Biomimetic Visual-Servoing Using A Partitioned Control Scheme

Bharat N. Shah and Paul Y. Oh *
Mechanical Engineering & Mechanics, Drexel University, Philadelphia PA
Email: bns23@drexel.edu and paul@coe.drexel.edu

Abstract

Presented is a partitioned controller for the visually-servoed tracking of objects using both image and the kinematic data. The resulting design shows remarkably similar performance to vision systems in biology. For example, people coordinate their eyes, head and body when tracking a moving target. Likewise, this controller couples a robot's degrees-of-freedom to point a camera at a moving target. Tracking experiments showcase the design on a robotic system featuring two revolute joints to mimic head-eye motions.

1 Introduction

For decades research in visually servoed tracking has been pursued. Much of the focus remains on developing robust and faster image processing algorithms along with controllers that overcome computational delays. For several years now, fast hardware and actuators have been available, but interestingly visual-servoing in real-time is not common. Eye-in-hand systems, where the camera is mounted on a robot end-effector, often fail in tracking fast moving targets [7] [3]. Stereo camera heads which mount on tripods are often engineered with fast motors in order to increase actuation rates [14]. A wide range of controller designs have been suggested to overcome delays introduced by image processing including feed-forward [4], adaptive [13], observers [9] and optimal mode-switching [6]. While such controllers work to overcome frame rate delays, they are still limited by servo dynamics; the slowest joint in the robot dictates the maximum speed a camera position can be updated. Oh and Allen in [10] demonstrated a controller they call *partitioning* which invokes both kinematic and visual servoing based on joint encoder and image data respectively. Such servoing modalities have been shown to be complementary [5] and partitioning appears to mimic head-eye motions found in biology. The underlying idea exploits joint redundancies to couple motions. As such, fast bandwidth actuators can be used to keep the target centered in the camera's field-of-view while slower degrees-of-freedom can increase range of motion. This is the case in people too. For example, a spectator tracks a ball during a tennis match by panning the eyes first and head soon after. Neuroscientists explain that such behaviors utilize both visual and kinematic data like accelerations measured by the inner ear and ocular muscle innervations to ascertain a target's location in space [1] [2]. Although head-eye behaviors have been mimicked on multi-DOF camera heads for many, gain tuning has often done *ad hoc.* This paper leverages analytical methods derived by the authors [11] that optimally use redundant DOF to avoid joint limits and track faster moving objects.

Motor	Precision deg/sec	Range deg	Max. Velocity deg/s
Turntable	5.49×10^{-4}	± 150	70
Pan-tilt unit	0.0514	± 75	135

Table 1: Motion range limits and speeds of the DOF

This paper couples rotationally redundant DOF in the visual-serving loop and analyzes the effects multiple joint kinematics and dynamics have on tracking performance. Figure 1 illustrates the experimental platforms which consists of a camera mounted on a turntable and pan-tilt unit; either or both can be invoked to keep a horizontally moving target's image centered in the camera's field-of-view. The redundancy, link lengths, joint limits and bandwidths can be varied to reveal their effects on tracking and define a suitable control system. This paper is arranged as follows. Section 2 introduces the apparatus. Section 3 describes the partitioned tracking. Experiments and results are provided in Section 4 where tracking results using single and redundant degrees-of-freedom are compared. Section 5 concludes and describes future work.

*Author for contact

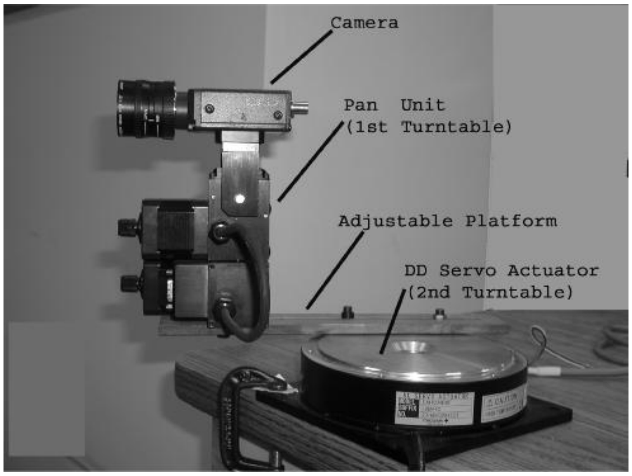

Figure 1: Experimental Platform: The camera has rotational redundancy and can be positioned either by the pan-tilt unit, turntable or both

2 Testbed

In people, both the eye and head rotate to direct the retina. Such rotational redundancy enables one to tracking moving objects over a wide range of speeds and motions. The eye can rapidly respond to sudden changes in target motion but has narrow joint limits. The slower moving head however, has a larger rotational range. The net effect is that head and eye leverage redundancy and work cooperatively. To understand this behavior, an experimental testbed was constructed. The visual servoing system is composed of a turntable, pan-tilt unit (PTU) and camera as shown in Figure 1.

The turntable and the PTU are computer-controlled motion devices and can independently rotate the camera. An slidable link was designed so that the offset between the rotational axes of the two DOF can be adjusted. The PTU is a light-weight device and can accelerate very quickly compared to the turntable. The relevant parameters of the 2-DOF turntable-pan system are given in Table 1. Image data is provided by a region-of-interest tracker implemented with Coreco's Sherlock vision package, an off-the-shelf hardware and software bundle.

3 Partitioned Visual-Servoing

Partitioned control uses both visual and kinematic servoing [10]. Image data is used to servo a degree-of-freedom. The resulting joint motion is then used to kinematically servo the remaining degree-of-freedom.

In image based visual-servoing, the error signal is expressed in the pixel space. The tracking task demands keeping the target's image centered in the camera's field-of-view. If off-centered a non-zero pixel error results. The control law's role is to bring this error to zero by servoing the camera. As such, an image Jacobian J is required to relate rates in image space, f, to those in task space, r

$$\dot{f} = J\dot{r} \tag{1}$$

In visual servoing applications, one is interested in finding $\dot{r}$ given $\dot{f}$. As such the pseudo-inverse of the image Jacobian is calculated since J is often a non-square matrix. Often, image features f are grouped into a vector s. An image centroid for example requires both the horizontal and vertical pixel locations. In regulator based tracking, a reference image is captured to define a set point vector (s^*). As the target moves, image features s are monitored at frame rate. An error function, which describes the geometrical relationship between the camera and the target, is defined as

$$e(\vec{r}(t), t) \doteq J^{(-1)}(s(\vec{(r)}(t), t) - s^*) \tag{2}$$

$\vec{r}(t)$ is a 6×1 vector at time t of the position and orientation of the target with respect to the camera. $e(\vec{r}(t), t)$ becomes

$$\frac{de}{dt} = \frac{\delta e}{\delta \vec{r}} T_c + \frac{\delta e}{\delta t} \tag{3}$$

where $T_c \doteq \frac{d\vec{(r)}}{dt}$. T_c is the change in camera velocity in response to the rates of changes in e. From (3)

$$T_c = (\frac{\partial e}{\partial \vec{r}})^{-1}(\frac{de}{dt} - \frac{\partial e}{\partial t}) = \frac{de}{dt} - \frac{\partial e}{\partial t} \tag{4}$$

where $\frac{\partial e}{\partial t} = J^{-1}J = I$. The error e will converge asymptotically by setting

$$\frac{de}{dt} \doteq -\lambda e \tag{5}$$

with $\lambda > 0$. With this the visual-servoing control law becomes

$$T_c = -\lambda e - \frac{\delta e}{\delta t} \tag{6}$$

In the above equation, the last term represents the target velocity. This results in

$$\frac{de}{dt} = T_c - \frac{\delta e}{\delta t} \tag{7}$$

If the camera's velocity is identical to the targets velocity then $s(\vec{r}(t)) \equiv s^*$and $\frac{de}{dt}$ would be zero. Thus,

$$T_c = \frac{\delta e}{\delta t} = T_{target} \tag{8}$$

Partitioning invokes both visual and kinematic servoing. This is illustrated in Figure 2 where the block diagram features two feedback loops. One loop uses image data to actuate a DOF. Joint encoders measure these changes and drive any remaining DOF. The net effect is coupled motion that does not require an explicit manipulator Jacobian. All partitioned DOFs work in concert to keep a target's image centered in the camera's field-of-view. To apply partitioning to the PTU/Turntable testbed in Figure 1, the PTU and turntable were respectively servoed visually and kinematically.

Visual-Servoing

The image Jacobian for a point is well known [7] and maps differential changes in pixel space, $\dot{s}$ to changes in task space T_c

$$\underbrace{\begin{bmatrix} \dot{u} \\ \dot{v} \end{bmatrix}}_{\dot{s}} = \underbrace{\begin{bmatrix} -\frac{f}{z} & 0 & \frac{u}{z} & \frac{uv}{f} & -\frac{f^2+u^2}{f} & v \\ 0 & -\frac{f}{z} & \frac{v}{z} & \frac{f^2+v^2}{f} & -\frac{uv}{f} & -u \end{bmatrix}}_{J} \underbrace{\begin{bmatrix} T_x \\ T_y \\ T_z \\ \omega_x \\ \omega_y \\ \omega_z \end{bmatrix}}_{T_c} \quad (9)$$

The above (2×6) matrix maps the velocity of a point (x, y, z) in the $\Re^3$ task space to a velocity of a point (u, v) in the $\Re^2$ (camera) image space. Here, f is the camera lens focal length. The subscript c denotes that the variable is with respect to the camera frame. The frame's origin is at the camera's lens. z_c is along the optic axis and points towards the target.

Tracking a horizontally moving target however only requires camera panning. As such, 9 reduces to a simple relationship for visually servoing the PTU

$$\frac{du}{dt} = \frac{uv}{f}\omega_x \quad (10)$$

Kinematic Servoing

Referring to Figure 2 any angular changes in the PTU invokes turntable motions. In other words

$$\frac{dq}{dt} = K_t(q_{pan}(t) - q^*{}_{set}) \quad (11)$$

where K_t is a proportionality gain constant. $q^*{}_{set}$ is a reference setpoint. The net effect is a coupled motion where both the PTU and turntable rotate in order to keep the target centered in the camera's field-of-view.

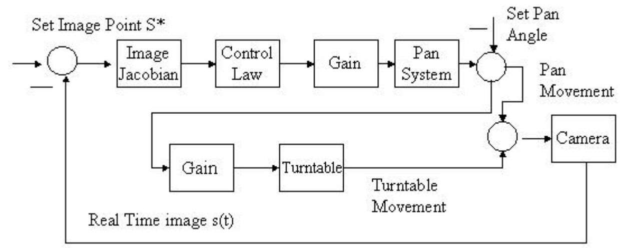

Figure 2: Partitioned Control Block Diagram.

4 Tracking Experiments

To contrast conventional and partitioned visually servoed tracking, step response experiments were performed. In conventional visual-servoing, only one DOF is needed to track a horizontally translating target. In the double-revoluted system shown in Figure 1 either the PTU or turntable can rotate the camera. The resulting tracking performance will be dictated by the DOF's dynamics. To illustrate limitations of visually servoing a single DOF, two cases were studied. The control law (6) and (10) were applied to one, the PTU and two, the turntable. The discrete-time closed-loop transfer functions for these two cases were derived for a sample time T and respectively are

$$G_{\text{PTU}} = \frac{T\lambda_p}{z^* z(z-1) + T\lambda_p(z^* + L)} = \frac{\theta_c}{x_t} \quad (12)$$

$$G_{\text{turntable}} = \frac{T\lambda_t(z^* + L)G_m}{T\lambda_t G_m + z^*(z-1)} = \frac{\theta_c}{x_t} \quad (13)$$

Here, λ_p and λ_t are the proportional gains for visually servoing the PTU and turntable respectively. The transfer functions relate target translation x_t with the resulting camera motion output θ_c. The PTU motors are steppers and hence modeled as a delay. The turntable is a brushless DC motor. Only the input and output relationship, $G_{\text{turntable}}$, was of interest and hence the exact DC motor transfer function G_m was not identified specifically.

To create a step input, the target was rapidly translated using a Mitsubishi PA-10 robot arm, seen in Figure 7. The target's horizontal displacement was $x_t = 0.1\ m$ while the distance between the lens and the target was $z^* = 0.25\ m$. The lens focal length f was 623 pixels.

Figures 3 and 4 show the step response and the pixel

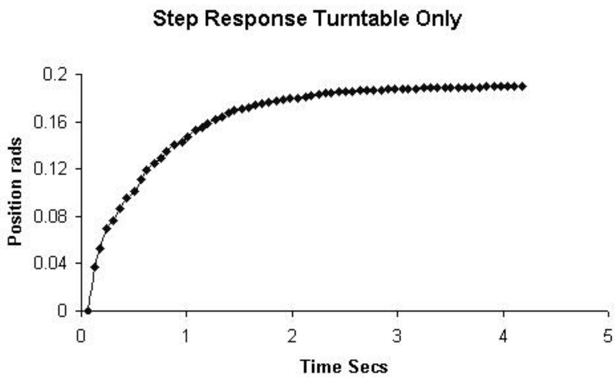

Figure 3: Step Response Turntable Only Tracking

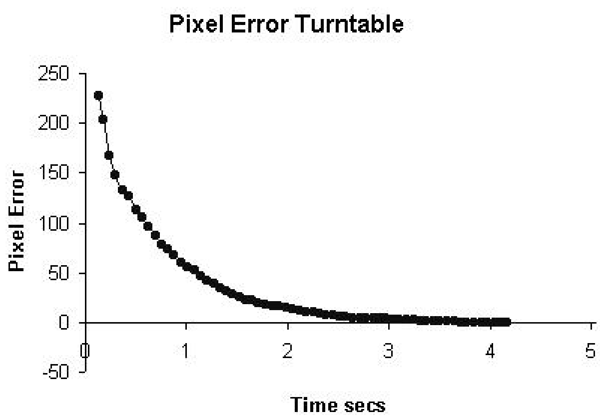

Figure 4: Pixel Error Turntable Tracking

error respectively for the visually servoed turntable. From these plots the rise time is approximately 0.8 seconds with the pixel error asymptotically converging. By contrast, the visually servoed PTU is faster. Figure 5 shows a rise time of approximately 0.3 seconds.

Partitioned Pan-Turntable Tracking

In contrast to visually servoing a single DOF, partitioning exploits any DOF redundancies to track. People also take advantage of DOF redundancy when coordinating both eye and head motions when tracking. Step response experiments for the system portrayed in Figure 2 were performed. Figure 6 (left) reveals a rise time of 0.17 seconds. In contrast to pan, speed almost doubled and to turntable, speed more than quadrupled.

Interestingly the plot bears a striking resemblance to human head-eye motions observed clinically when per-

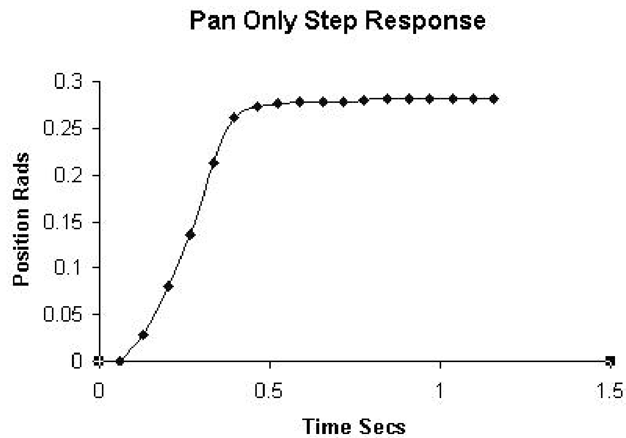

Figure 5: Step Response Pan Only Tracking

forming similar step input type tracking experiments, Figure 6(right). In partitioning, a stable and desired transient response can be obtained by choosing proportional velocity gains λ_p and λ_t that place closed-loop poles into desired positions [10].

5 Conclusions

People invoke both eye and head rotations when tracking. The eye can actuate the retina quickly but has limited range of motion. Head rotations help to overcome this limitation. This paper showed that invoking similar rotations for visually servoed tracking is possible. Experiments on a double-revoluted camera system composed of a pan-tilt unit and turntable were performed. Tracking using both actuators was faster than servoing either one alone.

The partitioning concept is currently being applied to camera motion devices like broadcast booms [12]. Often in such devices the DOF are completely controlled manually. As such, tracking performance is limited to operator skill. Future work is in formulating *human-in-the-loop visual-servoing* where some DOF are operated manually while others are controlled with visual-servoing. The net effect is a macro-micro manipulation scheme where an operator's tracking performance is augmented.

References

[1] Bernstein, N.I., Co-ordination and Regulation of Movement, New York : Pergamon Press 1967.

[2] Carpenter, R.H.S., Movements of the Eyes, 2nd Edition, Pion Publishing, 1988

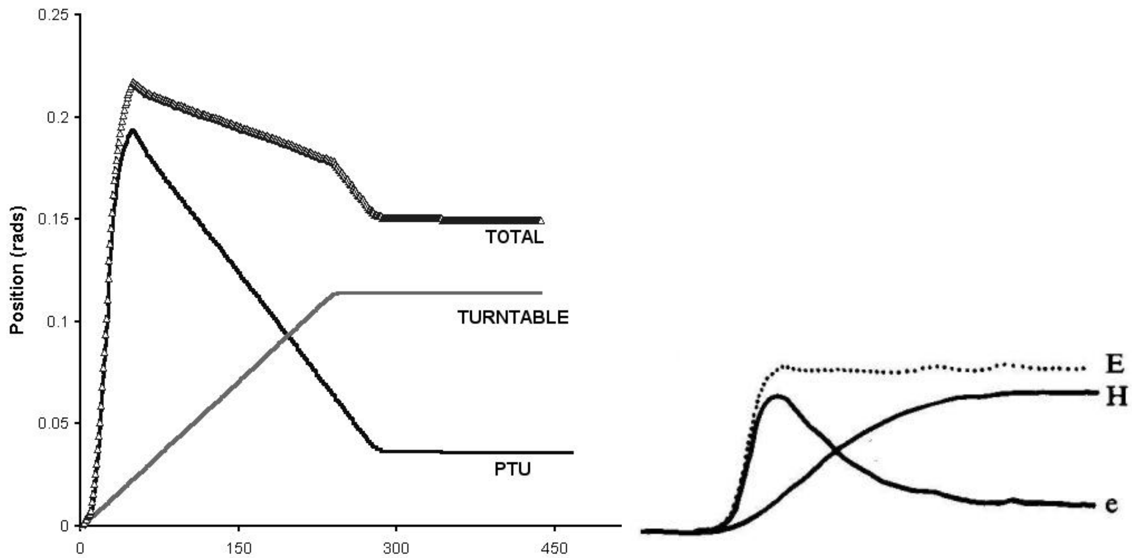

Figure 6: PTU and tutntable positions and their sum total for the experiment in Figure 7 (left). In clinical neurological experiments, Morasso recorded human head H, eye e and sum $E = H + e$ positions when a moving target is tracked [8]. His plot (right) was extracted from Carpenter's text p. 33 [2]. Both plots show large motions (turntable versus head) ramp slowly during which time there is a rapid fine rotation (PTU versus eye). Once the large motion reaches a non-zero velocity, there is a rapid counter-rotation.

Figure 7: A simple cross image was mounted on the end effector of a Mitsubishi PA-10 robot arm, The top row reveals the PTU panning first (left). Due to partitioning, the turntable is kinematically servoed simultaneously (middle). Finally, as the target returns to its initial position, the PTU and turntable also reach their home positions. The bottom row is the camera's point-of-view. Partitioning attempts to keep the target centered in the field-of-view. Lag is due to the dynamics of the actuators.

[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] 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.

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

[6] Hashimoto, K.; Nagahama, K.; Noritsugu, T.; "A mode switching estimator for visual servoing", *Int Conf Robotics and Automation (ICRA)*, V2, pp. 1610 -1615, 2002

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

[8] Morasso, P., Bizzi, E., Dichgans, J., "Adjustment of Saccadic Characteristics During Head Movements", *Experimental Brain Research*, V16, pp. 492-500, 1973.

[9] Nordlund.P , Uhlin. T., "Closing the Loop: Pursuing a moving object by a moving observer", *Proc Int conf Computer Analysis of Images and Patterns CAIP*, Springer LNCS V970, pp. 400-407, 1995.

[10] Oh, P.Y., Allen, P.K., "Visual Servoing by Partitioning Degrees of Freedom," *IEEE Transactions on Robotics and Automation*, V17, N1, pp. 1-17, February 2001

[11] Oh, P.Y., "A Biologically Inspired Approach to Visually Servoed Tracking", *CISST Motion Tracking and Object Recognition Conference*, Las Vegas, June 2002.

[12] Stanciu, R., Oh, P.Y., "Designing Visually Servoed Tracking to Augment Camera Teleoperators", *IEEE Conference on Intelligent Robotics and Systems (IROS)*, Lausanne Switzerland, Oct. 2002.

[13] Papanikolopoulous, N.P., et al., "Visual Tracking of a Moving Target by a Camera Mounted on a Robot: A Combination of Control and Vision," *IEEE Trans Automatic Control* V9 N1 pp. 14-35, Feb 1993.

[14] Sharkey, P.M., Murray, S.V., et al., "A Modular Head/Eye Platform for Real-Time Reactive Vision," *Mechatronics Journal*, V3 N4 pp. 517-535, 1993.