Enhancing Camera Operator Performance with Computer Vision Based Control

Paul Y. Oh and Rares I. Stanciu

Open Science Index, Mechanical and Mechatronics Engineering Vol:1, No:2, 2007 publications.waset.org/3370.pdf

***Abstract*—** Cameras are often mounted on platforms that can move like rovers, booms, gantries and aircraft. People operate such platforms to capture desired views of scene or target. To avoid collisions with the environment and occlusions, such platforms often possess redundant degrees-of-freedom. As a result, manipulating such platforms demands much skill. Visual-servoing some degrees-of-freedom may reduce operator burden and improve tracking performance. This concept, which we call human-in-the-loop visual-servoing, is demonstrated in this paper and applies a $\alpha - \beta - \gamma$ filter and feedforward controller to a broadcast camera boom.

***Keywords*—** Computer vision, visual-servoing, man-machine systems, human-in-the-loop control

I. Introduction

Human-in-the-loop systems involve an operator who manipulates a device for desired tasks based on feedback from the device and environment. For example, devices like rovers, gantries, and aircraft possess a video camera where the task is to maneuver the vehicle and position the camera to obtain desired fields-of-view. Such tasks have applications in areas like broadcasting, inspection and exploration. Such device-mounted camera systems often possess many degrees of freedom (DOF) because it is important to capture as many fields-of-view as possible. To overcome joint limits, avoid collisions and ensure occlusion-free views, these devices are typically equipped with redundant DOF. Tracking moving subjects with such systems is a challenging task because it requires a well skilled operator who must manually coordinate multiple joints. Tracking performance becomes limited to how quickly the operator can manipulate redundant DOF. Figure 1 for example, shows a typical broadcast boom and pan-tilt camera head. Here, the operator can push and steer the dolly, as well as boom, pan and tilt the camera. Our particular interest is to apply *visual-servoing* to augment an operator's ability to track moving targets; computer vision is used to control some DOF so that the operator has fewer DOF to manipulate.

The prototype shown in Figure 1 was constructed to capture data, implement controllers and assess performance. Hardware includes a 266MHz PC, an pan-tilt DC motor controller and quadrature encoders. The vehicle is a four wheeled dolly with gimbaled broadcast boom, a motorized pan-tilt head, color camera, wireless video transmitter and framegrabber. The boom pivots on the steerable dolly to sweep the camera horizontally and vertically. Both proportional [9] [10] and partitioned [7] controllers were designed that visually servo the pan-tilt motors to keep a moving target centered in the camera's field-of-view despite boom or dolly motions. Sample image stills acquired from videotaping tracking experiments are shown in Figure 3. The net effect is what we call *human-in-the-loop visual servoing* – the operator just focuses on safely manipulating the boom and dolly while computer-control automatically servos the pan-tilt camera.

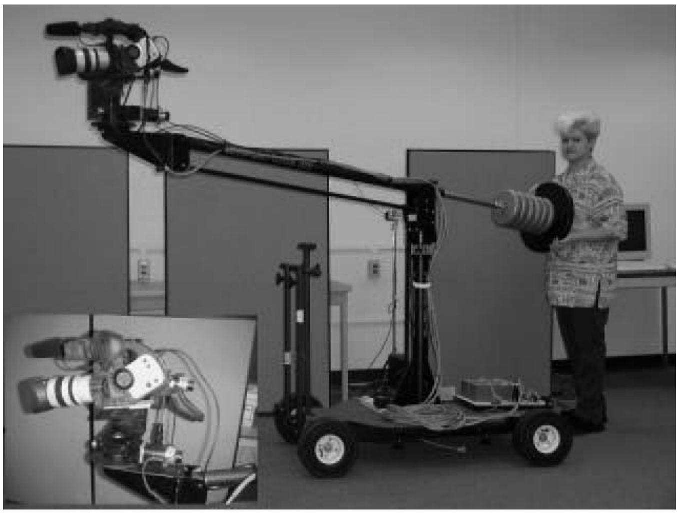

Fig. 1. The operator positions the camera by booming the arm horizontally and vertically. The pan-tilt head (inset) provides additional degrees-of-freedom.

A challenge underlined in [9] was the system's stability, especially when the target and the boom move 180 degrees out of phase. If boom motion data is not included, camera pose cannot be determined explicitly because there are redundant degrees-of-freedom. As a result, the system could track a slow moving target rather well, but would be unstable when the target or boom moves quickly. In this paper a feedforward controller is employed to improve stability. Section II describes the camera boom in more detail and provides its Denavit-Hartenberg configuration. Section III models the pan-tilt motors. The feedforward controller is presented in Section IV. Several experiments were performed to assess the performance of this controller. The results as well as some conclusions and a map of future work are presented in Sections V and VI respectively.

Paul Y. Oh (corresponding author) is the Director of the Autonomous Systems Lab at Drexel University, Mechanical Engineering & Mechanics, 3141 Chestnut Street, Philadelphia PA USA, 19038 Tel: 215-895-6396, Fax: 215-895-1478, Email: paulcoe.drexel.edu
Rares I. Stanciu was with the Drexel University Mechanical Engineering, 3141 Chestnut Street, Philadelphia PA USA, 19038 Email: ris22drexel.edu

Open Science Index, Mechanical and Mechatronics Engineering Vol:1, No:2, 2007 publications.waset.org/3370.pdf

II. System Description

The boom-camera system is composed of a 4-wheeled dolly, boom, motorized pan-tilt unit (PTU) and camera as shown in Figure 1. The 1.22 m long by 0.76 m wide dolly has four wheels and thus can be pushed and steered. The 1.2 m long boom is linked to the dolly via a 1.04 m cylindrical pivot, which allows the boom to sweep motions horizontally (pan) and vertically (tilt). Mounted on one end of the boom is a 2-DOF motorized PTU and video camera weighing 9.525 kg. The motors allow an operator to both pan and tilt the camera 360 degrees at approximately 90 deg/s. The PTU and camera are counterbalanced by 29.5 kg of dumbbell plates mounted on the boom's opposite end.

Normal broadcast use of this boom-camera system entails one or more skilled personnel: (1) With a joystick, the operator servos the pan-tilt head's DC motors to point the camera. A PC or small board computer motion control card, ISA or PC-104 bus respectively, allows for accurate and relatively fast camera rotations. (2) The operator physically pushes on the counterweighted end to boom the camera horizontally and vertically. This allows one to deliver a diverse range of camera views (like shots looking down at the subject), overcome pan-tilt head joint limits and capture occlusion-free views. (3) The operator can push and steer the dolly in case the boom and PTU are not enough to keep the target's image in the camera's field-of-view.

Our augmentation interests are to use machine vision to visually servo the pan-tilt camera and integrate computer control in the human-in-the-loop system. For the former, the target's image centroid can be measured from the real-time frame data to visually servo the 2-DOF PTU-camera. This can automatically keep the image centered in the camera's field-of-view and allow the operator to just focus on boom swings and dolly translations. For the latter, ultimately the pan-tilt head and boom motions redundantly orient the camera and can be problematic. For example during the visually servoing of the pan-tilt camera the operator conceivably can boom in the opposite direction. To compensate, the visually servoing must rotate the camera faster and if the two motions are out of phase by 180 degrees, they can conflict and visual-servoing will be unstable.

The control aspects of this latter problem are particularly interesting to us. The pan-tilt head is a fast bandwidth actuator but has limited range-of-motion. On the other hand, the boom can swing the camera over larger areas but its inertia limits swinging speeds. Such a system is an example of a manipulator with both fine and course ranges of motion. If a fine/course motion controller can be properly tuned then one can leverage the best performance each actuator has to offer. Such fine/course schema characterize many motion platform-mounted camera systems such as pan-tilt cameras mounted on helicopters and rovers; the vehicle provides large range of motion but fine pan-tilt motions are required to ensure the image remains centered in the camera.

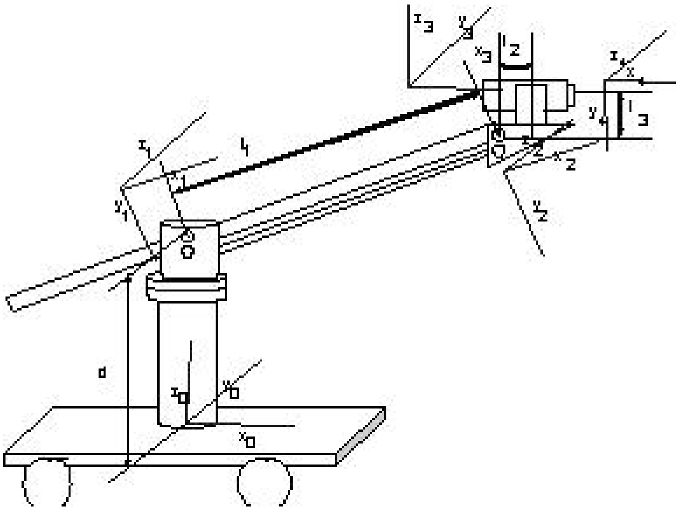

Fig. 2. Denavit-Hartenberg notation for joint frames

Joint	α	d	a	θ
1	$-\frac{\pi}{2}$	d	0	θ_1
2	0	0	l_1	θ_2
3	$+\frac{\pi}{2}$	0	l_2	θ_3
4	$-\frac{\pi}{2}$	l_3	0	θ_4

TABLE I
DENAVIT-HARTENBERG LINK AND JOINT PARAMETERS

To underscore the man-machine control issues of visually servoing redundant DOF systems, a simple but reliable vision system is used. Real-time image centroid measurements are performed using a Newton Cognachrome color tracker, which is an embedded microprocessor that serially transmits the centroid's pixel location of a colored target. Additionally a joint encoder to measure horizontal booming was installed.

To summarize, the boom-camera system's Denavit-Hartenberg reference frames and arm matrix are given in Figure 2 and Table I. Readers interested in the arm matrix are directed to [9] for a complete derivation.

III. Modeling the PTU

As shown in Figure 1, the camera is mounted on a 2 degree-of-freedom pan-tilt unit (PTU). Two DC motors are driven by a motion card installed in a PC. Like many commercial motion cards, the PID control gains are factory set, balancing transient response with minimal overshoot. Using a standard DC motor transfer function, one has

$$\begin{aligned} G_m(s) &= \frac{\dot{\theta}_m(s)}{E_a(s)} \qquad (1) \\ &= \frac{K_t}{K_v K_t + (sJ_m + D_m)(R_a + sL_a)} \end{aligned}$$

where $\dot{\theta}_m$ is motor speed, E_a is the applied voltage, K_t is the motor torque constant, K_v is the back EMF constant, R_a and L_a are the rotor resistance and inductance respectively and D_m

Open Science Index, Mechanical and Mechatronics Engineering Vol:1, No:2, 2007 publications.waset.org/3370.pdf

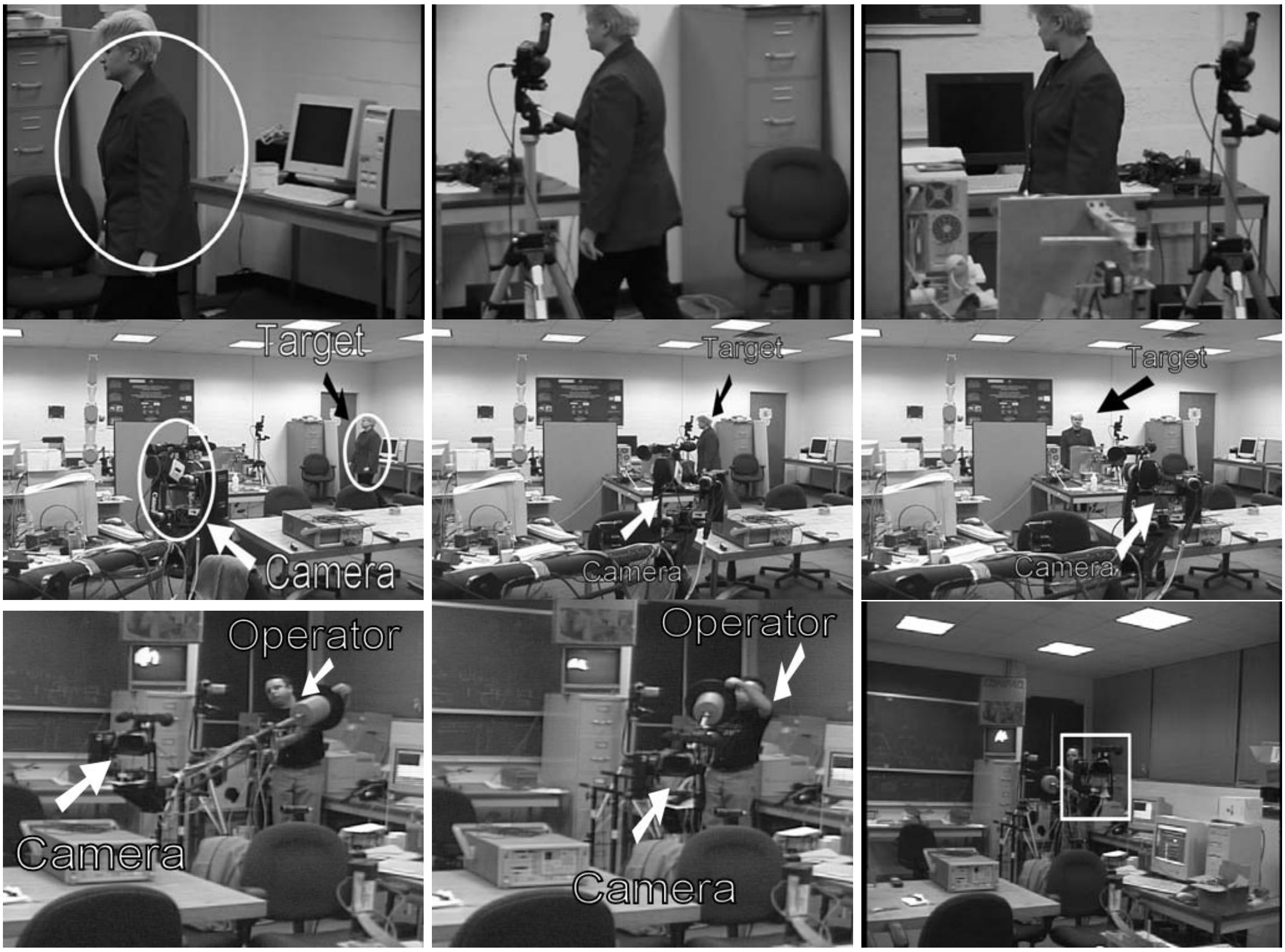

Fig. 3. Three sequential images from videotaping a tracking experiment. Camera field-of-view (top row) shows the target is kept centered in the image plane despite boom motions executed by the operator. Such motions are illustrated by the middle and bottom rows' images; two cameras placed in the room were used to record the experiment.

Motor Parameters	Value and Units
R_a, rotor resistance	1.15 Ω
L_a, rotor inductance	1.4 mH
K_t, torque constant	0.055 Nm/A
K_v, back EMF constant	5.8 $V/krpm$
J_a rotor moment of inertia	$1.33 \cdot 10^{-5}$ $kg \cdot m^2$

TABLE II
PTU MOTOR PARAMETERS.

is the armature viscous damping. Values for these parameters are given in Table II. J_m is the motor shaft's moment of inertia.

$$J_m = J_a + J_L(\frac{N_1}{N_2})^2 \qquad (2)$$

where, J_L is load moment of inertia, J_a is the rotor moment of inertia and $\frac{N_1}{N_2}$ is the gear ratio. The PTU's gear ratio and D_m are both small and were set to zero. As such, Equation 2 with values from Table II results in

$$G_m(s) = \frac{\dot{\theta}_m(s)}{E_a(s)} = \frac{5500}{0.001862s^2 + 1.295s + 31.9} \qquad (3)$$

Using a zero-order-hold to model a digital-to-analog converter, the discrete form of the transfer function can be calculated. Figure 4 gives the block diagram where v_{ref} is the command reference velocity, E is the error between the command and actual motor velocities and $K_e = 2000$ counts/rev is the encoder constant. The sampling time T was set at 1.25 msec. $D(z)$ is the factory tuned PID controller with proportional, integral and derivative gains set at $K_P = 15000$, $K_I = 40$ and $K_D = 20000$ respectively for the PTU pan motor. PID gains for the tilt motor were factory set at $K_P = 15000$, $K_I = 20$ and $K_D = 32000$. With $G_m(s)$ given by Equation 3, the discrete transfer function relating the command and actual velocities is given as

$$G_P(z) = \frac{0.704 - 0.787z^{-1} + 0.439z^{-2} - 0.055z^{-3} + 0.035z^{-4}}{1409.3 - 1575.36z^{-1} + 878z^{-2} - 11.02z^{-3} + 70z^{-4}} \qquad (4)$$

Equation 4 is validated experimentally as described in Section V.

IV. Feedforward Controller

As mentioned in Section I, the boom-camera system under proportional control [9] becomes unstable when tracking a fast moving target. The boom and PTU are redundant rotational

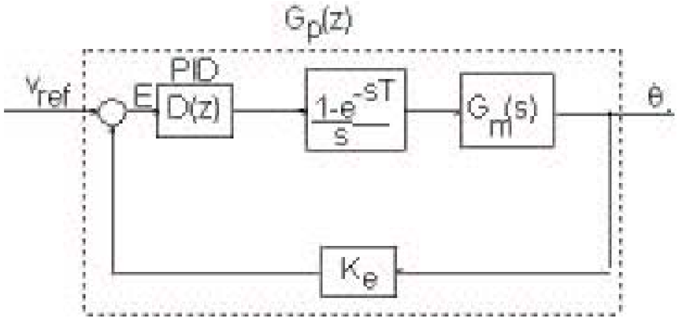

Fig. 4. The PTU Controller Block Diagram

Open Science Index, Mechanical and Mechatronics Engineering Vol:1, No:2, 2007 publications.waset.org/3370.pdf

DOF that at high frequencies can become 180 degrees out-of-phase. The net result is the boom and PTU rotations conflict rather than cooperate and tracking fails. To overcome such instabilities, a feedforward controller can be designed which provides target motion estimation [3]. Figure 6 depicts a block diagram with transfer function

$$\frac{^iX(z)}{X_t(z)} = \frac{V(z)(1 - G_p(z) \cdot D_F(z))}{1 + V(z) \cdot G_p(z) \cdot D(z)} \tag{5}$$

where $^iX(z)$ is the position of the target in the image, $X_t(z)$ is target position, $V(z)$ and $G_p(z)$, are respectively the transfer functions for the vision system and PTU. $D_F(z)$ and $D(z)$ are respectively the transfer functions for the feedforward and feedback controllers.

Clearly if $D_F(z) = G^{-1}_{PTU}(z)$ the tracking error will be zero, but this requires knowledge of the target position which is not directly measurable. Consequently the target position and velocity are estimated. For a horizontally translating target, its centroid in the image plane is given by the relative angle between the camera and the target

$$^iX(z) = K_{lens}(X_t(z) - X_r(z)) \tag{6}$$

where $^iX(z)$ and $X_t(z)$ are the target position in the image plane and world frame respectively. $X_r(z)$ is the position of the point which is in camera's focus (due to the booming and camera rotation) and K_{lens} is a constant mapping the world to image space. The target position prediction can be obtained from the boom and PTU as seen in Figure 5. Rearranging this equation yields

$$\hat{X}_t(z) = \frac{^i\tilde{X}(z)}{K_{lens}} + X_r(z) \tag{7}$$

where $\hat{X}_t$ is predicted target position.

A. $\alpha - \beta - \gamma$ Filter

Predicting target velocity requires a tracking-filter. Oftentimes a Kalman filter is used but is computationally expensive. Since Kalman gains often converge to constants, the simpler $\alpha - \beta - \gamma$ tracking filter can be employed which tracks both position and velocity without steady-state errors [6].

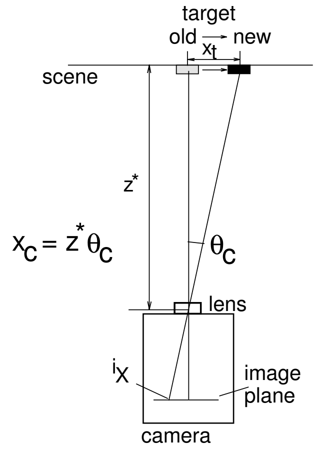

Fig. 5. A schematic of camera-scene.

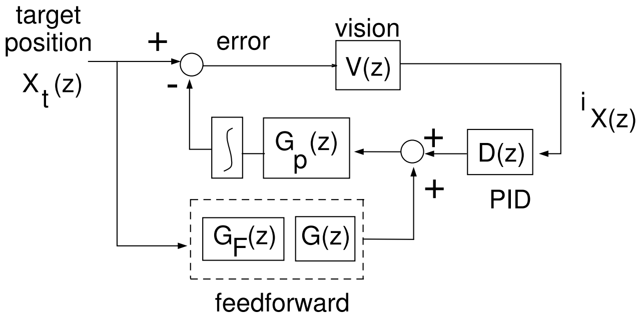

Fig. 6. The Feedforward Controller with Feedback Compensation.

Tracking involves a two step process. The first step is to predict target position and velocity

$$x_p(k+1) = x_s(k) + Tv_s(k) + T^2 a_s(k)/2 \tag{8}$$

$$v_p(k+1) = v_s(k) + Ta_s(k) \tag{9}$$

where T is the sample time and $x_p(k+1)$ and $v_p(k+1)$ are respectively the predictions for position and velocity at iteration $k+1$. $x_s(k)$, $v_s(k)$ and $a_s(s)$ are the corrected values of iteration k for position, velocity and acceleration respectively.

The second step is to make corrections

$$x_s(k) = x_p(k) + \alpha(x_o(k) - x_p(k)) \tag{10}$$

$$v_s(k) = v_p(k) + (\beta/T)(x_o(k) - x_p(k)) \tag{11}$$

$$a_s(k) = a_p(k-1) + (\gamma/2T^2)(x_o(k) - x_p(k)) \tag{12}$$

where $x_o(k)$ is the observed (sampled) position at iteration k. The appropriate selection of gains α, β and γ will

Open Science Index, Mechanical and Mechatronics Engineering Vol:1, No:2, 2007 publications.waset.org/3370.pdf

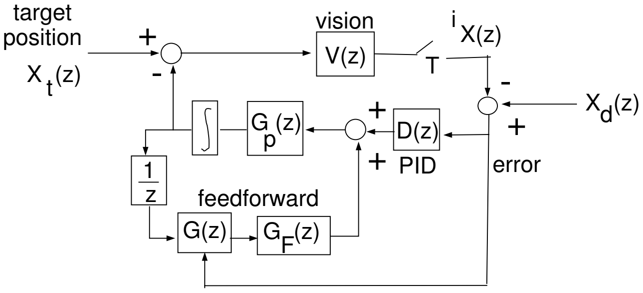

Fig. 7. The Feedforward Controller with Feedback Compensation as it was implemented.

determine the performance and stability of the filter [11].

An $\alpha - \beta - \gamma$ filter was implemented to predict target velocity in the image plane with gains set at $\alpha = 0.75$, $\beta = 0.8$ and $\gamma = 0.25$. This velocity was then used in the feedforward algorithm as shown in Figure 7.

Image processing in the camera system can be modeled as a $1/z$ unit delay which affects camera position x_r, and estimates of target position. In Figure 7, the block $G_{abg}(z)$ represents the transfer function of the $\alpha - \beta - \gamma$ filter, with the observed position as input and the predicted velocity as output.

The focal length, K_{lens}, was set a constant value and assumes a pinhole camera model that maps the image plane and world coordinates. This focal length was experimentally determined by comparing known lengths in world coordinates to their projections in the camera's image plane.

Taking the Z transform of the $\alpha - \beta - \gamma$ filter yields its discrete-time transfer function $G_F(z) = \frac{V_P(z)}{X_o(z)}$ or

$$G_F(z) = \frac{2.4z^4 - 8.47z^3 + 11.09z^2 - 3.81z}{z^5 - 3.3875z^4 + 4.2375z^3 - 2.1632z^2 - 0.7325z + 0.75} \quad (13)$$

where $V_P(z)$ is the predicted velocity and $X_o(z)$ is target's observed position.

V. Experimental Results

Experiments to validate the dynamic models and to compare the performance of feedforward and proportional control in *human-in-the-loop visual-servoing* were performed. A condensation algorithm was implemented to capture the target's position in image space. A 8.9×8.25 cm^2. wooden block was mounted in the end-effector of a 7-DOF Mitsubishi robot arm, Figure 8. The camera-to-target distance was 3.15 m and focal length of $K_{lens} = 500$ pixels.

To validate the dynamic model, Equation 4, a Bode plot was generated. Here, the input would be an oscillating target and the output would be the resulting PTU angle. As such, the robot arm oscillated the block horizontally over a range of frequencies and PTU output angles were recorded. As shown in Figure 9, the resulting magnitude and phase plots (top two) match well with a Matlab simulation on Equation 4 (bottom).

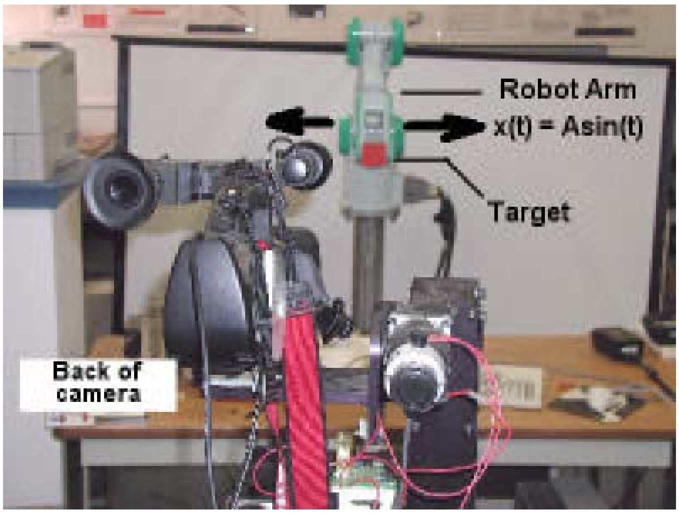

Fig. 8. A wooden block target was mounted in the end-effector of a Mitsubishi robot arm (background). The boom-camera system (foreground) attempts to keep the target's image centered in the camera's field-of-view.

Figure 10 shows the results tracking the target which oscillated at 0.08 Hz from -0.58 m to $+0.49$ m (top plot). While the controllers attempted to track the target, the boom was manually moved over from -15 to +25 degrees (second plot from top). The bottom two plots depict tracking errors resulting from such *human-in-the-loop visual-servoing*. Feedforward based control has a ±100 pixel peak-to-peak tracking error (bottom-most plot) compared to ±300 pixel errors in proportional-only control. Zero peak-to-peak pixel error reflect perfect tracking such that the target image always remains centered in the camera's field-of-view. As such, the results suggest that a feedforward strategy performs better than proportional control for *human-in-the-loop visual-servoing*.

VI. Conclusions and Future Work

This paper integrates visual-servoing for augmenting the tracking performance of camera teleoperators. By reducing the number of DOF that need to be manually manipulated, the operator can concentrate on coarse camera motion. Using a broadcast boom system as an experimental platform, the dynamics of a camera pan-tilt-unit were derived and validated experimentally. A feedforward controller with an $\alpha - \beta - \gamma$ filter was the formulated and implemented experimentally. Results comparing proportional and feedforward controllers were illustrated. Feedforward control yielded lower peak-to-peak pixel errors which suggest that estimating target position improves tracking performance despite a human-in-the-loop disturbances. Future work will look at increasing the bandwidth under which the boom-camera system can track stably. A multivariable controller approach is being considered.

References

[1] Canon, D.J. (1994). "Experiments With a Target-Threshold Control Theory Model for Deriving Fitts Law Parameters for Human-Machine

Open Science Index, Mechanical and Mechatronics Engineering Vol:1, No:2, 2007 publications.waset.org/3370.pdf

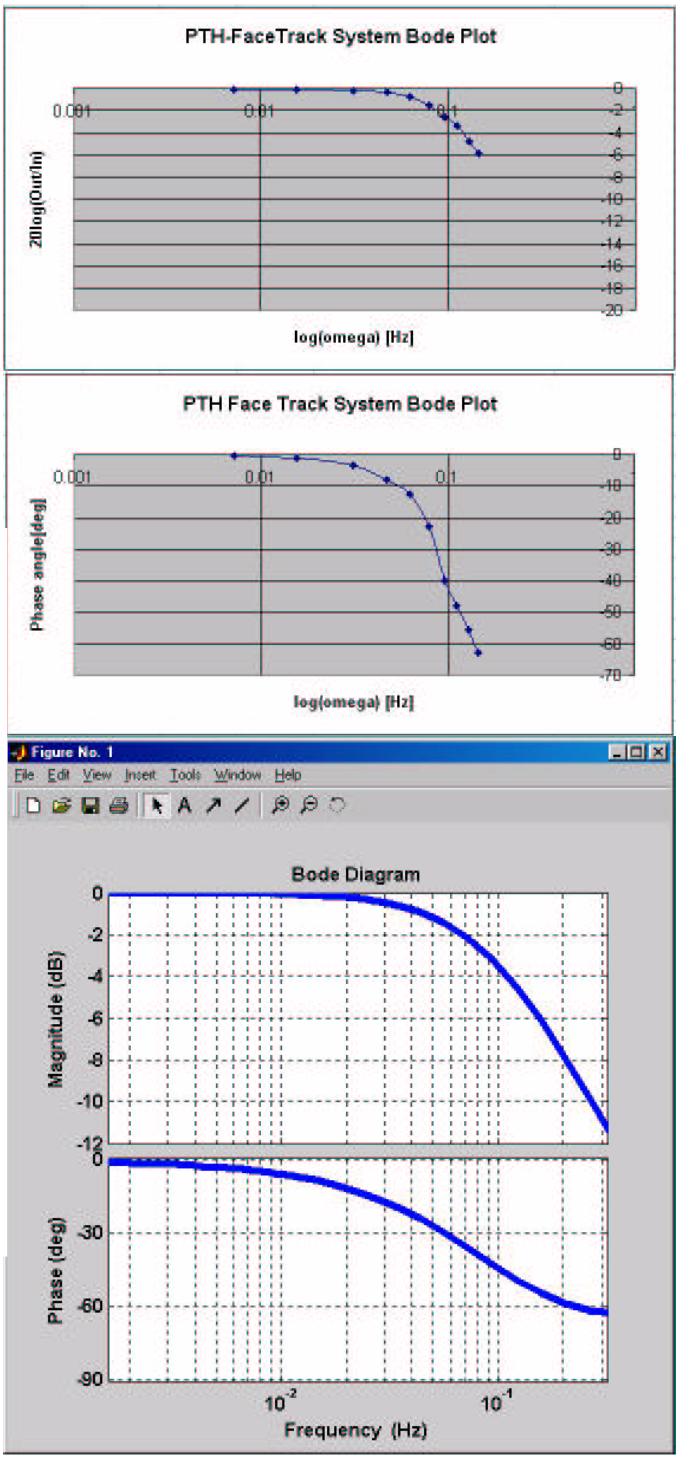

Fig. 9. The PTU Bode magnitude (top) and phase (bottom) plots

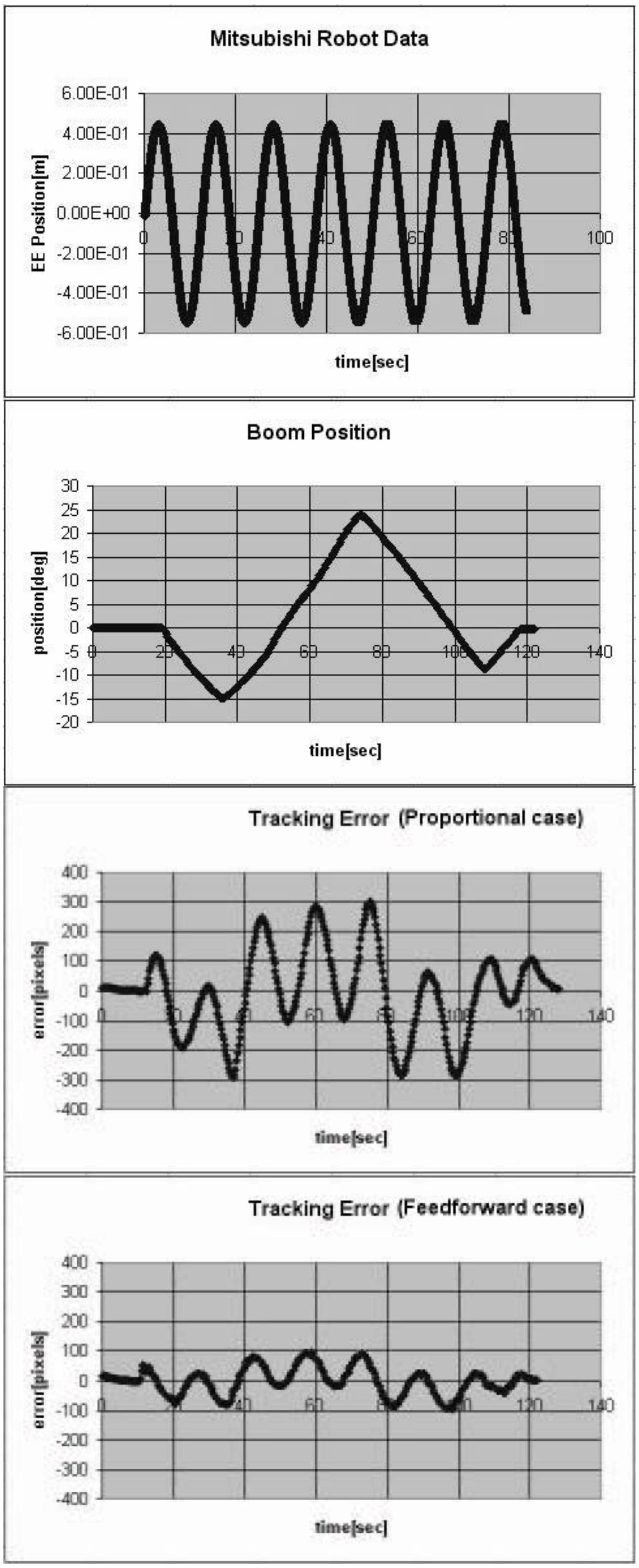

Fig. 10. Tracking errors comparing feedforward and proportional control in *human-in-the-loop* visual-servoing.

Systems", *IEEE Trans on Systems, Man and Cybernetics*, V24 N8, pp. 1089-1098.

[2] Chaumette F., Rives P., Espiau B. (1991). "Positioning of a robot with Respect to an Object, Track it and Estimating its Velocity by Visual Servoing", *IEEE Int Conf Robotics and Automation (ICRA)*, Sacramento, CA.

[3] Corke P.I., Good M.C. (1996). "Dynamic Effects in Visual Closed-Loop Systems" *IEEE Trans on Robotics and Automation* V12 N5.

[4] Hutchinson S., Hager G.D., Corke P.I. (1996). "A Tutorial on Visual Servo Control", *IEEE Trans on Robotics and Automation* V12 N5, pp. 651-670.

[5] Isard, M., Blake, A. (1998). "CONDENSATION – Conditional Density Propagation for Visual Tracking", *Int. J. Computer Vision*, V29, N1, pp. 5-28.

[6] Kalata, P.R., Murphy, K.M. (1997). '$\alpha - \beta$ Target Tracking with Track Rate Variations", *Proc of the Twenty-Ninth Southeastern Symposium on System Theory*, pp. 70-74.

[7] Oh, P.Y., Allen, P.K. (2001). "Visual Servoing by Partitioning Degrees of Freedom", *IEEE Trans on Robotics Automation* V17 N1, pp. 1-17.

[8] Sheridan T.B., Ferrell W.R. (1994). Man-Machine Systems: Information, Control, and Decision Models of Human Performance, MIT Press, Cambridge, Massachusetts.

[9] Stanciu R., Oh P.Y. (2002). "Designing Visually Servoed Tracking to Augment Camera Teleoperators" *IEEE Intelligent Robots and System (IROS)*, Lausanne, Switzerland, V1, pp. 342-347.

[10] Stanciu R., Oh P.Y. (2003). "Human-in-the-loop Visually Servoed Tracking" *International Conference on Computer, Communication and Control Technologies (CCCT)*, V5, pp. 318-323, Orlando, FL.

[11] Tenne, D., Singh, T. (2000). "Optimal Design of $\alpha - \beta - (\gamma)$ Filters", *Proc of the American Control Conference*, V6 pp. 4348-4352.

Open Science Index, Mechanical and Mechatronics Engineering Vol:1, No:2, 2007 publications.waset.org/3370.pdf