam, Amsterdam, The Netherlands, {rien.groen}@science.uva.nl

Abstract

In this paper we consider a single camera, attached to a moving vehicle. Based on the observations, we want to estimate the time-to-contact of vehicles around us. We propose three different approaches for the estimation of this parameter. The measurement input for these approaches respectively consists of the horizontal position in the image plane of a point observed on a vehicle, the vertical position and the observed vehicle width. This measurement data can robustly be obtained in practice. It is possible to automatically select the approach that provides the most accurate result for a given situation. The influence of disturbances in the observations in practice is suppressed by means of temporal filtering. Due to the similar structure of the three approaches, the same temporal filtering technique can be applied. The practical value of the algorithms is illustrated by means of real video data.

1 Introduction

For in-car intelligence systems we are interested in how the behaviour of other vehicles on the road limits our driving possibilities. In this paper we consider the application where a single camera is attached to a moving vehicle, typically driving on a highway. Based on observations with the camera, we want to determine the motion of other vehicles around us, relative to our own motion. In this paper we focus on the motion component, parallel to our own motion. In literature, this parameter is usually referred to as the time-to-contact. For many applications the time-to-contact is an important parameter. It expresses the "nearness" of an object relative to the observer.

The camera provides us with information about induced motion components in the image plane of the camera. This motion is the projection of the real world motion of a point on the moving image plane. In order to understand the observed motion of an object point (e.g. located on a vehicle), one needs to know which part of the induced motion is caused by the ego-motion of the camera and which part is the result of relative object motion. In other words:

to estimate the observed motion caused by relative motion of a vehicle, the image data should implicitly or explicitly be stabilized.

In the following section we briefly address the issue of camera stabilization. Section 3 presents algorithms for the estimation of the motion parameter time-to-contact. In section 4 we show several experimental results based on a real video data. The paper is concluded with a discussion in section 5.

2 Camera stabilization

In this section we briefly summarize our camera stabilization method. A more detailed description can be found in [7]. We start with the definition of 2 co-ordinate systems, attached to the nodal point of the camera: the real world co-ordinate system (X, Y, Z) and the observer co-ordinate system (x, y, z). Both are illustrated in figure 1. The Z -axis points in the motion direction of our own vehicle. The z -axis of the observer frame corresponds with the viewing-direction of the camera. In this paper, the position of a point in the image plane of the camera is expressed by the co-ordinate (r_x, r_y) (: observer co-ordinate (r_x, r_y, F) , with F the focal length of the camera).

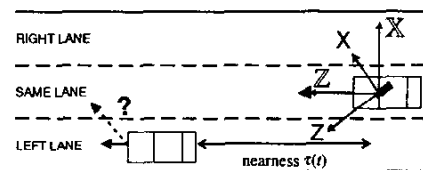


Figure 1: The behaviour of a vehicle is expressed relative to our own motion.

The motion of our vehicle is predominantly translational. Two aspects of the ego-motion of the camera are important for interpretation of the motion observed in the image plane: (i) the (average) angle between driving direction of the vehicle and the viewing direction of the camera and (ii) the (small) camera rotations around the mean viewing direction. Our estimation techniques employ data observed in the background. In order to be robust to the large

diversity in appearance of the background only moderate assumptions are made on the appearance of this data.

The dynamical characteristics of the mean viewing direction and the inter-frame rotations differ considerably, asking for different estimation techniques. The mean viewing direction can be estimated once in a while. It can be derived directly from the position 'in the image plane' of the so called Focus-of-Expansion (FOE). Points observed in the background seem to emerge from this point. Every time our own vehicle is moving in a straight lane, the observed motion of background points in the image plane is (approximately) along straight lines. These lines virtually intersection in the FOE. We apply least squares estimation techniques to estimate the position of the FOE.

The estimation of the (small) inter-frame rotations of the camera can only be based on a small number of images making the solution more sensitive for inaccuracies in its input. Therefore, we base our method on the estimation of the optical flow field of background points, and not on only a small number of tracks. The redundancy in the estimated motion field is exploited to obtain a robust estimation of the inter-frame rotations. An iteratively re-weighted orthogonal least squares technique is used.

The information about the viewing direction and the inter-frame camera rotations makes accurate stabilization of the video stream possible. With this information we can transform the observations into the situation where the camera is looking straight ahead (i.e. observer co-ordinate frame aligned with the real world co-ordinate frame). This justifies the simplified camera setup assumed in the next section.

3 Motion interpretation

3.1 The time-to-contact

Both our own motion and that of other vehicles is mainly in the Z-direction. The relative motion of a vehicle, measured in this direction, defines whether a vehicle approaches, retreats or stays at the same distance from the camera. The time-to-contact is defined in this direction. This parameter, denoted by $\tau(t)$, expresses how much time elapses until a vehicle reaches us. In case of an approaching vehicle, τ is positive. In case of retreating vehicles, τ is negative. τ is plus or minus infinity when a vehicle stays at approximately the same distance.

Different approaches to estimate τ have been described in literature. Examples of general estimation approaches for τ are optical flow based approaches [10][8], feature (point/line) based approaches [2][4][9] and closed contour based approaches [3]. However,

the kind of image features these general approaches rely on and/or the sensitive for camera rotations make them impractical for our application. Three alternative approaches have been proposed for our application. These approaches are based on:

1. the vehicle width ([12][5]).
2. the vertical position of the vehicle relative to the FOE ([1][5][6]).
3. the horizontal position of the vehicle relative to the FOE, when assuming parallel motion ([6]).

The first approach can be seen as a variation on the closed contour approaches. When the perspective projection model is applied, the following equation for $\tau(t)$ holds ([12][5]):

$$\tau(t) = w(t)/\dot{w}(t) \quad (1)$$

Equation 1 only provides us with an approximation for τ in case of deformation. In [7] we show that deformation can be neglected for our application. This approach based on the vehicle width is rather insensitive for camera rotations. Major advantage over the general closed contour approach is that the vehicle width is more easily determined than a closed contour. Draw back of this approach is its sensitivity for object occlusion.

Instead of the vehicle width, also the distance between headlights, the width of the license plate, etc. can be used. Notice that for practical reasons the vehicle height is less suited for our application. First of all the upper and lower border of a vehicle are generally more difficult to detect compared with the side borders. Furthermore, whereas camera pitch is more severe than yaw, estimation of τ based on vehicle height will be less accurate.

The second approach can be seen as a feature based approach combined with the assumption of planar motion. Knowledge about the position of the FOE together with the planar motion assumption provides enough information to relate the vertical position of a point in the image plane to its distance to the camera. Consider a camera, positioned at height H above the road surface and with viewing direction aligned to the driving direction. When a vehicle is observed above image line $r_y(t)$ in the image plane, we can estimate τ using the following relation:

$$r_y(t) = F \frac{H}{Z(t)} \rightarrow \tau(t) = -\frac{Z(t)}{\dot{Z}(t)} = \frac{r_y(t)}{\dot{r}_y(t)} \quad (2)$$

Advantage of this approach is that it is relatively easy to accurately determine the height at which the vehicle is observed. Furthermore, this approach is

less sensitive for deformation and occlusion, compared to e.g. the previous approach based on the vehicle width.

Draw back is its sensitivity for variations in camera pitch. As mentioned in section 2, we are able to transform our observations as if they were obtained by a camera with viewing direction aligned to its motion direction. Inaccuracies in this transformation yield to violation of the assumption of a perfectly aligned camera. Inaccuracy in the estimation of the camera pitch can propagate to large errors in the estimation of τ ([7]). This sensitivity aspect puts high demands on the ego-motion estimation procedure. Besides the accuracy of the estimated mean viewing direction of the camera, the impact of inter-frame rotations of the camera can not be underestimated. In order to be of practical use, this approach will require proper temporal filtering.

The third approach is similar to the previous approach based on $r_y(t)$. It employs the relation between τ and the horizontal position $r_x(t)$, under the assumption of parallel motion. Like $w(t)$ and $r_y(t)$, it's not difficult to extract a track $r_x(t)$ from the observations of a vehicle. In agreement with the previous approach, this approach is robust for deformation and partial occlusion of the observed vehicle. An important advantage over previous approach is that the observation of $r_x(t)$ is hardly affected by variations in camera pitch. In return, it depends more on variation in camera yaw. However, this ego-motion component is generally estimated with much higher accuracy.

For our application, the assumption of parallel motion is more often violated than the assumption of planar motion. Especially lane-shifts are a source for non-parallel motion. However, this doesn't completely eliminate the practical use of the third approach. We will return to this topic in the discussion.

3.2 General estimation scheme for $\tau(t)$

We present a general estimation scheme for $\tau(t)$, which can be used regardless the underlying approach. For the three approaches,

- the required input consists of tracks of features observed on vehicles;
- robust features are available for our application;
- the relation between measurement input $p(t)$ and $\tau(t)$ is given by the quotient of $p(t)$ and its time derivative, $\dot{p}(t)$ ¹;
- the measurement input satisfies $1/p(t) = \text{linear}$.

¹For each of the three approaches, $p(t)$ represents $w(t)$, $r_y(t)$ and $r_x(t)$ respectively.

Table 1: Approximated variance in $1/p_m$ due to propagation of the measurement inaccuracy in p_m . See [7] for details.

measurements p_m	measurement inaccuracy	variance σ^2 in function $1/p_m$
$r_{xm1} - r_{xm2}$ $= w + \eta_1 - \eta_2$	$\text{var}(\eta_1) = \text{var}(\eta_2)$ $= \sigma_w^2$	$2\sigma_w^2/w^4$
$r_{ym} = r_y + \eta$	$\text{var}(\eta) = \sigma_y^2$	σ_y^2/r_y^4
$r_{xm} = r_x + \eta$	$\text{var}(\eta) = \sigma_x^2$	σ_x^2/r_x^4
approximations assuming $\sigma_*^2 \ll p^2(t)$		

Both $p(t)$ and $\dot{p}(t)$ must be determined. An accurate estimation of $p(t)$ can be obtained from the observations. Being far too sensitive for the measurement inaccuracies in $p(t)$, the time-derivative, $\dot{p}(t)$, can't be obtained by means of direct differentiation. The property that $1/p(t) = \text{linear}$ is exploited to incorporate temporal smoothing in the estimation procedure for $\dot{p}(t)$.

We model the inaccuracy in the measurement sequence $p_m(t)$ as normally distributed zero mean noise:

$$p_m(t) = p(t) + \eta_p(t), \quad \text{with } \bar{\eta}_p = 0, \quad \text{var}(\eta_p) = \sigma_p^2 \quad (3)$$

We fit a linear model ($a_N \cdot t + b_N$) to the function $1/p_m(t)$ over a time-interval $[t_{k-N}, t_k]$. The model parameters $\{a_N, b_N\}$ are related to τ by

$$\tau(t_k)|_N = -t_k - a_N/b_N \quad (4)$$

Estimation of $\{a_N, b_N\}$ from the measurement data is a typical weighted least squares problem. Its solution is found by minimizing the following squared error:

$$E = \sum_{i=k-N}^k \left(\frac{a_N \cdot t_i + b_N - 1/p_m(t_i)}{\sigma_i} \right)^2 \quad (5)$$

The weighting factor σ_i^2 represents the variance in $1/p_m(t_i)$, given by $\sigma_i^2 = E \left\{ (1/p_m(t_i))^2 \right\} - E \{ 1/p_m(t_i) \}^2$. A practical approximation for σ_i^2 is given in table 1 for the three definitions of $p(t)$. This approximation assumes that the variance of the measurement inaccuracy in $p_m(t)$ is much smaller than $p(t)$. This can be guaranteed when the observed vehicle is not too far away. When the vehicle is observed near the FOE, this approximation can't be applied.

We incorporated temporal smoothing in our estimation procedure for $\tau(t)$ under the assumption of a linear time dependency of $1/p_m(t)$. However, there are several sources that introduce temporal nonlinearities in the function $1/p_m(t)$. Mainly, (temporal) acceleration or small steering actions will be to blame. When such an event occurs, we are by definition only interested in the linear extrapolation of

$1/p_m(t)$ after this event². In order to deal with these non-linearities we need to identify these events and adjust the length of the time interval, N , in a proper manner.

The optimal length of the time interval depends on the signal-to-noise ratio in $1/p_m(t)$ and the measure of non-linearity in this sequence. The optimal interval width corresponds to the maximum value for N for which the disagreement between the linear model and measurement data can be explained by measurement inaccuracy.

The expected measurement inaccuracy is given by:

$$\sigma_{\text{EXP}}^2(a_N, b_N, N, t_k) = \frac{1}{N} \sum_{i=k-N}^k \sigma_i^2 \quad (6)$$

The chi-square merit function indicates how well our linear model agrees with the measurement data. This function is given by ([11]):

$$\chi^2 = \frac{\sum_{i=k-N}^k \left(\frac{a_N \cdot t_i + b_N - 1/p_m(t_i)}{\sigma_i} \right)^2}{\sum_{i=k-N}^k 1/\sigma_i^2} \quad (7)$$

The smaller the value of χ^2 , the better the measurements are explained by the model. Due to the limited length of this paper we have to refer to [7] and [11] for more details.

4 Experiments

We illustrate the performance of our approach by means of a practical video sequence, contained 50 non-interlaced frames (568x768 pixels) per second. For 3 vehicles observed in this sequence we tracked 2 points: the lower left and lower right corner of a bounding box defined for each vehicle. The vehicles are labeled as A, B and C, as shown in figure 2.

During the sequence our own vehicle performs small steering actions. We estimated the inter-frame rotations for this sequence using the approach described in section 2. The estimations for roll, pitch and yaw are plotted in figure 3.

We will now present some estimation results of τ for each of the vehicles. When possible, the estimation results for τ will be compared with a best linear estimation for τ . This provides us with an indication for the ground truth. We divided the observations into time fragments for which the relative motion of the object is constant. The best linear representation

²By definition, we assume constant relative motion. Derivation of $\tau \approx p_m/\dot{p}_m$ by means of non-linear extrapolation of the function $1/p_m$ is inconsistent with the definition of τ .

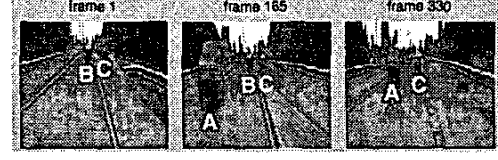


Figure 2: Three frames from the video sequence.

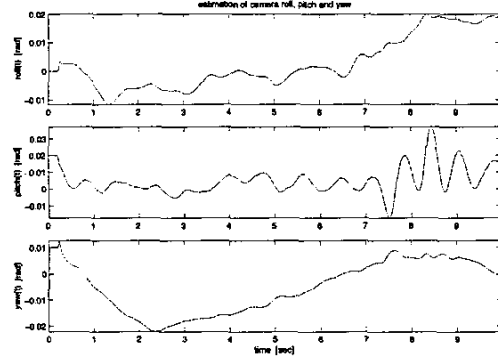


Figure 3: Estimated roll, pitch and yaw. The estimates are accumulated over time.

(ground truth indicator) for τ is then obtained by fitting a linear model to all measurements available for each time fragment.

The relative translational motion of vehicle A is not perfectly constant. After approximately 2 seconds the relative speed decreases somewhat, introducing non-linearity in τ . This non-linearity also appears in the estimation results for τ . Figure 4 shows the performance of the approach based on $r_x(t)/\dot{r}_x(t)$. The estimation results indicate that τ can accurately be estimated. The importance of ego-motion correction is illustrated. Furthermore, the results indicate that the algorithm correctly handles the non-linearity in τ .

Vehicle B slowly moves away from the camera. The vehicle is moving in the same lane as our own vehicle. As a result the signal-to-noise ratio of $1/w(t)$ is significantly better than for $1/r_x(t)$. Therefore, the approach based on $w(t)/\dot{w}(t)$ will provide the best estimations for τ for vehicle B. Figure 5 illustrates the estimation results for this approach. The limited sensitivity of this approach for inter-frame camera rotations is expressed by the results. The results confirm that the vehicle is slowly moving away from us.

Vehicle C initially moves in the acceleration lane and performs a slow lane shift in order to merge with our stream of traffic. At first we approach this vehicle, but after approximately 4 seconds it accelerates

in order to adapt to the general speed in our lane. Due to the significant translational motion component in horizontal direction, the approach based on $r_x(t)/\dot{r}_x(t)$ can not be applied. The approach based on $w(t)/\dot{w}(t)$ provided the best estimation for τ for vehicle C. Figure 6 illustrates the estimation results for this approach. Again, the results express the limited sensitivity of this approach for inter-frame camera rotations. The results confirm the motion pattern of the vehicle. First τ decreases. When vehicle C accelerates in order to adapt its speed to the general speed in our lane, τ increases.

5 Discussion

In this paper we combined 3 approaches for the estimation of the time-to-contact, $\tau(t)$, for our application. The combination is liable for accurate estimates for τ , as underlined by the experiments.

For our application, the position of these clues can accurately and robustly be extracted from the image data. We present a single filtering technique for the suppression of measurement noise. This technique exploits the linearity in the measurement function $1/p(t)$ (where $p(t)$ represents $r_y(t)$, $r_x(t)$ and $w(t)$ respectively), when observed over a period with constant relative motion. The optimal time-interval that is considered for the suppression of measurement noise is automatically selected.

Observing features on both the left and right side of the vehicle makes it possible to switch between the three approaches. In this way the estimation procedure is flexible to adapt to a specific case. It is mainly due to the transparency of the error propagation that we are able to select the most appropriate estimation approach. In general, the approach based on $w(t)/\dot{w}(t)$ will provide more accurate estimates for τ . It is almost insensitive for errors in the correction for inter-frame camera rotations. Also the influence of deformation of the observed vehicle during a lane-shift can be disregarded for our application. Major disadvantage of the approach based on $w(t)/\dot{w}(t)$ is the sensitivity for occlusion.

The approaches based on $r_y(t)$ and $r_x(t)$ are insensitive for partial occlusion of the vehicle. However, the function $r_y(t)/\dot{r}_y(t)$ features high sensitive for errors in the correction for camera pitch. In the same way, the function $r_x(t)/\dot{r}_x(t)$ is affected by inaccurate estimation of camera yaw. Besides this, the function $r_x(t)/\dot{r}_x(t)$ can only be applied for estimation of τ in case of parallel motion. Notice that, if the parallel motion assumption is violated during a lane-shift, an accurate update of τ is of less importance. Immediately after the lane-shift, estimates based on $r_x(t)/\dot{r}_x(t)$ can be used again.

References

- [1] T. Camus, "Calculating Time-to-Contact Using Real-Time Quantized Optical Flow", *National Institute of Standards and Technology NISTIR 5609*, March, 1995.
- [2] R. Cipolla, and P. Kovesi, "Determining Object Surface Orientation and Time-to-Impact from Image Divergence and Deformation", *University of Oxford (Memo)*, 1991.
- [3] R. Cipolla, and A. Blake, "Surface Orientation and Time to Contact from Image Divergence and Deformation", *Procs. 2nd European Conference on Computer Vision*, Santa Margherita, Italy, May, 1992, pp. 187-202.
- [4] D. DeMenthon, and L.S. Davis, "Exact and Approximate Solutions of the Perspective-Three-Point Problem", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 14(11), 1992, pp. 1100-1105.
- [5] T. Kalinke, and C. Tzomakas, "Objecthypothesen in Verkehrsszenen unter Nutzung der Kamerageometrie", *Internal report IR-INI 97-07*, Institut für Neuroinformatik, Ruhr-Universität, Bochum, Germany, 1997.
- [6] M.B. van Leeuwen, and F.C.A. Groen, "Motion Estimation in Image Sequences for Traffic Applications", *Procs. IEEE Instrumentation and Measurement Conference*, Baltimore (MD), USA, May, 2000, pp. 354-359.
- [7] M.B. van Leeuwen, *Motion Estimation and Interpretation for In-Car Systems*, Ph.D. thesis, University of Amsterdam, 2002.
- [8] M. Lourakis, and S. Orphanoudakis, "Using Planar Parallax to Estimate the Time-to-Contact", *Procs. IEEE Conf. on Computer Vision and Pattern Recognition*, Fort Collins (CO), USA, June, 1999, pp. 640-645.
- [9] F. Marmoiton, F. Collange, and L.P. Dérutin, "Location and Relative Speed Estimation of Vehicles by Monocular Vision", *Procs. IEEE Intelligent Vehicles Symposium*, Detroit (MI), USA, October, 2000, pp. 227-232.
- [10] F.G. Meyer, "Time-to-Collision from First-Order Models of the Motion Field", *IEEE Trans. on Robotics and Automation*, 10(6), December, 1994, pp. 792-798.
- [11] W.H. Press et al., "Chapter 15: Modeling of Data", *Numerical Recipes in C - The Art of Scientific Computing*, 2nd edition, Cambridge University Press, 1996, pp. 656-699.
- [12] T. Zielke, M. Brauckmann, and W. von Seelen, "Intensity and Edge-based Symmetry Detection with an Application to Car-following", *Computer Vision, Graphics, and Image Processing: Image Understanding*, 58(2), 1993, 177-190.

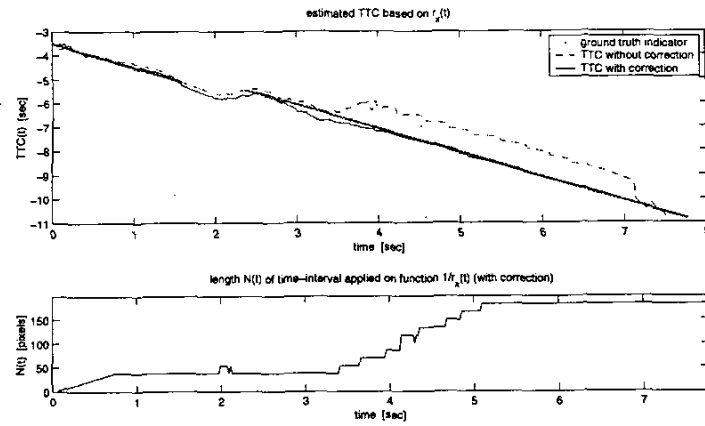


Figure 4: Estimated τ for vehicle A, based on $r_x(t)/\dot{r}_x(t)$. The upper figure shows τ , estimated with and without correction for inter-frame camera rotation. Also the ground truth indicator for τ is plotted. The lower figure shows the length of the time interval applied for the estimation.

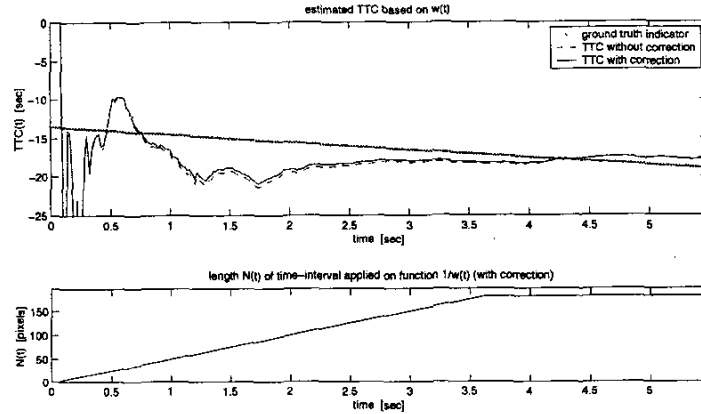


Figure 5: Estimated τ for vehicle B, based on $w(t)/\dot{w}(t)$.

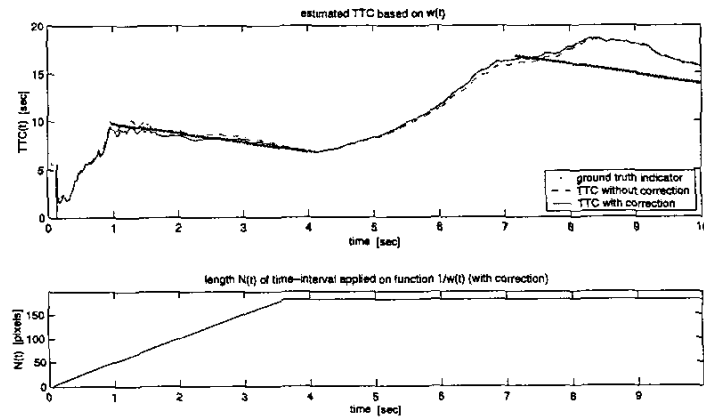


Figure 6: Estimated τ for vehicle C, based on $w(t)/\dot{w}(t)$.