

# A New Generic Model for Vision Based Tracking in Robotics Systems

Yanfei Liu, Adam Hoover, Ian Walker, Ben Judy, Mathew Joseph and Charly Hermanson  
Electrical and Computer Engineering Department

Clemson University  
Clemson, SC 29634-0915

{lyanfei, ahoover, iwalker, mathewj, bjjudy, cspaint} @clemson.edu

**Abstract**— Visual sensing for robotics has been around for decades, but our understanding of a timing model remains crude. By timing model, we refer to the delays (processing lag and motion lag) between “reality” (when a part is sensed), through data processing (the processing of image data to determine part position and orientation), through control (the computation and initiation of robot motion), through “arrival” (when the robot reaches “reality”). In this work we introduce a timing model where sensing and control operate asynchronously. We apply this model to a robotic workcell consisting of a Stäubli RX-130 industrial robot manipulator, a network of six cameras for sensing, and an off-the-shelf Adept MV-19 controller. We demonstrate some experiments to show how the model can be applied.

**Key words:** workcell, timing model, visual servoing

## I. INTRODUCTION

Figure 1 shows the classic structure for a visual servoing system [1]. In this structure, a camera is used in the feedback loop. It provides feedback on the *actual* position of something being controlled, for example a robot. This structure can be applied to a variety of systems, including eye-in-hand systems, part-in-hand systems and mobile robot systems.

In an eye-in-hand system [2], [3], [4], the camera is mounted on the end-effector of a robot and the control is adjusted to obtain the desired appearance of an object or feature in the camera. Gangloff [2] developed a visual servoing system for a 6-DOF manipulator to follow a class of unknown but structured 3-D profiles. Corke [3], [4] presented a visual feedforward controller for an eye-in-hand manipulator to fixate on a ping-pong ball thrown across the system’s field of view.

In a part-in-hand system [5], the camera is fixed in a position to observe a part which is grasped by a robot. The robot is controlled to move the part to some desired position. For example, Stavnitzky [5] built a system to let the robot align a metal part with another fixed part. Since the part is always grasped by the manipulator, we can also say that the part is something being controlled. Or in the other words, we can say that how the object appears in the camera is controlled.

In some mobile robot problems [6], the camera is mounted over an environment to sense the actual position of the mobile robot as feedback to the controller. For example, Kim [6] built a mobile robot system to play

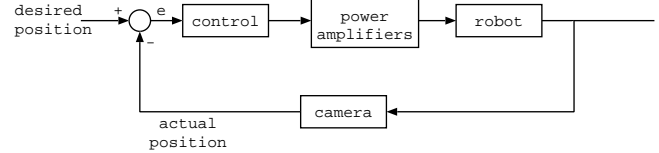


Fig. 1. Classical visual servoing structure.

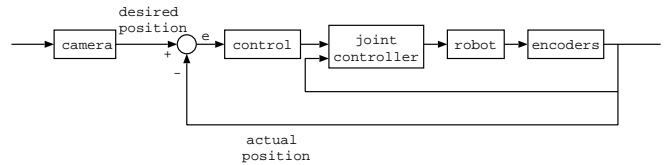


Fig. 2. Vision guided control structure.

soccer. The camera is fixed over the field and acts as a feedback position sensor. Here, the camera observes something which is directly controlled.

All of these systems, regardless of where the camera is mounted, use the camera in the same control structure. In each case the system regulates how the object appears in the camera.

In this paper, we consider the problem where the camera is used to provide the *desired* or *reference* position to the internal robot controller. Figure 2 shows the structure for this system. There are several types of problems that fit this kind of system, where the object of interest cannot be controlled directly. For example, imagine a robot trying to pick up live chickens, or a robot trying to manipulate parts hanging on a swaying chain conveyor. Similar problems have been investigated in some works. Allen [7] demonstrated a PUMA-560 tracking and grasping a moving model train which moved around a circular railway. Nakai [8] developed a robot system to play volleyball with human beings. From the above systems, we notice that the motion of the object was limited to a known class of trajectories. In this paper, we seek to extend this to let the robot follow an unstructured (completely unknown) trajectory. This will be enabled in part by providing a generic timing model for this kind of system.

Our timing model considers the problem where image processing and control happen asynchronously. We are

work	image processing rate (HZ)	control rate (HZ)	processing lag (ms)	motion lag (ms)
Corke and Good [3], [4]	50	70	48	–
Stavnitzy and Capson [5]	30	1000	–	–
Kim et. al. [6]	30	30	90	–
Allen et. al. [7]	10	50	100	–
Nakai et. al. [8]	60	500	–	–
this work	23	250	151	130

TABLE I  
SUMMARY OF RELATED WORK

faced with the following three problems:

- 1) The maximum possible rate for complex visual sensing and processing is much slower than the minimum required rate for mechanical control.
- 2) The time required for visual processing introduces a significant lag between when reality is sensed and when the visual understanding of that reality (e.g. image tracking result) is available. We call this the *processing lag*.
- 3) The slow rate of update for visual feedback results in larger desired motions between updates, producing a lag in when the mechanical system completes the desired motion. We call this the *motion lag*.

Table I presents a summary of how previous works have featured and addressed these three problems. From this table we note that the first two of the three problems have been addressed to some extent in previous works. However, no work appears to have explicitly considered problem #3. All of these works neglect the motion time (motion lag) of the robot. One work [7] noted this problem and used an  $\alpha - \beta - \gamma$  predictor to compensate for it instead of explicitly modeling it. None of these works has considered the generic modeling of this type of system.

Some works synchronized the image processing rate and control frequency for a more traditional solution. In [2], the frequency of visual sensing, processing and control were all set to 50 HZ. Basically, the control frequency was synchronized to the image processing rate for simplicity. Simulation results of high frequency control, i.e. 500 HZ, were also shown in [2]. Performance of the high frequency controller was, as expected, better than the low frequency version motivating a more thorough investigation of a generic timing model to solve the problem. Corke [4] and Kim [6] presented timing diagrams to describe the time delay. Based on the timing diagrams, these works tried to use discrete time models to model the systems. In order to do this, the authors simplify these asynchronous systems to single-rate systems. It is well known that the discrete time model can only be applied into single-rate systems or systems where the control rate and the vision sensing rate are very close. However, from Table I, we notice that most

real systems do not satisfy this condition. Therefore, in this paper we propose a continuous generic timing model to describe asynchronous vision-based control systems.

The remainder of this paper is organized as follows. In Section II, we describe our generic timing model, and then apply this model to an industrial robot testbed that uses a network of cameras to track objects in its workcell. In Section III, We demonstrate the importance of the application of our model by using it to derive a “lunge” expression that lets the robot intercept an object moving in an unknown trajectory. Finally, we conclude the paper in Section IV.

## II. METHODS

Figure 3 illustrates our timing model. From top to bottom, each line shows a component of the system in process order (e.g. sensing comes before image processing). The horizontal axis represents time. We use this model to quantify the processing lag and motion lag of the system. The processing lag is the time between reality (e.g. the actual position of an object) and when an estimate of that reality is available (e.g. the tracking result from processing the image of reality). Similarly, the motion lag is the time between when the control command is issued and when the mechanical system finishes the motion.

The sensing and control processes operate at fixed intervals  $\Delta s$  and  $\Delta c$ , where  $\Delta s > \Delta c$  (sensing is slower than control). The time required for all image processing and tracking operations is designated  $\Delta u$ . This processing starts when an input buffer is filled with image data (on a clock or sync signal defined by the sensing line). An input buffer cannot be filled with new image data until the processing of the previous image data in that buffer is completed. In Figure 3, this is why  $\Delta s_2$  starts on the next sync after the end of  $\Delta u_1$ .

Figure 3 depicts the case where  $\Delta u > \Delta s$  (the processing takes longer than the sensing interval) and when there is only one image buffer. Figure 4 depicts the case where two input buffers are used (commonly called “double buffering”). In this case, a second image buffer is being filled while the image data in the first buffer is

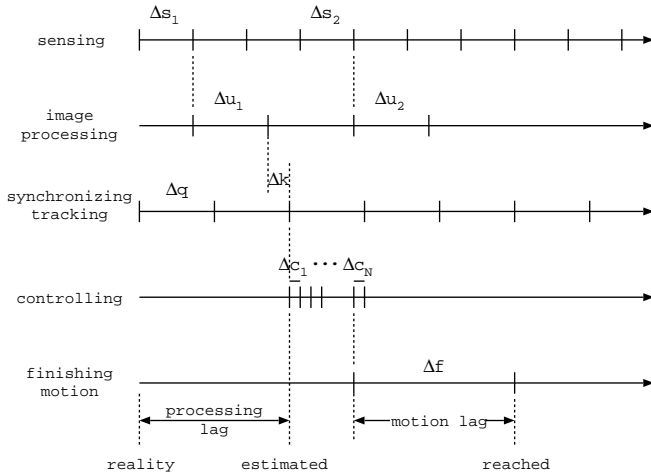


Fig. 3. Timing model for estimating the lag and latency.

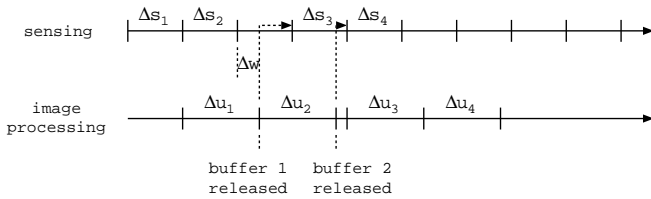


Fig. 4. Timing model using double buffering for processing image data.

being processed. Double buffering increases the lag (note the extra time  $\Delta w$  between the end of  $\Delta s_2$  and the start of  $\Delta u_2$ ) but increases the throughput (note the greater number of images processed in Figure 4 as compared to Figure 3).

In all of these cases, we have assumed that the processing takes longer than the sensing interval ( $\Delta u > \Delta s$ ). In the case where  $\Delta u < \Delta s$  (the processing is faster than the sensing rate), double buffering makes it possible to process every image. In any case, there is always a minimum lag of  $\Delta s + \Delta u$ , but depending on the buffering used and the relation of  $\Delta s$  to  $\Delta u$  the lag can be larger.

In order to handle all of these cases, we introduce a synchronous tracking process (line 3 in Figure 3) operating at a rate of  $\Delta q$ . The tracking line takes the most recent result from the image processing line and updates it for any additional delay (depicted as  $\Delta k$ ) using a Kalman filter. In general, we desire  $\Delta q \approx \Delta u$  so that tracking is updated approximately as fast as new results become available. A secondary benefit of the synchronous tracking line is that it satisfies standard control algorithm requirements that assume synchronous input data. Without this line, the results from image processing can arrive asynchronously (as in Figures 4).

The fourth and fifth lines in Figure 3 represent the control process and completion of motion. We consider



Fig. 5. Our prototype dynamic workcell.

the case where the distance traveled by an object between tracking updates is larger than a robot could safely or smoothly move during a single iteration of control. The control is therefore broken up into a series of  $N$  sub-motion commands occurring at a rate of  $\Delta c$ . Additionally, we expect the motion requested by any new iteration of control to take  $\Delta f$  time to complete. Figure 3 depicts the case where control commands are cumulative (each new control command is relative to the last commanded goal). In Section II-A we describe our prototype, which uses an off-the-shelf Adept controller that operates in this manner. For this controller, the motion is completed some time  $\Delta f$  after the last given control command.

Once values are known for the variables  $\Delta s$ ,  $\Delta u$ ,  $\Delta q$ ,  $\Delta c$  and  $\Delta f$ , it is possible to derive various expressions for controlling a robot to solve specific problems, for example, to intercept a moving object. In the next section, we describe our prototype workcell and derivation of the timing variables. In Section III, we derive an expression for implementing a “lunge” of the robot to intercept an object moving with an a priori unknown trajectory.

#### A. Prototype

Figure 5 shows a picture of our prototype workcell for this project. We use a Stäubli RX130 manipulator with its conventional controller, the Adept Corporation model MV-19. A network of six cameras surrounds the workcell, placed on a cube of aluminum framing. In [9], we detailed the workcell configuration, calibration, image differencing and real-time robot motion planning. In [10], we presented some tracking experiments to show that our system can track different kinds of objects using continuous visual sensing.

Figure 6 shows a timing diagram describing the flow of information through the system, along with how long each step takes. Some of these estimates were derived analytically from knowledge of the hardware, while other

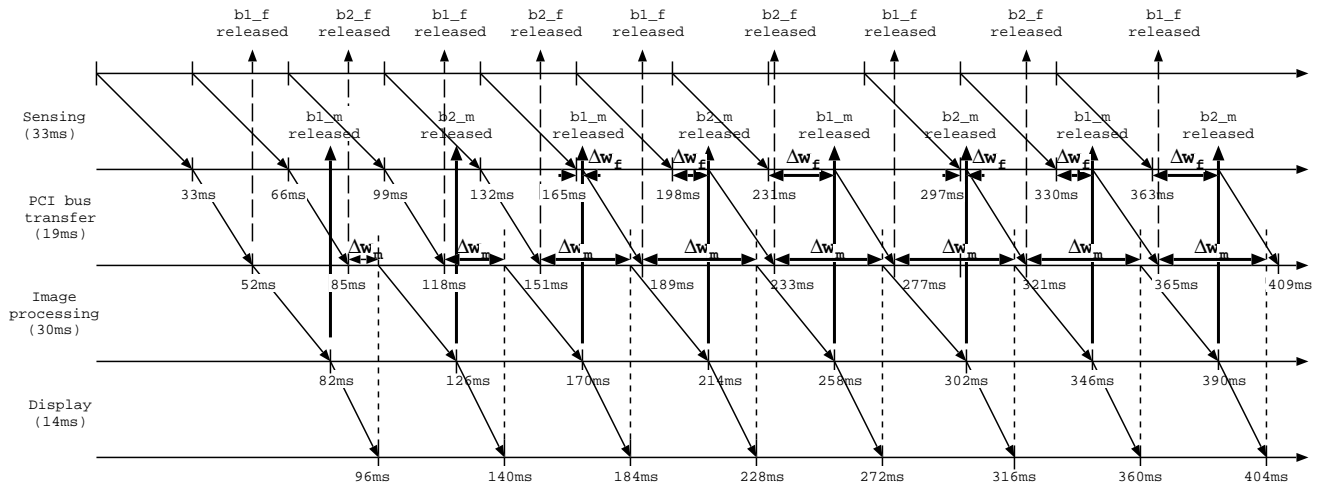


Fig. 6. Information flow through system.

estimates were derived from measurements taken while the system was operating. We will discuss each part in detail in the following paragraphs.

Figure 6 starts with an event that is happening in real time (e.g. an object moves). The cameras operate at 30 Hz using the standard NTSC (U.S. television) format, so that for this system  $\Delta s = 33$  ms. The reality that is imaged takes 33 ms to transfer from the camera to the framegrabber (an NTSC signal uses its full sampling interval for transmission).

The framegrabbers have microcontrollers that can operate as PCI bus masters, initiating transfers from framegrabber memory to main memory. Assuming negligible traffic on the PCI bus, the time for this transfer can be computed as the total number of image bytes divided by the bandwidth of the bus:  $(640 \times 480 \times 4 \times 2 \text{ bytes}) / (4 \text{ bytes} \times 33 \text{ MHz}) = 19$  ms, where the bandwidth is the theoretical maximum provided by the 32-bit 33 MHz PCI standard.

The image processing portion of our system creates a two-dimensional occupancy map of the space in a horizontal plane of interest in the workcell, locates the centroid of an object in this space, and displays the result on-screen [11]. Based upon empirical measurements of the run-time of the image processing methods on the Compaq, we observed them to take approximately 30 ms on average each iteration. This time can vary by approximately  $\pm 2$  ms depending upon the content of the images. We discuss the variances of our measurements in more detail at the end of this section. We also measured that the image display takes 14 ms on average. Therefore, the total processing time  $\Delta u$  for this system is  $19 + 30 + 14 = 63$  ms.

The images may wait in buffers before being processed, due to our use of double buffering. The waiting time can vary depending on the phasing of  $\Delta s$  and  $\Delta u$  as shown in Figure 6. The buffer in main memory is released right

after the occupancy map is available. We noticed that after several iterations, the waiting time repeats in a pattern. The main memory waiting time  $\Delta w_m$  becomes a constant (39 ms) after 4 iterations. The framegrabber waiting time  $\Delta w_f$  is repeating in 3 numbers, 5 ms, 16 ms, 27 ms. So we take the average waiting time  $\overline{\Delta w_f}$  as  $(5 + 16 + 27) / 3 = 16$  ms. Thus we can get the total average waiting time  $\overline{\Delta w} = 39 + 16 = 55$  ms. Adding up the appropriate terms ( $\Delta s + \Delta u + \overline{\Delta w}$ ), the processing lag ( $\Delta l$ ) for this system is  $33 + 63 + 55 = 151$  ms.

The process that synchronizes the tracking result uses a Kalman filter to update the actual  $\Delta k$  for the given iteration. This position is sent through a 10 Mbit ethernet link from the Compaq to the Adept. Based on empirical measurements, these operations were observed to collectively take less than 1 ms. For this system we set  $\Delta q = 40$  ms which is near but slightly under the occupancy map time plus the image display time ( $30 + 14$  ms).

At this point the Adept (robot) has a new goal. This goal is directly forwarded to the motor-level controller through the "Alter" command [9]. According to the manufacturer (Adept), the maximum issue rate for the Alter command is 500 Hz (once every 2 ms), but through experimentation we observed that this rate could not always be maintained. Therefore we set  $\Delta c = 4$  ms, issuing a new Alter every 4 ms. The precise details of the motor-level controller are proprietary to Adept Corporation and could not be determined. Therefore we determined  $\Delta f$  empirically through repeated measurements of the time it took to complete an Alter command. We varied the distance moved, direction moved, and initial condition (starting with the robot in motion at varying velocities, including at rest). Across all cases the time to complete an Alter was observed to vary from 120 to 150 ms, with a commonly occurring median value of 130 ms. We therefore set  $\Delta f$  to be 130 ms.

In our model we assume constants for most of the lag terms. It is important to note that all of these terms have variances, some of them having appreciable size in this context (more than 1 ms). For the sensing and processing terms, it would be ideal to timestamp each image upon acquisition and measure the terms precisely. However, in order to solve problems involving estimates into the future (for example to plan a catch of an object) it is necessary to have averages. Therefore we only note the variances here, and leave a more thorough understanding of their effect to future work.

### III. EXPERIMENTS

In order to test our methods, we experiment with the problem of catching a moving object. Figure 7 depicts the scenario. The object is moving at an unknown constant velocity in a straight line. In this example, the object is moving in one dimension; however, we formulate the solution using vectors to indicate that the solution is also applicable to 2D and 3D problems. Due to the processing lag, the most recently measured position of the object is where the object was  $(\Delta l + \Delta k)$  ms previously. We denote this location as  $\vec{x}_{t-\Delta l-\Delta k}$ . The current position of the object, i.e. the time when the robot starts to lunge towards the object, is denoted as  $\vec{x}_t$ . The current velocity of the object is denoted as  $\vec{v}_t$ , and is assumed to be equal to the last measured velocity  $\vec{v}_{t-\Delta l-\Delta k}$ . Therefore, the relationship between  $\vec{x}_t$  and  $\vec{x}_{t-\Delta l-\Delta k}$  is described in the following equation:

$$\vec{x}_t = \vec{x}_{t-\Delta l-\Delta k} + \vec{v}_{t-\Delta l-\Delta k}(\Delta l + \Delta k) \quad (1)$$

It will take some amount of time  $\Delta i$  ms for the robot to reach the point of impact where it will catch the object. We denote the impact location as  $\vec{x}_{t+\Delta i}$ . Therefore, We can describe the intercept problem as the following: if the robot desires to intercept the object at time  $t$  while following the object, how many control commands ( $N$ ) should be issued between time  $t$  and the time when the robot intercepts the object, and what is the constant distance ( $\Delta \vec{d}$ ) that each single control command should move.

Suppose that  $\vec{x}_{t-\Delta q}$  is the position where the robot was last commanded to move to, i.e. the position where the object is at time  $t - \Delta q$ ,  $\vec{n}$  is the number of alters which will be executed, and  $d$  is the maximum distance a single alter can move. The solution now involves two equations:

$$\vec{x}_{t+\Delta i}[i] = \vec{x}_t[i] + \vec{v}_t[i](\Delta c \times (\vec{n}[i] - 1) + \Delta f) \quad (2)$$

$$|\vec{x}_{t+\Delta i}[i] - \vec{x}_{t-\Delta q}[i]| = \vec{n}[i] \times d \quad (3)$$

Combining Equations 2 and 3, based on the assumption that the object does not change velocity direction between

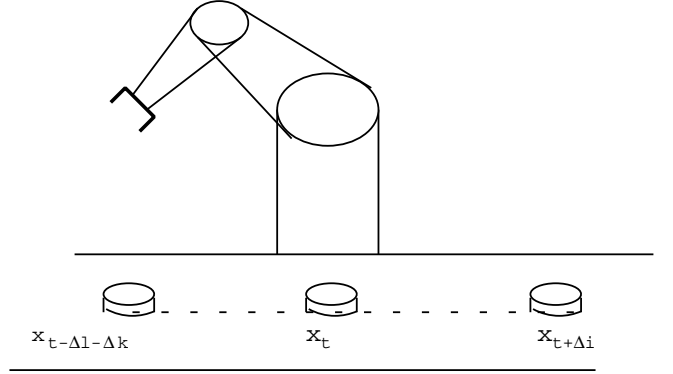
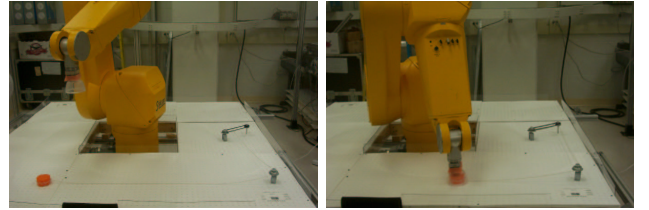


Fig. 7. Scenario for intercepting objects.



(a) initial position

(b) impact scene

Fig. 8. Experimental setup for catching the moving object.

$t \dots t - \Delta q$ , we obtain:

$$\vec{n}[i] = \left\lceil \frac{(\Delta f - \Delta c) |\vec{v}_t[i]| + |\vec{x}_t[i] - \vec{x}_{t-\Delta q}[i]|}{d - \Delta c |\vec{v}_t[i]|} \right\rceil \quad (4)$$

When we solve for  $\vec{n}$ , there is a constraint:

$$|\vec{v}_t[i]| < \frac{d}{\Delta c} \quad (5)$$

Then  $N$  is chosen as the maximum element of vector  $\vec{n}$ . Therefore,

$$\Delta \vec{d}[i] = \frac{\vec{x}_{t+\Delta i}[i] - \vec{x}_{t-\Delta q}[i]}{N} \quad (6)$$

#### A. Experimental setup and results

To verify that our model is effective, we design an experiment to let our industrial manipulator catch a moving object. A small cylindrical object is dragged by a string tied to a belt moving at a constant velocity. The belt is moved by a DC motor. The object moves along a straight line. Figure 8 shows our experimental setup. The robot follows the object with the end-effector pointing straight down approximately 300 mm above the object. When the robot is commanded to intercept the object, the robot will lunge and cover the object on the table with a modified end-effector, a small plastic bowl. The diameter of the object is 70 mm, the diameter for the small bowl is 90 mm. Therefore, the error should be less than 10 mm on each side in order to successfully catch the object.

$\Delta q = 40$			$\Delta q = 80$		
velocity (mm/s)	stdev (mm/s)	catch percentage	velocity (mm/s)	stdev (mm/s)	catch percentage
84.4 – 97.4	1.3 – 3.8	100%	85.9 – 95.1	2.5 – 3.7	100%
129.8 – 146.7	1.7 – 3.2	100%	126.1 – 137.7	1.7 – 3.3	100%
177.6 – 195.1	0.5 – 2.6	100%	175.8 – 192.8	1.1 – 2.7	100%

TABLE II  
EXPERIMENTAL RESULTS FOR CATCHING THE MOVING OBJECT

In order to test the applicability of our timing model, we conducted two sets of experiments. We set  $\Delta q$  to two different values, 40 and 80, in these two sets of experiments. We varied the voltage of the motor driving the conveyor to let the object move at three different velocities. For each velocity, we kept the voltage of the motor constant to make the object move in a relatively fixed velocity, and ran the experiment ten times. Table II shows the results. The velocity column is filled with the range of the average velocity in the ten experiments. The standard deviation column is the range of the standard deviation of the velocity of each experiment. The results indicate that if the object moves at a relatively fixed velocity, the robot catches the object 100% of the time independent of the velocity of the object and the position update time ( $\Delta q$ ).

#### IV. CONCLUSION

In this paper, we present a generic timing model for a robotic system using visual sensing, where the camera provides the *desired* position to the robot controller. We demonstrate how to obtain the values of the parameters in the model, using our dynamic workcell as an example. Finally, we show how this timing model can be used to solve problems, using as an example the problem of our industrial robot intercepting a moving object. The results of the experiments show that our model is highly effective, and generalizable.

#### V. ACKNOWLEDGMENTS

The South Carolina Commission on Higher Education seed funded this project during 2000-2001. The Stäubli Corporation partially donated a state-of-the-art RX-130 industrial manipulator. The U.S. Office of Naval Research currently funds the exploration of this technology to naval warehousing through the Expeditionary Logistics program. We gratefully thank all these organizations.

#### VI. REFERENCES

- [1] S. Hutchinson, D. Hager and P. Corke, "A Tutorial on Visual Servo Control," IEEE Trans. Robotics Automat., vol. 12, no. 5, pp. 651-670, Oct. 1996
- [2] J. Gangloff and M. F. de Mathelin, "Visual Servoing of a 6-DOF Manipulator for Unknown 3-D Profile Following," IEEE Trans. Robotics Automat., vol. 18, no. 4, pp. 511-520, August 2002
- [3] P. Corke and M. Good, "Dynamic Effects in Visual Closed-Loop Systems," IEEE Trans. Robotics Automat., vol. 12, no. 5, pp. 671-683, Oct. 1996
- [4] P. Corke and M. Good, "Dynamic Effects in High-Performance Visual Servoing," in Proc. IEEE Int. Conf. Robotics and Automation, pp. 1838-1843, Nice, France, May 1992
- [5] J. Stavnitzky and D. Capson, "Multiple Camera Model-Based 3-D Visual Servo," IEEE Trans. Robotics Automat., vol. 16, no. 6, pp.732-739, Dec. 2000
- [6] S. H. Kim, J. S. Choi and B. K. Kim, "Visual Servo Control Algorithm for Soccer Robots Considering Time-delay," Intelligent Automation and Soft Computing, Vol. 6, no. 1, pp. 33-43, 2000
- [7] P. Allen, A. Timcenko, B. Yoshimi and P. Michelman, "Automated Tracking and Grasping of a Moving Object with a Robotics Hand-Eye System," IEEE Trans. Robotics Automat., vol. 9, no. 2, pp. 152-165, April 1993
- [8] H. Nakai, Y. Taniguchi, M. Uenohara and T. Yoshimi, "A Volleyball Playing Robot," in Proc. 1998 IEEE Int. Conf. Robotics and Automation, Belgium, pp.1083-1089, May 1998
- [9] Y. Liu, A. Hoover and I. Walker, "Sensor Network Based Workcell for Industrial Robots," in IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 1434-1439, Hawaii, Oct. 2001
- [10] Y. Liu, A. Hoover and I. Walker, "Experiments Using a Sensor Network Based Workcell for Industrial Robots," in Proc. IEEE Int. Conf. Robotics and Automation, pp. 2988-2993, Washington, USA, May 2002
- [11] A. Hoover and B. Olsen, "A Real-Time Occupancy Map from Multiple Video Streams", in Proc. 1999 IEEE Int. Conf. Robotics and Automation, pp. 2261-2266, Detroit, May 1999