International Archives of Photogrammetry and Remote Sensing, Vol.30, Part 5W1, 1995, pp. 208-213.

GROUND PLANE OBJECT TRACKING UNDER EGOMOTION

Włodzimierz Kasprzak

Bavarian Research Center for Knowledge Based Systems
FORWISS, Am Weichselgarten 7, D-91058 Erlangen
E-mail: kasprzak@forwiss.uni-erlangen.de
Tel. +49 9131 691-194, Fax -185

KEY WORDS: Image Sequences, Kalman Filter, Motion Estimation, Object Recognition.

ABSTRACT:

An adaptive approach to moving object recognition in image sequences from a single low–cost camera under egomotion is proposed. The object motion is restricted by ground plane trajectories. Quantitative and qualitative results of vehicle tracking in traffic scenes are presented. It is demonstrated that acceptable results with a single low cost camera can be achieved for those objects, which are projected to image regions not smaller than 10x12 $pixel^2$.

1. INTRODUCTION

Recently more intensive work on object tracking in traffic scenes has been done, but mostly limited to a stationary camera case (without egomotion) (Koller, 93), (Tan, 93). The problem of vision systems for navigation purposes, as addressed for example in (Masaki, 92), consists of the recognition of road boundaries and moving obstacles within the road area. The detection of obstacles, that are very distant from the camera and a precise estimation of their positions, motions and orientations on the road plane is still a challenging problem (Regensburger, 94). Additionally to the non-parallel projection, there is a problem of stable camera orientation detection (overcoming the camera nodding movement) and of recognizing the road (if many obstacles exist in the scene). These are reasons why standard low level motion estimation (Schunck, 81) fails to provide us with discriminant object features in this case, even if the translational and circular velocities of the camera vehicle (egomotion) are known.

In this paper a model based approach to moving object recognition in image sequences under egomotion is described. A specific application in mind is a road scene analysis system (Kasprzak, 94). Due to large discretization errors in images of outdoor scenes, the scene and object domains are restricted by shape assumptions and the object motion trajectories are restricted to a ground plane circular movement. In section two an overview of the adaptive object recognition approach is given. The main computational steps of this approach – initialization and single object tracking – are discussed in section 3. Test results of moving object recognition in monocular image sequences of traffic scenes and a summary follow in subsequent sections.

2. THE RECOGNITION-BY-TRACKING APPROACH

2.1 Adaptive object recognition

There are at least two different approaches to the recognition of moving objects in image sequences. The first approach tries to initialize a nearly correct hypothesis, spending a lot of time on the estimation of an application–independent visual motion – either the optical flow is computed and segmented afterwards (Koller, 93) or discrete features are tracked and their image motion is interpreted in terms of object's motion and depth (Kasprzak, 93). This type of initialization corresponds to hypothesis dependent measurements, i.e. the judgment of new measurement is directly dependent o how well it fits the hypothesized object. This approach can be called *recognize–and–track.*

Here a different approach to model–based object recognition in image sequences is proposed. As the motions of tracked objects are not significantly different from the egomotion and the nodding of the camera is disturbing the estimated visual motion, a geometry and model–based object initialization is performed with default motion. No pre-computation of visual motion is necessary. Accordingly, the judgments of consecutive measurements are not dependent on the predicted hypothesis, but on how well they fit the global model expectations. Due to this statistical independence the Kalman filter (Wünsche, 88) for adaptive estimation of the tracked hypothesis can be applied. In this way the whole hypothesis can largely be modified during the consecutive analysis and the approach can be called *recognition–by–tracking.*

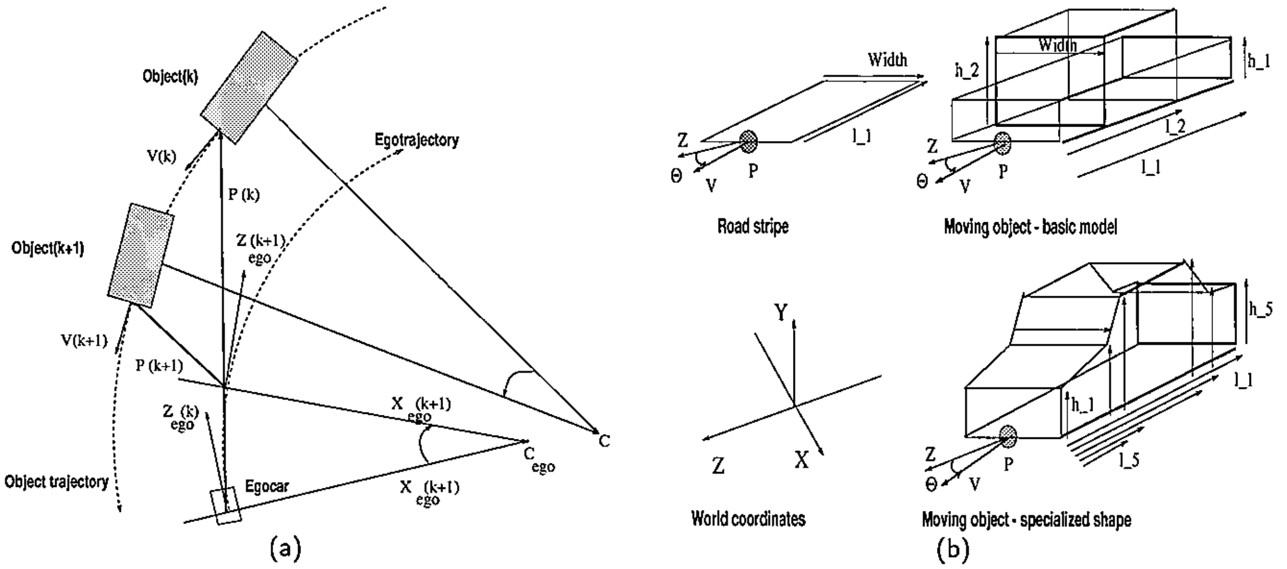


Figure 1: The object and ego-car trajectories on the ground plane (top view) (a) and the model shapes (b).

2.2 The dynamic model for a single hypothesis

A moving object is defined here by its parameter state vector: $s = [s^d, \xi]$, with

$$s^d = [P_X, P_Z, \Theta, V, \omega]^T; \qquad \xi = [Width, l_1, h_1, l_2, h_2]^T \tag{1}$$

Thus the object state consists of the dynamic part s^d and the shape parameters ξ. The dynamic part describes a curved trajectory on a planar road (**Figure** 1(a)). The dynamic state parameters are as follows: (P_X, P_Z) is the on–road position of the object's origin point, Θ is the object's direction angle, V and ω are the translational and rotational velocities respectively. As the observed object projections in the image are relatively small a simple volume model that consists of two parallelepiped is applied (**Figure** 1 (b)) allowing a classification of vehicles into cars and trucks. The shape parameters are as follows: $Width$ means the width of both parallelepipeds, l_1, l_2 are the lengths and h_1, h_2 are the heights of each parallelepiped respectively. A specialized object shape is not required in this application, but if necessary more vehicle classes can be handled by using more shape parameters. The specialized shape in **Figure** 1(b) is described by the width, 5 length and 5 height parameters.

Let $\boldsymbol{M}(k)$ be the *measurement vector*, associated with the $s(k)$ object's state vector at discrete time k. The vector $\boldsymbol{M}(k)$ contains the positions of object primitives, that can be observed in the k–th image. The time–dependent model of both vectors behavior is given by a stochastically disturbed nonlinear dynamic system with discrete time:

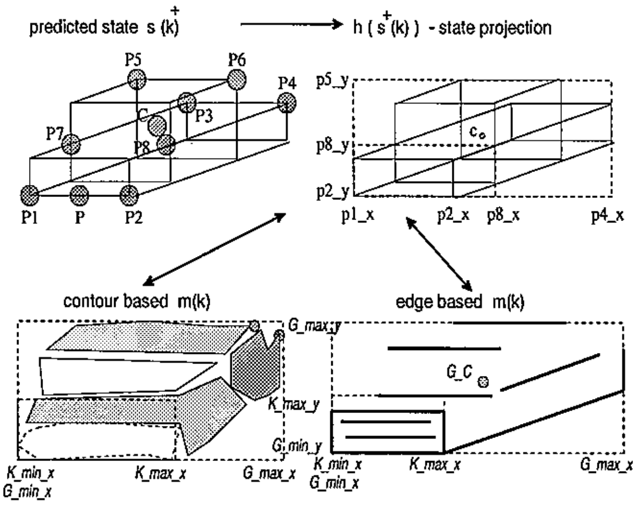
$$s(k+1) = f[s(k)] + v(k); \quad \boldsymbol{M}(k) = h[s(k)] + w(k) \tag{2}$$

where $f(.)$ is the state transition function, $h(.)$ is the state projection function, $v(k)$ is the system noise and $w(k)$ is the measurement error. As both state observation and measurement detection are disturbed by errors only an estimation of the state can be provided. The *estimation* task at time k is to give a state estimation $s^*(k)$ on the basis of real available measurements $\{\mu(i) \simeq \boldsymbol{M}(i) | i = 0, ..., k\}$. A consecutive solution is achieved due to *recursive* methods, where the old estimate $s^*(k)$ is updated after a new measurement $\mu(k+1)$ is available. Two matrices $\boldsymbol{E}(k), \boldsymbol{R}(k)$ are provided also – the error covariance matrices of current state estimation and measurement vector.

2.3 The object recognition procedure

The procedure for adaptive object recognition in many object scenes consists of following steps, that are performed for image segments SEGMENTS(k):

1. FOR each hypothesis $s^*(k)$ with $\boldsymbol{E}^*(k)$ DO: OBJ_TRACK($s^*(k)$,$\boldsymbol{E}^*(k)$, SEGMENTS(k))
2. Eliminate those segments from SEGMENTS(k) that contribute to a successfully tracked hypothesis
3. For remaining segments make new object initializations
4. Select consistent object hypotheses



$\{ \; C, Pi \; (i = 1,..,8) \; \}$: 3–D points
$\{ \; c, pi \; (i = 1,..,8) \; \}$: 2–D projections

$$h(s^*(k)) = \begin{pmatrix} p1_x \\ p2_x \\ c_x \\ c_y \\ p1_x \text{ or } p3_x \\ p4_x \text{ or } p2_x \\ p2_y \text{ or } p1_y \\ max(p3_y, p5_y) \text{ or } max(p4_y, p6_y) \\ p7_x \text{ or } p8_x \\ p7_y \text{ or } p8_y \end{pmatrix}$$

Figure 2: Alternative 2–D measurements are matched with the projected state vector

3. THE RECOGNITION STEPS

3.1 Initialization of an object hypothesis

A model–based initialization can be divided into two main steps – segment grouping and state initialization. The first step means a segment classification into vanishing point, road, surrounding area or horizon, followed by a backprojection of the segments onto the 3–D space over the road plane and by grouping the hypothesized 3–D segments into road stripes, car objects and obstacles in the road area.

The second step means the initialization of 3–D state parameters from the hypothesized group. Position coordinates P_X, P_Z are given from backprojection under the assumption $P_Y = 0$. Translational velocity V is set to a default value: $V = V_{ego}$ or $V = -V_{ego}$ The direction Θ is estimated from the relation between the widths of two boundary boxes in the image – the boundary box K of the assumed front or back part of the car and the boundary box G of the whole object projection – by assuming a default length. Other shape parameters are also given from backprojection of the two boundary boxes. The rotational velocity is set by default to current egomotion by considering the current road curvature: $\omega \simeq -\omega_{ego} + \omega_{road}$ or $\omega \simeq \omega_{ego} - \omega_{road}$

3.2 Object tracking with recursive state estimation

An extended Kalman filter (EKF) (Wünsche, 88) with sequential modification is applied for the recursive estimation of the hypothesis parameters. For each object hypothesis estimation $s^*(k)$ with its error covariance matrix $\boldsymbol{E}^*(k)$ and system error covariance $\boldsymbol{Q}(k)$ following steps are performed at the time k :

1. The prediction equations: $s^+(k) = f[s^*(k-1)]$; and $\boldsymbol{E}^+(k) = \boldsymbol{F}(k-1)\boldsymbol{E}^*(k-1)\boldsymbol{F}^T(k-1) + \boldsymbol{Q}(k-1)$; where $\boldsymbol{F}(k) = \frac{\delta f}{\delta s}|_{s^+(k)}$ is the *Jacobi* matrix of function $f(.)$.
2. Detect the measurement $\mu(k)$ with covariance matrix $\boldsymbol{R}(k)$.
3. The modification of object state $s^*(k)$ and its covariance matrix $\boldsymbol{E}^*(k)$

The measurement mode can be either based on object to object correspondence, due to repeated 3–D object initialization step in every image or it can be a 2–D measurement, that consists of 2–D image segment groups (like edge or contour groups).

In the first case the measurement $\mu(k)$ vector is equivalent to a state vector of reduced size. It consists of two parts:

$$\mu(k) = s_R(k) \cup s_g(k) = [P_X(k), P_Z(k), Width(k), h_1(k), h_2(k))]^T \cup [\Theta_g(k), l_{1g}(k), l_{2g}(k)]^T \qquad (3)$$

In relation to a full state vector no translational and rotational velocities are given, as they can be measured only indirectly from the differences in position and orientation of estimated hypotheses in images k and $k-1$. Additionally the orientation and length parameters are here indexed by g to indicate their geometry based nature. Measurements for these parameters of second type exist also, on the basis of the new measured object motion.

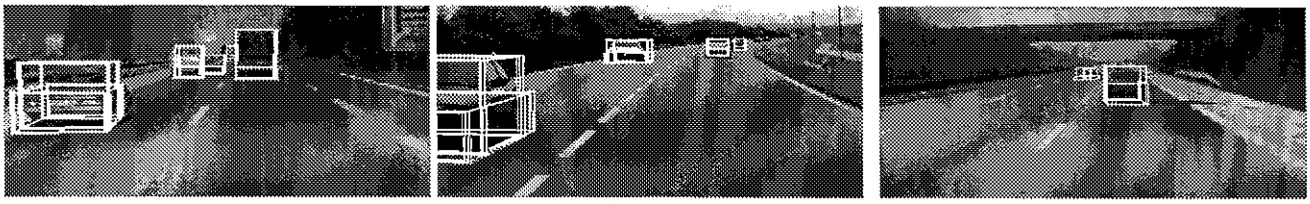


Figure 3: Object hypotheses in three images

In the 2–D measurement case there is a nontrivial predicted state projection transformation $h(s^+(k))$ (**Figure** 2). On the basis of predicted state several 3–D model points (the Pi-s, $i = 1, ..., 8$ and C) are selected and the visible points are projected onto the image plane defining two boundary boxes – of the front or back part K and of the whole object boundary box G – as well as the center point c. The projected points are matched with new measured 2–D points, that are contained in the following measurement vector:

$$\mu(k)^T = [K_{minx}, K_{maxx}, G_{Cx}, G_{Cy}, G_{minx}, G_{maxx}, G_{miny}, G_{maxy}, K_x, K_y]^T \qquad (4)$$

Various alternative methods for 2–D measurements can be applied. Two of them have been tested – line segment based measurements or closed contour (or region) based measurements.

It is evident that the predicted projection vector is not directly dependent on the motion parameters. Additionally, the measurements provide no way to estimate the direction Θ and the lengths l_1 and l_2 independently each other. This problem is common to both measurement cases. Thus, a two step modification of the state vector is performed. First of all the estimation of the reduced state $s^*_R(k)$ and the geometry–based measurement $s_g(k)$ are performed. From the vector $\delta s^*_R(k) = [s^*_R(k) - s_R*^(k-1)]$ current motions $V(k), \omega(k)$ are calculated and the motion based measurement $s_m(k) = [\Theta_m(k), l_{1m}(k), l_{2m}(k)]^T$ are performed. These synthetic measurements together with the first step measurement $s_g(k)$ are next used for the modification of the remaining state $s_E(k) = [V(k), \Theta(k), \omega(k), l_1(k), l_2(k)]^T$. Hence the modification process at time k for one hypothesis can be summarized as follows:

1. Modification of the reduced state and its covariance matrix
   $s^*_R(k) = s^+{}_R(k) + \boldsymbol{K}(k)\{\mu(k) - h(s^+(k))\}$; and $\boldsymbol{E}^*_R(k) = \boldsymbol{E}^+{}_R(k) - \boldsymbol{K}(k)\boldsymbol{H}(k)\boldsymbol{E}^+{}_R(k)$
   where $\boldsymbol{K}(k) = \boldsymbol{E}^+{}_R(k)\ \boldsymbol{H}^T(k)\Big\{\boldsymbol{H}(k)\boldsymbol{E}^+{}_R(k)\boldsymbol{H}^T + \boldsymbol{R}(k)\Big\}^{-1}$ is the Kalman gain matrix.
   and $\boldsymbol{H}(k) = \frac{\delta h}{\delta s_R}|_{s^+{}_R(k)}$ is the *Jacobi* matrix of function $h(.)$.
2. Detection of motion and motion based synthetic measurements from $s^*_R(k) - s^*_R(k-1)$.
3. Modification of remaining state $s^*_E(k)$ on the basis of predicted remaining state $s^+{}_E(k)$ and synthetic measurements (motion parameters, $s_g(k), s_m(k)$).

4. RESULTS

The approach has been tested on several monocular image sequences of road scenes (see for example **Figure** 3). For image acquisition a low cost camera (its focal length to pixel size ratio was 708) was located in the ego-car at the height of 1.67 m over the road plane. Up to 6 moving cars have occured in one image. From them up to five cars have been properly detected and tracked nearly all the time (i.e. in 95–100% of images). The sizes of their image projections have ranged from 20x20 $pixel^2$ to 50x70 $pixel^2$. Only small and partially hidden cars located very far from the camera have been detected with a rate below 50% (their image sizes were about 10x12 $pixel^2$).

Quantitative results of parameter measurement and estimation for two moving objects – the left car and the middle truck – are presented in **Figures** 4–10. These results can be summarized as follows:

- the depth estimation error is < 10 % and the direction estimation error is < 0.2 *rad*
- the estimation of translational velocity is of good performance for the near car (error up to ± 2.5 m/s), for the truck the error is much higher (up to ± 7.5 m/s).
- a rotational velocity should not occur, as the objects are moving approximately straight ahead, but small velocities are estimated up to ± 0.5 rad/s.
- the error of car width estimation is < 0.2 m for the car and < 0.5 m for the truck; as the Θ angle is nearly π the length can be only weekly measured (in the range of $3 - 8m$);

The presented approach was simulated on a workstation with 25 MIPS. The processing time of a complete scene analysis was $4 - 5$ s, but for the object recognition procedure the required cpu time was only $0.6 - 0.8$ s.

5. CONCLUSION

Quantitative and qualitative results of object tracking on the ground plane under egomotion of the camera have been presented. By applying a single low cost camera, located at the driver's height, high quality position and velocity estimations have been achieved for objects within the depth of up to 60 m or 20 m respectively. The ratio of the ego-camera height over the road to the depth of the truck is very small, hence the measurement errors are growing fast with increased depth. There are two obvious ways to increase the measurement quality for distant objects. Either the same camera should be located higher over the road or a camera with great focal length should be used. It is also evident that the quality of velocity estimation could be increased while working with short–term averaged velocity measurements instead of velocities between two consecutive frames (i.e. synthetic velocity measurement is done from a short–term tracking of position and direction in 3–5 frames instead of 2).

Acknowledgments: The support from the "Deutsche Forschungsgemeinschaft", Bonn, Germany, is gratefully acknowledged (Grant Ni-191/8-2). The real images are by courtesy of the BMW AG., Munich, Germany.

REFERENCES

[Kasprzak, 1993] Kasprzak, W. (1993). Modellunabhängige Schätzung von 3–D Attributen während der Bildfolgensegmentierung. In: *Mustererkennung 1993*, Informatik aktuell, 51–58. Springer, Berlin etc.

[Kasprzak et al., 1994] Kasprzak, W., Niemann, H., and Wetzel, D. (1994). Adaptive estimation procedures for dynamic road scene analysis. In: *Proceedings ICIP–94, IEEE Int. Conference on Image Processing*, 563–567, IEEE Computer Society, Los Alamitos, CA.

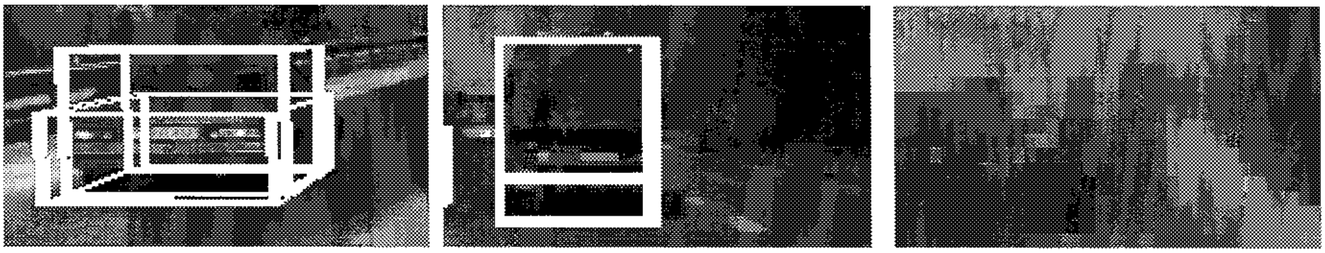
[Koller et al., 1993] Koller, W., Daniilidis, K., and Nagel, H.-H. (1993). Model–based object tracking in monocular image sequences of road traffic scenes. *International Journal of Computer Vision*, 10(3):257–281.

[Masaki, 1992] Masaki, I. (1992). *Vision-based Vehicle Guidance*. Springer, New York etc.

[Regensburger & Graefe, 1994] Regensburger, U. and Graefe, V. (1994). Visual recognition of obstacles on roads. In: *IROS '94. Proceedings of the IEEE/RSJ/GI International Conference on Intelligent Robots and Systems*, Munich, Germany, 980–987.

[Tan et al., 1993] Tan, T., Sullivan, G., and Baker, K. (1993). Recognizing objects on the ground plane. *Image and Computer Vision*, 12(3):164–172.

[Wünsche, 1988] Wünsche, H.-J. (1988). *Bewegungssteuerung durch Rechnersehen*. Springer, Berlin.



(a) (b) (c)

Figure 4: Vehicles and their hypotheses: (a) left car; (b) truck; (c) badly detectable car.

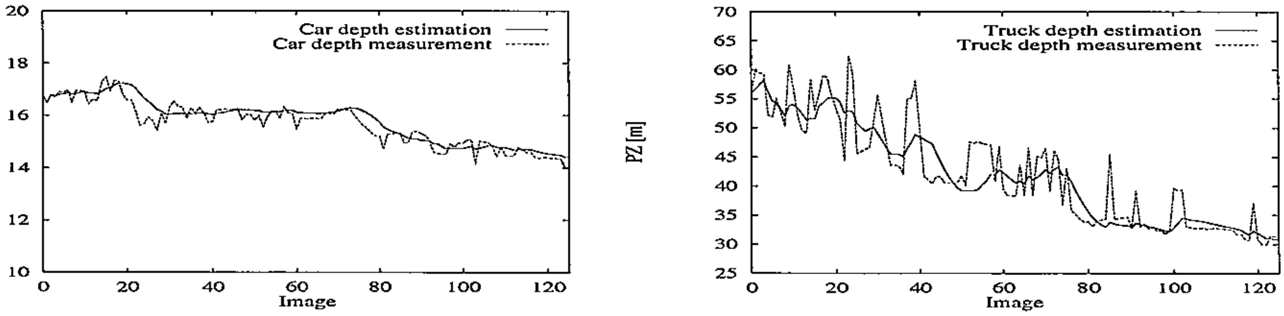


Figure 5: The measured and estimated depths (PZ) of the left car (left drawing) and the truck (right drawing).

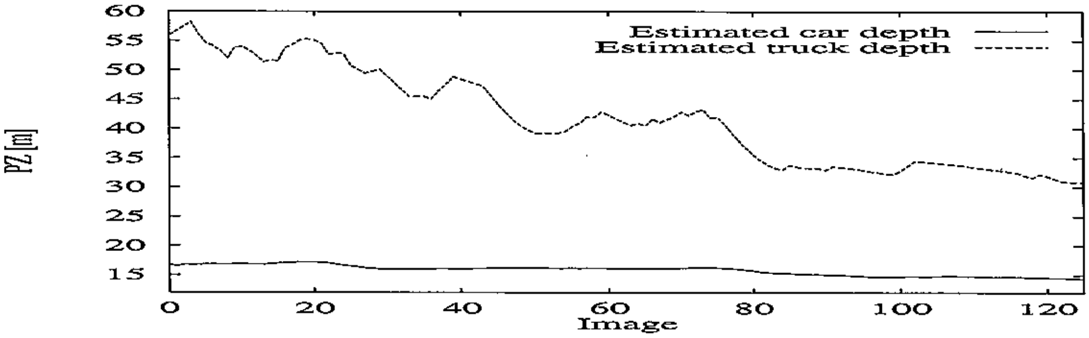


Figure 6: The estimated depths of the left car and truck.

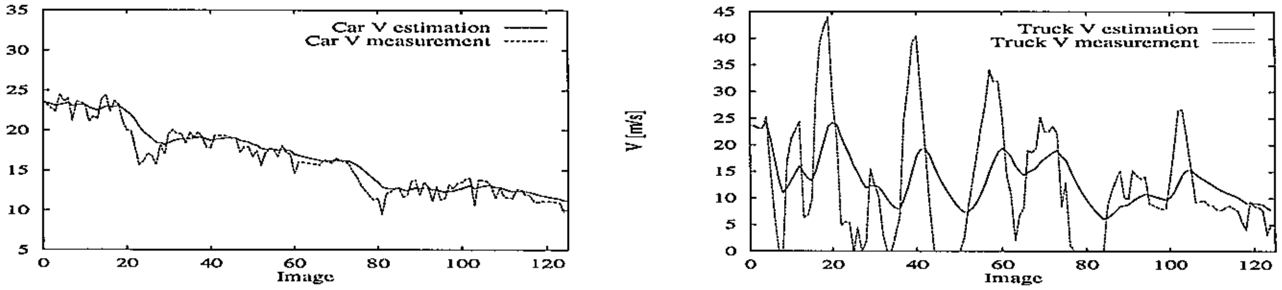


Figure 7: The measured and estimated velocities V of the left car (left) and the truck (right).

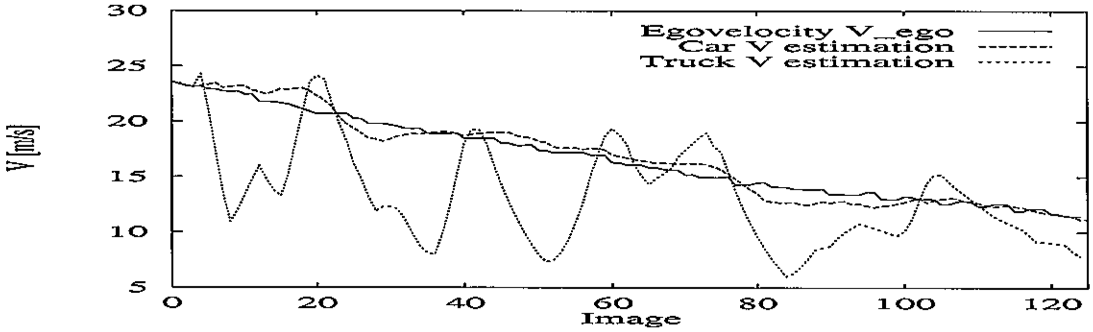


Figure 8: The estimated velocities V vs. the egocar's translational velocity.

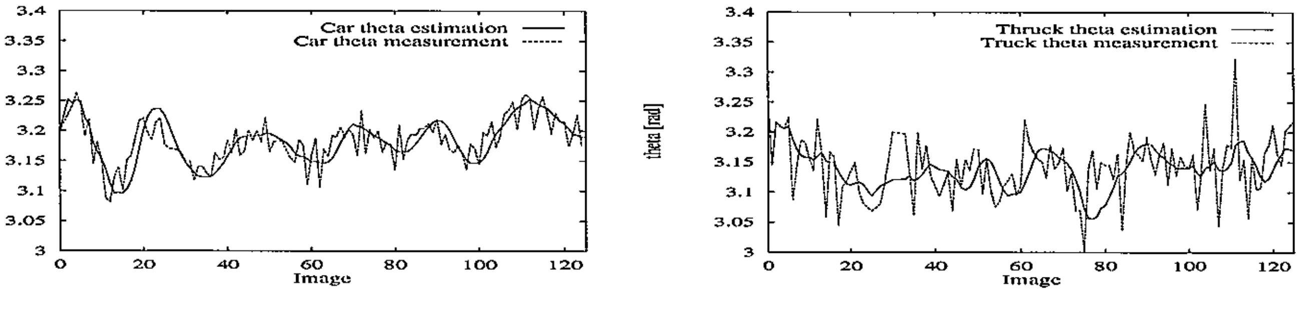


Figure 9: The measured and estimated direction Θ of the two objects (Θ should be ≃ 3.14).

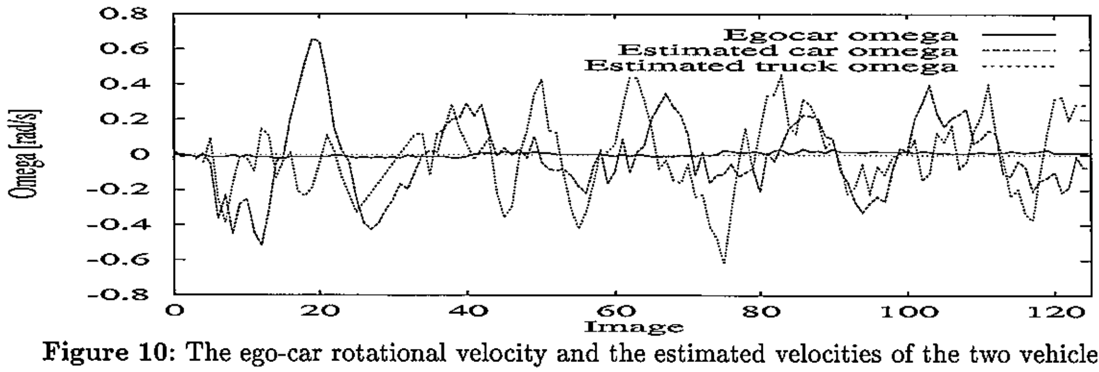


Figure 10: The ego-car rotational velocity and the estimated velocities of the two vehicles.