

Joint-coupled Compensation Effects in Visually Servoed Tracking

Paul Y. Oh and Peter K. Allen

Dept. of Computer Science, Columbia University, New York, NY 10027 *

Abstract

Humans have degrees-of-freedom (DOF) of varying bandwidths and one casually observes that we coordinate these DOF while visually tracking. This suggests that joint interplay aids tracking performance. In a control scheme we call partitioning, both image and kinematic data are used to visually-serve a 5-DOF robot by defining a joint-coupling among the rotational and translational DOF. Analysis of simulations and experiments reveal that a robot's fast bandwidth joints physically serve as lead compensators when coupled to slower joints thus reducing tracking lag.

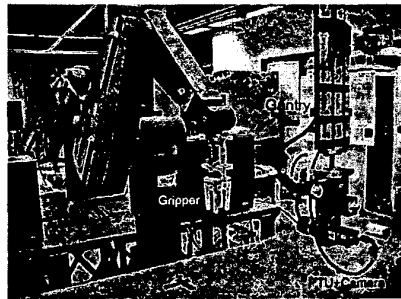


Figure 1: PTU and camera mounted on gantry end-effector tracks targets like grippers in workcell

1 Introduction

Compensators are often added to the control law to makeup for performance deficiencies. For example, a common objective in eye-in-hand tracking is to robotically servo the camera such that its image plane remains centered over the moving target. But, joint bandwidth and saturation limits not permitting, the camera will lag behind the target. One can then mathematically design feedforward [4] or lead compensators to add phase and cancel out camera lag. *Skeet shooting* is a good example of lead compensation. Often a marksman will aim ahead of the target, taking its flight path, speed and distance into account. He thereby physically adds phase to the target's bearing angle.

People display interesting behaviors which also suggest that we employ compensation. For example the eyes lead (i.e. move before) the neck, when we track a moving target. One possible explanation is that the eyes, which have fast response times, are physically compensating for the neck's slower motions. Beyond the mathematical abstractions of compensator design, joints with fast bandwidths can physically act as lead compensators when their kinematic encoder data is

added to the visual-servoing loop. This will be illustrated in this paper.

Our particular vision interests are in visually monitoring a large $3.6 \times 6.4 \times 1 m^3$ assembly workcell with a 5-dof robot (Fig. 1). This robot is a custom-built 3-dof Cartesian gantry mounted with a Directed Perceptions 2-dof pan-tilt-unit (PTU) and a camera on its end-effector. For our initial tracking efforts, we designed a pose regulator [6], [9], [2]. The control law in such designs rely exclusively on image data to update camera motions using an image Jacobian to maintain a desired camera-to-target pose. This initial effort met with several limitations associated with joint bandwidth and saturation. Target translations forced gantry motions which are marked by large inertial loads and slow responses. Oftentimes the target would accelerate past the camera's field-of-view (FOV) before the gantry got up to speed thus losing visual contact. By contrast, the small, lightweight PTU responds quickly and easily. Another problem was that abrupt target starts and stops forced large gantry accelerations that generated camera end-point vibrations. Kalman filtering [10], [1] was implemented to improve robustness but it requires *a priori* knowledge of the target's motion dynamics, which in general, isn't always available.

To overcome these limitations we designed a control

*This work was supported in part by an ONR/DARPA MURI award ONR N00014-95-1-0601 and NSF grants CDA-96-25374 and EIA-97-29844.

law we call *partitioning* that couples joints [7], [8]. The control law uses both image and kinematic joint data. The pan and tilt degrees-of-freedom (DOF) are visually servoed using image data while gantry translations are kinematically servoed using pan and tilt angle data. In this paper we present an analysis of partitioning to illustrate that joint coupling improves visually servoed tracking. We found that fast bandwidth joints, when visually servoed, physically act as lead compensators. Section 2 presents the partitioning control architecture and linearized input-output relationships. Simulations and experiments in the frequency domain reveal phase characteristics and are given in Section 3. Conclusions follow in Section 4.

2 Coupling DOF

Tracking can be defined as the task of centering the camera's image plane over a moving target. For a horizontally translating target, tracking requires only visually servoing one DOF; the camera can be panned or translated. In either case, tracking is achieved by controlling camera velocity in proportion to pixel position changes of the target's image centroid. Tracking in this manner is known as *piloting* [3] or *steering* [4].

Tracking by camera pan requires mapping pixel positions in the image space to target bearing angles in the task space. The radial distance between the camera and target, as measured along the optical axis, factors into this mapping. This distance cannot be measured using a single image centroid, but it can be approximated if the bearing angle is small and the initial camera-to-target distance is known. The net effect is lower pan tracking performance as the target horizontally translates away from the camera. Tracking by camera translation is much simpler. The camera-to-target distance remains constant when the target horizontally translates in a plane parallel to the camera's image plane and simple perspective is used for the image-to-task space mapping. Camera translations however invoke a gantry DOF which is marked by large inertias. In other words, the gantry's long time response limits tracking performance to slow moving targets.

Another way to track is to invoke both pan and translational DOF and design a control law that exploits the advantages each has to offer. The PTU has a short time response and can be visually servoed to keep a fast moving target in view, whereas the gantry has the ability to transport the camera anywhere in the workcell. Thus better pan tracking performance can

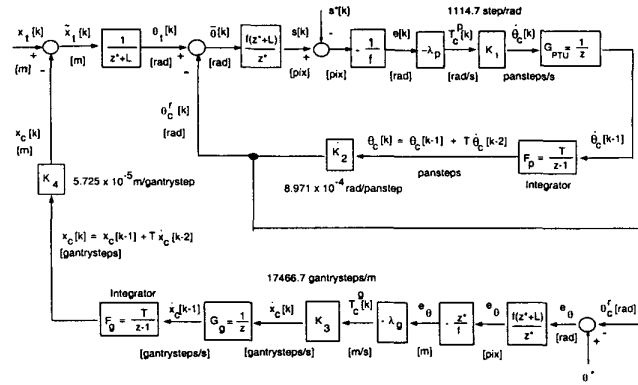


Figure 2: Coupled pan-gantry block diagram.

be achieved if the gantry simultaneously translates to reduce the camera-to-target bearing angle. To realize simultaneous pan and translation, a joint-coupling can be defined in the underlying control law, and is illustrated in the discrete-time block diagram (Fig. 2). Both sample instants k and units are given for additional clarity.

The coupling is achieved with two feedback loops and the target position input, x_t [m], results in both pan angle, θ_c^r [rad], and gantry translation, x_c [m], outputs. Pan velocity $\dot{\theta}_c$, is visually servoed using pixel differences $s[k] - s^*$ and gain λ_p and the gantry translates using camera pan angle differences $\theta_c^r - \theta^*$ and gain λ_g . $s[k] = (u, v)$ is the center-of-gravity pixel position of the target's image centroid with u and v being the horizontal and vertical pixel coordinates respectively. $s^* = (u^*, v^*)$ is a user-defined set point, with $s^* = (0, 0)$ being the pixel position of the camera's image plane center.

To visually servo camera pan, u must be mapped into a camera-to-target bearing angle, θ_t . This mapping is modeled using Fig. 3 and described as follows. If the target is not centered in the camera's image plane, then a non-zero angle difference $\tilde{\theta}$ exists between θ_t and camera angle θ_c . Assuming small angles, $\tilde{\theta}$ can be approximated by

$$\tilde{\theta} = \theta_t - \theta_c \approx \frac{x_t}{z^* + L} - \frac{x_c'}{z^* + L} \quad (1)$$

where z^* and L are the initial camera-to-target distance and camera's radius of rotation respectively. ϕ is the angle the target makes with respect to the lens

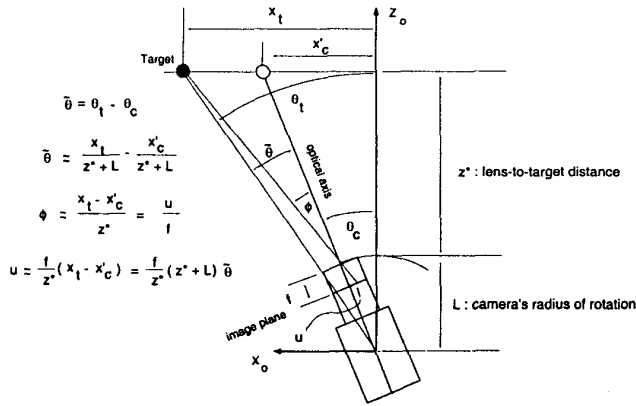


Figure 3: Relevant notation used for tracking by camera pan

center and using perspective

$$\phi \approx \frac{x_t - x_c'}{z^*} \equiv \frac{u}{f} \quad (2)$$

where f is the lens focal length in pixels. Combining these two equations yields

$$u = \frac{f}{z^*} (x_t - x_c') = \frac{f}{z^*} (z^* + L) \tilde{\theta} \quad (3)$$

or the following input-output relationships

$$\frac{x_t}{\theta_t} = \frac{1}{z^* + L} \quad \text{and} \quad \frac{\tilde{\theta}}{u} = \frac{f(z^* + L)}{z^*} \quad (4)$$

These two relationships are represented as block elements before and after the top-left adder in Fig. 2. The $\frac{f(z^*+L)}{z^*}$ term is a camera-to-target distance dependent gain and models a camera's lens. The net effect is that tracking performance by visually servoed pan is limited to small target bearing angles.

The gantry DOF is invoked by kinematically servoing translations in response to camera pan angle differences $\theta_c^r - \theta^*$. θ^* is a user-defined setpoint camera orientation and without loss of generality, θ^* is set to zero. The net effect is that the constant gain, λ_g , actuates camera translations at speeds proportional to its pan angle.

Both the gantry and PTU stepper motors are modeled as unit delays, $1/z$. Forward-rectangular integration, $T/(z-1)$, with sample time T is used to yield the gantry position, $x_c[k]$ and pan angle, $\theta_c[k]$, in steps. Unit conversion constants, K_1 , K_2 , K_3 and K_4 yield

these outputs in meters and radians. Target translation input x_t [m] thus yields the following closed-loop transfer functions:

$${}^cG_\theta(z) = \frac{\theta_c \text{ [rad]}}{x_t \text{ [m]}} = \frac{\lambda_p T z^2 - \lambda_p T z}{\text{CE}} \quad (5)$$

$${}^cG_g(z) = \frac{x_c \text{ [m]}}{x_t \text{ [m]}} = \frac{(z^* + L) T^2 \lambda_g \lambda_p}{\text{CE}} \quad (6)$$

$$\begin{aligned} \text{CE} = & z^* z^4 - 2z^* z^3 + (z^* + (z^* + L)\lambda_p T) z^2 \\ & - (z^* + L)\lambda_p T z + (z^* + L) T^2 \lambda_p \lambda_g \end{aligned}$$

Superscript c in ${}^cG_\theta$ and cG_g denotes "coupled" and subscripts θ and g describe pan and gantry respectively. The net effect of the coupling in Fig. 2 is that tracking behavior (i.e pan and gantry servo action) depend on two gains λ_p and λ_g . This is because both ${}^cG_\theta$ and cG_g have the same characteristic equation (CE) and the closed-loop pole locations depend on the product $\lambda_p \lambda_g$. Tracking behavior can also be seen by the fact that if $\lambda_g = 0$ then there is no camera translation (since ${}^cG_g = 0$) and all tracking is done exclusively by panning. The gain dependence on pan and gantry responses can be seen more clearly in the ensuing root locus diagrams and step responses.

2.1 Root Locus & Time Response

Reexpressing (5) and (6) into root locus forms yields two plots (Figs. 4 and 5). The former plot is of varying λ_p and fixed $\lambda_g = 0.5$ and the latter is of varying λ_g and fixed $\lambda_p = 1.0$. T was set at 0.12 s in light of the PTU's serial latency. New gantry velocity updates are issued at this loop time and its serial latency (6 ms) is comparatively small. These two plots complement each other. Fig. 4 reveals double poles at both $z = 0$ and 1, and a pair of zeros at 0.0641 and 0.936. The loci breaks away at $z = 0.5$ and pan response will be more oscillatory as λ_p increases above 1.958 with instability at $\lambda_p > 6.242$. Fig. 5 reveals yields poles at $z = 0, 1, 0.786$ and 0.214 and no zeros. Gantry overshoot will occur for $\lambda_g > 0.35$ and settling time will increase. As λ_g is increased, oscillations will increase with unstable response occurring for $\lambda_g > 2.548$.

The net effect of kinematic coupling pan and gantry DOF together is graphically shown in Fig. 6. Step response experiments were videotaped and a sequence of image stills are shown in Fig. 7. Fig. 8 is the experimentally obtained (dash) step input responses for θ_c and x_c with $\lambda_p = 1.0, \lambda_g = 0.5$. Simulation results

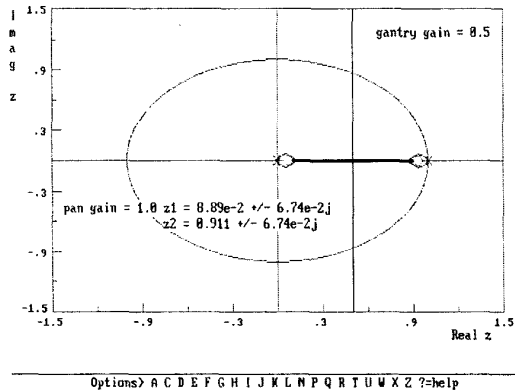


Figure 4: $G_{\lambda_g}^{RL}$ root locus with $\lambda_p = 1.0$

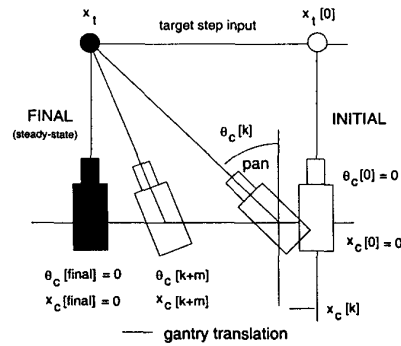


Figure 6: Coupled pan-gantry step response

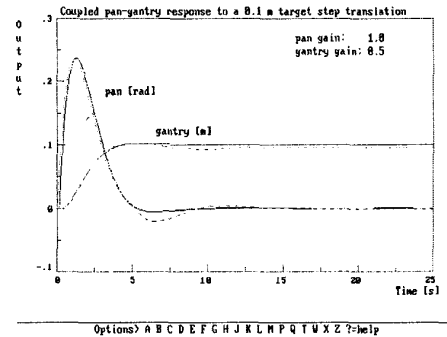


Figure 8: Coupled pan-gantry step response. Experimental (dash) and simulations (solid) shown. $\lambda_p = 1.0$ and $\lambda_g = 0.5$. Pan accelerations result in gantry translations.

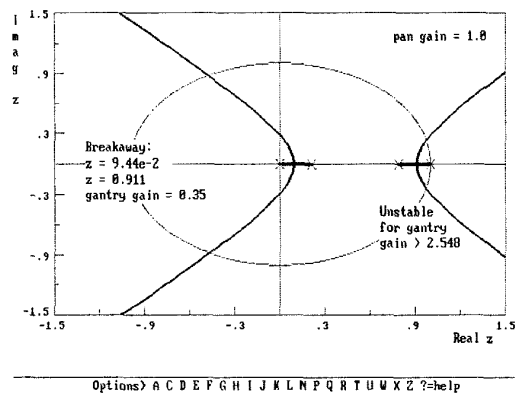


Figure 5: $G_{\lambda_p}^{RL}$ root locus with $\lambda_g = 0.5$

(solid) using equations (5) and (6) match these results closely.

3 Coupling Effects

Rise time is often used as a performance metric and is defined as the time it takes the system to go from 10 to 90 percent of its steady-state value. The rise times of visually servoing a gantry DOF ($\lambda_g = 0.5$) was experimentally measured to be 4.29 s. Visually servoing the pan DOF ($\lambda_p = 1.0$) yielded a 1.13 s rise time. Coupling the pan and gantry DOF at these gains yielded a 2.50 s rise time.

Visually servoing the gantry results in sluggish response because its large inertial load forces using a low gain. Partitioning introduces coupling and takes advantage of the pan's fast rise time to translate the gantry quickly. As mentioned in the introduction, the

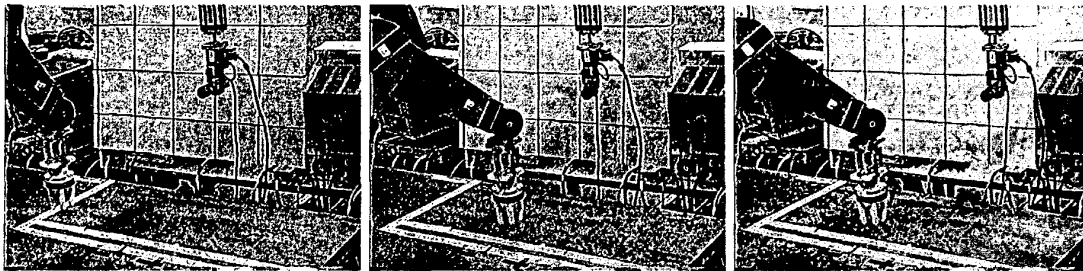


Figure 7: Coupled pan-gantry step response: The target is a robot hand mounted on a Puma end-effector (left). To create the step input, the hand first translates 0.1 m and then the camera is allowed to servo. The camera simultaneously pans and translates (middle) under partitioning to the final camera pose (right).

pan DOF serves as a lead compensator and is analogous to skeet shooting. The pan adds positive phase when aiming at the target. This can be seen analytically from the Bode phase and magnitude plots of (5) and (6) given in Fig. 9.

The pan's Bode plot reveals a resonant peak at 0.189 Hz (12.95 dB). Below this value, there is phase lead with amplified gain. At higher frequencies, gain is attenuated, filtering high-frequency signals. At low frequencies (e.g. 0.02 Hz) there is little pan and larger gantry motion. In other words, most of the tracking action is done by the gantry. At higher frequencies (e.g. 0.80 Hz) the reverse is true with pan motions being more prevalent. In mid-frequencies, both pan and gantry motions contribute towards tracking. The net effect is that the pan introduces phase lead to compensate for the lower gantry bandwidth. Fig. 10 shows this pan phase lead ($+55.88^\circ$ and 6.37 dB gain) for a 0.1 Hz sinusoidal input.

Peak-to-peak pixel error in response to a sinusoidal input is another performance metric. A small pixel error means that the target is centered in the camera's image plane. Fig. 11 illustrates the improved phase and peak-to-peak pixel error that joint-coupling offers. Tracking by visually servoing the gantry alone ($\lambda_g = 1.0$) results in a peak error of 133 pixels (5.3 cm). Tracking by coupling the pan and gantry DOF ($\lambda_g = 1.0$, $\lambda_p = 1.0$) results in a peak error of 73 pixels (2.9 cm). By increasing λ_p , even smaller pixel errors are possible (Fig. 12). For $\lambda_g = 1.0$ and $\lambda_p = 2.0$, the peak error is 33 pixels (1.3 cm).

4 Conclusions

Casual observance reveals that humans coordinate their DOF while tracking and the resulting motions

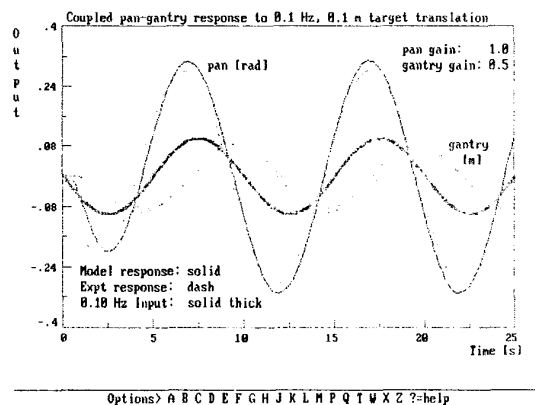


Figure 10: Coupled pan-gantry response to a 0.1 Hz, 0.1 m amplitude sinusoidal target translation x_t (thick solid line). Both experimental (dash) and simulated (solid) responses shown. $\lambda_p = 1.0$ and $\lambda_g = 0.5$.

suggest a joint-coupling. The partitioned scheme defines a joint-coupling in the control law that takes advantage of both image and kinematic data. Simulations and experiments reveal that coupling smaller peak-to-peak pixel errors, and hence improved tracking performance. Like skeet shooting, the partitioned pan DOF adds phase to the gantry's dynamics due to its faster response and physically acts like a lead compensator. This suggests that improved lag performance can be achieved by kinematically coupling a robot's faster joints to its slower joints under a partitioned scheme. The analysis reveals that partitioned tracking performance depends on two gains which can be tuned using standard root locus techniques. We are currently exploring linear quadratic cost functions as an alternative approach for gain tuning. Such an approach would prescribe the amount of pan and gantry control effort and ensure both stability and optimal

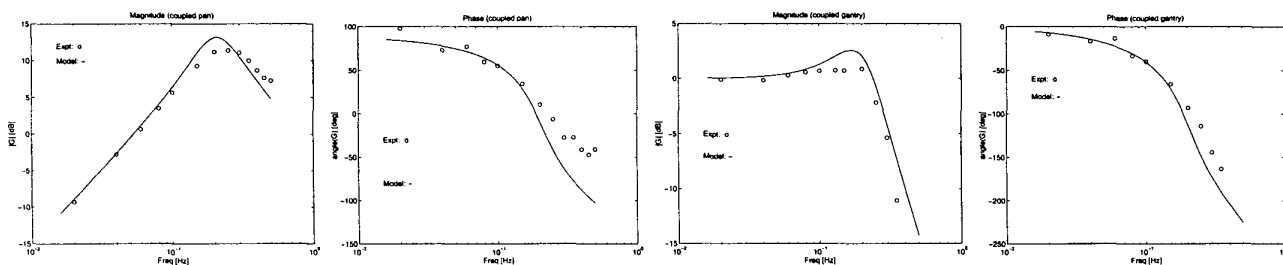


Figure 9: Bode plots of coupled pan-gantry (left to right): pan magnitude and phase, gantry magnitude and phase for $\lambda_p = \lambda_g = 1.0$. Both Matlab-generated (solid) and experimental (dot) results shown. The pan's magnitude peaks at 0.189 Hz with a 12.95 dB gain and 0° phase angle. The gantry 3 dB down occurs at 0.293 Hz and -166° phase angle.

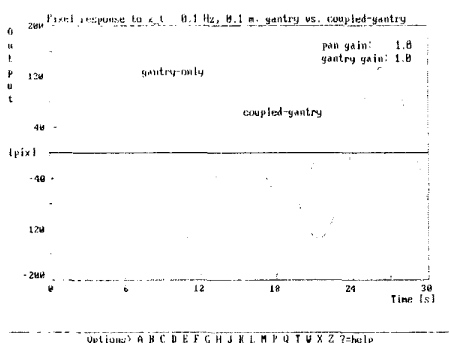


Figure 11: Pixel errors for both gantry-only ($\lambda_g = 1.0$) and coupled pan-gantry ($\lambda_g = \lambda_p = 1.0$) systems to a 0.1 Hz, 0.1 m amplitude target translation x_t . Note how coupling improves phase and has less peak-to-peak pixel error.

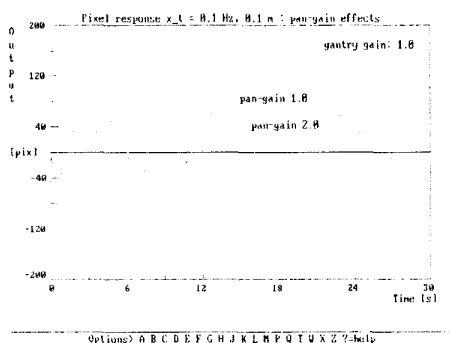


Figure 12: As pan gain λ_p increases, the pixel error decreases but can yield gantry end-point vibrations due to larger accelerations

tracking performance.

References

- [1] Bensalah, F., Chaumette, F., "Compensation of Abrupt Motion Changes in Target Tracking by Visual Servoing", *Proc IROS '95* p 181, Pittsburgh, 8/95.
- [2] Chaumette, F., Rives, P., Espiau, B., "Positioning of a Robot with respect to an Object, Tracking it and Estimating its Velocity by Visual Servoing", *IEEE Int Conf Robotics & Automation*, p 2248, 1991.
- [3] Chaumette, F., Santos, A., "Tracking a Moving Object by Visual Servoing", *Proc. of 12th World Congress IFAC*, Vol 9 pp 409-414, Sydney, July 1993.
- [4] Corke, P. I., "Dynamics of Visual Control", in *IEEE Int Conf Robotics & Automation*, Workshop M-5, San Diego, 5/94.
- [5] Corke, P., "Design, Delay and Performance in Gaze Control: Engineering and Biological Approaches", in

The Confluence of Vision and Control, Kriegman et al, Eds. Springer-Verlag pp. 146-158, 1998.

- [6] Feddema, J.T., Lee, G.C.S., "Weighted Selection of Image Features for Resolved Rate Visual Feedback Control", *IEEE Trans Robotics & Automation*, V7 N1 2/91.
- [7] Oh, P.Y., Allen, P.K., "Design of a Partitioned Visual Feedback Controller", *IEEE Int Conf Robotics & Automation*, Leuven, Belgium, pp 1360 5/98.
- [8] Oh, P.Y., Allen, P.K., "Performance of a Partitioned Visual Feedback Controller", *IEEE Int Conf Robotics & Automation*, Detroit, Michigan 5/99.
- [9] Papanikolopoulos, N.P., Khosla, P.K., "Adaptive Robotic Visual Tracking: Theory and Experiments", *IEEE Trans Automatic Control* V38 N3 3/93.
- [10] Wilson, W.J., "Visual Servo Control of Robots Using Kalman Filter Estimates of Robot Pose Relative to Work-pieces", in *Visual Servoing*, World Scientific, Koichi Hashimoto Editor, 1993