FEEDBACK CONTROL ANALYSIS OF A STEREO ACTIVE VISION SYSTEM

G. MĂCEŞANU[1]   S.M. GRIGORESCU[1]

Abstract: *The position and orientation estimation of a stereo camera system is of key importance in 3D scene reconstruction. In this paper, the authors present a feedback analysis of an active vision system used to control the pan, tilt and zoom of an active stereo camera. The proposed procedure considered both a time and frequency domain analysis. The main outcome of the research presented here is the analytical computation of the maximum dead-time value which the vision system can introduce.*

Key words: *active vision, dead-time, digital image processing.*

1. Introduction

In many real world applications, video acquisition systems are placed in constant poses (*position and orientation*), this fact leading to physical limitation in the scene understanding process. To enhance the visual analysis procedure of a machine vision platform, the camera parameters (i.e. pose and zoom) must be able to adapt according to the environment [9].

Real world scenes are usually dominated by moving objects that change their pose stochastically. One of the main objectives of active vision system is to track such objects. In order to do this, the camera's viewpoint must adapt along with the poses of the objects. Such a system is traditionally called an active vision system [8], which aims at increasing the performances of the visual understanding architecture.

In this paper, a stability analysis of a stereo visual feedback system is proposed. The basic block diagram of the considered architecture can be seen in Figure 1a. The idea of visual feedback control in active vision systems has been tackled previously in the computer vision community in papers such as [2] or [10]. One of the first comprehensive papers on the usage of feedback information at image processing level can be found in [9], where visual feedback is used to control a robot manipulation task. This process is also encountered under the name of visual servoing [3]. The process of visual servoing is divided into two categories [9], [3]: position-based and image-based. Recent studies on visual feedback control are related to camera zoom adaptation, which consider the adjustment of a camera's focal length in order to keep a constant-sized image of an object moving along the camera's optical axis [4]. Based on the human biological model of neural pathways, in [7] an active vision system able to keep an object fixed during smooth pursuit is presented. A multi-view active vision system is presented in [12], where multiple objects are located using the feedback information provided by a humanoid head.

---

[1] Automation and Information Technology Dept., *Transilvania* University of Braşov.
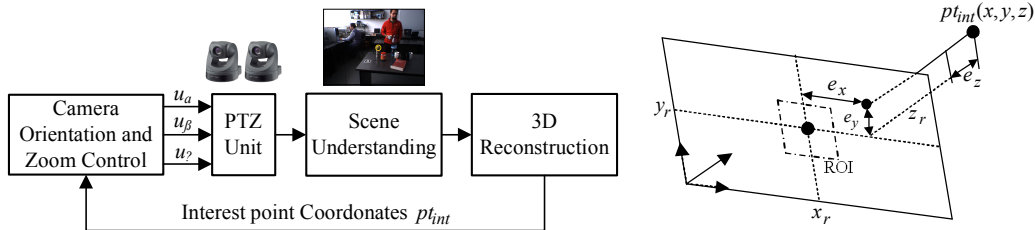
Fig. 1. *3D camera gaze orientation control system:* (*a*) *Visual control structure;*
(*b*) *ROI definition within the 2D image plane*

The main goal of the research presented in this paper is to analytically compute the maximum value of the dead-time which ensures the stability of the overall machine vision architecture used in object tracking.

One other concept used in active vision is the definition of an image Region of Interest (ROI) in which image processing algorithms are to be applied. An advantage of this approach is the reduction of the imaged scene complexity, that is, the reduction of the object's search area from the whole imaged scene to a smaller region containing the object(s) to be recognized. In Figure 1a, the block diagram of the proposed active vision structure used to control the orientation of a stereo camera system is presented. Based on robust feature extraction of boundary and region segmented objects, position of interest points pt_{int} (x, y, z) on the 2D image plane can be calculated. Once the coordinates of pt_{int} (x, y, z) have been determined, the orientation angles of the camera's Pan-Tilt-Head (PTH) can be adapted via the control signals u_α, u_β and u_γ, and, which correspond to the yaw, pitch and zoom of the camera. In Figure 1b, the position of a ROI on the 2D image plane and the position of pt_{int} (x, y, z) in the 3D Cartesian space are illustrated. The goal of the camera orientation controller is to minimize the errors e_x, e_y and e_z in order to superimpose the ROI on the interest point and to obtain a central zoom value, thus ensuring a proper visualization of the object of interest.

This paper is organized as follows. Section II presents the mathematical modelling of the considered active vision system, followed in Section III by its stability analysis. Finally, before conclusions, the implementation of a 3D scene understanding framework, based on the results from Section III, is described.

2. Modelling an Active Vision System

The goal of the visual control system used in this paper is to control a stereo camera, which, as the name suggests, consists of two vision sensors. For the sake of clarity, we consider the feedback control modelling of only a camera sensor, the second one being analogous. The machine vision system used is composed of two digital Sony-Evi-D70P® video cameras. Each of them has an internal module controlling the pan, tilt and zoom. The purpose of this system is to maintain a constant distance between the camera and an interest object and, at the same time, to control the gaze orientation of the system in such a way that the object of interest stays in the central part of the image.

The input control signals for pan-tilt-zoom blocks are extracted from the video analysis block. The output signal of each block (p, t, z) represent the current position of that controlled element. Using, such a configuration and based on the three control signals, a stable object tracking in 3D space can be achieved. In the configuration, a closed-loop is used for

each module (i.e. pan, tilt and zoom). The closed-loop control system is illustrated in Figure 2 using the block diagrams algebra.

The system in Figure 2, represented in the continuous time domain, is composed by a proportional controller, $G(s) = k_p$, used to convert the controlled signal y_0 from the pixel camera metric to real world coordinates. The input signal e_p represents the object 3D position error, with respect to the camera coordinate's system, while r is the reference position, in real world coordinates. e_p is composed of 3 errors, one for each axis of the Cartesian space, that is e_x, e_y and e_z, as can be seen in Figure 1b. $R(s) = K_c(1 + T_c s)/T_c s$, is the controller for the inner position control loop, which includes the positioning process modelled as $P(s) = K_m/(1 + T_m s)$. The input signal e represents the error between the object's of interest current pose and the camera pan-tilt-zoom values. The goal of the closed-loop system is to maintain the Cartesian error at zero. The process is modelled as a PT1 element (the 1st order lag element) and controlled by a PI regulator (Proportional-Integral). Physically, the process is a DC (*Direct Current*) motor, used to control the camera's pan, tilt and zoom. The parameters of the inner control loop were chosen based on the available technical documentation [11]. The final block of the diagram from Figure 2 is represented by the image processing component $V(s)$. Mathematically, the image processing system can be expressed as a time-delay transfer function, represented by $V(s) = e^{-s\tau}$. The delay τ of this function represents the time needed to process the acquired video information in

order to extract the 3D pose of an interest object. The reference signal, r_i, represents the image ROI coordinates and central zoom value.

Using the above description, the transfer function of the inner loop is defined as:

$$G_{cl}(s) = \frac{R(s)P(s)}{1 + R(s)P(s)}$$
$$= \frac{K_c(1 + T_c s/T_c s)(K_m/(1 + T_m s))}{1 + K_c(1 + T_c s/T_c s)(K_m/(1 + T_m s))}. \tag{1}$$

Using Equation (1), the open loop transfer function of the entire system can be written:

$$G_{ol}(s) = G(s) \cdot G_{cl}(s) \cdot V(s), \tag{2}$$

or:

$$G_{ol}(s) = k_p \cdot G_{cl}(s) \cdot e^{-s\tau}. \tag{3}$$

In order to analyse the stability of the system from Figure 3, the dead-time component was expressed as a Pade approximation [1]:

$$e^{-s\tau} \approx \frac{B(s\tau)}{A(s\tau)}, \tag{4}$$

where,

$$B(s\tau) = \sum_{k=0}^{r} \frac{(2r-k)!}{k!(r-k)!}(-s\tau)^k,$$
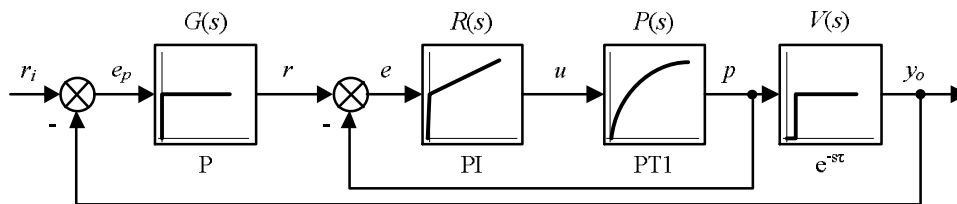$$A(s\tau) = \sum_{k=0}^{r} \frac{(2r-k)!}{k!(r-k)!}(s\tau)^k. \tag{5}$$

Fig. 2. *Detailed block diagram of the active vision closed-loop control system*
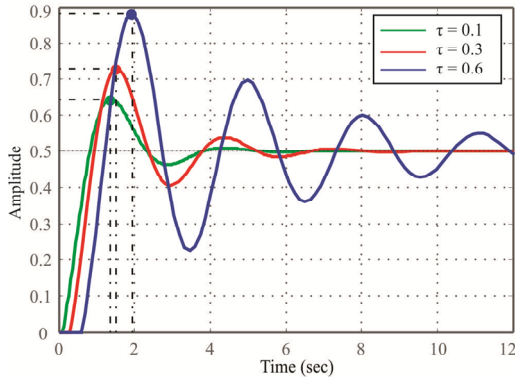
Fig. 3. *Step response for different values of the dead-time*

Using Equations (1-5), and a Pade approximation of order three, the time delay can be written as:

$$e^{-s\tau} \approx \frac{-s^3\tau^3 + 12s^2\tau^2 - 60s\tau + 120}{s^3\tau^3 + 12s^2\tau^2 + 60s\tau + 120} . \quad (6)$$

2. Stability Analysis

In order to determine the stability of the system, the step responses for the closed-loop transfer function (3) were considered, together with an open-loop Bode diagram stability analysis.

The step response is presented in Figure 3. In the figure, the variation of the output signal is presented, with the following values of the dead-time component: $\tau = 0.1$, $\tau = 0.3$ and $\tau = 0.6$. As shown in Figure 3, the delays, introduced by the image processing system $V(s)$, reduces the performance of the system in a way that the settling time increases and the overshoot gets higher. As can be seen, the delay has a major role in the system stability. Based on the step response, it has been established that the image processing system, that is the introduces dead time, should not have a value larger than $\tau = 0.6$.

In Figure 4 and Figure 5, the Bode diagrams for the proposed control system are presented. The Bode plot is one of the

most important graphical tools for analyzing and designing control systems. One of the advantages of using the Bode diagram is that the stability of the system can be evaluated based on the open-loop transfer function. The diagrams are only for a time delay value of $\tau = 0.1$ and $\tau = 0.6$. For every described case, the phase margin and gain margin were calculated. Based on this information the system's stability could be determined. As can be seen, the phase margin 88.3 and 38.7 deg, respectively, stays in the same range, while the gain margin has an important decrease, from 27.8 to 3.74 dB. These values contain a negative sign when the dead-time is more than $\tau = 0.6$. For values between $\tau = 0.1$ and
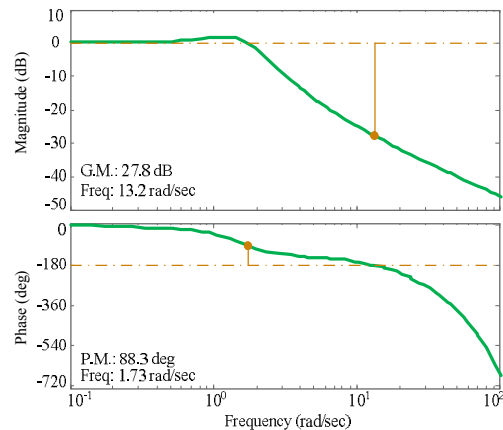

Fig. 4. *Bode diagram for τ = 0.1*

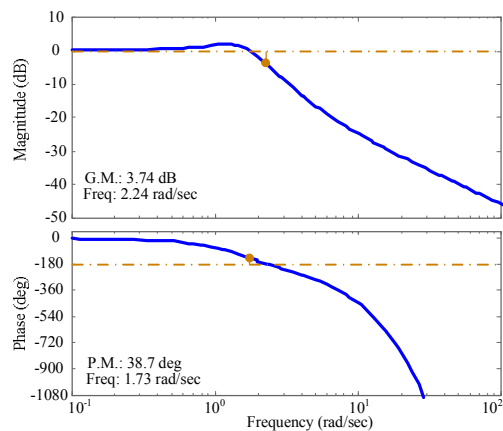Fig. 5. *Bode diagram for τ = 0.6*

and $\tau = 0.6$ the phase margin and gain margin have positive values, thus the system can be considered stable.

3. 3D Position of the Interest Object

One important problem to be solved in obtaining the position of interest object is estimating the distance between the camera and the object. This distance is determined using a calibrated stereo camera. In order to determine the object's 3D position, the images used must be rectified. The rectification process transforms each image plain in such a way that pairs of conjugate epipolar lines become collinear and parallel to one of the image axes [5].

The model of the stereo camera used is illustrated in Figure 6. A real-word point $P = [X\ Y\ Z\ 1]^T$ is projected onto the image planes of a stereo camera as the homogeneous 2D image points:

$$\begin{cases} p_L = [x_L\ y_L\ 1]^T, \\ p_R = [x_R\ y_R\ 1]^T, \end{cases} \tag{7}$$

where, p_L and p_R represents the projection of a P point which has the (x_L, y_L) and (x_R, y_R) coordinates. The projected points are situated on the left I_L and right I_R images, respectively. The O_L and O_R represent the camera's optical centres. The plane formed by the optical centres and the P point represents the epipolar plane. This plane intersects the image pane in p_L and p_R. The image, or the principal plane, is located at a distance f from the optical center of the camera, where f is the focal length. The z axis of the coordinate system is referred to as the principal ray, or optical axis. The intersection between the image plane and the principal ray at the image center (c_x, c_y) is known as the principal point. The top-left corner of the image is origin of the coordinate system.
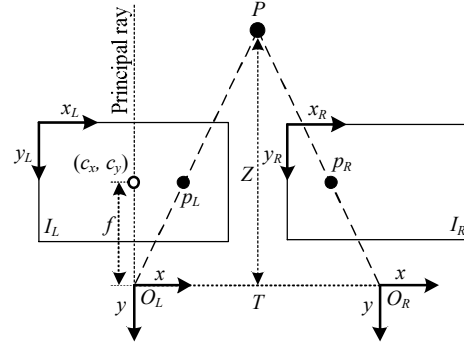

Fig. 6. *Depth estimation of a point P on a pair of rectified stereo images*

In order to determine de distance Z from the stereo camera to point P the distance T between the optical centers of the two cameras and the projection points p_L and p_R has to be known. Knowing these parameters, the 3D position of P with respect to the camera is determined using the following Equations [6]:

$$X = x_L \cdot \frac{T}{d}, \tag{8}$$

$$Y = y_L \cdot \frac{T}{d}, \tag{9}$$

$$Z = f \cdot \frac{T}{d}, \tag{10}$$

where, d is the disparity of the projected point P:

$$d = x_L - x_R \tag{11}$$

The image processing algorithms used for 3D scene reconstruction are out of the scope of this paper. Further documentation can be found in [6].

4. Conclusions

In this paper, the analysis of an active vision system was presented. The main goal was to determine the maximum value of the dead-time component introduced by

the image processing system. The performance of the considered model was studied using time and Bode analysis. As future work, the authors consider the overall modelling of the visual understanding process, taking into account also the dynamics of the image processing algorithms.

References

1. Coman, S., Comnac, V., Boldisor, C., et al.: *Fractional Order Control for DC Electrical Drivers in Networked Control Systems*. In: Proc. of the 12th Int. Conf. on Optimization of Electrical and Electronic Equipment (OPTIM), Braşov, Romania, 2010, p. 804-810.
2. Corke, P.I.: *Design, Delay and Performance in Gaze Control: Engineering and Biological Approaches*. In: The Confluence of Vision and Control **237** (2007) No. 1, p. 146-158.
3. Corke, P.I.: *Visual Control of Robots: A High-performance Visual Servoing*. Great Britain. Research Studies Press Ltd., 1996.
4. Fayman, J.A., Sudarsky, O., et al.: *Zoom Tracking and its Applications*. In: Machine Vision and Applications **13** (2001), p. 25-37.
5. Fusiello, A., Trucco, E., Verri, A.: *A Compact Algorithm for Rectification of Stereo Pairs*. In: Machine Vision and Applications **12** (2000) No. 1, p. 16-22.
6. Grigorescu, S.M., Moldoveanu, F.: *Controlling Depth Estimation for Robust Robotic Perception*. In: 18th World Congress of the International Federation of Automatic Control (IFAC), Milano, Italy, 2011, p. 3136-3141.
7. Gu, Y., Sato, M., Zhang, X.: *An Active Stereo Vision System Based on Neural Pathways of Human Binocular Motor System*. In: J. of Bionic Engineering **4** (2007) No. 4, p. 185-192.
8. Jimenez, E.V.C., Navarro, D.Z., Rojas, R.: *Intelligent Active Vision Systems for Robots*. Goettingen. Cuvillier Verlag, 2007.
9. Sharma, R.: *Role of Active Vision in Optimizing Visual Feedback for Robot Control*. In: The Confluence of Vision and Control **237** (1998) No. 1, p. 27-40.
10. Sundareswaran, V., Bouthemy, P., Chaumette, F.: *Active Camera Self-Orientation Using Dynamic Image Parameters*. In: Proc. of the Third European Conference on Computer Vision, Stockholm, Sweden, 1994, p. 110-116.
11. Sul, S.K.: *Control of Electric Machine Drive System*. New Jersey, USA. Wiley-IEEE Press, 2011.
12. Welke, K., Asfour T., Dillmann, R.: *Active Multi-View Object Search on a Humanoid Head*. In: IEEE Inter. Conf. on Robotics and Automation (ICRA), Kobe, Japan, 2009, p. 417-423.