Comparison of Kinematic and Visual Servoing for Fixation

James L. Crowley, Mouafak Mesrabi
LIFIA (IMAG), 46 Ave Felix Viallet, 38031 Grenoble, France

Francois Chaumette
IRISA, INRIA Rennes, Campus Beaulieu, 35042 Rennes Cedex, France

Abstract[1]

This paper describes experiments in estimation and control of the fixation point with a 10 degree of freedom head-neck-body system. A system has been constructed which uses standard kinematic techniques to estimate the fixation point. Two approaches have been investigated to control the fixation point: 3D kinematic estimation and 2D visual servoing. Kinematics is shown to be efficient provided that there is a sufficiently precise model of the kinematic chain. In particular, large errors in the kinematic model can cause the system to oscillate. In contrast, visual servoing is extremely robust with respect to errors in the kinematic model, but require a much larger (factor of 10) number of cycles to converge. The two approaches are found to be complementary. Kinematics can be used to perform a ballistic saccade to the fixation point, while visual servoing serves to correct small errors with a "micro-saccade". This approach mimics saccadic control found in the human visual system.

1. Introduction

In the human vision system, saccadic movements are ballistic [16]. Information from diverse sources are fed into the Superior Colliculus, where they are fused in a retinotopic map [13]. This fusion process results in a single dominant activation center expressed in retinotopic coordinates which then becomes the reference for a rapid saccadic placement of the fovea. The eye is rotated at a speed which is so high that the retinal signal has a motion component which exceeds the spatio-temporal transfer function of the visual system. Thus the movement is necessarily guided by proprioceptive control. Once the saccadic movement has terminated, micro-saccades displace the fovea by small amounts using visual feedback.

Approaching the problem of fixation control from a purely engineering standpoint led us to a similar architecture. In an experimental comparison of saccadic control based on pro-prioception (from motor encoders) and visual servoing [8], we have obtained the following results:

Kinematic control of saccadic movement leads to a system which very rapidly (2 or 3 cycles) converges to a specified fixation. However, such control depends on a reliable model of the kinematics of the head-eye system. The resulting fixation point is often slightly misplaced because of errors in the calibration of the opto-kinematic model (head-eye plus focal length).

Visual servoing is extremely tolerant of errors in the opto-kinematic model. The convergence condition requires only that the terms in the Jacobian be estimated with the correct sign. However, even with a proper opto-kinematic model, this approach can take 20 cycles to converge.

The conclusion is that these two approaches are complementary. Kinematic fixation control permits the system to execute rapid saccadic movements placing the fovea at approximately the desired location in the visual field. Final placement is performed by visual servoing, using as yet, poorly understood measures in the visual field.

2 Kinematic Estimation of the Fixation Point

Estimation of the cartesian location of the endpoint of a sequence of kinematic members with rotational links is a standard problem in robotics. The standard solution to the kinematic estimation problem is to represent the chain of links as a sequence of 4 x 4 homogeneous coordinate transformations. The kinematic model of the robot is the product of the coordinate transformations. The parameters of the model are the parameters (length and rotation axes) used in each of these transformations. This approach can be easily extended to cover kinematic estimation of the fixation point defined by the interesection of the optical axes from a stereo pair of cameras.

The LIFIA head is mounted on a six axis manipulator, as illustrated in figure 1. This manipulator was furnished without a Cartesian controller. A six axis controller was

[1]Funded by the CEC DG III, ESPRIT Basic Research project EP 7108 Vision as Process

0-8186-7108-4/95 $4.00

programmed using classical kinematics. This controller was then extended by the addition of two axes:

Axe 7: The transformation from the end of the arm to head coordinates. The head coordinate system has its origin at the mid-point between the camera rotation centers (figure 2). The X axis is co-incident with the stereo base line. The Z axis is vertical and aligned with the rotational axes of the cameras. The Y axis corresponds to a cyclopean axis and is used to define the plane of the optical axes of the cameras. We call this plane the ocular plane.

Axe 8: The fixation point is located within the ocular plane at a distance, D_C, and an orientation, α_C, determined by the vergence angles of the cameras.

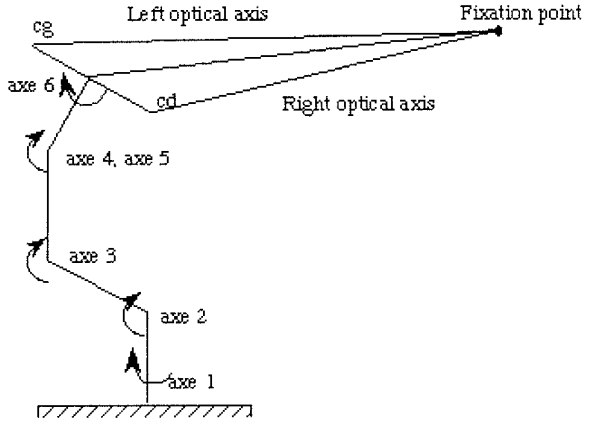

Figure 1. The kinematic model of the arm-head system.

The estimation of the fixation point in the ocular plane has been described in [5]. The vergence angles α_L and α_R are measured from the motor encoders. Given the offset of each camera center along the base line by simple geometry we can derive expressions for the position of the fixation point in the ocular plane.

$$X = B\frac{\mathrm{Sin}(\alpha_l-\alpha_r)}{\mathrm{Sin}(\alpha_l+\alpha_r)} \qquad Y = 2B\frac{\mathrm{Sin}(\alpha_l)\mathrm{Sin}(\alpha_r)}{\mathrm{Sin}(\alpha_l+\alpha_r)}$$

By transforming to polar coordinates, we obtain the length, D_C , and angle, α_C, of the cylopian axis.

$$\alpha_C = \mathrm{Tan}^{-1}\left(\frac{Y}{X}\right) = \mathrm{Tan}^{-1}\left(2\frac{\mathrm{Sin}(\alpha_l)\mathrm{Sin}(\alpha_r)}{\mathrm{Sin}(\alpha_l-\alpha_r)}\right)$$

$$D_C = \sqrt{X^2 + Y^2}$$

These parameters, together with the kinematic model of our robot, define the transformation the kinematic transformation from the joint angles to the fixation.

There are a number of sources of imprecision of such a kinematic model. These include

Retinal Mounting: on many cameras, the solid state retina is mounted on a computer board which is attached to the camera body by springs. This means that the retina can rotate by very small amounts with respect to the lens. The result is a change in the transformation from scene to image.

Camera Optics: We have demonstrated dilation effects due to changes in the focus and aperture settings of the lenses [5].

Image Acquisition: Electronic noise can perturb the start time of the sampling process, which changes the position of the line of an image with respect to the projection on the retina.

Image processing: The filter operations necessary for image processing can blur the image and provoke a loss of precision.

The mechanical links of our system are rigid. The design specifications are sufficient to specify the parameters of the kinematic model. Unfortunately, this is not true for the optical relationship between the retina and the lens. This relationship must be calibrated.

Unlike most work in calibrating cameras, we do not need to deduce an image origin from the projective transformation. It is sufficient to designate a pixel in each image somewhat near the image center, such that the ray emanating from this pixel is in aligned with the Y axis. These pixels are obtained by observing a horizontal edge when the head is placed so that its X-Y axes are horizontal. They are then used as the origins of the left and right images. We calibrate the encoder count which corresponds to the rays being parallel to the Y axis, by looking at a pair of parallel bars whose separation is the same as the cameras. The cameras are verged until these bars fall exactly on the chosen pixels and the encoder counts are noted to be used as the zero offset.

In our binocular head we have mounted the rotational center of vergence under the approximate location of the projection (or stenope) point. This alignment can be tuned by hand by a knob which moves the cameras along the optical axis. When this position is correct there is no translational component to a vergence rotation. A rotation in vergence is modelled by a linear shift in the image. This shift is measured by simply turning the camera vergence angle while watching a point in the world. These experiments permitted the determination once and for all of a ratio of pixels to degrees for vergence. This ratio permits fixation to be servoed by image measurements of disparity. The pan and the tilt of the head correspond to the axes 5 (tilt) and 6 (pan) of the arm. These axes are not coincident with the camera. None the less, we have verified experimentally that the translational components are so small as to be negligible.

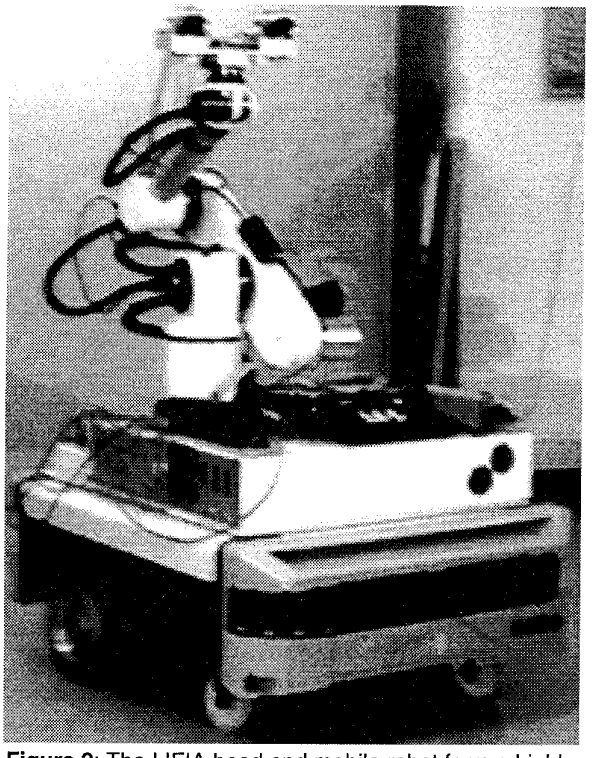

Figure 2: The LIFIA head and mobile robot form a highly redundant set of axes.

As shown in figure 2, the LIFIA head is mounted on a six axis arm mounted on a mobile base. This provides 10 degrees of freedom with which to move the 3 degree of freedom fixation point. The control of such a redundant system is an active research area in robotics control.

We have experimentally compared two approaches for control of this redundant system. The first solution used standard inverse kinematics. The axes were organised in a hierarchy so that a non-redundant subset could be selected for a specified fixation point. This corresponds to a saccadic movement controlled by pro-prioception. The second approach was based on the use of a hierarchy of control loops and visual servoing.

4 Kinematic control of fixation

In this first method, a module in the vision system specifies a position in the scene where fixation is to be placed. Because the current fixation point is always available from the fixation controller, this displacement is usually specified incrementally. The fixation controller then selects a non-redundant subset of the axis to use to obtain the specified fixation. The final joint angles are computed, and the joints are simultaneously servoed to the required angles. Because this method does not use vision to guide the movement, it is often referred to as "open loop". In fact the control loop is closed at the level of motors by encoders. We refer to this as kinematic or proprioceptive control.

In order to obtain a non-redundant set of axes the axes are organised in a hierarchy of subsets as follows.

1) $\{\alpha_L, \alpha_R\}$.
2) $\{\alpha_L, \alpha_R, q_6\}$.
3) $\{\alpha_L, \alpha_R, q_6, q_4\}$.
4) $\{\alpha_L, \alpha_R, q_6, q_5, q_4, q_3, q_2, q_1\}$.
5) $\{\alpha_L, \alpha_R, q_6, q_5, q_4, q_3, q_2, q_1$, rotational axes of the mobile base$\}$

When ever possible, the specified configuration is obtained with the lowest level set of axes in the hierarchy. When a configuration would drive an axis to a limit, the next subset in the hierarchy is used, and lower level axes are placed in the center of their range. This control scheme has, in practice, proved to be both reliable and efficient. It has been placed in a C library and is regularly used by students who do not need to know the particular of how it is programmed. The user is provided with a function called "fixate" with parameters (x, y, z).

5 Some experimental results with kinematic control

Figure 3 shows the result of a purely kinematic command of fixation. The x and y positions of the reference point are shown for both the left and right images as a function of time. Because of slowness of the digitizer used in these experiments, the system was driven with very slow speeds. What is relevant is to see that the system converged after three cycles. These appear as three straight line trajectories in pixels as a function of time. Each cycle is commanded by making a measurement in the image.

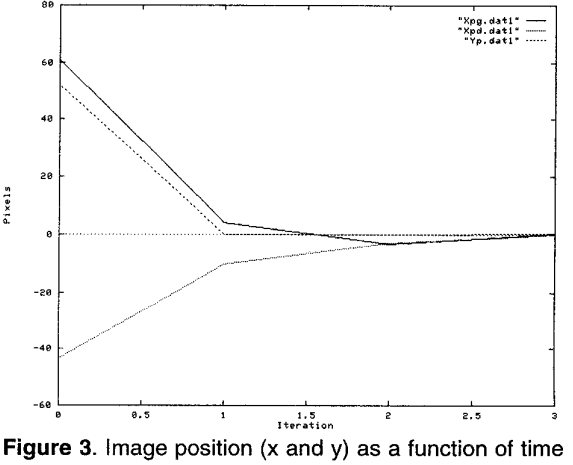

Figure 3. Image position (x and y) as a function of time for reference point in the left and right cameras for a purely kinematic command.

Figure 4 shows the vergence angles of the left and

right camera as a function of time for the same example. The angles where sufficiently similar as to be overlayed on the graph. To test the robustness of the system to errors in the kinematic model, we repeated the same experiment with the focal length of the cameras over estimated by 80%. This corresponds to specifying a gain which is larger than unity. Not surprisingly the system over corrects and exhibits a serious oscillation, as shown in figure 5. Figure 6 shows the camera vergence angles as a function of time for the same experiment. The model was also tested in the case where the focal length was under-estimated by 80%. This case corresponds to setting a very small gain. Figure 7 shows the position of the reference point (x and y) in the left and right cameras for this case. Figure 8 shows vergence angle as a function of time.

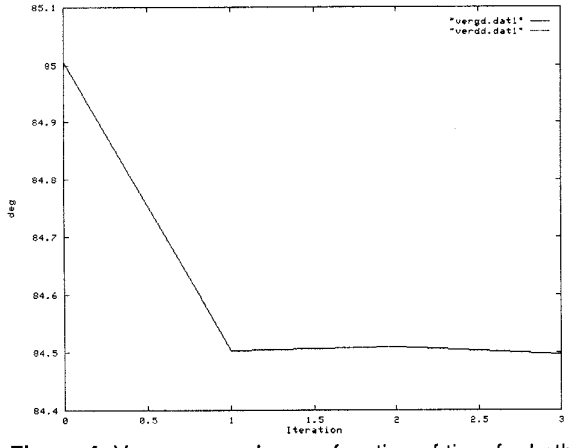

Figure 4. Vergence angle as a function of time for both the left and right cameras.

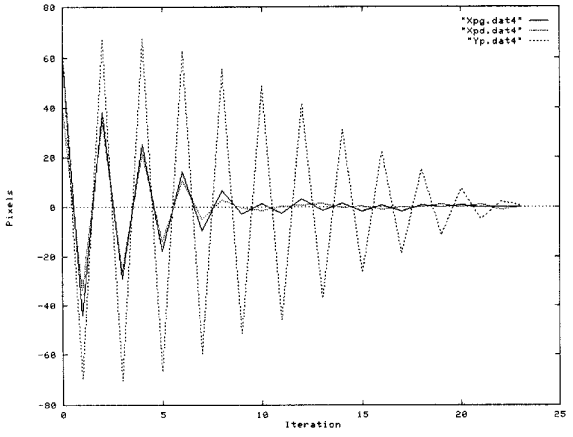

Figure 5 Position of the reference point when the focal length of the two cameras is over estimated by +80%.

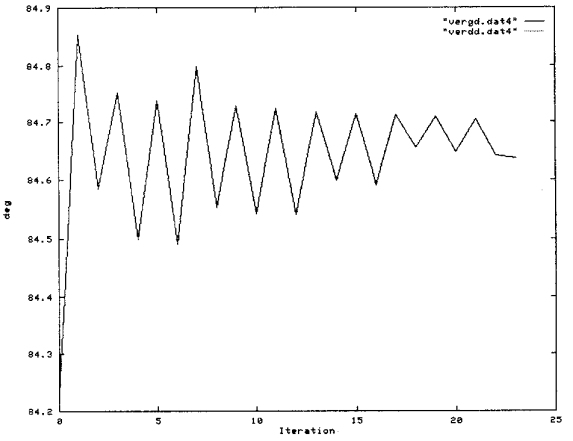

Figure 6 Camera vergence angle for focal length over-estimated by +80%

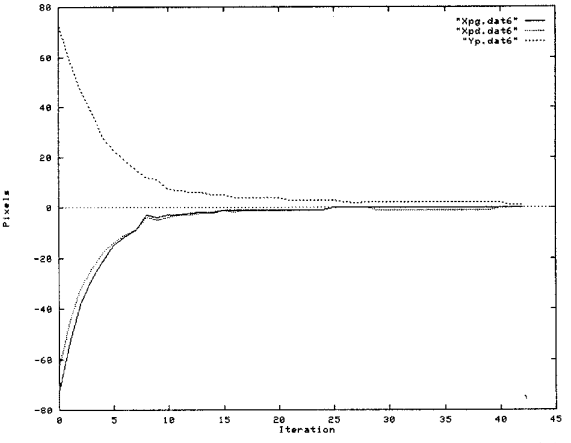

Figure 7. The position of the reference point (x and y) in the left and right camera, when the kinematic error was corrupted by focal length being under-estimated by 80%.

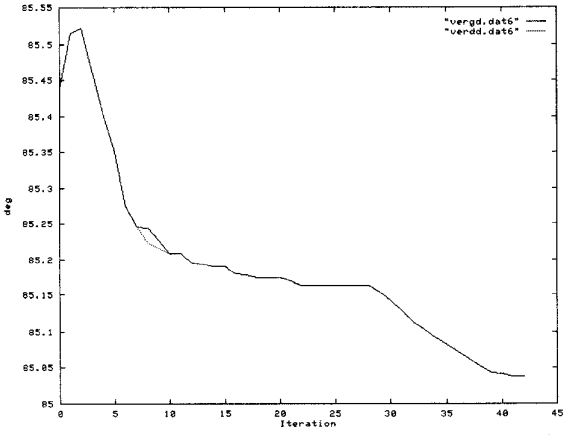

Figure 8. The value of the the camera vergence angle as a function of time when the kinematic error was corrupted by focal length being under-estimated by 80%.

6 Fixation control by visual servoing.

Visual servoing [17] [3] uses measurements made in the image in order to command the axes of a robot. The approach requires estimating a Jacobian transformation from the image to the axes. This Jacobian specifies the movement in the image as a function of the movement of each robot axis. By using this approach with a velocity based control, the system turns out to be remarkably insensitive to the coefficients in the Jacobian. These coefficients acts as a gain term, and the system will converge if the gain terms are simply of the correct sign. Of course, the convergence is fastest for coefficients as precise as possible (gain = 1) and is smooth if the coefficients are not over-estimated (gain $\leq$ 1).

The mechanical redundancy may be handled by defining a hierarchy of servo loops. Some control theorists [7] [14] call a task. We prefer to use the term task in the sense of planning systems: "A goal to be accomplished".

Let us define $\vec{q}$ to be the vector of axis which can be controlled.

$$\vec{q} = [\alpha_L, \alpha_R, q_1, q_2, q_3, q_4, q_5, q_6, w_R, w_L]^T \quad (1)$$

where w_R, w_L are the rotational axes of the mobile base. Let us define s to be a property measured in the image. S is a function of time, t, and of the vector of axes which can be controlled.

$$\vec{s} = s(\vec{q},t) \quad (2)$$

A vision-based error vector, $\vec{S}$, can be defined as the difference between a specified vector of properties in the image and a measured set.

$$\vec{e_1} = \mathbf{C}\,[s(\vec{q}, t) - \vec{s}^*] \quad (3)$$

Where C is the inverse Jacobian of $\vec{e_1}$. This jacobian can be obtained from the classical robot Jacobian J, (such that the camera veiocity T is related to the velocity of each robot axis throught $T = J\,\dot{\vec{q}}$) and from the interaction matrix L^T (which relates the velocity $\dot{\vec{S}}$ in the image to the camera velocity T through $\dot{\vec{S}} = L^T T$). Thus $C = (L^T J)^+$ where the symbol + denotes the pseudo inverse of a matrix.

A servo loop is admissible if it is possible to drive the error to zero, and if the Jacobian is regular and invertable. A mechanically redundant system gives a non-invertable Jacobian and thus a servo loop which is not admissible. To make the servo admissible, it is necessary to add conditions. One possibility is to add an evaluation function based on the cost of a movement. A second solution is to add a number of secondary servo loops which will drive some axes toward "nominal" values. The addition of a secondary servo loop, $\vec{e_2}$, converts the error term into a global vector $\vec{e}$,

$$\vec{e}(\vec{q},t) = \begin{pmatrix} \vec{e_1}(\vec{q},t) \\ \vec{e_2}(\vec{q},t) \end{pmatrix} \quad (4)$$

However $\vec{e_1}$ and $\vec{e_2}$ have to be compatibale in order to be both achieved. A better solution consists in building a hybrid servo loop that permits the system to be controlled by minimizinf $\vec{e_2}$ "as far as possible", under the constraint that the term $\vec{e_1}$ is realized [14].

$$\vec{e}(\vec{q},t) = W^+ \vec{e_1}(\vec{q},t) + (I_n - W^+W)\,\vec{e_2}(\vec{q},t) \quad (5)$$

where W^+ is the pseudo-inverse of W and $(I_n - W^+W)$ is an operator of orthogonal projection (see [7] for details). Using this method, $\vec{e_1}$ is considered as priority, and only the components of $\vec{e_2}$ which are compatible with $\vec{e_1}$ are considered in the global error $\vec{e}$ to be minimized.

More precisely, for our fixation task using a redundant system, the visual features $\vec{s}$ used in the vision based error $\vec{e_1}$ are simply the coordinates of the target position in the images successively acquired by the two cameras. We have also specified several secondary goals in order to solve the redundancy problem [8]. For instance, ew have designed $\vec{e_2}$ to maintain symmetric vergence ($\alpha_L(t) = \alpha_R(t)$) after convergence of $\vec{e_1}$.

A velocity control law which realizes an exponential decrease in $\vec{e}(\vec{q},t)$ without tracking errors takes the form:

$$T_d = \lambda \vec{e} - W^+ \frac{\partial \vec{e_1}}{\partial t} - (I_n - W^+W) \frac{\partial \vec{e_2}}{\partial t} \quad (6)$$

where λ determines the speed of the decay, and $\frac{\partial \vec{e_1}}{\partial t}$ serves to compensate possible target motion.

7 Experiments with control based on velocity mode visual servoing.

As a comparison, the experiment described in section 4 has also been performed for the case of control based on visual servoing. Servoing a reference point to the image center is illustrated in figures 10 (image position of point

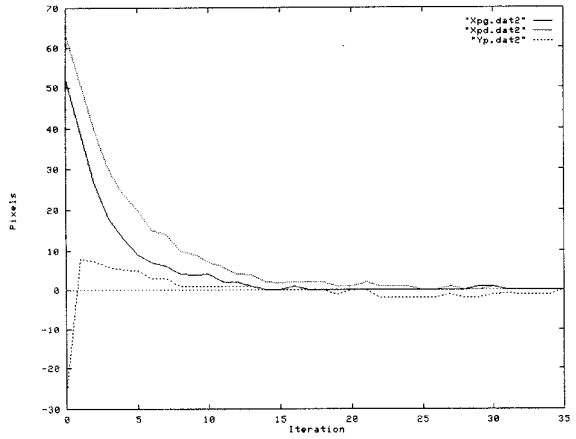

Figure 10. Image position (x and y) as a function of time for reference point in the left and right cameras for visual servoing.

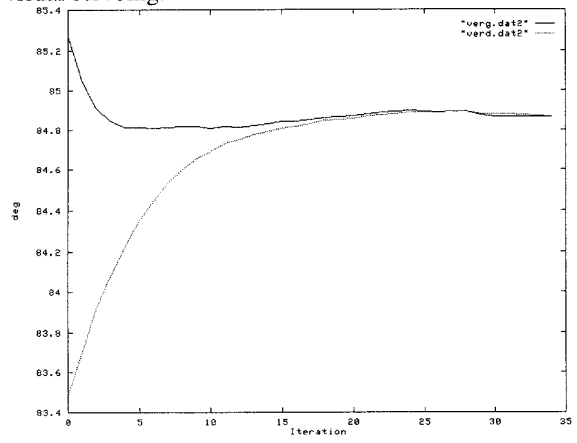

Figure 11. Vergence angle as a function of time for both the left and right cameras during visual servoing.

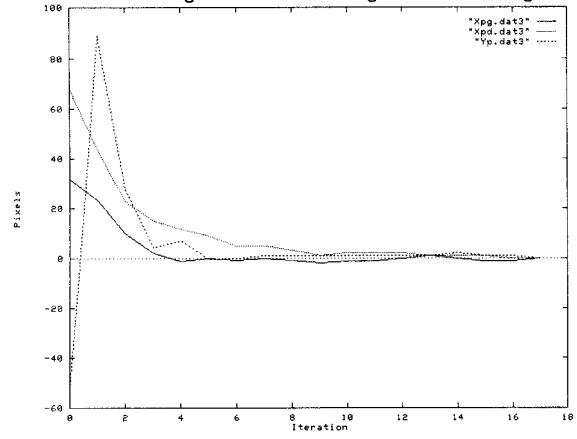

Figure 12 Position of the reference point as a function of time when the focal length of the two cameras is over estimated by +80% for visual servoing of velocity.

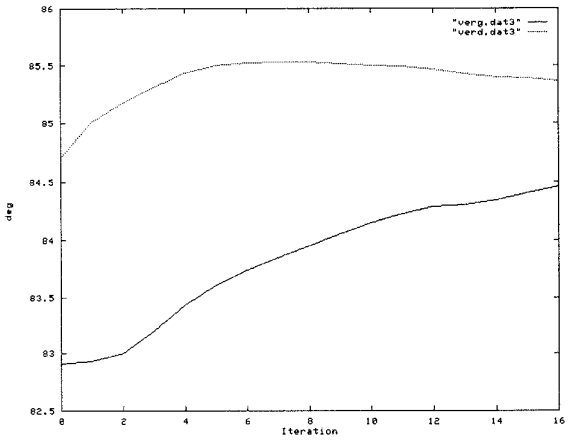

Figure 13. Camera vergence angle for visual servoing of velocity when the focal length is over-estimated by 80%.

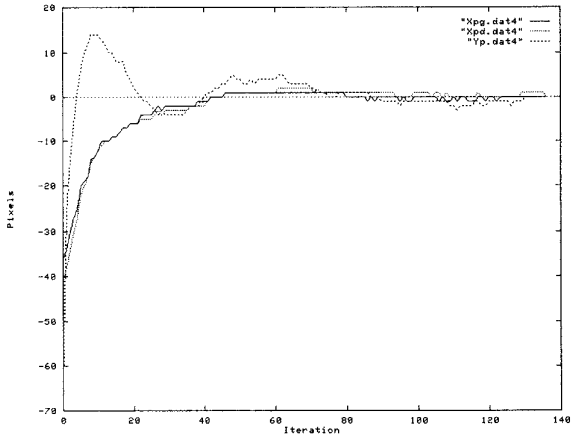

Figure 14. Position of the reference point (x and y) in the left and right camera, when the kinematic error was corrupted by focal length being under-estimated by 80% for visual servoing.

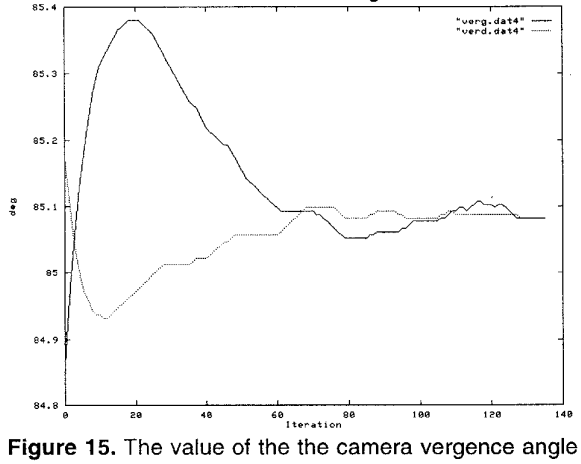

Figure 15. The value of the the camera vergence angle as a function of time when the kinematic error was corrupted by focal length under-estimated by 80%.

in the two cameras) and 11 (camera vergence angles). It can be seen that velocity based visual servoing requires a much larger number of cycles then position based kinematics control.

The kinematic error was then overestimated by 80% producing the curves shown in figure 12 and 13. Despite this error the system smoothly converged the observed points to the reference point. Figure 14 and 15 show the same curve for the case where the focal length was underestimated by 80%. Once again the system smoothly converged.

8 Conclusions concerning the two control methods.

From the experiments presented in sections 5 and 6 it is possible to draw a number of conclusions concerning kinematic control of fixation position and visual servoing of fixation velocity.

Energy: Kinematic command of position consumes considerably less energy than visual servoing of fixation velocity for a single saccadic movement. This is because in position control only a minimum number of axes are moved, while in velocity servoing all of the axis are commanded, and the convergence rate is much slower.

Convergence: Kinematic control of fixation position converges much more rapidly than visual servoing of fixation velocity PROVIDED that an accurate kinematic model is available. Errors in the kinematic model act as gain factors, and can result in oscillatory behaviour and even divergence if such errors drive the system gain to larger than 1.

Occlusion: Although not shown in the experiments for reasons of space, visual servoing of fixation velocity was found to be more robust with respect to occlusions in the scene.

Computational Cost: Visual servoing of fixation velocity was found to have a much higher computational cost.

Robustness to errors in kinematics: As mentioned above, visual servoing of fixation velocity is much more robust with respect to errors in kinematics than kinematic control of fixation position.

The overall conclusion is that these two methods are complementary, with kinematic control of fixation position being most appropriate for large saccadic changes in fixation and visual servoing of fixation velocity appropriate for micro-saccadic movements.

Bibliography

[1] C.M. Brown, "Prediction and cooperation in gaze control", Biological Cybernetics, 63,61-70, 1990.

[2] R.H.S. Carpenter, "Movements of the Eyes", Pion Press, London, 1988

[3] F. Chaumette, "La Relation Vision-Commande : Théorie et Application à des Tâches Robotique", Thèse de l'Université de Rennes I, Juillet 1990.

[4] J. Craig, Introduction to Robotics Mechanics & Control. Addison Wesley,1986.

[5] J. L. Crowley, P. Bobet and M. Mesrabi, "Camera Control for a Active Camera Head", Pattern Recognition and Artificial Intelligence, Vol 7, No. 1, January 1993.

[6] J. L. Crowley and H. I. Christensen, Vision as Process. Springer Verlag Basic Research Series,1994.

[7] B. Espiau, F. Chaumette, and P. Rives. "A New Approach to Visual Servoing in Robotics", IEEE Trans. on Robotics and Automation, 8(3):313-326, June 1992.

[8] M. Mesrabi, "Contrôle du Régard pou un Système de VIsion Active", Doctoral Thesis, I.N.P.Grenoble, March, 1994.

[9] R. Prajoux, "Visual Tracking", chapter 3, 9th Report on Machine Intelligence Research Applied To Industrial Automation, tecnical report, pp. 17-37/145-150. SRI International, Menlo Park (California) August. 1979.

[10] D. A. Robinson, "The Mechanics of Human Saccadic Eye Movements", Journal of Physiology, London, 1964, 174, pp. 245- 264.

[11] D.A. Robinson 1965. "The mechanics of human smooth pursuit eye movement J. Physiol.", 180: 569-591, 1965.

[12] D.A. Robinson, "Why Visuomotor Systems Don't Like Negative Feedback And How They Avoid It", in M.A . Arbib and A.R. Hanson, Eds, Vision, Brain, and Cooperative Computation. Cambridge, MA: MIT. 1988.

[13] A. Roucoux and M. Crommelink, "Control of head movement during visual orientation", in Control of Head Movements, B. W. Peterson and F. J. Richmond eds., Oxford University Press, 1988.

[14] C. Samson, B. Espiau, M. Le Borgne, "Robot Control: the task function approach", Oxford University Press, 1991.

[15] O. J. Smith (1958), Feedback Control Systems. McGraw-Hill, New York

[16] L. Stark, G. Vossius and L. R. Young, "Predictive Control of Eye Movements", M.I.T. Research laboratory for Electronics, Quarterly Progress Report, No. 62, July 61.

[17] L.E. Weiss, A.C. Sanderson, and C.P. Neuman, "Dynamic sensor-based control of robots with visual feedback", IEEE Journal of robotics and automation,Vol. RA-3, No. 5, October 1987, pp. 404-417.

[18] A. Yarbus, "Eye Movements and Vision", Plenum Press, New York, 1967.