Time-delay Analysis of a Robotic Stereo Active Vision System

Gigel Macesanu, Sorin M. Grigorescu, *Member*, *IEEE*,
and Vasile Comnac, *Member*, *IEEE*

Abstract— The time-delay introduced by an image processing system is a crucial factor which plays an important role in the robustness and stability of an active vision system. In this paper, the authors present a feedback control analysis of a stereo active vision platform used to control the pan, tilt and zoom of the active stereo camera. The main goal of the analysis is to determine the maximum time-delay, or delay margin, which can be introduced by the image processing chain without affecting the stability of the overall machine vision system. The proposed procedure considers the investigation of the frequency domain using two indicators of relative stability: gain and phase margin. The obtained delay-margin is used to design the feedback control using the Hermite-Bihler theorem.

I. Introduction

REAL world scenes are usually characterized by objects moving in cluttered environments, which change their *position and orientation* (pose) stochastically. One of the main objectives of a stereo active vision system, with which nowadays robots are equipped, is to track such objects and reconstruct their poses in a virtual 3D space. In order to track these objects, the stereo camera's viewpoint and zoom must adapt along with the poses of the objects.

In many robotics applications, the image acquisition system is used with a constant orientation and zoom, this fact leading to physical limitations in the scene understanding process. In order to enhance the visual perception capabilities of an autonomous robot, the camera's extrinsic parameters (i.e. pose and zoom) must be adapted according to the imaged environment.

In this paper, a time-delay and feedback control analysis of a robotic stereo active vision system is presented. The idea of *visual feedback control*, or *active vision*, has been heavily investigated in the computer vision community [12], [15]. One of the first comprehensive papers on the usage of feedback information at the image processing level can be found in [13], where visual feedback is used to control a robot manipulation task. This process is also encountered under the name of *visual servoing* [3]. Visual servoing is usually divided into two categories [3], [13]: *position* and *image based*. Recent investigations on active vision involve the camera's zoom adaptation, which considers the problem of controling the focal length in order to keep a constant-sized image of an object moving along the camera's optical axes [5], [17]. Also, a multi-view active object search is presented in [16], where multiple objects are detected using

the feedback information provided by a humanoid head.

The main goal of the research presented in this paper is to analytically compute the maximum value of the time delay which can be introduced by the image processing chain within an active vision system, without affecting the stability of active object tracking. Despite its large usage, to the best of our knowledge, there is no known publication that deals with the maximum values of the time delay component introduced by the machine vision software in the robot tracking system. For comparison, in [9], a design and optimization process for a visual control system is proposed for solving problems caused by large delays introduced by image processing algorithms. The closed-loop controller for time-delay systems can be chosen as is explained in [14]. The obtaining controller is tuned so that the closed-loop control system remains stable. In [1], a complete analytical description of all stabilizing parameters for a *Proportional-Integer* (PI) controller used over a communication network link is presented.

The rest of the paper is organized as follows. In Section II, the object recognition and 3D reconstruction methods used by our active vision system are introduced, along with the variables that have to be adjusted by the control system. The mathematical modeling of the considered active vision architecture is given in Section III, followed in Section IV by its stability analysis. Finally, conclusions are stated in Section V.

II. 3D Object Detection and tracking

A. Object Detection

In 3D vision, one of the most important issues to be solved is the distance computation between the imaged object and the stereo camera. This distance is determined through a calibrated stereo camera. The block diagram of the proposed object detection and 3D reconstruction chain is presented in Fig. 1. Firstly, the left and right images are acquired from two *Pan-Tilt-Zoom* (PTZ) cameras which form the hardware of the stereo vision system. Both images are then segmented and the object of interest recognized. Once the corresponding points of the object have been determined in the left and right images, its 3D location can be calculated using the stereo triangulation method [8]. The obtained 3D pose is further used as feedback information to control the orientation and zoom of the two PTZ cameras.

In the 2D image plane, the object of interest is segmented using a robust region-based segmentation method [7], while its recognition is performed using invariant Hu moments that are invariant to object rotation, translation and scaling. Based on the determined 2D image position of the object,
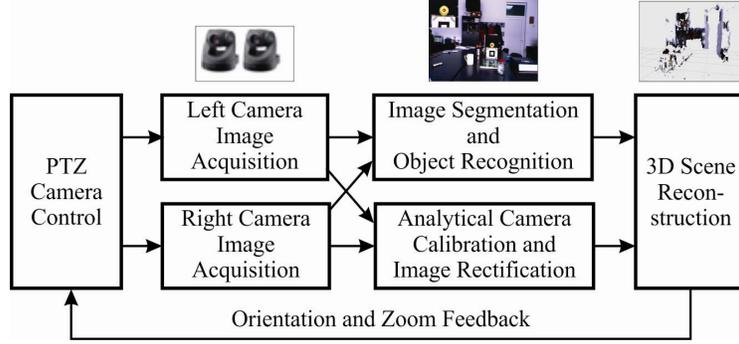
Fig. 1. The block diagram of the image processing chain.

the 2D correspondence points are calculated as the object's center of gravity in both images. Through the 2D positions of the correspondence points and the epipolar geometry constraints the 3D pose of the imaged object can be calculated [8].

The 3D pose of an object can only be obtained from rectified stereo images. Namely, the rectification process transforms each image plain in such a way that the epipolar lines become collinear and parallel to one of the image's axes [6].

B. Control Objectives

The goal of the active vision system is to maintain the image object in the middle of the 2D image plane, as well as within the image boundaries. In surveillance applications, in which wide areas have to be monitored, the camera's *Field of View* (FOV) must be automatically adjusted in order to cover a wider scene as possible. In such an active vision control system, only the adjustment of the pan and tilt of the camera is insufficient, since it cannot cope with objects that are too close or too far from the vision sensor. The problem can be solved by adding an extra *Degree of Freedom* (DoF) to the camera, that is, of the zoom control system, which aims at controlling the focal length through with the environment is sensed.

The errors which have to be compensated by the active vision system are illustrated in Fig. 2, where the position of a real world point, $pt_{int}(x, y, z)$ in the 3D Cartesian space, along with its 2D location in the image plane $pt_{proj}(x, y)$ is
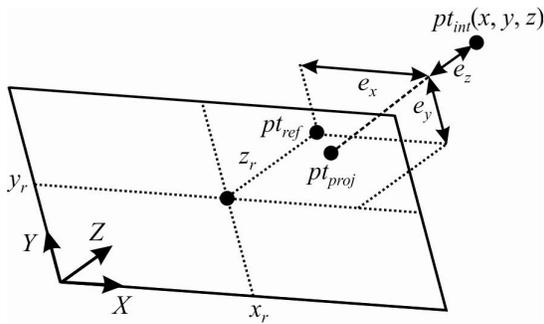

Fig. 2. The position error for a 3D point.

presented. $p_{ref}(x_r, y_r, z_r)$ represents the desired, or *reference*, coordinates of the object of interest in pixels. We consider x_r and y_r as the central 2D coordinates in the left image of the stereo camera system. The third coordinate, that is z_r, is the distance, or depth, from the middle point on the baseline between the two cameras and the imaged object. All the considered measurements are in pixel values. As will be explained in the next section, their conversion to real world meters is performed by the PTZ controllers.

Having in mind the above notations, a position error vector $e = \lfloor e_x, e_y, e_z \rfloor$ can be defined, where the errors are the differences between the position of the object of interest and $p_{ref}(x_r, y_r, z_r)$ along the three considered axes. The goal of the visual controller is to minimize the e_x and e_y errors by adjusting the pan and tilt values of the cameras and e_z by controlling the focal length of the cameras.

The proposed approach aims as imaging the full shape of the object in the middle of the FOV of the stereo camera system, while maintaining its boundaries within the left image. The object boundaries are maintained by the zoom controller, as will be further explained.

III. ACTIVE VISION SYSTEM MODELING

The goal of the proposed active vision system is to control the orientation and zoom of the stereo camera. Since a stereo camera is considered, the control system has to adapt the two PTZ cameras which make up the stereo vision robotic platform, thus ending up with a 6-DoF system (e.g. 2x pan, 2x tilt and 2x zoom). Since, in both cameras, the pan, tilt and zoom are modified using the same time of servo-drive, we consider the modeling and analysis of only a camera's DoF, that is, of the left camera's pan. The control of the other 5 DoF is analogous.

The control system, with the block diagram of a single DoF shown in Fig. 3, consists of two Sony-Evi-D70P® PTZ video cameras. In the diagram, the system is represented in the continuous time domain, being composed of a *Proportional* (P) controller $C(s) = k_c$, with the choice for k_c detailed in next sections, a system element $M(s) = k_m$,
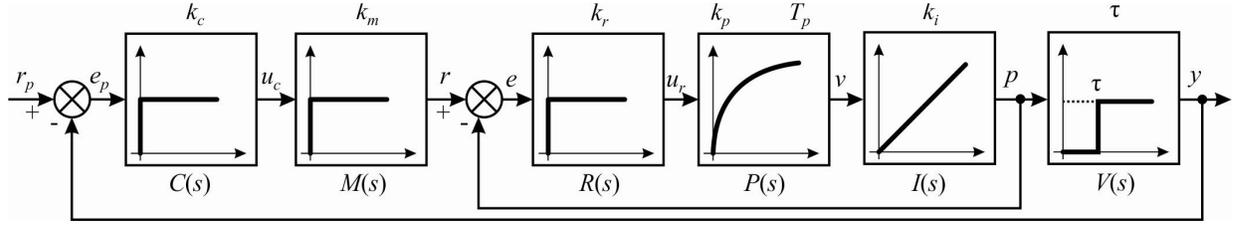
Fig. 3. Block diagram of the pan control system within the proposed active vision architecture.

used to convert the controlled signal u_c from the pixel camera metric to real 3D world coordinates. r_p is the reference position signal, for the pan case, being the central image plane coordinate.

The input signal e_p represents the object's 3D position error with respect to the camera coordinate's system, while r is the reference position, in real world coordinates. $R(s) = k_r$, is the controller transfer function of the inner position control loop. The process, or servo-drive, is modeled as a first order lag element $P(s) = k_p / (1 + T_p s)$. Physically, the process is a *direct current* (DC) motor, used to adapt the camera's pan. The output v represents the camera velocity. Since the inner loop should control the pan position, an extra integrator element, $I(s)$, is added to integrate the position v over time. The signal e represents the error between the current pan orientation of the camera and the position along the x image axis of the object of interest. The goal of the closed-loop system is to maintain the x axis Cartesian error at zero. The last block shown in the diagram from Fig. 3 is represented by the image processing chain $V(s)$. Mathematically, the image processing system can only be expressed as a time-delay transfer function, written as $V(s) = e^{-s\tau}$ [3]. The delay τ represents the time needed to process a pair of images in order to extract the 3D pose of the object of interest.

Having in mind the above explanations, the transfer function of the inner-loop can be express as:

$$G_{il}(s) = \frac{R(s) \cdot P(s) \cdot I(s)}{1 + R(s) \cdot P(s) \cdot I(s)} =$$
$$= \frac{k_r k_p k_i}{s(s \cdot T_p + 1) + k_r k_p k_i}, \qquad (1)$$

where the parameters values have been determined using the Matlab® identification toolbox. The identification process was needed because of the lack of an analytical process model. The model for the internal camera was chosen as in [4]. The transfer function parameters are estimated based on these two known parameters: the input reference signal r and the output camera position p. Using the reference position, considered as a step signal, we have evaluated the output camera pan positions. Using Eq. (1) as process model, the

obtained plant's transfer function has the gain product $k_r \cdot k_p \cdot k_i = 1.74$ and the time constant $T_p = 0.84$ sec.

From Eq. (1), the open-loop transfer function of the entire system can be written as:

$$G_{ol}(s) = C(s) \cdot M(s) \cdot G_{il}(s) \cdot V(s), \qquad (2)$$

or:

$$G_{ol}(s) = k_c \cdot k_m \cdot G_{il}(s) \cdot e^{-s\tau}, \qquad (3)$$

where $k_m = 1.37$. The numerical domain values for k_c are obtained using Theorem 1 (see Appendix).

Although the camera model, $G_{il}(s)$ and the $M(s)$ elements are linear, the time delay introduced by the image processing chain changes the system into a nonlinear one. The process modeled by $G_{ol}(s)$ is time-delay dependent, since it is always influenced by the processing time needed by the vision component.

IV. STABILITY ANALYSIS

A. Problem Formulation

To evaluate the stability of the proposed control system we have used the *Nyquist criterion* with the *gain margin* K_g and *phase margin* γ for the frequency response method [4]. The basic idea is to determine the behavior of the open-loop transfer function in the frequency domain, that is $G_{ol}(j\omega)$. Having in mind the Nyquist criterion, the locus of $G_{ol}(j\omega)$ represent a measure of the system's relative stability. The objective of this analysis it to determine the gain and phase margins in order to observe the system's stability reserve.

A system containing a time-delay τ has a stability which can be classified into two possible situations [10]:

1) *delay-independent stability*, if the locus of $G_{ol}(j\omega)$ is on the right side of the $(-1, +j0)$ point, for all positive and finite values of the delay, with $\tau \in [0, \infty)$;

2) *delay-dependent stability*, if the locus of $G_{ol}(j\omega)$ is on the right side of the point $(-1, +j0)$ for a finite number of delays $\tau \in [0, \tau_{max})$, while for $\tau \in [\tau_{max}, \infty)$ the locus of $G_{ol}(j\omega)$ is on the left side of the point

$(-1, +j0)$.

The performance of a system which contains time delays, that is stable in closed-loop form $\tau = 0$, can be evaluated using the so-called *delay-margin*. The delay-margin can be defined as be the largest time-delay τ_{max} for which the closed-loop system remains internally stable [2]. Hence, for any $\tau < \tau_{max}$ a closed-loop system is stable. In the next section, we presented the computation of the maximum time-delay τ_{max} which can be introduced by the machine vision component without affecting the stability of the pan control system.

B. Time-delay Stability Analysis

Through the stability analysis, that is delay dependent stability, of the system we aim to determine de value of τ_{max} for which the system remains stable.

The analysis is performed in an open-loop loop manner on the transfer function $G_{ol}(s)$, without considering the time-delay and the system controller, $C(s)$. The system stability with $C(s)$ controller is presented in next section. The goal of the approach is to determine the open-loop stability and to observe the closed-loop system's evolution when a time-delay is introduced. For this purpose, the phase margin and its associated gain crossover frequency ω_c has to be determined [4]. When a time-delay exists within the system, it affects the phase margin and the phase crossover frequency.

From Eq. (1-3) and the model's parameters the delay free system has been represented as a second order element, with its frequency domain transfer function defined as:

$$G_{ol}'(j\omega) = \frac{2.38}{j\omega(0.84 \cdot j\omega + 1) + 1.74}. \tag{4}$$

The element from Eq. (4) is stable, since the gain margin and phase crossover frequency, ω_p, have infinite values [11]:

$$K_g \to \infty, \ \omega_p \to \infty, \tag{5}$$

and, according to the Nyquist stability criterion, the locus of (4) resides on the right side of the critical point $(-1 + j0)$.

The gain crossover frequency is determined using the following equation:

$$\left| G_{ol}'(j\omega_c) \right| = 1. \tag{6}$$

After solving the above equation, a value of $\omega_c = 1.93 \, rad/sec$ is obtained. The phase margin for the time-delay free open-loop transfer function $G_{ol}'(j\omega_c)$ can be written as:

$$\gamma = 180° + \arg G_{ol}'(j\omega_c) = $$
$$= 180° - \text{arctg}(0.48 \cdot \omega_c) - \text{arctg}(0.57 \cdot \omega_c) = 53.84°. \tag{7}$$

The phase margin is the amount of additional phase lag, at the gain crossover frequency, which can bring the system to the stability limit [11]. Hence, the characteristic shows the open-loop system's stability reserves, before the insertion of the time-delay component. Since the time-delay component does not actually modify the gain crossover frequency, the time-delay phase margin can be computed as:

$$\gamma = 180° + \arg G_{ol}'(j\omega_c) - \omega_c \cdot \tau, \tag{8}$$

where, τ represents the time-delay in seconds.

When the system is at stability limit, the time-delay phase margin is zero:

$$\gamma_{TD} = 180° + \arg G_{ol}'(j\omega_c) - \omega_c \cdot \tau = 0. \tag{9}$$

Using the previous expression, we can evaluate the delay margin, that is, the maximum value of the time-delay after which the overall closed-loop system becomes unstable:

$$\tau_{max} = \frac{[180° + \arg G_{cl}'(j\omega_c)]_{rad}}{\omega_c} = \frac{\gamma}{\omega_c} = 0.48 \, sec. \tag{10}$$

The obtained time-delay margin from Eq. (10) indicates that *the image processing chain from Fig. 1 should process a pair of stereo images (e.g. image acquisition, segmentation, object recognition and 3D reconstruction) in less than* $0.48 \, sec$. Larger time-delay values will render the active vision system unstable.

C. Designing the system k_c controller

As before, the stability of the proposed active vision system has been analyzed in the frequency domain. The goal is to determine the delay margin, or the maximum time-delay, which renders the system unstable.

The basic idea here is to consider the transfer function of the open-loop system, containing the time-delay, and to find the $C(s)$ controller parameters which can stabilize the closed-loop process. In this case, the process is described by the following transfer function:

$$G_{ol}(s) = \frac{2.38}{0.84 \cdot s^2 + s + 1.74} \cdot e^{-s\tau} = $$
$$= \frac{2.83}{s^2 + 1.19 \cdot s + 2.07} \cdot e^{-s\tau}. \tag{11}$$

where τ represent the time-delay in seconds. Having in mind [14], we propose a P controller, $C(s) = k_c$, which could stabilize the system for $\tau = \tau_{max}$. The main goal of this method is to analytically determine the region in the k_c

parameter space, using Theorem 1 [14] (see Appendix), for which the closed-loop system remains stable.

Since the considered open-loop system from Eq. (5) is stable [11], using Theorem 1 we can determine the region for the k_c parameter which delivers a stable close-loop system. Firstly, the computation of parameter α is needed:

$$\alpha = \tau\sqrt{a_0 - (a_1^2/2)} = 0.88. \tag{12}$$

Further, the *od* and *ev* parameters need to be determined, as in (A.6) and (A.7) from Theorem 1. In the considered case they have the values $od = 1$ and $ev = 2$. The stable region of the control can be obtained by computing the roots z_l, for $l = 1, 2, 3, 4$, of equation:

$$\cot(z) = \frac{z^2 - 0.47}{0.57 \cdot z}, \tag{13}$$

where $z_1 = 0.9301$, $z_2 = 3.319$, $z_3 = 6.3734$ and $z_4 = 12.5663$. The z_l roots are further used to compute parameter *A* (see Appendix):

$$A = \frac{1.19 \cdot z_l}{2.83 \cdot 0.48 \cdot \sin(z_l)}, \tag{14}$$

for each z_l previously determined. Using relation (A.1):

$$\max_{l=ev,ev+2}\{A\} < k_c < \min_{l=od,od+2}\{A\}. \tag{15}$$

The stable values of k_c can be obtained:

$$-16.4754 < k_c < 1.0164. \tag{16}$$

In order to check the stability of the obtained interval of k_c parameter, we use a generalized form of the Hermite-Biehler Theorem, as presented in [1] and [14]. A random value of the controller's gain has been chosen in the determined interval, such as $k_c = 0.2$. The characteristic quasi-polynomial $\delta^*(s)$ of the system is given by [14]:

$$\delta^*(s) = 0.566 + (s^2 + 1.19 \cdot s + 2.07) \cdot e^{0.48 \cdot s}. \tag{17}$$

Substituting $s = j\omega$ we end up with:

$$\delta^*(j\omega) =$$
$$= [0.56 + (2.07 - \omega^2)\cos(0.48\omega) - 1.19\omega\sin(0.48\omega)] + \quad (18)$$
$$+ j[(2.07 - \omega^2)\sin(0.48\omega) + 1.19\omega\cos(0.48\omega)].$$

In Fig. 4 the behavior of the real and the imaginary parts of $\delta^*(j\omega)$ is presented. As can be seen, the conditions of

Hermite-Biehler Theorem [14] are satisfied, since the roots of the real and imaginary parts interlace.
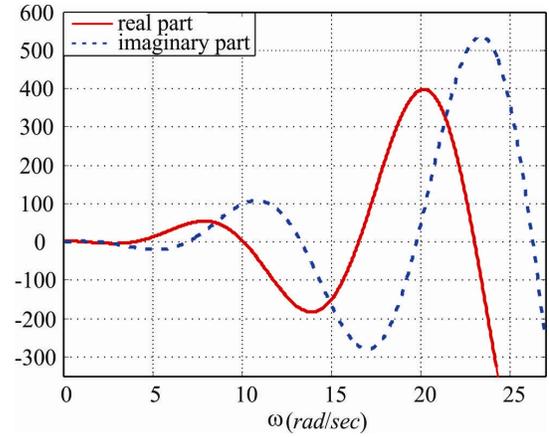

Fig. 4. Plot of the real and imaginary parts of $\delta^*(j\omega)$.

V. CONCLUSION

In this paper, the analysis of an active vision system was presented. The main goal was to determine the maximum value of the time delay component introduced by the image processing system (so called performance characteristic: delay margin) and to analytically calculate the parameters of the closed-loop controller for the considered system. The performance of the considered active vision model was studied using the frequency characteristics, phase margin and gain margin. Also, we use the generalized theorem of Hermite-Biehler to study the proposed closed-loop system. As future work, the authors consider the overall modeling of the visual understanding process, taking into account also the dynamics of the image processing algorithms..

APPENDIX

The Theorem 1 uses a constant gain feedback controller for stabilizing time-delay systems. Consider systems with a time delay that can be mathematically described by the transfer function:

$$G(s) = \frac{k}{s^2 + a_1 s + a_0} \cdot e^{-Ls}. \tag{A.1}$$

Theorem 1: The set of all stabilizing gains k_c for a given open-loop stable plant with a transfer function $G(s)$ as in (A.1) is given by:

1) If $a_1^2 \geq 2a_0$, then

$$-\frac{a_0}{k} < k_c < \frac{a_1 \cdot z_1}{k \cdot L \cdot \sin(z_1)}, \tag{A.2}$$

where, z_1 is the solution of equation

$$\cot(z) = \frac{z^2 - \tau^2 \cdot a_0}{\tau \cdot a_1 \cdot z}, \tag{A.3}$$

in the interval $(0, \pi)$.

2) If $a_1^2 < 2a_0$, then

$$\max_{l=ev,ev+2} \{A\} < k_c < \min_{l=od,od+2} \{A\}, \tag{A.4}$$

where, $A = \dfrac{a_1 \cdot z_l}{k \cdot \tau \cdot \sin(z_l)}$ and z_l is the solution of the equation:

$$\cot(z) = \frac{z^2 - L^2 \cdot a_0}{L \cdot a_1 \cdot z}, \tag{A.5}$$

in the interval $((l-1)\pi, l\pi)$; $\alpha = \tau\sqrt{a_0 - (a_1^2/2)}$; od represent an odd number, defined as:

$$od = \arg\min_{l=1,3,5,\dots} \{\alpha - z_l\}, \text{ subject to } \alpha - z_l \geq 0, \tag{A.6}$$

and ev represent an even number, defined as:

$$ev = \arg\min_{l=0,2,4,\dots} \{\alpha - z_l\}, \text{ subject to } \alpha - z_l \geq 0. \tag{A.7}$$

REFERENCES

[1] N. A. Almutairi, M. Y. Crow, Stabilization of Networked PI Control System Using Fuzzy Logic Modulation, In *Proc. of the American Control Conference*, Colorado, USA, 2003, pp. 975-980.

[2] V. D. Blondel, A. Megretski, *Unsolved Problems in Mathematical Systems and Control Theory*, Princeton University Press, New Jersey, USA, 2004

[3] P.I. Corke, *Visual Control of Robots: A High-performance Visual Servoing*, Research Studies Press Ltd., John Wiley & Sons Inc, Great Britain, 1996.

[4] R. C. Dorf, R. H. Bishop, *Modern Control Systems*, 12th ed., Prentice Hall, New Jersey, USA, 2010.

[5] J. A. Fayman, O. Sudarsky, E.Rivlin, Michael Rudzsky, "Zoom Tracking and its Applications", In *Machine Vision and Applications*, Springer, vol. 13, no. 1, 2001, pp. 25-37.

[6] A. Fusiello, E. Trucco, A. Verri, "A Compact Algorithm for Rectification of Stereo Pairs", In *Machine Vision and Applications*, vol. 12, no.1, 2000, pp. 16-22.

[7] S.M., Grigorescu, "Robust Machine Vision for Service Robotics", Ph.D. Thesis, Faculty of Physics and Electrical Engineering, University of Bremen, Bremen, Germany, 2010.

[8] R., Hartley, A., Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge Univ. Press, Cambridge, UK, 2000.

[9] M. B. Khamesee, M. Rapolti, A. Amirfazli, "Controller Design and Optimization for Large-delays Image Processing in Visual Closed-loop Systems", In *Mechatronics*, Elsevier, vol. 18, 2008, pp. 251-261.

[10] T. Mori, "Criteria for Asymptotical Stability of Linear Time-delay Systems", In *IEEE Trans Automatic Control*, vol. 30, 1985, pp. 158-160.

[11] K. Ogata, *Modern Control Engineering*, 4th ed., Prentice Hall, New Jersey, USA, 2002.

[12] H. Rotstein, E. Rivlin, "Control of a Camera for Active Vision: Foveated Vision, Smooth Tracking and Saccade", In *Proc. of the 1996 IEEE Int. Conf. on Control Applications*, Dearborn, USA, 1996, pp. 691-696,

[13] R. Sharma, "Role of Active Vision in Optimizing Visual Feedback for Robot Control", In *Collection: The confluence of vision and control*, Springer, 1998.

[14] G. J. Silva, A. Datta, S.P. Bhattachaiyya, *PID Controllers for Time-Delay Systems*, Birkhauser, Boston, USA, 2005.

[15] V. Sundareswaran, P. Bouthemy, F. Chaumette, "Active camera self-orientation using dynamic image parameters", In *Proc. of the Third European Conference on Computer Vision*, Stockholm, Sweden, 1994, pp. 110-116.

[16] K. Welke, T. Asfour, Rudiger Dillmann, "Active Multi-View Object Search on a Humanoid Head", In *IEEE Inter. Conf. on Robotics and Automation* (ICRA), Kobe, Japan, 2009, pp. 417-423.

[17] Y. Yao, B. Abidi, M. Abidi, "3D Target Scale Estimation and Target Feature Separation for Size Preserving Tracking in PTZ Video", In *Int. Journal of Computer Vision*, Springer, vol. 82, 2009, pp. 244-263.