12^{th} *World Congress IFAC, Sydney, Australia, Vol.9, pp. 409-414, July 1993*

TRACKING A MOVING OBJECT BY VISUAL SERVOING

F. CHAUMETTE and A. SANTOS

IRISA / INRIA Rennes, Campus de Beaulieu, 35042 Rennes Cedex, France.

Abstract. This paper deals with target tracking by visual servoing. We first briefly describe the visual servoing approach and the application of the task function concept to vision-based tasks. Next, we present a complete control scheme which explicitly enables to pursue a moving object. This control scheme is based on the estimation and prediction of the target motion in the image through Kalman filtering. Finally, real-time experimental results using a camera mounted on the end effector of a six-d.o.f. robot are presented.

Key Words. Visual servoing; target tracking; task function; Kalman filtering.

1 INTRODUCTION

Recent advances in vision sensor technology and image processing now allow the effective use of visual data in the control loop of a robot. In robotics applications, this enables us to handle uncertainties and/or changes in the environment (for example, to compensate for small positioning errors, to grasp objects moving on a conveyor belt, etc.). Concerning vision aspects, it becomes possible to control the camera motion in order to improve recognition, localization or inspection of the environment.

In this paper, we present an adaptive and predictive vision-based control scheme which facilitates various robotics tasks such as positioning a camera with respect to a moving object. This control law computes the components of the camera velocity ensuring an exponential decrease of the task function error. In order to take into account the unknown motion of the target, which generally induces tracking errors, an estimation scheme of the resulting motion in the image is proposed. To compensate for this motion, its prediction, obtained through augmented Kalman filtering with a constant acceleration state model, is introduced in the control law. The experimental results described at the end of this paper show the robustness of the proposed control scheme with respect to noise and mismeasurement.

2 VISUAL SENSING AND TASK FUNCTION

The work described in this paper is based on the visual servoing approach. In this approach (Weiss and Sanderson, 1987; Feddema and Lee, 1989; Papanikolopoulos *et al.*, 1991; Espiau *et al.*, 1992), vision data is modeled as a set $\underline{s}$ of elementary visual signals which only depends on the relative position and orientation between the camera and the scene (for example, $\underline{s}$ may be chosen as the coordinates of a point, or the parameters describing a straight line, an ellipse, etc.). For a given vision-based task, a desired target image is identified, consisting of a chosen set of values, $\underline{s}^*$ (corresponding to a good achievement of the defined task). Considering the desired target image and the image currently observed by the camera, the control problem is then reduced to the regulation in the image of $(\underline{s} - \underline{s}^*)$.

Referring to earlier developments (Espiau *et al.*, 1992), the time variation of $\underline{s}$ can be modeled through:

$$\dot{\underline{s}} = L_{\underline{s}}^T \, T \ , \qquad (1)$$

where

- $T = (V, \Omega)$ is the velocity screw quantifying the relative motion between the scene and the camera (V and Ω respectively represent the translational and rotational components of T);
- $L_{\underline{s}}^T$, called the interaction screw related to $\underline{s}$, completely characterizes the interaction between the sensor and its environment and can be explicitly computed for the parameters describing the projection in the image of geometrical primitives such as points, lines, circles, etc.

More precisely, let us now consider the situation $\underline{r}$ of the camera with respect to a reference frame,

and the camera velocity screw, T_c, with $T_c = \frac{d\underline{r}}{dt}$. We have $\underline{s} = \underline{s}(\underline{r}(t), t)$, thus $\dot{\underline{s}}$ has the following form:

$$\dot{\underline{s}} = L_{\underline{s}}^T \; T_c + \frac{\partial \underline{s}}{\partial t} \; , \tag{2}$$

where $\frac{\partial \underline{s}}{\partial t}$ represents the contribution of the object motion in the image.

Applying the task function approach developed by Samson *et al.* (1990) to the case of visual sensor, we define a vision-based task function $\underline{e}(\underline{r}(t), t)$ of the form (Espiau *et al.*, 1992):

$$\underline{e} \; = \widehat{L_{\underline{s}}^T}^+ \, (\underline{s}(\underline{r}(t), t) \; - \; \underline{s}^*) \; , \tag{3}$$

where $\widehat{L_{\underline{s}}^T}^+$ is the pseudo-inverse of a chosen model of $L_{\underline{s}}^T$ ($\widehat{L_{\underline{s}}^T}^+ \widehat{L_{\underline{s}}^T} = \mathbb{I}$) and can be considered as the inverse Jacobian matrix related to the task.

For simplicity, we do not present here the redundancy framework of the task function approach, using which it is possible to combine a vision-based task described by (3) with another task (such as, for example, a desired camera motion using the non-constrained camera d.o.f.), even if all the results developed further remain valid for such a framework (see Santos and Chaumette, 1992).

3 CONTROL SCHEME

Considering the control problem as a closed loop regulation of $\underline{e}$, we can ensure that the task $\underline{e}$ is perfectly achieved if, at each time t, $\underline{e}(\underline{r}(t), t) = 0$. In order that $\underline{e}$ exponentially decreases, and then behaves like a first order decoupled system, the desired evolution of $\underline{e}$ takes the form:

$$\dot{\underline{e}} = -\lambda \; \underline{e} \;\; \text{with} \;\; \lambda > 0 \; . \tag{4}$$

Since $\underline{e}$ is function of $\underline{r}$ and t, we have:

$$\dot{\underline{e}} \; = \; \left(\frac{\partial \underline{e}}{\partial \underline{r}} \right) \; T_c \; + \; \frac{\partial \underline{e}}{\partial t} \; . \tag{5}$$

Furthermore, the camera velocity T_c is related to the position of the robot joints $\underline{q}$ through the relation $\dot{\underline{q}} \; = \; J^{-1}(\underline{q}) \; T_c$, J^{-1} representing the inverse Jacobian matrix of the robot. We will assume that this Jacobian is known and is non-singular, and that the accessible parameter to control the robot is the robot desired joint kinematics $\dot{\underline{q}}_c$, as it is generally the case in most industrial applications. Considering the desired camera velocity vector T_c as the pseudo-control term, from (4) and (5), we have:

$$T_c \; = \; \left(\widehat{\frac{\partial \underline{e}}{\partial \underline{r}}} \right)^+ \left(-\hat{\lambda} \; \underline{e} \; - \; \widehat{\frac{\partial \underline{e}}{\partial t}} \right) \; , \tag{6}$$

where

- $\widehat{\frac{\partial \underline{e}}{\partial \underline{r}}}$ can be taken as the identity matrix $\mathbb{I}$ under certain conditions described by Samson *et al.* (1990), particularly in our vision-based task case;
- $\hat{\lambda}$ is a proportional coefficient involved in the exponential convergence of $\underline{e}$, and which has to be tuned in order to preserve the stability of the system and to optimize the time to convergence of the task function (see Santos and Chaumette (1992) for more details on this tuning); and
- $\widehat{\frac{\partial \underline{e}}{\partial t}}$, on which we will now focus, is the estimate of the contribution of the target motion to the control law.

Chaumette *et al.* (1991) describes such an estimator which only requires the measurement of the successive visual data. However, this estimation scheme enables to pursue a moving target without tracking errors only if that target has a constant velocity. In order to consider more complex situations, let us now assume that the camera motion T_c, due to the applied control law, can also be measured through the successive position of the robot joints. In that case, using equation (5), we can easily compute an estimate of the target motion in the image and we obtain:

$$\left(\widehat{\frac{\partial \underline{e}}{\partial t}} \right) = \widehat{\dot{\underline{e}}} - \left(\widehat{\frac{\partial \underline{e}}{\partial \underline{r}}} \right) T_c \; . \tag{7}$$

After discretization, this relation becomes:

$$\left(\widehat{\frac{\partial \underline{e}}{\partial t}} \right)_{(k)} = \frac{\underline{e}_{(k)} - \underline{e}_{(k-1)}}{\Delta t} - \left(\widehat{\frac{\partial \underline{e}}{\partial \underline{r}}} \right)_{(k)} T_{c(k-1)} \; . \tag{8}$$

Remark:. Let us assume that the object is motionless and that the camera observes $\underline{s}_{k-1}$ at the sampling period $(k-1)\Delta t$. Using a first order approximation leads to:

$$\underline{s}_{(k)/(k-1)} = \underline{s}_{k-1} + \dot{\underline{s}} \; \Delta t \; . \tag{9}$$

Using (2) and (3), we obtain the following prediction of the task function value:

$$\underline{e}_{(k)/(k-1)} = \underline{e}_{(k-1)} + \widehat{L_{\underline{s}}^T}^+_{(k)} \widehat{L_{\underline{s}}^T}_{(k)} T_{c(k-1)} \Delta t \tag{10}$$

The estimator of the target motion can thus be written:

$$\left(\widehat{\frac{\partial \underline{e}}{\partial t}} \right)_{(k)} = \frac{\underline{e}_{(k)} - \underline{e}_{(k)/(k-1)}}{\Delta t} \; , \tag{11}$$

which represents (modulo the sampling period) the discrepancy between the actual measure of the

task function value and the predicted one. We note that this discrepancy is null if the target is motionless, and constant if the target velocity is constant. Its evolution thus has the same model that of the target motion.

Furthermore, a prediction of the position $\underline{s}$ of the visual features in the image is given by the relation (9). Its benefit comes from on the fact that searching for image features is generally time-consuming. This searching time is highly reduced using the obtained prediction for the next image, whose processing is then reduced, after the necessary recognition stage, to a simple verification process (Feddema and Lee, 1990). Furthermore, in case of mismeasurement, the predicted position in the image enables to pursue the target tracking by ensuring that it is a sufficiently good estimation.

Let us now come back to the control point of view and let us consider robustness issues. Two different sources for noise are possible in our estimation scheme: it can be either introduced through the extraction of the visual data or due to robot joint position measurement errors.

According to several works investigating the field of filtering for target tracking (Blackman, 1986; Hunt and Sanderson, 1982), two common approaches are employed: the first consists in using fixed tracking coefficients ($\alpha - \beta$, $\alpha - \beta - \gamma$ trackers), and the second, Kalman filtering, generating time-variable tracking coefficients that are determined by *a priori* models of target dynamics. While the first approach has computational advantages, the second one seems much more appealing, thanks to the adaptability of its coefficients for tracking highly maneuvering targets. However, implementing a Kalman filter requires first to define, or estimate, the state model evolution of the parameters, the simplest cases for motion parameters being the constant speed and constant acceleration models.

When a target maneuvers (for example when abrupt changes in its acceleration occur), a tracking filter should respond. Such maneuvering may be detected by a rapid increase in the normalized discrepancy. The recommended methods for dealing with those situations are numerous (Blackman, 1986; Brown *et al.*, 1989) and we have chosen, for robustness issues, to consider model maneuvers as "colored noise". In particular, target acceleration can be considered as a zero-mean first order Markov process (directly if referring to Singer's model (Singer, 1970), or indirectly through the Kalman augmented filter (Hunt and Sanderson, 1982)). We have implemented the latter approach, with a constant acceleration state model, the equations of which are given by:

$$\begin{cases} \left(\frac{\partial \underline{e}}{\partial t}\right)_{(k+1)} = \left(\frac{\partial \underline{e}}{\partial t}\right)_{(k)} + \Delta t \left(\frac{\dot{\partial \underline{e}}}{\partial t}\right)_{(k)} + \underline{\nu}_{(k)}, \\ \underline{\nu}_{(k+1)} = \rho\, \underline{\nu}_{(k)} + \underline{v_1}_{(k)} , \\ \left(\frac{\dot{\partial \underline{e}}}{\partial t}\right)_{(k+1)} = \left(\frac{\dot{\partial \underline{e}}}{\partial t}\right)_{(k)} + \underline{v_2}_{(k)} , \end{cases} \tag{12}$$

where ρ is the degree of correlation between successive accelerations and can range from 0 to 1 (0.3 in the experiments described below), $\underline{v_1}$ and $\underline{v_2}$ are the zero-mean gaussian white noises on the chosen model. Furthermore, the relation involved in the Kalman filter relating the observed data to the chosen model is given by:

$$\left(\widehat{\frac{\partial \underline{e}}{\partial t}}\right)_{(k)} = \left(\frac{\partial \underline{e}}{\partial t}\right)_{(k)} + \underline{w}_{(k)} , \tag{13}$$

where $\widehat{\frac{\partial \underline{e}}{\partial t}}$ is the estimated value of the target motion obtained with (8), and where $\underline{w}$ is the zero-mean gaussian white noise on the observations.

Finally, let us note that the control law given by (6) is insufficient to compensate for possible tracking errors due to non-zero target accelerations. To overcome this problem, the prediction of the target motion, provided by the Kalman filter, can be used; this leads to the following relation defining our complete adaptive predictive control law:

$$T_{c(k)} = -\hat{\lambda}_{(k)}\, \underline{e}_{(k)} - \left(\widehat{\frac{\partial \underline{e}}{\partial t}}\right)_{(k+1)/(k)} . \tag{14}$$

4 EXPERIMENTAL RESULTS

The chosen task for validating the proposed method is a classical one and consists in the gaze control of a camera in order to pursue a moving target (see Vieville and Faugeras (1991) for the description of a reactive vision system used for stereo gaze control and more complex sensing behaviors). More precisely, the visual data used in the vision-based task are the coordinates of the center of gravity (CG) of the projection of the target in the image: $\underline{s} = (X_c, Y_c)^T$; the desired image position is such that the object lies on the optical axis of the camera: $\underline{s}^* = (0, 0)^T$ (approximatively, in the middle of the image), and the two controlled d.o.f. are the camera pan and tilt. The corresponding interaction matrix $L_{\underline{s}}^T$ takes the form (see Santos and Chaumette (1992)):

$$L_{\underline{s}}^T = \begin{pmatrix} X_c Y_c & -(1 + X_c^2) \\ (1 + Y_c^2) & -X_c Y_c \end{pmatrix} . \tag{15}$$

In this case, the model of the interaction matrix $\widehat{L_{\underline{s}}^T}$ involved in (3) can be chosen equal to the real $L_{\underline{s}}^T$, so the corresponding task function is defined by:

$$\underline{e} = L_{\underline{s}}^{T^{-1}} \begin{pmatrix} X_c \\ Y_c \end{pmatrix} = \begin{pmatrix} Y_c/(1+X_c^2+Y_c^2) \\ -X_c/(1+X_c^2+Y_c^2) \end{pmatrix} . \quad (16)$$

Experiments have been split into two cases: in the first one, the non-constrained camera d.o.f. are used to simulate target motions (at constant acceleration with abrupt changes in speed or acceleration). In the second one, combined with the previous motions, the target lying on a record player (see Fig. 1) has a sinusoidal projected motion.

Due to the simplicity of the considered scene and the prediction of the next target position, let us note that it was possible to realize the presented results at video rate (50Hz).

4.1 Simple target motions (see Fig. 2)

In this case, the camera first performs a translation along the optical axis in the forward and then backward (Fig. 2.f, iteration 0 to 500), followed by rotations around the optical axis in the two different senses (Fig. 2.e, it. 500 to 1000), and then simultaneous translations parallel to the image plane (Fig. 2.f, it. 1000 to 2000). At the beginning of the experiment, the camera was correctly positioned with respect to the object. That is why there is no initial error on Fig. 2.a where the time variation of $(\underline{s} - \underline{s^*})$ (that we want to always vanish) is plotted.

As expected, there is no measured motion (Fig. 2.c, it. 0 to 1000) and no task function error when the camera is translating along or rotating around the optical axis. Indeed, these motions does not modify the position in the image of the CG of the target. When translations parallel to the image plane are performed, the estimation of $\widehat{\frac{\partial \underline{e}}{\partial t}}$ is done well, and correctly filtered (see Fig. 2.c and 2.d). Furthermore, the task function is regulated to zero after few iterations due to the reconfiguration of the Kalman filter parameters. This reconfiguration time is due to the fact that, when a jump in the target motion occurs, an update of the effective on-line acceleration (to be estimated) is not immediately taken into account.

It is worth noting that the prediction of the position of the visual features, which are generally within ± 2 pixels (see Fig. 2.b where the discrepancy between the measured and predicted values are plotted), is sufficiently accurate to reduce significantly the searching area in the image, thus reducing the required time for image processing. Besides, the reconfiguration period of the Kalman filtering parameters does not affect it, due to the fact that this prediction only works with on-line measurements.

4.2 Complex target motions (see Fig. 3)

In this experiment, the target is moving with a sinusoidal motion due to rotations of the record player (20 tpm) and complex translational motions in different directions are simultaneously performed on the camera to simulate extra target motions. Since the target is moving before the beginning of the gaze control, the initial target position in the image is far away from the desired one (about 200 pixels in the presented example).

The obtained results are satisfactory, even if not perfect. Firstly, the convergence phase, similar to a saccade (iteration 0 to 40, i.e. 800 ms) is perfectly achieved in spite of the simultaneous target motion. After the convergence, the task function error, which is the best indicator of the achievement of the target tracking, remains within ± 10 pixels (see Fig. 3.a) in spite of the complex target motion in the image (see Fig. 3.c where the time variation of its filtered value is plotted). The observed residual errors are due to the fact that the target motion follows a non-linear model, especially due to the sinusoidal motion. The constant acceleration state model chosen for the Kalman filter is thus insufficient in that particularly bad case to perfectly compensate for target motion with continuous changes in accelerations. The stability of the system is however preserved and the prediction of the position of the target in the image again remains within ± 2 pixels (see Fig. 3.b).

Improving these results would consist in identifying a non-linear model, thus leading to an extended Kalman filter based on the estimation of the pulsation and the amplitude of the sinusoidal motion. Such a method should bring better results for the presented case, but can be expected to be more time-consuming and less general than the proposed one.

Let us finally note that more detailed results are presented in Santos and Chaumette (1992).

5 CONCLUSION

We have presented in this paper a visual servoing scheme using the task function approach. This scheme specifies the task problem in terms of a regulation in the image using 2D visual data. We have proposed a new adaptive and predictive con-

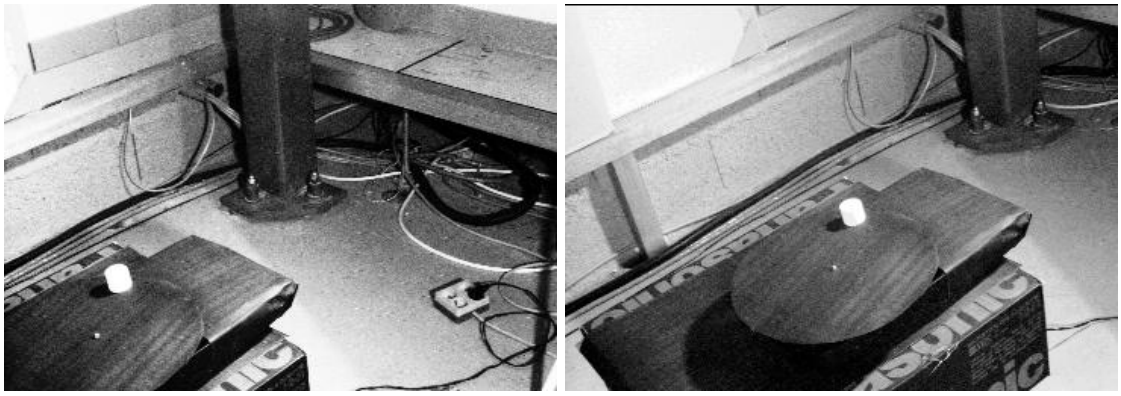
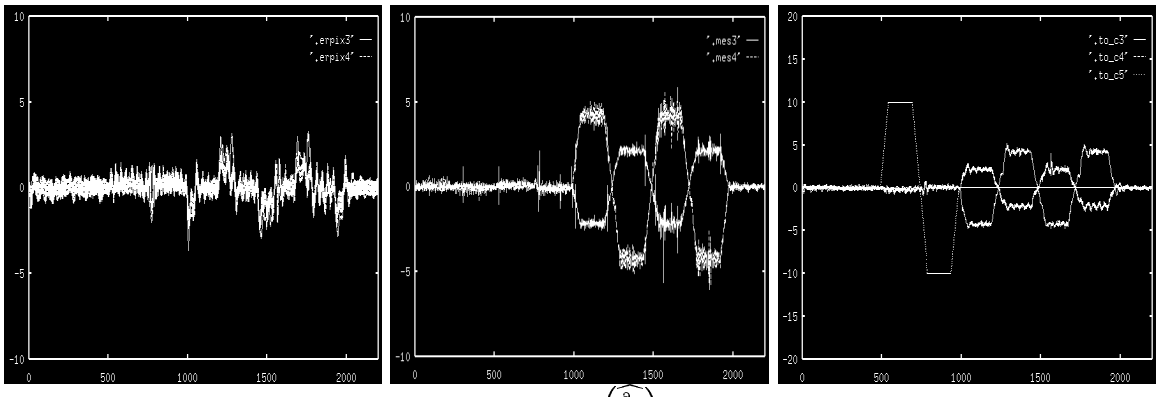
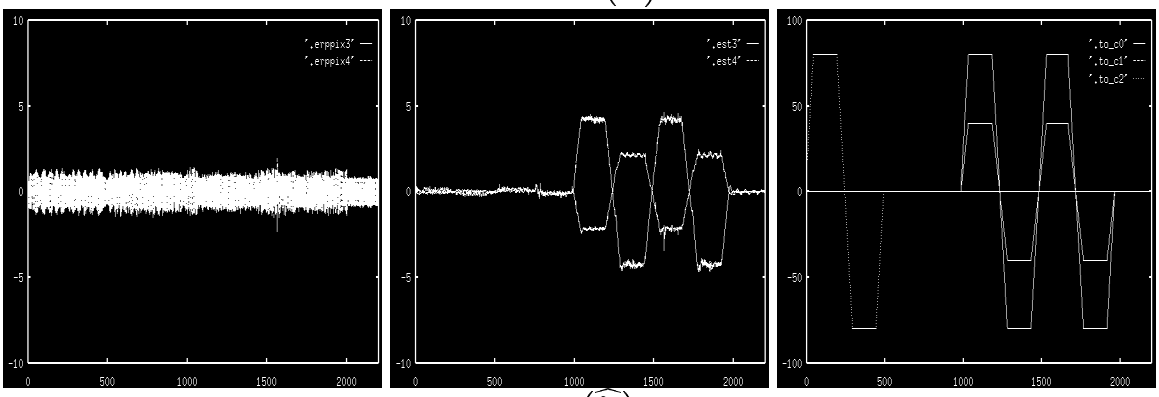


Fig 1: Initial and final images (512 × 730 pixels)

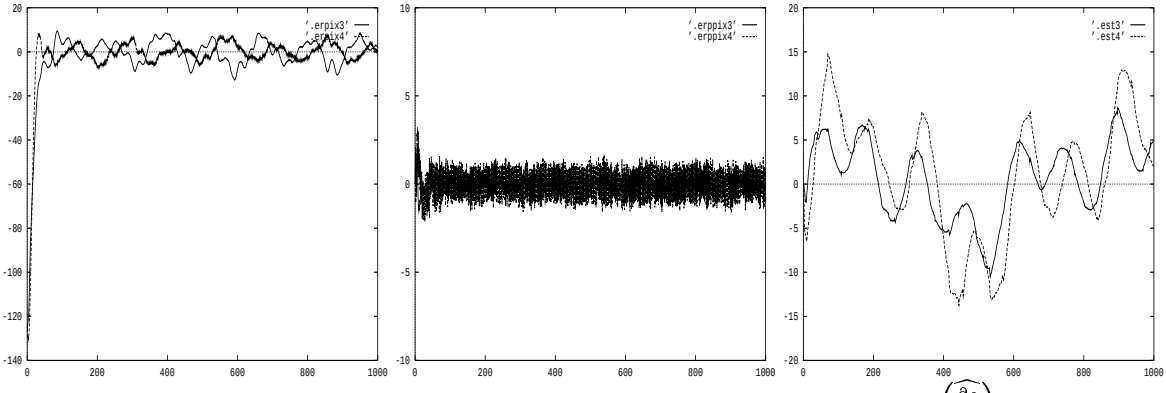


(a) $\underline{s} - \underline{s}^*$ (pixels) (c) estimated $\left(\widehat{\frac{\partial \underline{e}}{\partial t}}\right)$ (deg/sec) (e) Ω_c (deg/sec)



(b) $\underline{s}_{pred.} - \underline{s}_{meas.}$ (pixels) (d) filtered $\left(\widehat{\frac{\partial \underline{e}}{\partial t}}\right)$ (deg/sec) (f) V_c (mm/sec)

Fig 2: Tracking with simple target motions



(a) $\underline{s} - \underline{s}^*$ (pixels) (b) $\underline{s}_{pred.} - \underline{s}_{meas.}$ (pixels) (c) filtered $\left(\widehat{\frac{\partial \underline{e}}{\partial t}}\right)$ (deg/sec)

Fig 3: Tracking with complex target motions

trol law based on this approach which enables to track an object with unknown motion. For doing that, a robust estimation and prediction scheme of the target motion in the image has been presented and introduced in the control law.

Experimental results outline the fact that the estimation through Kalman filtering based on a constant acceleration state model with colored noise is sufficiently efficient to track a highly maneuvering target, even in the presence of noise or mismeasurement.

Nevertheless, when target abruptly maneuvers, a reconfiguration of the filter parameters actually necessitates a few iterations of the control law. The probabilistic jump in the parameter detection approach (Basseville and Benveniste, 1983) may improve the behavior of the estimator, mainly through Hinkley's Cumulative Sum for the jump detection and Willsky's Generalized Likelihood Ratio algorithm (Willsky *et al.*, 1982) for the estimation of the jump magnitude. Indeed, this approach facilitates the detection of changes in the nature of the target motion (for example, a constant velocity followed by a constant acceleration) and the automatic selection of the new adequate filter parameters. Thus it should be possible to pursue a hypothesis tree of parallel state models, based on both constant speed and constant accelerations models (and more complicated ones, if necessary), in order to combine their respective advantages and select the one which is the more appropriate for the current target motion. That will be the future step of our study.

Aknowledgements: This work has been partially realized under Esprit Project P5390/RTGC (Real Time Gaze Control) in collaboration with INRIA Sophia Antipolis, the University of Oxford, GEC and SAGEM.

REFERENCES

Basseville, M. and A. Benveniste (1983). Design and Comparative Study of Some Sequential Jump Detection Algorithms for Digital Signals. *IEEE Trans. on Acoustics, Speech, and Signal Processing,* Vol. 31, n. 3.

Blackman, S. (1986). Multiple Target Tracking with Radar Application. *Artech House Inc.*.

Brown, C., H. Durrant-Whyte, J. Leonard, B. Rao, and B. Steer (1989). Centralized and Decentralized Kalman Filter Techniques for Tracking, Navigation, and Control. *University of Rochester Computer Science Technical Report,* **277**.

Chaumette, F., P. Rives and B. Espiau (1991). Positioning of a Robot with respect to an Object, Tracking it and Estimating its Velocity by Visual Servoing. *IEEE Int. Conf. on Robotics and Automation*, Vol. 3, 2248-2253.

Espiau, B., F. Chaumette and P. Rives (1992). A New Approach to Visual Servoing in Robotics. *IEEE Trans. on Robotics and Automation*, Vol. 8, n. 3, 313-326.

Feddema, J.T., and G.C. Lee (1990). Adaptive Image Feature Prediction and Control for Visual Tracking with a Hand-Eye Coordinated Camera. *IEEE Trans. on Systems, Man, and Cybernetics,* Vol. 20, n. 5.

Hunt, A.E., and A.C. Sanderson (1982). Vision-Based Predictive Robotic Tracking of a Moving Object. *Carnegie-Mellon University Technical Report*, **82.15**.

Papanikolopoulos, N., P.K. Khosla and T. Kanade (1991) Vision and Control Techniques for Robotic Visual Tracking. *IEEE Int. Conf. on Robotics and Automation*, Vol. 1, 857-864.

Samson, C., B. Espiau and M. Le Borgne (1990). Robot Control: the Task Function Approach. *Oxford University Press.*

Santos, A., and F. Chaumette (1992). Target Tracking by Visual Servoing. *IRISA Research Report,* **683**.

Singer, R.A. (1970). Estimating Optimal Tracking Filter Performance for Manned Maneuvering Targets. *IEEE Trans. on Aerospace and Electronic Systems*, Vol. 6, n. 4.

Vieville, T., and O. Faugeras (1991). Real-Time Gaze Control : Architecture for Sensing Behaviors. *Proc.* 5^{th} *Workshop on Computational Vision*, Rosenön, Sweden.

Weiss, L. E. and A. C. Sanderson (1987). Dynamic Sensor-Based Control of Robots with Visual Feedback. *IEEE Journal of Robotics and Automation*, Vol. 3, n. 5, 404-417.

Willsky, A.S., J. Korn and S.W. Gully (1982). Application of the Generalized Likelihood Ratio Algorithm to Maneuver Detection and Estimation. *ACC*, 792-798.