Vision-based Control for Car Platooning using Homography Decomposition *

S. Benhimane, Ezio Malis, P. Rives, J.R. Azinheira

Vision-based Control for Car Platooning using Homography Decomposition*

Selim Benhimane, Ezio Malis and Patrick Rives
INRIA
2004 route des Lucioles, BP93
06902, Sophia Antipolis cedex, France
Firstname.Secondname@sophia.inria.fr

José Raul Azinheira
DEM/IDMEC
Instituto Superior Tecnico
av Rovisco Pais, 1 - 1049-001 Lisboa, Portugal
jraz@dem.ist.utl.pt

Abstract— In this paper, we present a complete system for car platooning using visual tracking. The visual tracking is achieved by directly estimating the projective transformation (in our case a homography) between a selected reference template attached to the leading vehicle and the corresponding area in the current image. The relative position and orientation of the servoed car with regard to the leading one is computed by decomposing the homography. The control objective is stated in terms of path following task in order to cope with the non-holonomic constraints of the vehicles.

Index Terms— Car platooning, vision-based control, real-time visual tracking, non-holonomic vehicles.

I. Introduction

Many researches on automatic driving are currently done in the laboratories working on Intelligent Vehicles. Among these researches, a major issue is devoted to studying the platooning problem. A key point to be addressed is the robust detection and tracking of the ahead vehicle. At the present time, the most operational systems are based on distance sensors such as radar or lidar. This kind of sensors has the advantage of providing a measurement which can be directly exploited in the feedback control. Alternative or complementary sensors could be linear [4], standard [2] or omnidirectional [3] cameras. Although they are sensitive to weather and lighting conditions during an outdoor use, they provide a richer information which can be exploited to make the algorithms more robust. Moreover, recent developments on high dynamic range CMOS cameras [6] specifically dedicated to automotive applications could provide real solutions in terms of robustness with regard to the lighting changing conditions.

In this paper, we present a complete vision-based platooning approach running in an outdoor environment. The visual tracking is achieved by directly estimating the projective transformation (in our case a homography) between a selected reference template attached to the leading vehicle and the corresponding area in the current image [1]. The relative position and orientation of the servoed car with regard to the leading one, is computed by decomposing the homography [5]. The control objective is stated in terms of path following task in order to cope with the

non-holonomic constraints of the vehicles. A virtual frame rigidly linked to the leading vehicle is defined and the path tracking error is expressed in this Cartesian space using the relative pose computed through the data given by the camera at video rate.

The paper is organized as follows. A first part deals with the modeling and the control aspects. A second and a third parts are dedicated to the visual tracking. Finally, experimental results are presented and discussed.

II. Control Aspects

A. Kinematic modeling

Let us consider a car-like vehicle with a pure rolling without slippage assumption (i.e. the velocity vector is always perpendicular to the rear wheels axis, as depicted in fig. 1). The kinematic model of the vehicle is given by :

Fig. 1

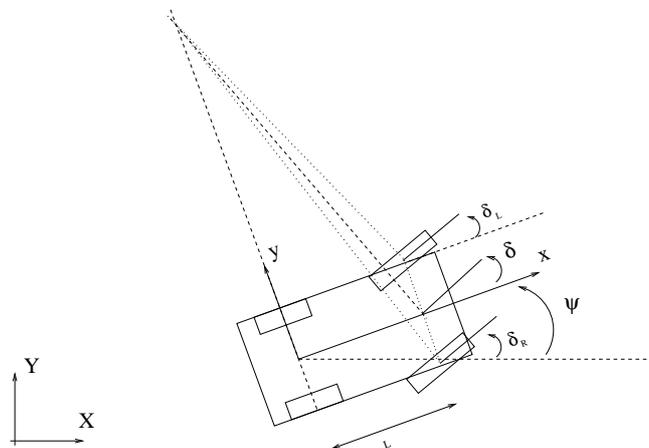ABSOLUTE (GROUND) AND LOCAL (VEHICLE) FRAMES

$$
\begin{cases}
\dot{x} = U\cos(\psi) \\
\dot{y} = U\sin(\psi) \\
\dot{\psi} = \frac{U}{L}\tan(\delta)
\end{cases} \tag{1}
$$

where (x, y) are the absolute Cartesian coordinates, U is the longitudinal velocity, ψ is the heading angle, and δ is the steering angle. The state of the system is represented by the (3×1) vector (x, y, ψ).

B. Path tracking

Consider the problem of tracking a moving reference frame $\mathcal{F}^*$, attached to a leading car, by a servoed car. The origin of the reference frame $\mathcal{F}^*$ is the moving point $\mathbf{q}^* = (x^*, y^*)$, while the point $\mathbf{q} = (x, y)$ is the origin of the current frame $\mathcal{F}$ (see figure 2). The reference frame is moving along a given trajectory. The $\vec{x^*}$ axis is always tangent to the curve (i.e. $\dot{y}^* = 0$). Let $\dot{x}^* = U^*$ be the longitudinal velocity of the reference frame. The angular velocity $\dot{\psi}^*$ is linked to the steering angle δ^* of the leading car. The path tracking error $\mathbf{e} = (e_x, e_y)$ is the difference
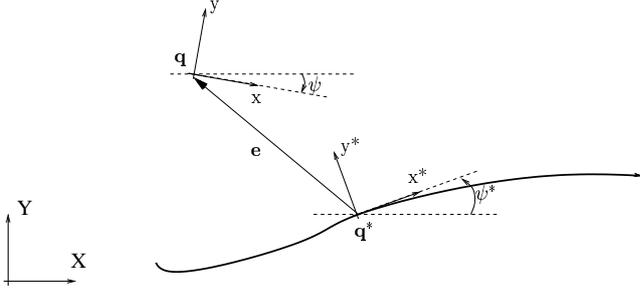

Fig. 2

PATH TRACKING

between the current position $\mathbf{q}$ and the desired position $\mathbf{q}^*$ of the vehicle, and it is defined in the reference frame as follows:

$$\mathbf{e} = \mathbf{R}_{\psi^*}^\top (\mathbf{q} - \mathbf{q}^*) = \mathbf{R}_{\psi^*}^\top \begin{bmatrix} x - x^* \\ y - y^* \end{bmatrix} \quad (2)$$

where $\mathbf{R}_{\psi^*}$ is the rotation of angle ψ^*:

$$\mathbf{R}_{\psi^*} = \begin{bmatrix} \cos(\psi^*) & -\sin(\psi^*) \\ \sin(\psi^*) & \cos(\psi^*) \end{bmatrix}$$

The error in the orientation between the current and the reference frames is $e_\psi = \psi - \psi^*$. Computing the derivative of the equation (2) around the origin ($x^* = 0$, $y^* = 0$ and $\psi^* = 0$) yields:

$$\begin{bmatrix} \dot{x} \\ \dot{y} \end{bmatrix} \approx \begin{bmatrix} \dot{e}_x - \psi^* \dot{e}_y + U^* \\ \dot{e}_y + \psi^* \dot{e}_x \end{bmatrix} \quad (3)$$

Using the equations (1) and (3), a local kinematic model can be deduced:

$$\begin{cases} \dot{e}_x \approx U - U^* = u \\ \dot{e}_y \approx U (\psi - \psi^*) = U e_\psi \end{cases}$$

For the vehicle dynamics, a simplified first-order model can be used for the control design:

$$\begin{cases} \dot{u} = (U_c - U)/\tau_u \\ \dot{\delta} = (\delta_c - \delta)/\tau_\delta \end{cases}$$

where (U_c, δ_c) are the vehicle inputs, respectively the longitudinal speed and steering angle commands and (τ_u, τ_δ) are

positive scalars. The linearized 5th order dynamic model used for the control design is:

$$\begin{cases} \dot{e}_x = u \\ \dot{e}_y = U e_\psi \\ \dot{e}_\psi = U\delta/L \\ \dot{u} = (U_c - U)/\tau_u \\ \dot{\delta} = (\delta_c - \delta)/\tau_\delta \end{cases}$$

This model appears indeed decoupled into a longitudinal model, with state (e_x, u) and a lateral model, with state (e_y, e_ψ, δ), and a simple feedback, computed with the standard LQR tool, leads to:

$$\begin{cases} U_c = U^* - \mathbf{k}_u^\top \begin{bmatrix} e_x \\ u \end{bmatrix} \\ \delta_c = \delta^* - \mathbf{k}_\delta^\top \begin{bmatrix} e_y \\ e_\psi \\ \delta - \delta^* \end{bmatrix} \end{cases} \quad (4)$$

where $\mathbf{k}_u = (k_{u1}, k_{u2})$ and $\mathbf{k}_\delta = (k_{\delta 1}, k_{\delta 2}, k_{\delta 3})$ and k_{u1}, k_{u2}, $k_{\delta 1}$, $k_{\delta 2}$, $k_{\delta 3}$ are positive scalars.

C. Distance-based lateral tracking

As an alternative more robust solution, a distance-based design was used for the lateral path tracking. The approach is based on the assumptions that no slippage occurs and that the actuator loop time-rate is shorter than the overall closed-loop time-rate. Thus, the path tracking problem can be formulated as a "speed-independent" problem. The curvilinear distance s is related to the vehicle speed by:

$$U = \frac{ds}{dt}$$

We perform a change of variables in order to consider the curvilinear distance s as the independent variable (instead of the time variable t):

$$\frac{dx}{ds} = \frac{dx}{dt}\frac{dt}{ds} = \frac{1}{U}\frac{dx}{dt} \quad (5)$$

Using equation (5), the kinematic expression (equation (1)) can be reformulated:

$$\begin{cases} x' = \cos(\psi) \\ y' = \sin(\psi) \\ \psi' = \frac{1}{L}\tan(\delta) \end{cases}$$

where x', y' and ψ' are the notations for the derivatives $\frac{dx}{ds}$, $\frac{dy}{ds}$ and $\frac{d\psi}{ds}$. The new distance-based model is formally similar to the previous one, with the advantage of having the lateral model independent to the vehicle longitudinal speed U. Since the controller is a pure feedback gain, without dynamics, a classical tool can be used for the control design. The lateral model with state (e_y, e_ψ, δ), is independent on the velocity U:

$$\begin{cases} e_y' = e_\psi \\ e_\psi' = \delta/L \\ \dot{\delta} = (\delta_c - \delta)/\tau_\delta \end{cases}$$

We use again the control law (4) for controlling the vehicle.

D. Platooning

In order to adapt the previous control approach for a platooning application, a new control objective is defined in terms of tracking a virtual reference frame $\mathcal{F}^*$ rigidly linked to the leading vehicle as shown in the figure 3. The virtual reference frame is obtained by translating the
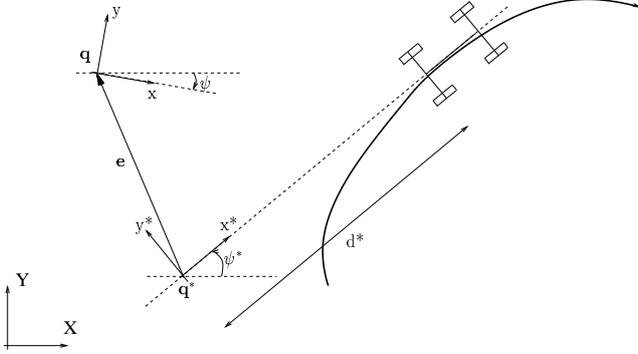

Fig. 3

VEHICLE TRACKING

frame of the leading vehicle by a distance d^*. In practice, the virtual reference frame and the camera frame of the follower vehicle coincide at the beginning of the tracking. The errors $\mathbf{e} = (e_x, e_y)$ and e_ψ measured by the vision system are the relative positions of the reference and current frames expressed in the virtual reference frame. The relative position of the leader vehicle can be measured at video rate by decomposing a homography matrix which defines the projective transformation of a plane attached to the leader vehicle. The distance-based lateral tracking described in the previous section is applied to control the measured error. The longitudinal speed U^* of the leader vehicle is estimated on-line from vision data.

III. CARTESIAN STATE RECONSTRUCTION

We suppose that we observe a planar object and we suppose that a reference template corresponding to the frame $\mathcal{F}^*$ has been selected during an off-line step. The current image acquired at each iteration of the control scheme corresponds to frame $\mathcal{F}$. The reference template is related to the current image by a homography matrix $\mathbf{G}$. Indeed, a point in the reference template, with homogeneous coordinates $\mathbf{p}^*$, is projected onto a point in the current image by $\mathbf{p} \propto \mathbf{G}\mathbf{p}^*$. Knowing the upper triangular matrix $\mathbf{K}$ containing the camera intrinsic parameters, we can extract the camera displacement by decomposing the homography matrix:

$$\mathbf{G} = \mathbf{K}(\mathbf{R} + \frac{\mathbf{t}}{d^*}\mathbf{n}^{*\top})\mathbf{K}^{-1}$$

where $\mathbf{R} \in SO(3)$ and $\mathbf{t} \in \mathbb{R}^3$ are respectively the rotation matrix and the translation vector between the frames $\mathcal{F}$ and $\mathcal{F}^*$, $\mathbf{n}^*$ is the unit vector normal to the plane expressed in $\mathcal{F}^*$ and d^* is the distance between the plane and the center of the frame $\mathcal{F}^*$. In general, there are two possible solutions to the homography decomposition [5]. In order to distinguish the right solution, an estimation of the normal $\mathbf{n}^*$ should be known. The relative position and orientation of the leader needed for the control law can be extracted from $\mathbf{R}$ and $\mathbf{t}$. The homography matrix $\mathbf{G}$ can be estimated using the template-based visual tracking technique described in the next section.

IV. VISUAL TRACKING

We consider an on-board camera mounted on the servoed vehicle looking at a planar target attached on the back of the leading vehicle. The visual tracking can be achieved by directly estimating the projective transformation (in our case a homography) between a selected reference template and the corresponding area in the current image. The core of the visual tracking method is the Efficient Second-order Minimization (ESM) algorithm proposed in [7]. The application of the ESM algorithm to the visual tracking allows an efficient real-time homography estimation and a template-based tracking with high inter-frame displacements. The figure 4 gives a general overview of the method. A detailed description of the tracking method can be found in [1].

Since the homography matrix $\mathbf{G}$ is defined up to a scalar factor, without loss of generality, it can always be considered as an element of the SL(3) group (i.e. the group of (3×3) matrices with the determinant equal to 1). Indeed, if $\det(\mathbf{G}) = 0$ then the plane passes through the optical center and all the points on the plane project on a line. Starting from an initial prediction of the homography, we iteratively estimate the optimal homography which minimizes the SSD (the sum of squared differences) between a reference pattern $\mathbf{T}$ and the current pattern $\mathbf{W}$ reprojected using the current homography $\mathbf{G}$. If an initial prediction of the homography is not available, we start with $\mathbf{G}$ equal to the identity matrix. Both the image derivatives of the template $\nabla\mathbf{T}$ and the image derivatives of the current pattern $\nabla\mathbf{W}$ are used to obtain an efficient second-order update. It is an efficient algorithm since only first image derivatives are used and the Hessians are not explicitly computed.

In order to improve the tracking algorithm, we use a multiresolution method which makes it possible to deal with large motion [8]. A change of the resolution of the image can be obtained by an affine transformation and thus by a homography. Let ${}^i\mathbf{S}_j$ be the homography matrix which allows to warp the image from resolution j to resolution i. When the reference pattern $\mathbf{T}_0$ is selected in the reference image at full resolution, we warp it n times (using the homographies ${}^1\mathbf{S}_0, {}^2\mathbf{S}_0, ..., {}^n\mathbf{S}_0$) until we reach a minimum resolution (e.g. the size of the pattern at resolution n should be greater than 20×20 pixels). We obtain once and for all $n+1$ reference patterns $\mathbf{T}_0, \mathbf{T}_1,..., \mathbf{T}_n$. If the homography transforming an area of the current image in the reference template at scale j is ${}^j\mathbf{G}$, then the homography transforming that area of the current image in
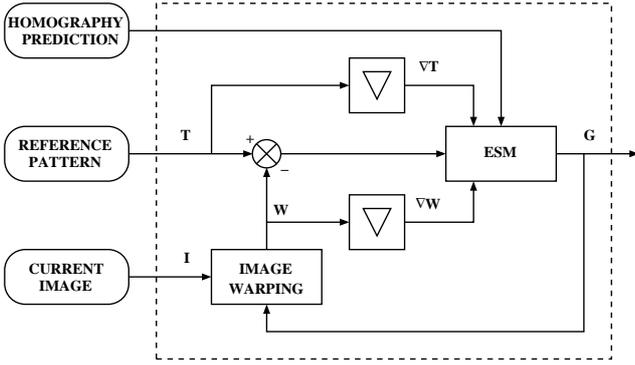
Fig. 4

VISUAL TRACKING BASED ON AN EFFICIENT SECOND-ORDER MINIMIZATION METHOD.

the reference template at scale i is:

$$^{i}\mathbf{G} = {}^{i}\mathbf{S}_{j}\ {}^{j}\mathbf{G}$$

The multiresolution ESM tracking is started at the scale n using an initial estimation homography $^{n}\mathbf{G}$ (if a prediction is not available, we set $^{n}\mathbf{G}$ equal to the matrix $^{n}\mathbf{S}_0$). Once the ESM algorithm has computed the homography $^{n}\mathbf{G}$, we simply obtain the homography $^{n-1}\mathbf{G}$ by changing the scale $^{n-1}\mathbf{G} = {}^{n-1}\mathbf{S}_n\ {}^{n}\mathbf{G}$. The ESM algorithm is repeated $n+1$ times and at the scale 0, we obtain the homography $^{0}\mathbf{G}$. The loop is repeated by rescaling the homography at the higher scale $^{n}\mathbf{G} = {}^{n}\mathbf{S}_0\ {}^{0}\mathbf{G}$. The algorithm can also be stopped at scale $k > 0$ of the pyramid (e.g. if the computation time is limited) by rescaling the homography $^{0}\mathbf{G} = {}^{0}\mathbf{S}_k\ {}^{k}\mathbf{G}$. The multiresolution approach improves the visual tracking a lot since it permits to have a coarse-to-fine strategy.

V. EXPERIMENTAL RESULTS

In this experiment, a vision-based car platooning is performed in a real outdoor environment. Two electric vehicles of type "Cycab" (see the figure 5) are used one as a leading car and the other as the servoed car. The leading car is manually driven. The servoed car, in automatic control mode, is equipped with a camera mounted on a pan-tilt turret situated behind its front windshield (see the figure 6) and with two computers (Pentium III 700 MHz). The first one is used for the vision, the control law computations and the pan-tilt turret control which is done via a serial communication RS232. The second computer is devoted for the low-level control of the "Cycab" velocity and the wheel steering. An internal network is set between the two computers and the data transfer is done thanks to a TCP/IP socket communication process. The control scheme presented in the previous sections takes into account that the vehicle is non-holonomic and tries to keep the distance between the two vehicles constant and equal to the initial distance. The relative position is computed from the vision at video rate. The pan-tilt turret is controlled in order to

keep the leading car in the field of view of the camera during the experiment.

Fig. 5

THE ROBOTIC SYSTEM USED IN OUR EXPERIMENTS.

Fig. 6

THE PAN-TILT TURRET SITUATED BEHIND THE FRONT WINDSHIELD OF THE SERVOED "CYCAB".

Since the leading car back windshield is transparent, a poster is sticked on it in order to be used as a target for the visual tracker. In the starting situation, when the leading car is in front of the servoed car, a window of (100×100) pixels is selected to be the reference pattern. In order to have a metric reconstruction and reasonable gains, the camera has been roughly calibrated and the distance between the two cars is given to the control process. All of the vision sensor, the image tracking, the pan control and the position estimation act as a unique sensor that provides the displacement (e_x, e_y, e_ψ) between the current vehicle frame and the virtual reference frame. The visual tracking algorithm computes the displacement $(x_t, y_t$ and $\psi_t)$ between the current camera frame and the virtual

reference frame. After measuring the pan angle ϕ, the displacement (x_v, y_v, ψ_v) of the vision frame is computed as follows:

$$\left\{ \begin{array}{l} \left[\begin{array}{c} x_v \\ y_v \end{array} \right] = \mathbf{R}_\phi \left[\begin{array}{c} x_t \\ y_t \end{array} \right] \\ \qquad \psi_v = \psi_t - \phi \end{array} \right.$$

The vision frame does not coincide with the vehicle frame. Thus, the error between the virtual reference frame and the current vehicle frame is obtained as follows:

$$\left\{ \begin{array}{l} \left[\begin{array}{c} e_x \\ e_y \end{array} \right] = \mathbf{R}_{-\frac{\pi}{2}} \left[\begin{array}{c} x_v \\ y_v \end{array} \right] \\ \qquad e_\psi = -\psi_v \end{array} \right.$$

Finally, we apply the control described in section II.

Filtering vision measurements : The vision sensor is characterized by some specific issues:

1) sampling rate reduced to $25Hz$ or $40ms$, with a further delay around $40ms$ due to image processing and position computing;
2) a pan device was introduced in the vision system, regulating the position x_v and allowing to keep the target inside the image, but the speed limitation of this pan device limited the range of the allowable angular speed during the visual tracking process;
3) intrinsic noise, reduced in static conditions, but increasing with speed;
4) isolated noise peaks, probably when vision hasn't fully finished its optimization process;
5) high sensitivity to the lighting conditions;
6) oscillations due to possible "ambiguity" between the real physical position solution and an erroneous second solution when decomposing the homography.

In order to reduce the effects of issues 3 and 4, a filtering was introduced with a non-linear part, limiting the measurement rate of change:

$$\left\{ \begin{array}{l} |x_{v,k} - x_{v,k-1}| < \dot{x}_{v,max} \\ |y_{v,k} - y_{v,k-1}| < \dot{y}_{v,max} \\ |\psi_{v,k} - \psi_{v,k-1}| < \dot{\psi}_{v,max} \end{array} \right.$$

and a linear part, consisting of a simple first-order low-pass filter:

$$\left\{ \begin{array}{l} x_{v,k} = c x_{v,k-1} + (1-c) x_{v,k} \\ y_{v,k} = c y_{v,k-1} + (1-c) y_{v,k} \\ \psi_{v,k} = c \psi_{v,k-1} + (1-c) \psi_{v,k} \end{array} \right.$$

with $c = 0.6$, corresponding to a time of constant $20ms$. Figure 7 shows the experimental results of a car platooning application. The top-right curve exhibits the relative position measurement as produced by the vision system, showing a longitudinal distance (in green) smoothly regulated around $3m$. This distance is reduced during the first part of the turn, due to the tracking kinematics, and then at the end, when the vehicle slows down and stops. The bottom-right curve shows the filtering effect on the vision output. Finally, the bottom-left curve shows the control variables: in the speed demand (in blue) a component of noise is clearly present, and the $1m/s$ saturation is reached at the
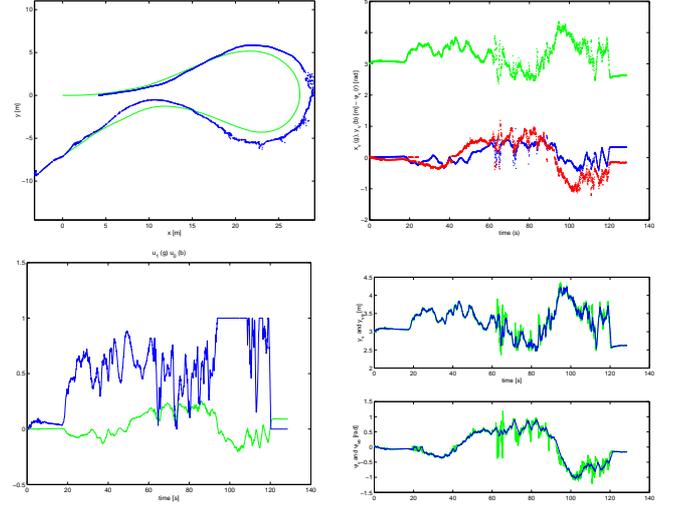

Fig. 7

CAR PLATOONING EXPERIMENT. TOP-LEFT: HORIZONTAL PATH, WITH ODOMETRY (GREEN) AND VIEWED TARGET (BLUE); TOP-RIGHT: VISION OUTPUT WITH x_v(BLUE), y_v(GREEN) AND ψ_v(RED); BOTTOM-LEFT: CONTROL, WITH SPEED DEMAND (BLUE) AND STEERING ANGLE (GREEN); BOTTOM-RIGHT: VISION FILTERING, WITH y_v ABOVE AND ψ_v BELOW.

end of the turn (this saturation level is very low and should be put at $2m/s$ by instance in future experiments). The steering angle (in green) is less noisy.

The figure 8 illustrates another experiment where the leading car is driven in a 100 meter long closed loop. In the first and the third rows, the relative position between the leading car and the servoed car can be seen. In the second and the fourth rows, the corresponding images grabbed during the experiment and used as input for the ESM visual tracking algorithm are shown. The blue square indicates the tracked region. Thanks to the pan-tilt turret servoing, the pattern tracked remains in the center of the image during the experiment. The tracking algorithm performs well although the experiment takes place outdoor and sun reflection on the tracked region occurs.

VI. CONCLUSION

The vision-based control system presented in this paper is able to perform car platooning in an outdoor environment. The control of the follower vehicle is achieved by reconstructing the position of the leading vehicle using an on-board camera. The non-holonomic constraints are taken into account in the control law. The visual tracking algorithm is able to track a target without any particular texture or evident features. Thanks to its flexibility, the visual tracking system can be modified in order to track directly the back of the leading vehicle. Indeed, a robust estimation algorithm (e.g. M-estimators) can be used instead of the standard least square minimization. Robust algorithms can also help to handle illumination changes.
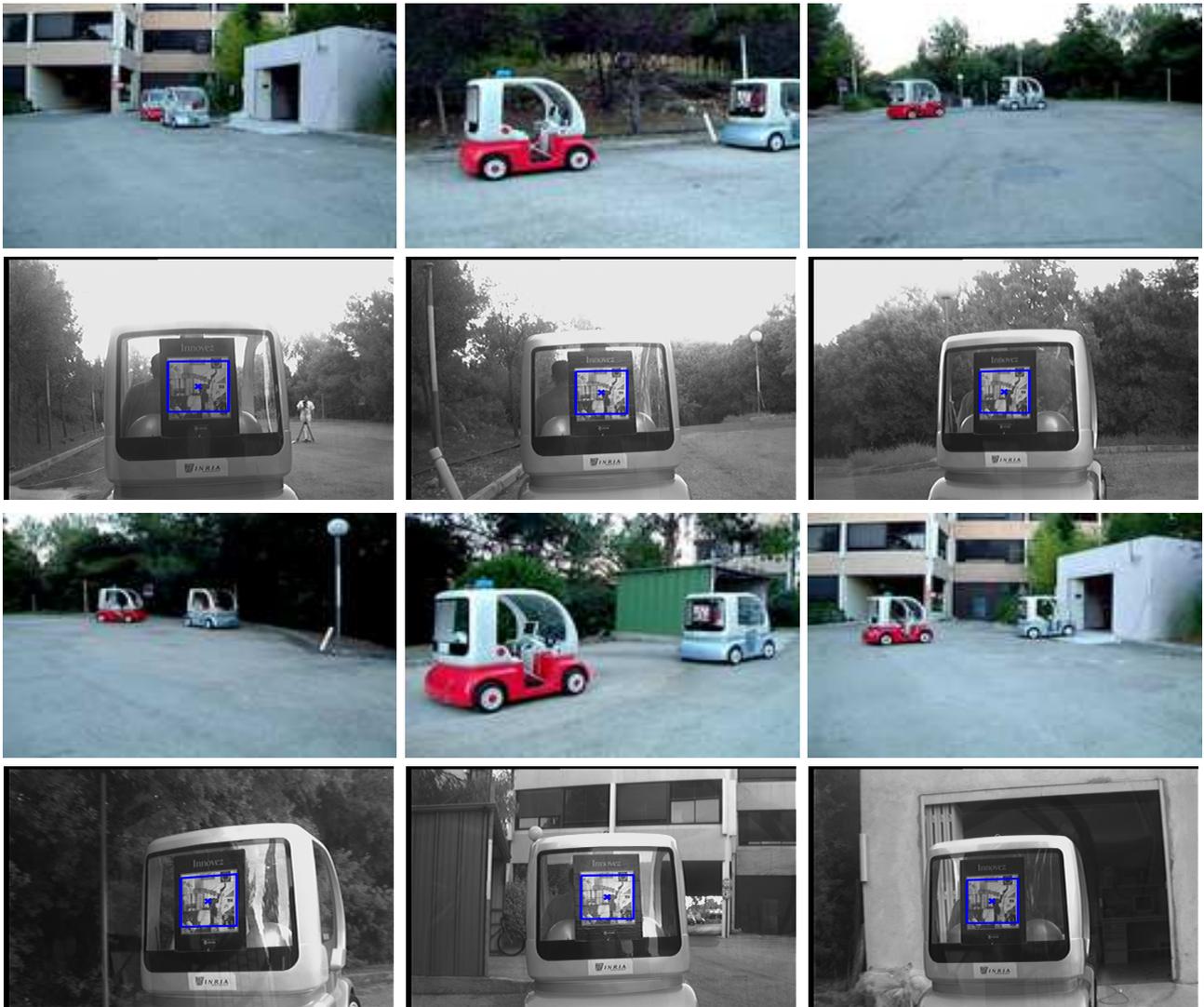
Fig. 8

IMAGES OF A CAR PLATOONING APPLICATION. THE IMAGES IN THE FIRST AND THE THIRD ROWS ARE TAKEN WITH AN EXTERNAL CAMERA. THE IMAGES IN THE SECOND AND THE FOURTH ROWS ARE TAKEN WITH THE CAMERA ON-BOARD AND SHOW THE REGION TRACKED IN THE CURRENT IMAGE (SEE THE BLUE SQUARES).

REFERENCES

[1] S. Benhimane and E. Malis. Real-time image-based tracking of planes using efficient second-order minimization. In *IEEE/RSJ Int. Conf. on Intelligent Robots Systems*, pages 943–948, 2004.

[2] X. Cindy, F. Collange, F. Jurie, and P. Martinet. Object tracking with a pan-tilt-zoom camera: application to car driving assistance. In *IEEE Int. Conf. on Robotics and Automation*, pages 1653–1658, 2001.

[3] A. K. Das, R. Fierro, V. Kumar, J. P. Ostrowski, J. Spletzer, and C. J. Taylor. A framework for vision based formation control. *Multi-Robot Systems: A Special Issue of IEEE Transactions on Robotics and Automation*, 18(5):813–825, 2002.

[4] F. Daviet and M. Parent. Platooning for small public urban vehicles. In *4th International Symposium on Experimental Robotics*, pages 345–354, 1997.

[5] O. Faugeras and F. Lustman. Motion and structure from motion in a piecewise planar environment. *Int. Journal of Pattern Recognition and Artificial Intelligence*, 2(3):485–508, 1988.

[6] S. Kavusi and A. El Gamal. Quantitative study of high dynamic range image sensor architectures. In *Proceedings of the SPIE Electronic Imaging Conference*, 2004.

[7] E. Malis. Improving vision-based control using efficient second-order minimization techniques. In *IEEE Int. Conf. on Robotics and Automation*, pages 1843–1848, 2004.

[8] J. M. Odobez and P. Bouthemy. Robust multiresolution estimation of parametric motion models. *Jour. of Visual Comm. and Image Representation*, 6(4):384–365, 1995.