Tracking with a pan-tilt-zoom camera for an ACC system

Xavier Clady, François Collange, Frédéric Jurie, Philippe Martinet

To cite this version:

Xavier Clady, François Collange, Frédéric Jurie, Philippe Martinet. Tracking with a pan-tilt-zoom camera for an ACC system. 12th Scandinavian Conference on Image Analysis (SCIA '01), Jun 2001, Bergen, Norway. pp.561–566, 2001. <inria-00548284>

HAL Id: inria-00548284

https://hal.inria.fr/inria-00548284

Submitted on 21 Dec 2010

TRACKING WITH A PAN-TILT-ZOOM CAMERA FOR AN ACC SYSTEM

X. Clady, F. Collange, F. Jurie and P. Martinet

LASMEA - UMR 6602 du CNRS - Blaise Pascal University
24 avenue des Landais 63177 Aubière Cedex FRANCE
e-mail: {*Firstname.Name*}*@lasmea.univ-bpclermont.fr*

ABSTRACT

In this paper, visual perception of frontal view in intelligent cars is considered. A Pan-Tilt-Zoom (PTZ) camera is used to track preceding vehicles. The aim of this work is to keep the rear view image of the target vehicle stable in scale and position. An efficient real time tracking algorithm is integrated. It is a generic and robust approach, particularly well suited for the detection of scale changes. The camera rotations and zoom are controlled by visual servoing. The methods presented here were tested on real road sequences within the VELAC demonstration vehicle. Experimental results show the effectiveness of such an approach. The perspectives are in the development of a visual sensor combining a PTZ camera and a standard camera. The standard camera has small focal length and is devoted to an analysis of the whole frontal scene. The PTZ camera gives a local view of this scene to increase sensor range and precision.

1. INTRODUCTION

Computer vision for driving assistance systems has attracted considerable attention over the last decade. These systems offer potential solutions to the problems caused by the increase in world road traffic (particularly in industrialized countries): pollution, traffic congestion, safety... They aim to help drivers prevent and avoid collisions and to alleviate traffic congestion.

Within this framework, several devices[13] for intelligent vehicles are being developed by research laboratories and car manufacturers:

- “Stop and Go” devices, stop and start vehicles in congestion.
- “Platooning”, to create trains of vehicles (several vehicles follow the trajectory of a car-leader).
- “Adaptive Cruise Control” (ACC)[1, 14], automatically controls the speed of a vehicle in order to maintain safety distances relative to preceding vehicles.
- “Collision Warning or Avoidance”, informs or replaces the driver when accident risks are detected.

The LASMEA is involved in such projects and research has been undertaken in order to implement an ACC device on the VELAC[1] experimental vehicle. Within this framework, algorithms for detecting and tracking the roads, with or without lanes, and the preceding vehicles provided with visual landmarks, have been developed using monocular vision, and tested successfully in real conditions [1]. However these solutions present some limitations. One of these relates to the use of visual land-marks on the detected vehicles. Moreover, the sensor range is limited to 60 meters by the resolution of the camera (which must have a short focal distance to capture all the frontal scene). This limitation means that distant obstacles cannot be detected, in particular those with slow kinematics. Therefore, the addition of a camera provided with a zoom could solve this problem. Moreover, this camera can be focused on a particular vehicle, the most dangerous, in order to determine its kinematic characteristics more precisely.

This latter application is the main subject of this article. It describes a sensor system to track the rear view of a vehicle with a PTZ camera. This camera, controlled for zoom and pan tilt angles, represents the hardware part of the system. It provides image sequences of the car to a tracking algorithm. This algorithm is the software part of the sensor system. It computes the position, the orientation and the size of the target in the image. This data is used to control the PTZ camera, in order to keep the target centered in the image with a constant size.

The originality of the algorithm is its ability to track any kind of patterns in real time video, on low cost hardware. These characteristics make the sensor system very efficient for real time applications like ACC. Moreover, it runs without land-marks.

This article comprises three sections. The first section describes the context of the application. It explains how the PTZ camera is combined with a standard camera to make up the sensor equipment. The second section presents the tracking algorithm and the control law. The experiments were conducted from inside the demonstration VELAC vehicle and are presented in the third section.

---

[1] VELAC, Véhicule Expérimental du LASMEA pour l’Aide à la Conduite (see fig. 9).

2. CONTEXT OF THE APPLICATION

Three perception functions are essential in ACC application: road and lane detection (in order to locate the experimental vehicle), preceding vehicles detection and location and finally, tracking of the most dangerous vehicle[2] [1]. The kinematic characteristics (i.e. speed and relative position) of dangerous vehicles can be computed by combining these three functions.

These perception tasks can be efficiently achieved using video cameras. Despite the fact that cameras cannot directly perceive the 3D environment, they have numerous advantages. They can provide very useful information about the structure of the environment like the position of the lanes or the position of obstacles and surrounding vehicles. Such equipment is generally not very heavy and not very expensive, and can therefore be easily embedded inside a vehicle.

PTZ cameras have a variable focal length. This is a great advantage for A.C.C. applications. Indeed, the average braking distance for a vehicle running at a speed of 100 km/h is 100 meters and it can reach 200 meters at 140 km/h. Thus, for applications on motor-ways, it is necessary to detect vehicles as far ahead as possible. An acceptable limit would be around 300 meters (in a straight line). The PTZ camera is the only equipment satisfying these constraints at the moment.

But an increase in the camera focal length limits the visual field. Targets can leave the field of view, especially in turns or declivities of the road.

The originality of the approach presented is to combine a short fixed focal length camera with the PTZ camera to overcome these limitations. The PTZ camera offers a local but directional vision while the standard camera offers a global vision of the frontal scene.

The first camera will be used to detect lanes, to locate the vehicle and to detect vehicles stated as potentially dangerous (by taking into account the distances of the frontal vehicles and their positions on the road lanes) within the limits of its visual field.

The PTZ camera is focused on the detected vehicle (see fig. 1). However, when no vehicle is detected with the first camera, the PTZ camera is used to detect distant vehicles (see fig. 2). The switch between these 2 functions of the PTZ camera is decided according to traffic evolution.

This work requires an intelligent module, called *Decision Module* (see fig. 3). It analyzes the information given by both cameras (and proprioceptive sensors) and computes the speed of the vehicle, in order to respect the safety distances [14]. This speed will be reached by acting on the vehicle actuators (brakes, accelerators).

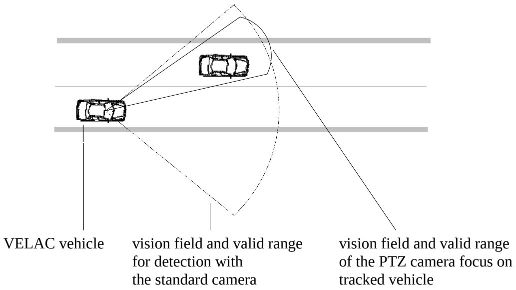

Fig. 1. PTZ camera used for tracking.

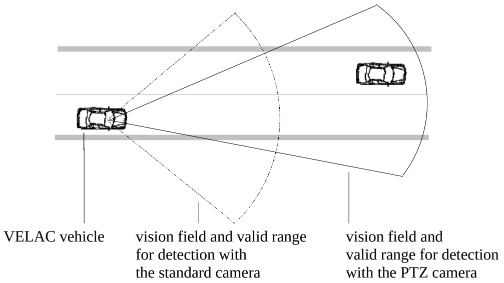

Fig. 2. PTZ camera used for detection.

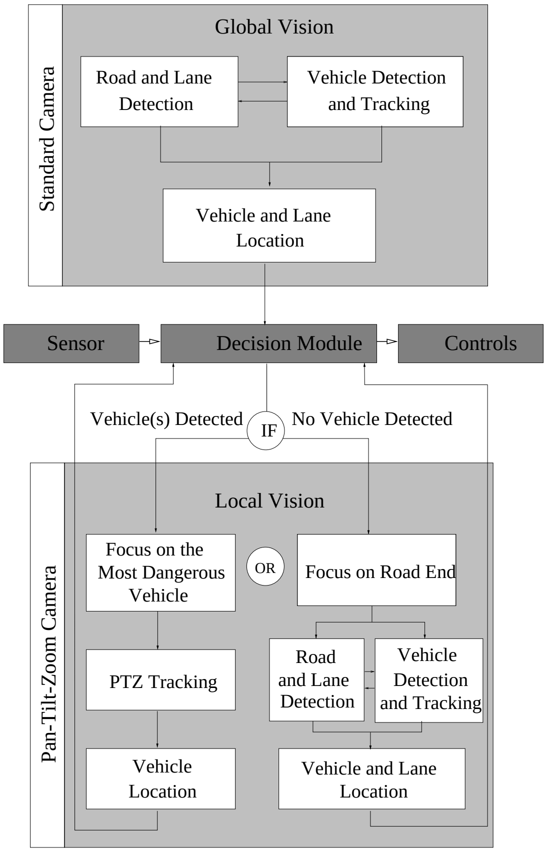

Fig. 3. Application General Synopsis

---

[2] Generally, the most dangerous vehicle is the closest vehicle in the same lane the host vehicle is driving in. But this notion can change according to the traffic situation.

3. PTZ TRACKING

In this application, tracking an object means tracking a 2D pattern in image sequences, in order to keep it along the optical axis of the camera with a constant scale. In this section, the tracking algorithm is first presented. Second, control of the camera is explained.

3.1. The tracking algorithm

The algorithm related in this section makes it possible to estimate the displacement (location, scale and orientation) of a 2D image pattern. It is based on the general framework proposed by Hager and Belhumeur [6].

It consists in relating the displacement of a *reference pattern* to intensity variations in a region of interest (ROI) centered on the pattern. This relation, supposed to be linear, is included in a matrix called the*interaction matrix*. This matrix is computed during a learning stage. The following paragraphs show how this matrix is defined and used to track the pattern. More details and results can be found in [2].

3.1.1. *Definitions*

To represent the position of the ROI (including pattern) and its content, two vectors are defined: a *parameters vector* **E** and a *pattern vector* **I**. In this article, ROIs are elliptic and their position vectors contain five parameters: the position of the center (X_c, Y_c), the orientation (θ) and the largest/smallest axis (R_1,R_2) lengths (see Fig. 4). Vector **I** is obtained by sampling the pattern included in the ROI. The samples are distributed on concentric ellipses deduced by scaling the ROI edges, in order to keep their density constant throughout the ROI (in Fig. 5, dots represent the points where samples are taken). This representation of patterns in a local reference is independent of their position, orientation and scale in the image.

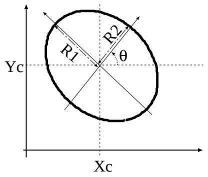

Fig. 4. The five parameters

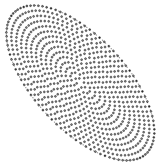

Fig. 5. Pattern vectors

With these definitions, a reference pattern vector corresponding to the samples of the reference pattern is computed. By writing $\mathbf{\Delta I} = \mathbf{I_{ref}} - \mathbf{I_{cur}}$, $\mathbf{\Delta I}$ is null when the current pattern is equal to the reference pattern. An optimal parameter for $\mathbf{E_{opt}}$ leading to this configuration needs to be estimated.

The knowledge of a function f defined by:

$$\begin{aligned} \mathbf{\Delta E} = \mathbf{E_{opt}} - \mathbf{E_{cur}} = f(\mathbf{\Delta I}) \\ \Leftrightarrow \mathbf{E_{opt}} = \mathbf{E_{cur}} + f(\mathbf{\Delta I}) \end{aligned} \quad (1)$$

is a powerful tool to track patterns. This function can be approximated by the first order term of its Taylor development, and can be written under the following matrix form:

$$\mathbf{\Delta E^T} = \mathbf{A}.\mathbf{\Delta I^T} \quad (2)$$

where $\mathbf{E^T}$ denotes the transposition of **E**.
This matrix **A** is called the *interaction matrix*.

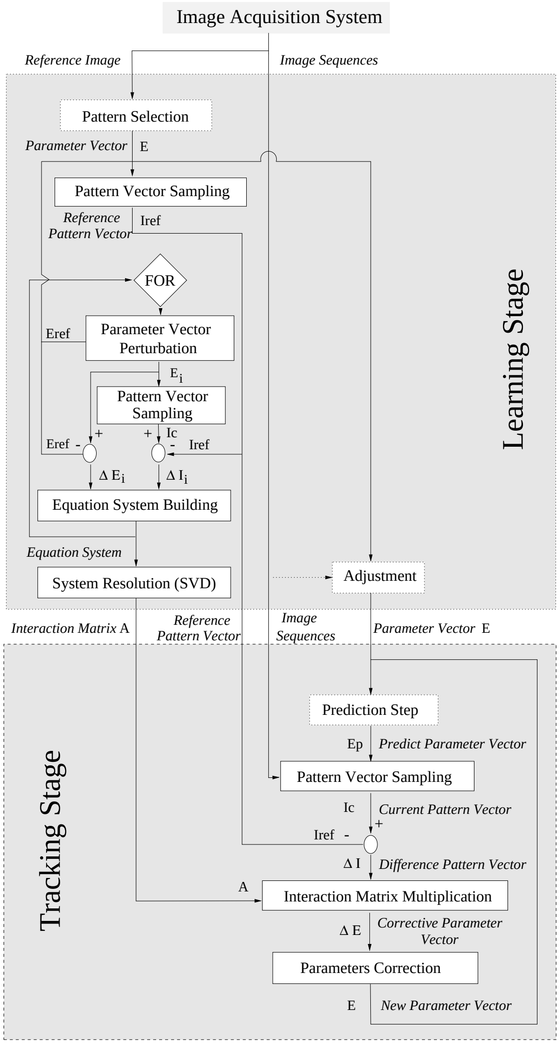

Fig. 6. Tracking Process Scheme

3.1.2. *Estimation and use of the Interaction Matrix*

The interaction matrix is obtained in a learning stage. At the beginning of this stage the ROI, including the reference pattern, is (manually or automatically) selected. Then, it is randomly modified: small perturbations are applied to each of its five parameters. Each time, the pattern vectors corresponding to this new ROI are sampled and kept for further treatments.

When a number of perturbations superior to the number of samples (giving a system with more equations than unknowns), a least square approximation of $\mathbf{A}$ can be estimated by using a Singular Values Decomposition.

Then, the use of the interaction matrix is very straightforward. Multiplied by the difference between the current and the reference pattern vectors, it gives an estimation of the transformation aligning the current ROI with the location of the reference pattern in the current image (see eq.2).

3.1.3. Conclusions about the tracking algorithm

This tracking algorithm presents the following advantages:

- it is a generic approach: any kind of pattern can be tracked after being learnt (in a relatively short time - see section 3.3),
- it runs in real time: the computation consists in a single matrix multiplication,
- it tracks small displacements in location, scale and orientation with great precision,
- it could be enlarged to large displacements by adding a prediction step.

This algorithm is well suited for the application presented where any kind of vehicle needs to be tracked in real time. Current vehicles, apart from motor-cycles, often present a 2D surface in their rear view. A precise computation of small displacements is needed to control the PTZ camera.

3.2. Control law

The control of the PTZ camera obtained through by a conventional visual servoing approach. The information provided by the tracking algorithm concerns the ellipse parameters. One way is to use the work conducted in [11] where the ellipse parameters are used to control 5 degrees of freedom (d.o.f). As the sensor equipment has only 3 d.o.f, a simpler method has been chosen.

Here, the sensor signals are fixed to:

$$\mathbf{s} = \begin{pmatrix} u & v & h \end{pmatrix}^T \tag{3}$$

where $\begin{pmatrix} u & v \end{pmatrix}$ represents the center of the ellipse in image space, and h the length of a particular segment tracked by the ellipse (In practice, the ellipse englobes a particular segment of the rear preceding vehicle. So, the size of the ellipse in x direction is used as a measurement of the tracked segment length). Let ${}^e\mathbf{v}$ be the kinematic screw applied to the sensor equipment.

Building the *Task function* $\mathrm{e} = C.(\mathbf{s}-\mathbf{s}^*)$, a proportional control law is defined by the relation [4]:

$${}^e\mathbf{v} = -\lambda.\widehat{\mathbf{L}_{s^*}}^{+}(\mathbf{s}-\mathbf{s}^*) \tag{4}$$

where $\widehat{\mathbf{L}_{s^*}}^{+}$ is the pseudo-inverse of an estimation (at the equilibrium, when $\mathbf{s} = \mathbf{s}^* = (0, 0, h^*)^T$) of the image jacobian. Then, the sensor signal vector $\mathbf{s}$ decreases exponentially.

Computation at the image jacobian at equilibrium shows that each d.o.f can be controlled separately as was the case in [10]. The pan and tilt angles of the camera are controlled with the center of the ellipse, and the focal length with the size of the segment.

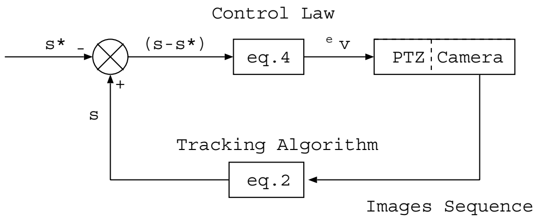

Fig. 7. PTZ Control Synopsis

Then, control of the pan-tilt camera is achieved by:

$$\begin{pmatrix} \Omega_x \\ \Omega_y \end{pmatrix} = -\lambda \begin{pmatrix} 0 & \frac{1}{F_v} \\ -\frac{1}{F_u} & 0 \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} \tag{5}$$

where Ω_x and Ω_y are the angular velocity.

Control of the zoom is more rudimentary. An ideal length for the tracked segment h^* is fixed. The zoom is increased (respectively decreased) if the axis length is below (above) a given threshold. In fact, control of the zoom is achieved through a low speed range (hardware limits).

Thus the command proposed here is a basic solution for the problem of target tracking, using a PTZ camera. In future development, improvements of the control law can be made using the modelling presented in [3, 5, 9]. Therefore as will be shown in the next section the first results obtained in real conditions are highly satisfactory.

4. APPLICATION TO VISUAL SENSOR

In this section, the hardware of the sensor equipment is described. Then results obtained on real motor-ways with the VELAC vehicle (see fig. 9) are presented. These experiments consist in tracking a preceding vehicle in real condition.

4.1. PTZ Camera

The PTZ camera is a Sony camera EVI-G21 (see fig. 8). The following table presents its main characteristics:

focal length	4.5 - 13.5mm (×3 zoom)
pan angle	$\pm 30°$
tilt angle	$\pm 15°$

Fig. 8. SONY camera

Fig. 9. VELAC

The tracking process needs a calibration stage in order to evaluate the position of the principal point (u_0,v_0) of the image and the focal lengths (in pixels). The focal lengths (F_u,F_v) are functions of the *zoom* and *focus* camera values through the relation: $a_* \cdot focus + b_* \cdot zoom + c_* \cdot F_* + d_* = 0$, ($* = u, v$). By using the calibration method given in [12], the results show:

u_0	379.97 ± 0.05 pixels
v_0	295.53 ± 0.05 pixels

-	$F_u(pixels)$	$F_v(pixels)$
a_*	-0.05049	-0.07326
b_*	0.1208	0.1207
c_*	-0.9914	-0.9900
d_*	1028	1055

4.2. Results

The algorithm has been tested on a SGI Indy Station embedded in the experimental vehicle VELAC. During these tests in real conditions on motor-ways, the reference patterns are manually selected on the rear view of vehicles.

Results obtained show that the training phase of the tracking algorithm is rather constraining because of its excessive computation time (approximately 7 seconds). So, the camera position has to be readjusted before starting the tracking.

Furthermore, the current algorithm cannot track specular surfaces. Fortunately, the non-specular surfaces are rare on standard cars and have small size (the license plate, for example). It is necessary to have a camera equipped with a powerful zoom. This is not the case of the sensor equipment presented in this article. Thus, the tests in real conditions were only efficient with vehicles of significant size and less prone to local variations of illumination, like trucks or camping-car.

On the other hand, the tracking procedure is completely suited to real time (20 ms for the tracking algorithm and less than 25 ms for the acquisition and control stage) and to real conditions. The first results obtained are presented in Fig.10, 12 and 11 (left: standard camera view, right: PTZ camera view). They show the abilities of the proposed new sensor:

- to keep a constant image-size of the tracked vehicle (Fig.10),
- to track a preceding vehicle in different situations. For example, in Fig.12, the PTZ camera tracks the camping-car during an overtaking procedure whereas the camping-car disappears from the standard camera view.

The algorithm process is relatively robust because:

- the tracking resists to global illumination changes (passing under a bridge, see Fig.11),
- the tracking time lasts for several minutes (until the target vehicles quit the motor-way).

Fig. 10. Tracking a truck

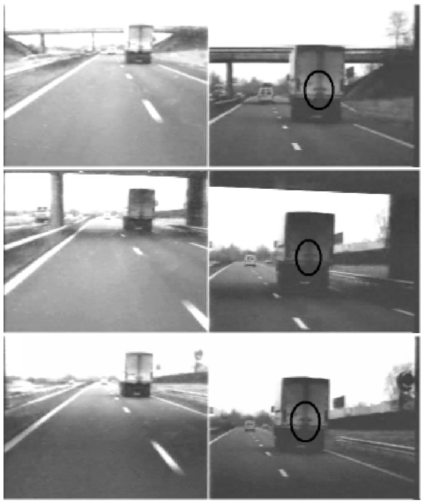

Fig. 11. Tracking a truck under a bridge

5. CONCLUSION AND FUTURE WORK

This research represents the first step in the development of a new visual sensor for car driving assistance systems. A PTZ camera is combined with a standard camera having a

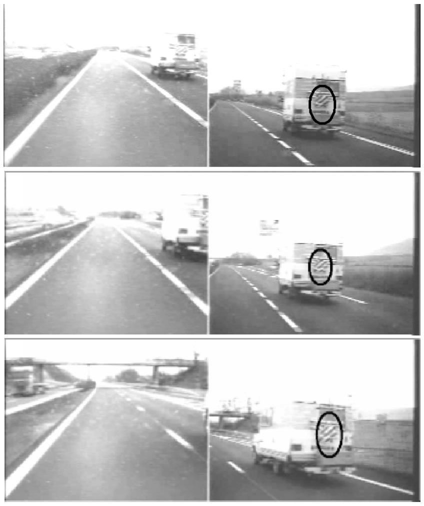

Fig. 12. Tracking a camping-car during overtaking

small focal length. This sensor equipment makes it possible to track preceding vehicles. It is based on a 2D pattern tracking algorithm and a PTZ camera controlled by visual servoing.

The results obtained in real conditions on motor-ways are very encouraging. The visual servoing process means the PTZ camera can be focused on a preceding vehicle along image sequences. The tracking algorithm runs at the video rate (20ms) and can track any kind of pattern. However, a comparaison with other tracking algorithms (like, for example, the one used in [7, 8]) will be necessary to demonstrate the effectiveness of such algorithms.

Recent research into zoom control [5] considers complex models of PTZ cameras. These models should improve the visual servoing process. Further developments of the tracking algorithm will make it robust with regard to occultations. The sensor equipment will be complete when the *Decision Module* is fully implemented.

6. REFERENCES

[1] R. Chapuis, F. Marmoiton, and R. Aufrere. Road detection and vehicles tracking by vision for an on-board acc system in the velac vehicle. In *Proceedings of the 3rd International Conference on Information Fusion*, volume 2, pages WeB5:11–18, Paris, France, July 2000. FUSION'2000.

[2] F. Duculty, M. Dhome, and F. Jurie. Tracking of 3d objects from appareance. In *Proceedings of the 12th Scandinavian Conference on Image Analysis*, Bergen, Norway, June 2001. SCIA2001.

[3] B. Espiau. Visual servoing with zoom control. Technical Report 2613, INRIA, July 1995.

[4] B. Espiau, F. Chaumette, and P. Rives. A new approach to visual servoing in robotics. *IEEE Transactions on Robotics and Automation*, 8(3), 1992.

[5] J.A. Fayman, O. Sudarsky, and E. Rivlin. Zoom tracking. In *Proceedings of the IEEE International Conference on Robotics and Automation*, volume 4, pages 2783–2788, Leuven, Belgium, May 1998. ICRA'98.

[6] G.D. Hager and P.N. Belhumeur. Efficient region tracking with parametric models of geometry and illumination. *PAMI*, 20(10):1025–1039, October 1998.

[7] E. Hayman, I. D. Reid, and D. W. Murray. Zooming while tracking using affine transfer. In *Proceedings of the British Machine Vision Conference*, pages 395–404, 1996.

[8] E. Hayman, T. Thorhallson, and D. W. Murray. Zoom-invariant tracking using points and lines in affine views : an application of the affine multifocal tensors. In *Proceedings of the International Conference on Computer Vision*, pages 269–276, Corfu, Greece, September 1999.

[9] K. Hosoda, H. Moriyama, and M. Asada. Visual servoing utilizing zoom mechanism. In *Proceedings of the IEEE International Conference on Robotics and Automation*, volume 1, pages 178–183, Nagoya, Japan, May 1995. ICRA'95.

[10] F. Huynh. *Vision-based manipulation for supervision and grasping tasks on a static or mobile object*. PhD thesis, Paul Sabatier University, Toulouse, France, 1998.

[11] D. Khadraoui, G. Motyl, P. Martinet, J. Gallice, and F. Chaumette. Visual servoing in robotics scheme using a camera/laser-stripe sensor. *IEEE Transactions on Robotics and Automation*, 12(5):743–750, October 1996.

[12] M. Li and J.M. Lavest. Some aspects of zoom lens camera calibration. *Transactions on Pattern Analysis and Machine Intelligence*, 18:1105–1110, November 1996.

[13] H. Pfannschmidt. Safety measures in transportation systems. In *Proceeding of the World Engineers'Convention : Professional Congress Mobility*, pages 153–176, Hannover, Germany, June 2000.

[14] S. Schultz, J.W. Rozenblit, and K. Buchenrieder. Towards an application of model-based codesign: An intelligent autonomous cruise controller. In *Proceedings of the 1997 IEEE Conference and Workshop on Engineering of Computer Based Systems*, volume 1, Monterey, Canada, May 1997.