Proceedings of the 1999 IEEE
International Conference on Robotics & Automation
Detroit, Michigan • May 1999

# Performance of a Partitioned Visual Feedback Controller

Paul Y. Oh and Peter K. Allen
Dept. of Computer Science, Columbia University, New York, NY 10027 *

## Abstract

*We present a novel approach we call partitioning where the robot's degrees-of-freedom (DOF) are categorized into two classes based on joint kinematics and dynamics to design a coupled multi-input control system. We use image data to visually servo the first class of joints that have quick response time. Position-based data is used to kinematically servo the second class of joints that have large kinematic range. The net effect is an active-vision system that synergistically tracks a diverse range of targets (without using CAD-based models) over a wide bandwidth of motion dynamics.*

## 1 Introduction

Active vision systems typically use a robot to position and orient a camera. This mobility provides the camera with variable pose, field-of-view and resolution that cannot be achieved with stationary cameras. Furthermore a robot's programmability allows for tasks to be automated by using image-based visual servoing.

Our particular interest in active vision is for the automated monitoring of an assembly workcell. We have custom built a 3 translational DOF cartesian gantry robot. At the gantry's end-effector is a 2-DOF pan-tilt unit (PTU). The net effect of this is a 5-DOF hybrid robot (Figure 1) that can position and orient a camera anywhere in our $3.6 \times 6.4 \times 1\ m^3$ workcell. Our goal is to visually servo the camera to automatically track a moving part, robot-mounted tool or gripper as it moves in the workcell.

By its design, this robot has 2 complementary control systems that can be partitioned among its DOF. The $X, Y, Z$ gantry translational DOF's are marked by large masses, slow responses and low velocities. The gantry motors' velocity bandwidths are limited,

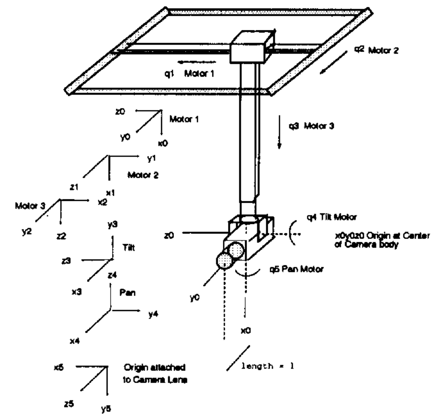


Figure 1: 5-DOF Hybrid Robot

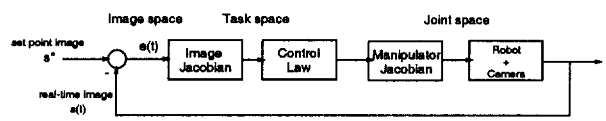


Figure 2: Conventional Tracking Block Diagram

and are much less than those of the PTU's motors. The PTU is a small, lightweight device that only needs to support a camera. It is able to achieve high accelerations and responds faster (and more accurately) than the much larger gantry. The net effect of this is that performance for tracking moving targets with this robot is related to which degrees of freedom are invoked in the tracking task. For example, we can track a target at high velocities using the PTU alone (fixed gantry position); however the range of the PTU pan-tilt is limited, and arbitrary pose configurations of the camera-to-target cannot be satisfied. If we allow all 5-DOF to be used, we then limit our tracking velocities. Assuming we had a single control system for all 5 DOF, the conventional tracking approach would be to design an image Jacobian with a control law as described by Figure 2.

Here, the servo effort maintains a strict camera-to-

*This work was supported in part by an ONR/DARPA MURI award ONR N00014-95-1-0601, DARPA AASERT awards DAAHO4-93-G-0245 and DAAH04-95-1-0492, and NSF grants CDA-96-25374 and IRI-93-11877.

target pose constraint by minimizing the error in the image. Feddema [7] and Papanikolopoulos [11] and Chaumette [3] have used this approach to track blocks (plane projections) and gaskets (circular projections) in real-time.

In our own experiments [10] with this approach, we encountered bandwidth problems while tracking fast moving targets. First, abrupt robot motions as the large gantry moved (stops, starts and turning) led to end-point vibrations which corrupted image data. Kalman-based filters [15], [1] were implemented and improved image robustness somewhat. Mechanically, we added mass to reduce the frequency of vibrations but did not eliminate them completely.

A second (and related) problem was that at fast accelerations the target would leave the camera's field-of-view before the robot would accelerate to speed. In other words, the motors did not have the acceleration bandwidth to position the camera quickly enough. One can use stronger motors but larger accelerations (and consequently larger torques) increase end-point vibrations.

Third, some of our targets we are tracking, such as robotic grippers, are non-rigid, i.e. the finger positions are configurable. Our other targets are also geometrically complex. Capturing their pose would require a larger number of feature points from which to compute and image Jacobian which is both computationally expensive and image noise sensitive.

These three problems forced us to come up with an alternative approach by considering that using a full image Jacobian to maintain a strict camera-to-target pose regulation may not be necessary. Alternatively, we have come up with a method that allows us to partition our control between the 2 different control systems.

## 2 Partitioning Approach

People can visually track fast moving objects and we rarely maintain a fixed pose relative to the target. Instead, we coordinate our visual DOF (eyes, head, torso etc) by keeping the target in field-of-view as it moves. When required (as in inspection) we *localize* the DOF to establish a desired pose when the target is motionless.

Furthermore, our visual DOF move synergistically. For example, our neck tends to rotate in the same direction as our eyes. Our machine vision system

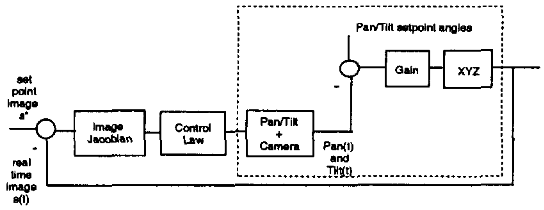


Figure 3: Partitioned Control Block Diagram

mimics this sort of behavior by *coupling* visually and kinematically servoed motions. In other words, we use image data to visually servo certain DOF and position-based encoders to kinematically servo other DOF. This multi-input control approach is shown in Figure 3.

Most robots today employing closed-loop control rely exclusively on position (or derivative) based feedback, such as joint-encoders. On the other hand conventional vision tracking systems rely entirely on image data feedback. Our approach integrates these two methods. We partition our robot's DOF such that vision is only used to servo joints (pan/tilt) with fast response time. We use the PTU joint-encoder positions in a separate feedback loop to kinematically control the gantry translational joints. This allows us to *localize* the camera based upon the net movement of the object being tracked. Figure 3 is a block diagram of this method which produces a synergistic tracking motion.

### 2.1 Coupled Servoing

In a previous work [10] we used a $2 \times 2$ image Jacobian $L^T$ to visually servo the pan and tilt DOF:

$$\begin{bmatrix} \frac{du}{dt} \\ \frac{dv}{dt} \end{bmatrix} = \begin{bmatrix} \frac{uv}{f} & -\frac{f^2+u^2}{f} \\ \frac{f^2+v^2}{f} & -\frac{uv}{f} \end{bmatrix} \begin{bmatrix} \omega_x \\ \omega_y \end{bmatrix} \tag{1}$$

The synthesis of $L^T$ has been well documented in the literature [7], [13], [11], [9].

Since only pan and tilt DOF are visually servoed, the typical $2 \times 6$ image Jacobian (for a single point) is reduced to its $2 \times 2$ form given in (1). $\omega_x$ and $\omega_y$ are the rotational velocities of a point in the camera's task space and are mapped to velocities $du/dt$ and $dv/dt$ in the camera's image space through the image Jacobian and camera focal length $f$.

Our visually-servoed control law follows [4] and uses

$\left[L^T\right]^{-1}$ and a camera task-to-joint space mapping. Using only pan and tilt our camera is able to keep a target centered in its field of view. These motors have large acceleration bandwidths compared to the gantry's motors and generate almost no endpoint vibrations. The net effect is that using this $2 \times 2$ form of $L^T$ we can effectively maintain visual contact over a wide range of target motion dynamics. $\omega_x$ and $\omega y$ in (1) are mapped to the camera's joint space $q_{\text{pan}}$ and $q_{\text{tilt}}$ with a simple coordinate transformation. The pan and tilt motor encoders are then used in a feedback loop to kinematically servo the gantry's DOF with proportional gain $K_g$.

$$\begin{bmatrix} \dot{q}_{\text{horiz}} \\ \dot{q}_{\text{vert}} \end{bmatrix} = K_g \begin{bmatrix} q_{\text{pan}} - q^*_{\text{pan}} \\ q_{\text{tilt}} - q^*_{\text{tilt}} \end{bmatrix} \tag{2}$$

This follows our previously mentioned human-vision analogy with certain DOF responding in the direction of other DOF. In (2), the gantry's DOF translates the camera at a speed $\dot{q}_{\text{horiz}}$ and $\dot{q}_{\text{vert}}$ in the direction of pan and tilt. Here, $q^*_{\text{pan}}$ and $q^*_{\text{tilt}}$ are the desired pan and tilt setpoint angles respectively. In essence (2) prescribes a *coupling* which localizes the camera (via translation) while keeping the target centered in its field of view (via pan and tilt). Video stills of this coupling (Figure 4) illustrate this coupling.

# 3 Experimental Results

Two experiments were conducted to demonstrate the partitioned controller's performance. Tracking was done in real-time, with a K2T framegrabber and Sparc 20 at video update rates (30 frames per second). Image data acquisition was implemented in software with a single SSD region tracker using Hager's XVision tool package [8]. $40 \times 40$ or $80 \times 80$ pixel windows were used.

## 3.1 Experiment 1: People Tracker

The partitioned controller was used to track a person (Figure 5) walking around the assembly workcell. The results highlight several points. First, under our partitioned approach the camera could effectively track the person using a single region-based SSD focused on the head. We were also able to track other geometrically complex targets such as workpieces, tools and robotic hands. No CAD-based models were needed.

People tracking is a good example of where fixed camera-to-target pose may be an overly rigid constraint. Furthermore it would require designing an

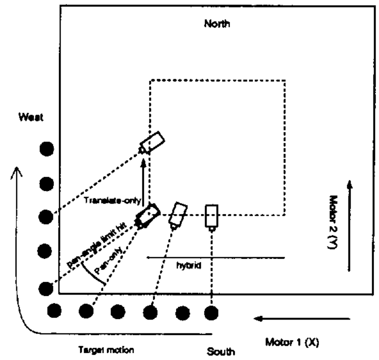


Figure 6: Kinematic Constraints Handling

image Jacobian that captures salient image features that define a head's pose (e.g. eyes and mouth). Head motion dynamics are often non-deterministic (bobbing, jerks and sudden turns). The net effect would make tracking people difficult with traditional regulator-style systems.

The second point is how DOF can be coupled using kinematic data. Figure 6 depicts the case where the camera should switch from one translational DOF (Motor 1) to another (Motor 2) while tracking a cornering target. Since pan angle $q_{\text{pan}}$ is monitored constantly, we can easily determine what quadrant it is in by comparing the size of $|\sin(q_{\text{pan}})|$ versus $|\cos(q_{\text{pan}})|$. The sign of $\sin(q\text{pan})$ determines which direction Motor 1 or 2 should translate and use in (2).

The third point is that pan/tilt servo limits can sometimes be compensated for using the translational DOF. Figure 6 illustrates this point. Here, the camera reaches its pan limit ($\pm 150$ deg) as the target passes the southwest corner. Although no more panning is possible, Motor 2 can be engaged to maintain the target in its field-of-view. In regulator-style tracking systems all DOF are servoed purely by image data. In this cornering case, the failure to pan beyond its limit would make preserving a rigid camera-to-target pose impossible and tracking would then fail. By contrast, the partitioned scheme has kinematic servoing. The translational DOF respond directly to pan/tilt motor encoder data as given by (2). In essence by having kinematic servoing we can take advantage of the redundant DOF. The paper highlights these three points by a sequence of video stills taken while tracking a moving person (Figure 5). The switch from Motor 1 to Motor 2 and translational compensation for pan limit can be seen.

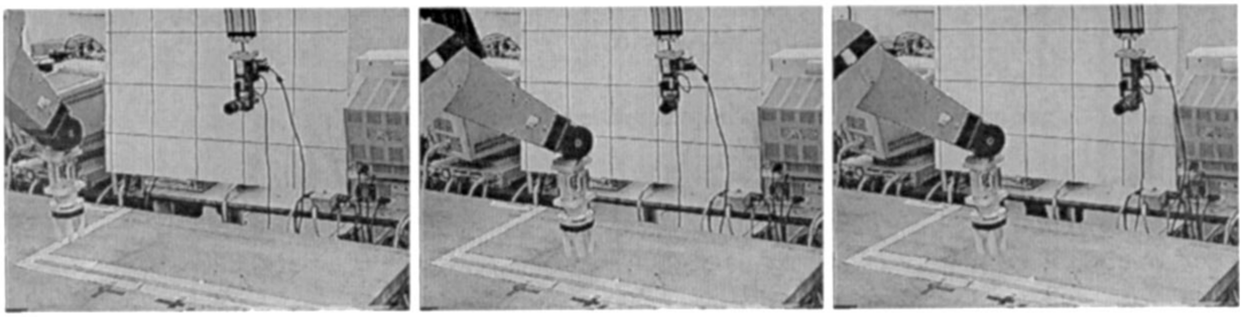

Figure 4: Partitioned controller tracking a Puma end-effector mounted gripper (foreground): As the gripper moves, (Left) pan is visually-servoed and keeps target in field-of-view. (Middle) Kinematically servoed translation occurs in direction of pan. (Right) Target comes to a stop

Figure 5: People Tracker: 6 sequenced images captured by a handheld video camera (rows 1 & 3) and the gantry-ptu camera (rows 2 & 4). A single SSD region-based tracker (white box) captures real-time position of head in the image. The camera (mounted on gantry's end-effector) pan, tilts and translates under partitioned control while tracking a moving person. As the person corners (rows 3 & 4), the pan hits its maximum angle. At this point the redundant translational DOF is engaged and tracking is maintained. One can note the change in camera position relative to the small grid in the background.

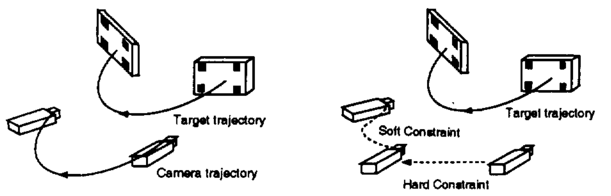


Figure 7: Regulator-style tracking (left): Bandwidths of camera servoing motors can effectively track small target accelerations (radial and/or tangential). Temporal Constraints (right): Partitioned control can track faster target speeds. The rigid camera-to-target pose constraint is reestablished under regulator control when the target is slow or motionless.

## 3.2 Experiment 2: Temporal Constraints

Figure 7 (left) is the camera trajectory as it tracks the moving target under regulator controller. As mentioned before, motor bandwidth becomes a problem if the target moves too fast. As mentioned previously, a rigid camera-to-target pose may only be needed at *critical* times. For example, in tracking tools or workpieces, pose may be most important during part pickup or tool alignment. At these times, speeds tend to be relatively slow and thus can be adequately handled by a conventional regulator tracking approach. At other times, such as during part or tool travel (when speeds are relatively fast) partitioned control can be used to keep it in view.

We thus introduce the idea of temporal constraints whereby the camera-to-target pose is time-dependent. We define *soft* and *hard* constraints in Figure 7 (bottom). In the former, partitioned control is used during fast target motions and is kept in field-of-view. When slow target velocities are detected pose is reestablished with a regulator controller. The net effect is a hybrid partition-regulator controller.

We highlight this concept in one experiment (Figure 8). Here, the target is a block with four fiducial marks with known sidelengths and is tracked with four SSDs. It quickly (10 $cm/s$) translates, slowly curves and stops. With only regulator control, the target would leave the camera's field-of-view before the servoing motors could get up to speed and tracking would fail. We then added partitioned control to the regulator to handle the fast target translation. The camera-to-target pose was reestablished under regulator control when the block moved slowly and stopped as seen by the similarity in the initial and final pose (Figure 8

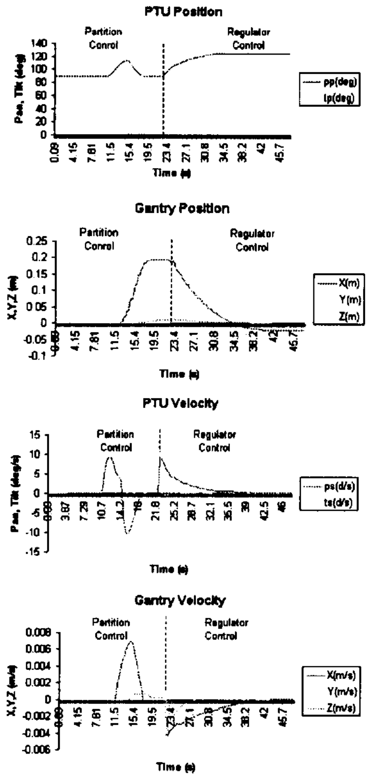


Figure 9: PTU and gantry position (top two) and velocity (bottom two) responses. The dashed vertical line emphasises the time when the robot switches from partition to regulator control

top left and bottom right images respectively). Figure 9 shows the gantry and PTU position and velocity responses. The dashed line (added) is when camera servoing switches from partition to regulator control. Asymptotic convergence can be observed as pose is reestablished under regulator control.

# 4 Conclusion

Tracking fast moving, geometrically complex targets with non-deterministic motion dynamics presents a challenging problem for conventional machine-vision systems. The results suggest that using a partitioned controller may be a viable method for improving tracking performance. It also suggests that the wealth of past kinematic servoing research can be potentially integrated into visual servoing efforts. The results presented here lend support to the idea that kinematic and visual servoing are complementary [6]. Lastly, we showed how a partitioned controller can be retrofitted to a conventional tracking system by defin-

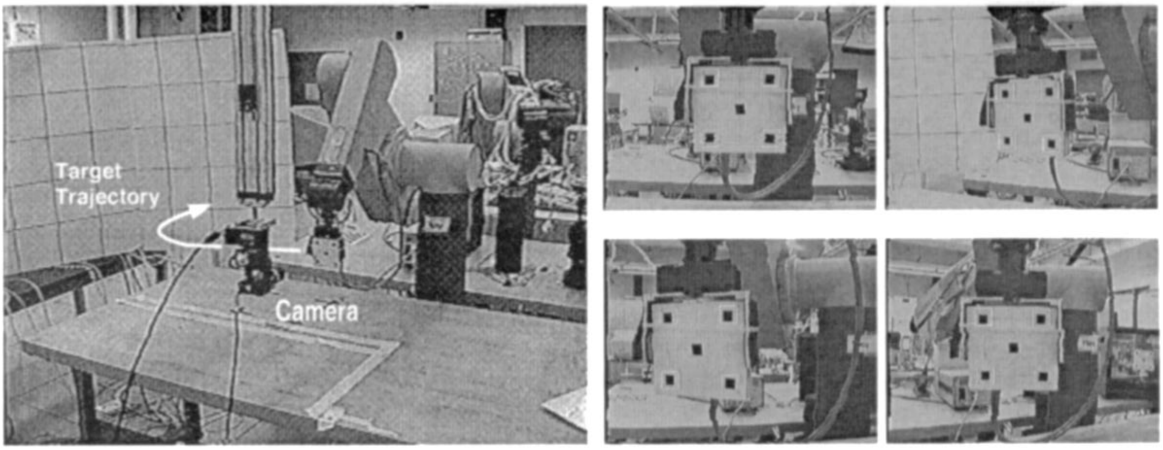


Figure 8: Temporal Constraints: A Puma-mounted block moves along a trajectory drawn by a white arrow in the large left photo. The block translates at 10 $cm/s$, curves slowly then stops. The block's sidelengths are known and 4 SSDs track each corner. The 4 smaller photos on the right are sequenced images of the block captured by the camera while tracking it. The initial image (top left) before tracking begins defines the desired camera-to-target pose. Partitioned control tracks the accelerating block (top right). As the block curves and stops, regulator control begins (bottom left) and reestablishes pose (bottom right).

ing temporal constraints. At present we are trying to implement depth-handling so that a desired image resolution can be maintained. Furthermore, we are trying to correlate the position and velocity responses between the coupled DOF. This may lead to path planning schemes by using key response points for dead-reckoning.

## 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] Castano, A., Hutchinson, S., "Visual Compliance: Task-Directed Visual Servo Control", *IEEE Trans Robotics & Automation,* V10, N3, p 334, 6/94.

[3] 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.

[4] 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.

[5] Corke, P. I., "Visual control of robot manipulators - A review," in *Visual Servoing*, K. Hashimoto, Ed. Singapore: World Scientific, 1993, pp 1.

[6] Crowley, J.L., Mesrabi, M., Chaumette, F., "Comparison of Kinematic and Visual Servoing for Fixation", *IEEE Int Conf Robotics & Automation,* pp 335, 1995.

[7] 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.

[8] Hager, G.D., Toyama, K., "X vision: A Portable Substrate for Real-Time Vision Applications", Tech Report, Dept of Computer Science, Yale Univ, 1995.

[9] Hutchinson, S., Hager, G., Corke, P., "A Tutorial on Visual Servo Control", *IEEE Trans Robotics & Automation,*, V12 N5, 10/96.

[10] Oh, P.Y., Allen, P.K., "Design of a Partitioned Visual Feedback Controller", *IEEE Int Conf Robotics & Automation,* Leuven, Belgium, pp 1360 5/98.

[11] Papanikolopoulos, N.P., Khosla, P.K., "Adaptive Robotic Visual Tracking: Theory and Experiments", *IEEE Trans Automatic Control"* V38 N3 3/93.

[12] Rizzi, A.A, Koditschek, D.E., "A Dynamic Sensor for Robot Juggling" in *Visual Servoing*, K. Hashimoto, Ed. Singapore: World Scientific, 1993, pp 1.

[13] Samson, C., Espiau, B., Le Borgne, M., "Robot Control: The Task Function Approach", Oxford University Press, 1990.

[14] Sharma, R., Sutanto, H., "A Framework for Robot Motion Planning with Sensor Constraints", *IEEE Trans Automatic Control"* V13 N1 2/97

[15] 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