Human-in-the-Loop Visually Servoed Tracking

Rares I. Stanciu and Paul Y. Oh

Drexel University, Dept. of Mechanical Engineering, Philadelphia PA USA
Email: ris22@drexel.edu and paul@coe.drexel.edu

Abstract

There are many features to take into consideration when designing servoed vision systems especially when redundant degrees-of-freedom (DOF) are present. Motion platforms mounted with camera systems usually have multiple joints. Example platforms include rovers, booms, gantries, aircrafts and submersibles. Teleoperating such systems to track moving objects is particularly challenging. The operator is part of the feedback loop and must the associated dynamics and delays must be taken into consideration. Together DOF redundancy one must resolve any potential motion conflicts arising from the shared man-machine control. This paper identifies such dynamics and designs an appropriate control system that leverages redundant DOF in the visual-servoing loop. A simulation and several experiments were performed to assess its performance.

Keywords: visual-servoing, tracking, redundancy, degrees-of-freedom

1 Introduction

The term *human-in-the-loop* refers to systems where an operator controls a device with a desired task. The operator acts on the device depending on information received from it and the environment. Some of these devices like rovers, gantries, aircrafts or submersibles, possess a video camera. The task is to maneuver the camera to obtain desired fields-of-view. Such tasks have applications in areas like broadcasting, inspection and exploration.

These devices 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 a moving subject is a challenging task because it requires a well skilled operator who must manually coordinate multiple joints. Hence, tracking performance is limited by 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*; computer vision is used to control some DOF so that the operator has fewer DOF to manipulate. A description of visual servoing can be found in [3]. An important aspect of visual servoing is the delay which is inherently involved [2]

As hardware we are using a 266MHz PC, an MEI ISA card to drive the pan-tilt's head DC motors, a Cognachrome color tracker and a US Digital ISA board to read the two encoders which retrofits the boom. Our system consist of a four wheeled dolly, a boom, a motorized pan-tilt head and a color camera. The boom pivots on the steerable dolly to sweep the camera horizontally and vertically. The mass of the pan-tilt head, 9.5 kg, and the camera, 1.7 kg, are counterbalanced by a dumbbell of mass 29.5 kg.

Automated image centering with the boom can be achieved by visually servoing the pan-tilt camera [1]. This would reduce the number of DOF the operator must manipulate. The net effect would be *human-in-the-loop visual servoing*. An experiment of tracking a person was performed. A subject was asked to walk back and forth in front of the camera and an operator was booming at 5 degrees/sec. The distance between camera and subject was about 5 meters. Some pictures taken during the experiment are shown in Figure 2. A challenge in [1] was the system's stability, especially when the target and the boom were moving 180 degrees out of phase. The main reason is that the vision system had no information about boom movement. As a result, the system could track a slow moving target rather well, but would be unstable when the target moves quickly.

In this paper Section 2 model the human-controlled boom. Inspired by [4] and dividing the system into two parts, human-in-the-loop and pan-tilt unit, serves to overcome instability challenges and lay the foundation for designing a coupling algorithm in Section 3. By taking advantage of the fact that the motion of the human-in-the-loop is much slower then the motion of the pan-tilt-head, this algorithm is able to counter-rotate the camera and keep a stationary target in camera's field of view when operator booms. Simulations and experimental results as well as conclusions with a map of future work are presented in Sections 4 and 5 respectively.

2 Human-in-the-loop Modeling

Human-in-the-loop systems are characterized by delays, which affect their performance. Sheridan and Farrell [6] describe such parameters that need to be taken into consideration to obtain an accurate model. The first parameter is reaction-time delay t_r also known as the refractory-period. This parameter includes neural synaptic delays, and both nerve conduction and central processing times. This period is about 0.15 seconds. The second parameter is the dimensionless gain K which varies between 2 and 20 at low frequencies. The third parameter is neuromuscular lag. When a muscle is commanded to move, its inherent viscosity and inertia, combined with the asynchrony of the fiber contraction, might be expected to result in an exponential response. The muscle

moves with a time constant t_n of 0.1 to 0.2 seconds. Combining these parameters, a transfer function model of the human operated boom is

$$Y_H(s) = \frac{F}{\dot{\theta}_{REF} - \dot{\theta}} = \frac{K \cdot e^{-s \cdot t_r}}{1 + s \cdot t_n} \tag{1}$$

The input is the difference between the reference and actual boom's angular velocity $\dot{\theta}_{REF}$ and $\dot{\theta}$ respectively. The output is the force F applied on the boom by the operator. Assuming no friction and a rigid structure for the boom we can write

$$M(t) = J \cdot \ddot{\theta}(t) \tag{2}$$

where, $\ddot{\theta}$ is the boom angular acceleration, $M(t)$ is the torque acting on the boom by the human operator and J is the moment of inertia for the boom. The transfer function of the boom is then

$$Y_{BA}(s) = \frac{\ddot{\theta}(s)}{M(s)} = \frac{1}{J} \tag{3}$$

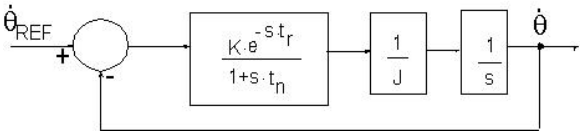

Figure 3: The human-in-the-loop system

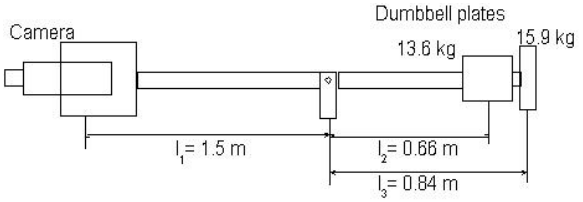

Figure 4: Boom side view

The block diagram of the human-in-the-loop boom is shown in Figure 3. The closed-loop function for this model will then be

$$G(s) = \frac{\dot{\theta}(s)}{\dot{\theta}_{REF}} = \frac{K \cdot e^{-s \cdot t_r}}{K \cdot t_n \cdot s^2 + J \cdot s + K \cdot e^{-s \cdot t_r}} \tag{4}$$

where the input is the reference angular velocity of the boom $\dot{\theta}_{REF}$ and the output is the actual value of it. Replacing $e^{-s \cdot t_r}$ with its Taylor series, and neglecting higher order terms (4) becomes

$$G(s) = \frac{\dot{\theta}(s)}{\dot{\theta}_{REF}} = \frac{K - K \cdot s \cdot t_r}{K \cdot t_n \cdot s^2 + (J - K \cdot t_r) \cdot s + K} \tag{5}$$

In order to derive the moment of inertia of the boom depicted in Figure 1, the moment arm lengths were measured. The results are sketched in Figure 4 and result in the following

$$J = (m_{cam} + m_{PTH}) \cdot l_1^2 + m_2 \cdot l_s^2 + m_3 \cdot l_3^2 \tag{6}$$

where m_{cam} is the mass of the camera, m_{PTH} is the mass of the pan-tilt unit $m_2 = 11.7\ kg$, $m_3 = 15.8\ kg$, $l_1 = 1.5\ m$, $l_2 = 0.66\ m$ and $l_3 = 0.84\ m$. From (6), $J = 65.249\ kg \cdot m^2$. For $t_r = 0.12\ sec$, $t_n = 0.18\ sec$ and $K = 20$, (5) becomes

$$G(s) = \frac{\dot{\theta}(s)}{\dot{\theta}_{REF}} = \frac{20 - 2.4 \cdot s}{3.6 \cdot s^2 + 62.849 \cdot s + 20} \tag{7}$$

To validate this model, a simulation and some experiments were performed. There is good correspondence between the simulated and experimental curves

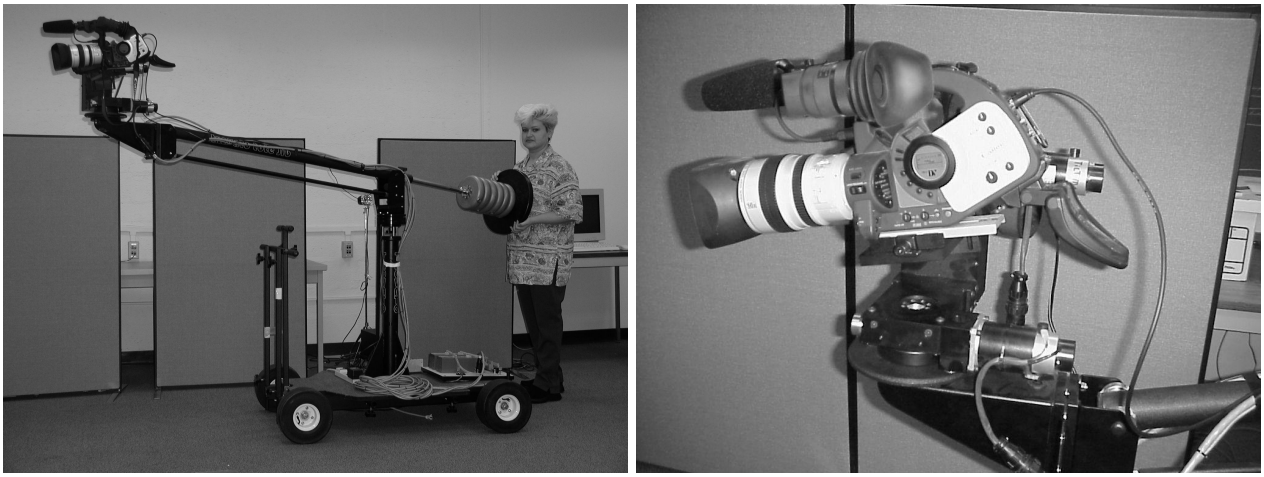

Figure 1: Left: The human operator can boom the arm horizontally and vertically. Right: The 2 DOF motorized pan-tilt head servos the camera.

Figure 2: Three sequential images from videotaping the experiment. Top row: camera field-of-view shows target is tracked. Middle row: boom manually controlled. Bottom row: view from another camcorder

as seen in Figure 5. A step input of 30 degrees/second was given to the system model. The angular velocity of the boom is shown in Figure 5 (bottom). Several people were asked to boom the camera and the angular velocity was recorded. After the transient regime dies out, the angular velocity of the boom was almost constant. This was expected, considering that the heaviest parts are mounted at the end of the boom. Observing that the transfer function poles are $s_1 = -0.324$ and $s_2 = -17.133$ there is a dominant pole. At low bandwidth, the human-in-the-loop boom behaves like a first order system. The operator's closed-loop transfer function in (7) can thus be approximated as

$$G(s) = \frac{e^{-s \cdot \alpha}}{T_0 \cdot s + 1} \tag{8}$$

where $T_0 = 3.1$ seconds and $\alpha = 0.27$ from [5].

3 The Coupling Controller

Previous experiments in [1] underlined a stability challenge because the vision system had no information about booming. To sidestep this and improve performance, a coupling algorithm was designed by taking advantage of the fact that the pan-tilt head moves faster than the boom, approximately 90 deg/sec versus 20 deg/sec respectively. The algorithm's goal is to servo the camera when the target moves or if there is any booming. Towards this, one observes that the operator booms at a velocity given by

$$\dot{\theta}(t) = (1 - e^{-\frac{t-\alpha}{T_0}}) \cdot \dot{\theta}_{REF} \tag{9}$$

where $\dot{\theta}(t)$ is boom's angular velocity, $\dot{\theta}_{REF}$ is the reference angular velocity of the boom, T_0 is human-in-the-loop time constant and α is a specific delay. In Figure 5 there is a delay before the boom actually moves. This is the delay associated with the refractory period discussed in Section 2.

A schematic depicting the top view of the boom is shown in Figure 6. Assuming a stationary target one can calculate the camera velocity needed to compensate for boom rotation.

In Figure 6, L is the distance between the scene and boom's pivot, l_{BA} is the length of the boom, γ is the angle of the pan-tilt head rotates with respect to the boom. Assuming that the operator will boom such that Equation (9) is satisfied, then the value of the pan-tilt head velocity which will compensate for

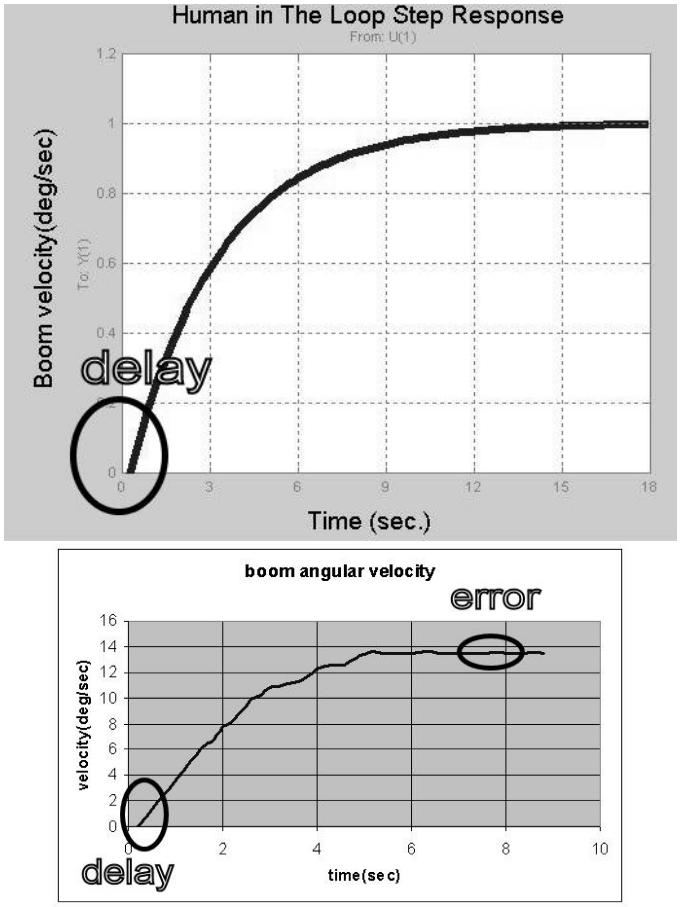

Figure 5: Human-in-the-loop step response simulation (top) and experimental results (bottom). There is a slight delay before motion begins and a 10 percent steady-state error. Experiements show good correlation with simulation.

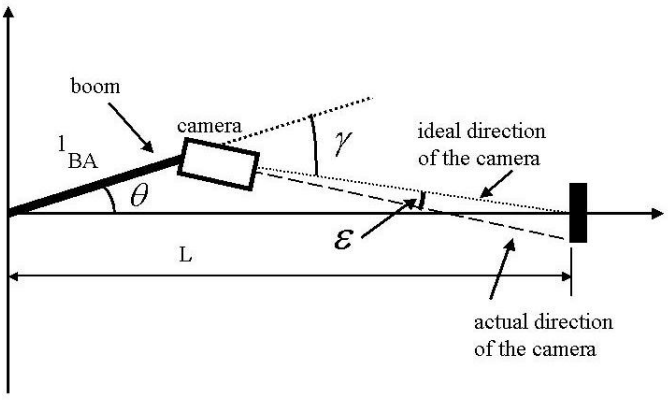

Figure 6: Boom top view

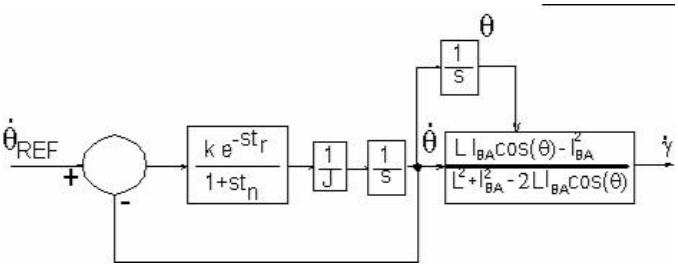

Figure 7: The coupling block diagram. Based on the boom's angular velocity, the camera velocity is calculated

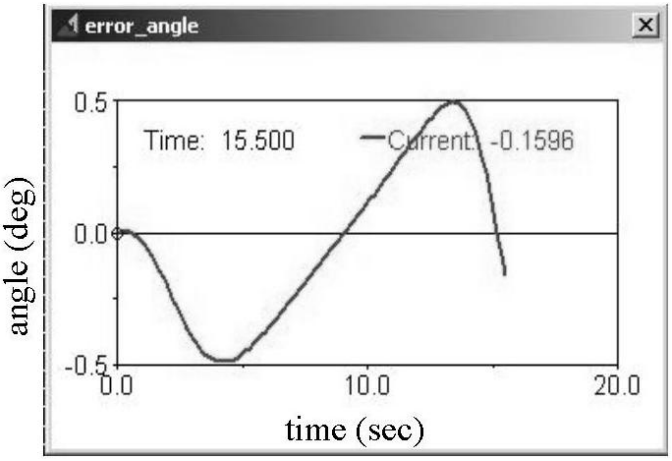

Figure 8: The simulated error angle for a boom complete revolution

booming is given by

$$\dot{\gamma}(t) = \left(\frac{Ll_{BA}\cos(\theta) - l_{BA}^2}{L^2 + l_{BA}^2 - 2Ll_{BA}\cos(\theta)} + 1\right)\dot{\theta}(t) \qquad (10)$$

where θ is the angle of the boom with respect to its initial position. A block diagram of this is shown in Figure 7. Here, the input is the boom's angular velocity (Figure 3) and the output is the camera angular velocity. Figure 10 represents the coupling controller. A dynamic simulation was performed to assess its performances. The angle between the camera's axis and the line between the camera and the target's center (see Figure 6) was referred to as error angle ϵ. The values of this angle were plotted versus time during the simulation of a complete boom revolution. The curve can be seen in Figure 8 and shows a maximum ϵ of 0.5 degrees.

Taking into account that the scene was 4.5 m away from the boom's pivot, 0.5 degree represent an error

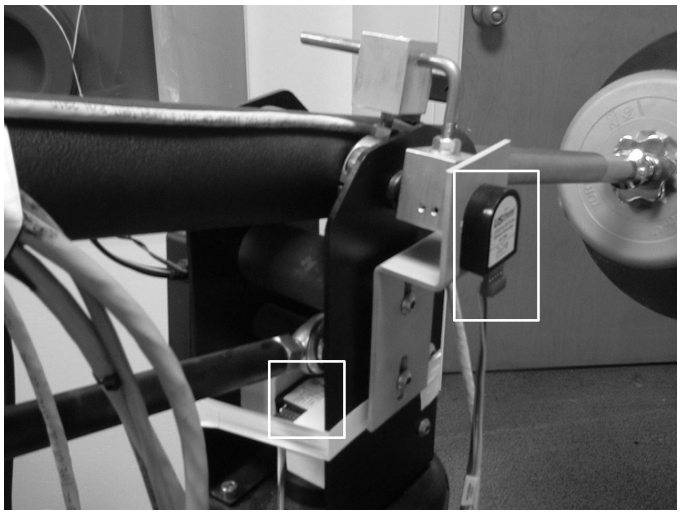

Figure 9: Boom Encoders

of about 15 pixels. This 0.5 degree error can be easy compensated by visual-servoing. To check whether the assumption of booming such that Figure 9 is satisfied, several experiments were performed. The conclusion was that after the transient regime dies, the boom's steady-state angular velocity is constant with an error in the range of 10% (Figure 5 bottom). This value is still too big, especially because this value will directly affect the camera's angular velocity. Another problem was to generate repetitive motion with constant angular velocity. Experiments revealed that this is difficult for operators to do. To implement the controller, the boom was retrofitted with two encoders for panning and tilting as shown in Figure 9.

4 Experimental Results

The coupling algorithm was implemented and several experiments were performed to assess its performances. The program reads the two encoders and computes the new reference velocity of the pan-tilt camera. The operator was booming while the target, located at 4.5 m, was stationary. Pan and tilt angles as well as pan and tilt errors were recorded during booming. The plots are shown in Figure 10. It can be seen that despite an error of 100 pixels, the target remains in the camera's field-of-view. Delays and the difference in the boom's actually angular velocity and the one predicted by (9) contribute towards the large error. The pan-tilt head will not be able to keep the target in the camera's field-of-view when there are large accelerations. The coupling algorithm was able to maintain the target in the camera's field-

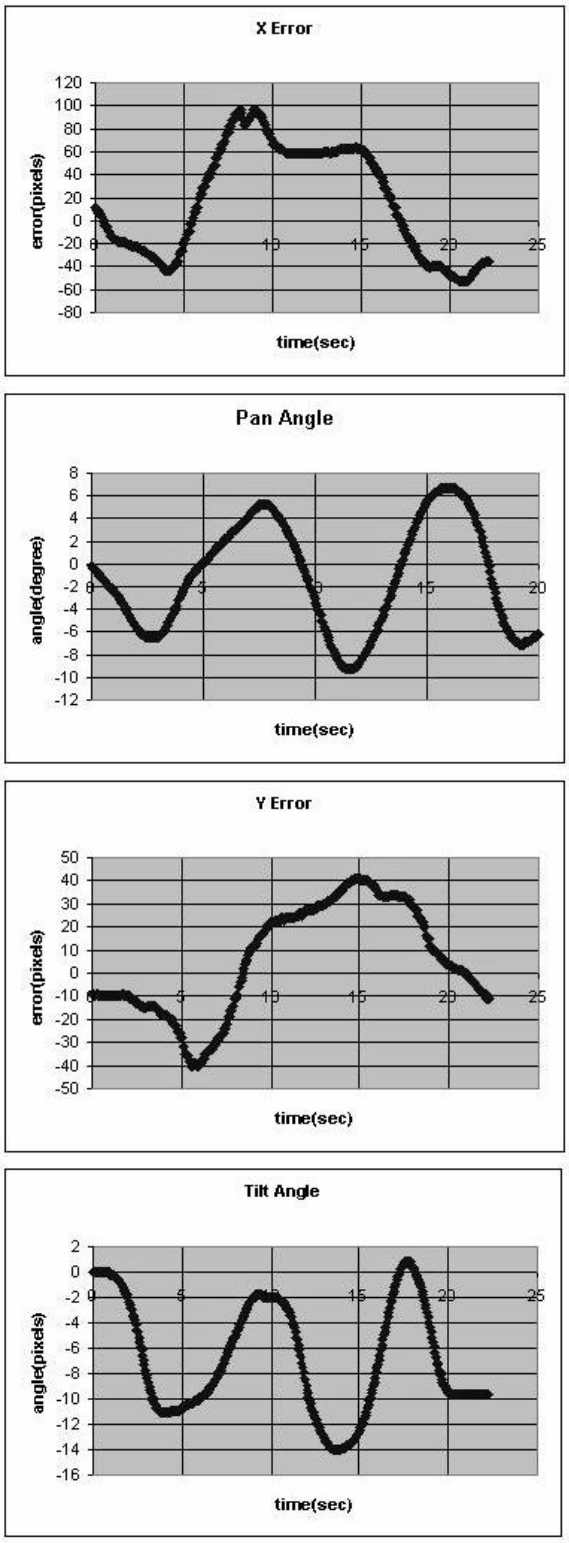

Figure 10: Pan and Tilt errors and angles

of-view when the angular velocity was small (about 5 degrees/sec). It follows that the vision system must compensate for the movement of the target. The desired camera velocity will be generated by the coupling algorithm and image-based visual-servoing.

5 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 motions. A coupling controller was developed and both simulations and experiments were performed to estimate performances. Results are promising under certain speed limits. Relaxing these limits will demand alternative control techniques. Multivariable discrete-time controller may be most suitable.

References

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

[2] Corke P., "Design, Delay and Performance in Gaze Control: Engineering and Biological Approaches" in The Confluence of Vision and Control Springer Verlag, pp 146-158, 1998.

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

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

[5] Canon, D.J., "Experiments With a Target-Threshold Control Theory Model for Deriving Fitts Law Parameters for Human-Machine Systems", *IEEE Transactions on Systems, Man and Cybernetics*, V24 N8, pp. 1089-1098 August 1994.

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