Towards Saliency-based Gaze Control in a Binocular Robot Head

André Filipe Roos , Hugo Vieira Neto (Advisor)

Graduate Program in Electrical and Computer Engineering
Federal University of Technology - Paraná (UTFPR)

andrefroos@gmail.com, hvieir@utfpr.edu.br

Abstract – Robots that emulate biological visual systems must solve the perceptual problem of when and where to direct gaze. In this paper, a gaze control scheme for a specific binocular robot head is described and assessed. A well-known computational model of visual attention is used to find attractive target locations and a proportional-integral position control is used to drive pan and tilt servomotors of a dominant camera. Experiments demonstrate that the system is capable of tracking conspicuous moving targets in real-time (30 fps) in non-cluttered environments without any previous knowledge or strong assumptions. Addition of vergence control for the non-dominant camera is the next step towards a complete binocular control framework.

Keywords: visual attention, gaze control, stereo vision, robot head.

1. Introduction

Robot heads are classic examples of biologically-inspired active vision systems. The presence of actuated cameras allows a robot to sense a broad portion of its surroundings, while the incoming visual input may act as feedback to influence the next motion to be performed [1]. This direct link between perception and action makes this kind of agents suitable to interact with highly dynamic environments.

Several robot heads have been successfully developed since the late 1980's, such as the Rochester [4], Harvard [6] and MEDUSA [9] heads. All of them had to address the gaze problem: where to look next in response to external stimuli. Although the solution to this question may depend on top-down cues like the task being executed and previously acquired knowledge, studies indicate that very basic visual information contributes to eye fixation. In fact, the investigation of human psychophysical data has led researchers to propose models of low-level (bottom-up) attention that solely rely on primitive features of the scene, for example color and orientation [10]. Some of these models have been translated to the computational domain, yielding accurate predictions when correlated with real human fixations recorded by eye-tracking equipment [3].

By analyzing the past work in visual attention and active vision, two facts become evident. First, attention models have been predominantly experimented on static images, rather than on dynamic and interactive scenes [2]. Hence, the effects of motion, time and top-down processes in attention are still trending research topics not completely understood. Second, much effort has been devoted to emulate the complex saccadic and smooth pursuit eye movements of the Human Visual System (HVS) [5, 8].

In this paper, we investigate a simpler gaze control scheme based on a proportional-integral (PI) position controller and a static bottom-up visual saliency model [7] that seems sufficient to direct the camera movements of a robot head in various scenarios. In the following sections, detailed information about the proposed control architecture and the experiments conducted for its quantitative assessment are given.

2. The Dexter Robot Head

A stereo robot vision head called Dexter was built at the Computer Vision Laboratory of the Federal University of Technology - Paraná (UTFPR) to serve as a visual navigation module for autonomous mobile robots. Figure 1 illustrates Dexter, which is currently available as a research platform for the investigation of robot vision concepts such as depth estimation, obstacle detection and avoidance, and simultaneous localization and mapping in real environments.

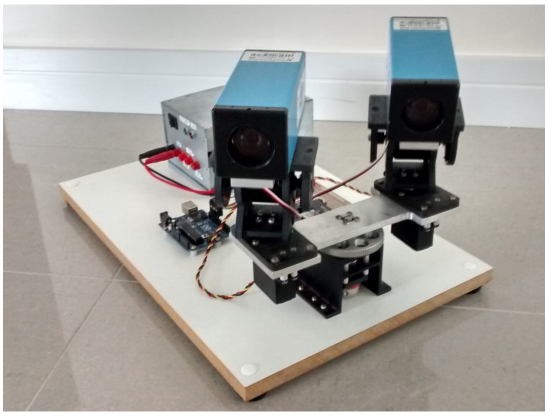


Figure 1. The Dexter Robot Head physical implementation, disconnected from the control desktop computer.

From a mechanical perspective, Dexter is a kinematic chain of four rotational degrees of freedom: left eye pan (θ_l), right eye pan (θ_r), eye tilt (θ_t) and head pan (θ_h), as shown in Figure 2. The mechanism has five actuated joints because each camera bears its own tilt motor. Independent vertical eye movement was eliminated in software, based on the constraints of the HVS.

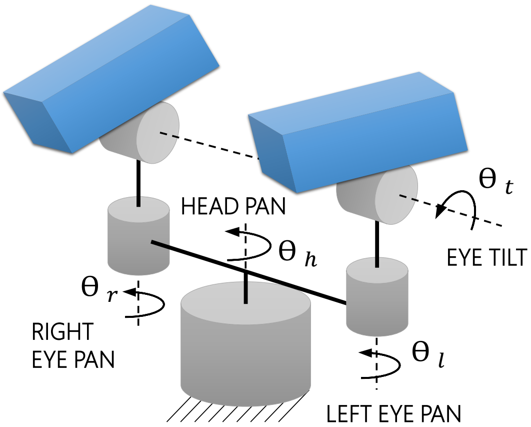


Figure 2. Kinematic chain of the Dexter Robot Head, composed of five actuated joints, but only four degrees of freedom, as eye tilt motors are virtually coupled.

3. Control System Architecture

The complete system is composed of four subsystems: the robot head mechanism, an Arduino R3 board used for servo control, a power supply unit and a control desktop computer running the Ubuntu 14.10 Linux operating system.

The main goal of the complete control system is to maintain both cameras directed to the most visually attractive location at every instant. This problem involves two issues: deciding where to look (gaze control) and pointing the two cameras to the same location (vergence control). In this work, only the first issue is addressed.

The approach followed in this work assumes that there is a dominant (left) camera, which selects the target and drives both cameras towards it. The cameras periodically send color image frames to the control computer, from which a saliency map [7] is computed by the visual attention module. A peak detector selects the most salient location (x, y) in the map, which acts as the feedback source for the gaze controller. The saliency detection algorithm was implemented in C++, using the OpenCV library framework and following the formulations of Walther [11] – complete details of the saliency model can be found in [7].

The gaze controller setpoint is defined as the central coordinates of the image frame, since each camera should maintain the target at a "virtual fovea". Thus, the error signal is the displacement between the detected salient location and the center of the frame. The controller aims at minimizing the error signal in the shortest period of time by driving the servomotors.

A discrete PI controller corrects the left pan and tilt motor positions in response to the target location error, as depicted in Figure 3. Given that the input image frame has width w and height h, the error vector $e[k]$ is computed as the difference between a constant setpoint ($x_d[k] = w/2$ for pan, $y_d[k] = h/2$ for tilt) and the target location given by the saliency algorithm ($x[k]$ for pan, $y[k]$ for tilt).

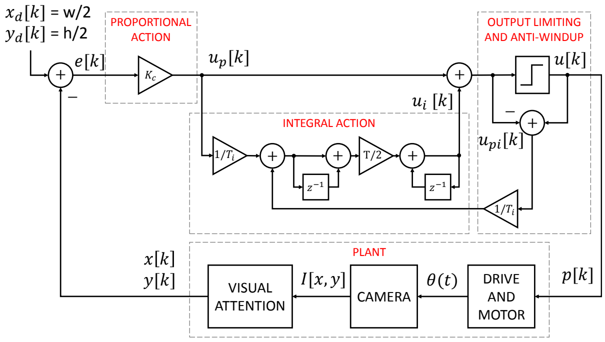


Figure 3. Discrete PI controller implementation using trapezoidal integration, output saturation and anti-windup. The same controller structure is used for both pan and tilt.

The integral action is discretized by means of trapezoidal integration (Tustin) in order to avoid abrupt motion when there is a sudden change in the target location. Controller output $u[k]$ is limited between $u_{min} = 0.0$ and $u_{max} = 1.0$ to protect servomotors against out-of-range inputs. Finally, an anti-windup algorithm prevents the error from being permanently integrated to large values when the output is saturated.

4. Experimental Setup

A first experiment was conducted to identify, by means of step response analysis, the pan and tilt plant blocks, in which both motors were individually excited with a step input while the camera captured a reference mark at 30 fps. The position of the reference mark in each frame was manually collected and plotted against time. The resulting signal was then resampled at 30 Hz using cubic interpolation to compensate for non-uniform sampling and then subjected to a least-squares estimation routine. Finally, the selected continuous estimated models were discretized at 30 Hz using zero-order-hold sampling and the controllers were manually tuned by assessing the effect of changes in gain.

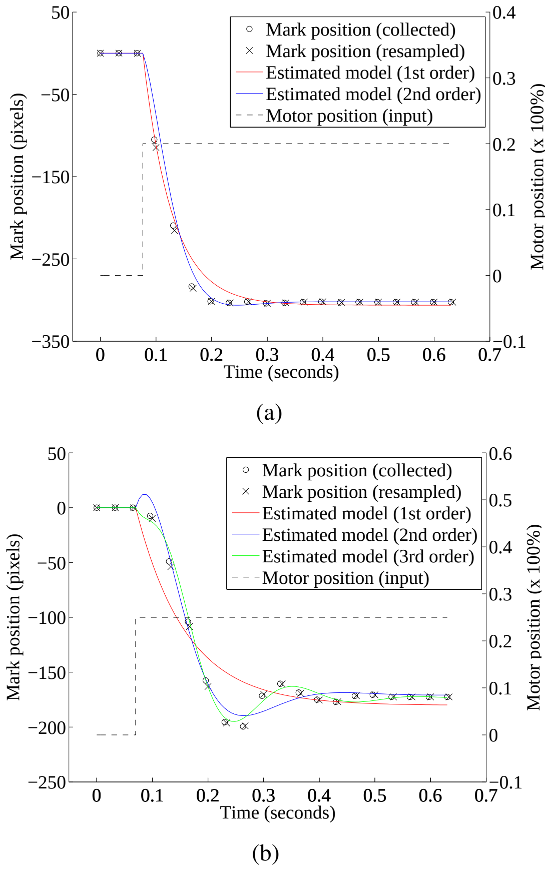
The second experiment assessed quantitatively the gaze controller's ability to pursue a conspicuous moving target without previous knowledge of its existence. A pendulum connecting a conspicuous red ball to the lab ceiling was built and made to oscillate in harmonic motion for several seconds, while the system recorded current error and motor positions.

The optimized OpenCV implementation allowed real-time performance with acceptable frame sizes (320 × 240 pixels). The processing of one image frame took less than 15 ms, leaving enough time available for further computational tasks.

5. Results and Discussion

The outcomes of the system identification experiment are depicted in Figure 4, where one can visually notice that the pan plant can be modelled by a second order system and that the tilt plant can be modelled by a third order system. This conclusion is corroborated by the smaller Mean Squared Error (MSE) of each estimated model, as Table 1 shows. The tilt system is naturally more oscillatory because of the effects of gravity and the mass distribution of the camera.

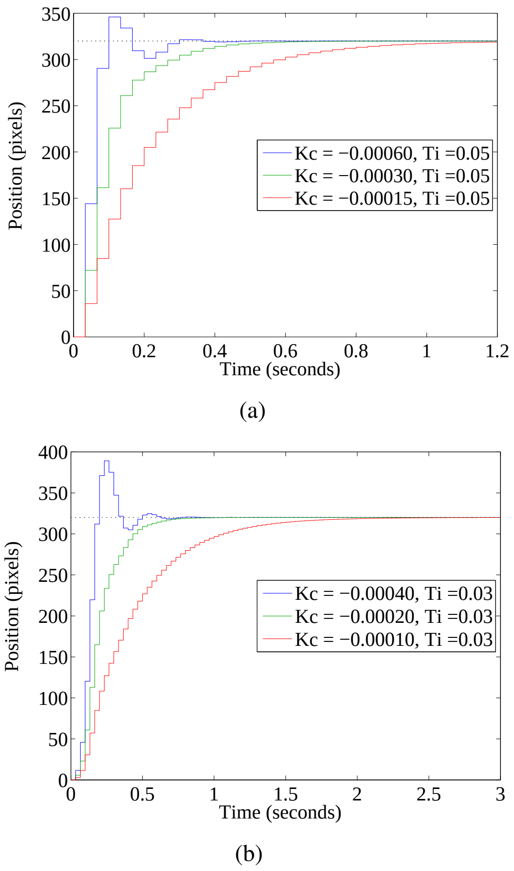
Extracted model parameters such as the gain relating joint and image plane displacements, as well as time constants, aided simulation and tuning of the PI controllers, as shown in Figure 5. As expected, the presence of the integral time T_i eliminated steady state errors. The increase of controller gain Kc, while holding T_i constant, decreased rise time and increased overshoot, degrading overall stability. The best trade-off between tracking performance and robustness was obtained by selecting intermediate gains (green lines in Figure 5).



(a)

(b)

Figure 4. Model identification via least squares estimation for (a) pan and (b) tilt.



(a)

(b)

Figure 5. Controller tuning by simulation for (a) pan and (b) tilt.

Table 1. Mean Squared Error of each estimated model (smaller values indicate a better fit).

Model	Mean Squared Error		
Order	1	2	3
Pan	110.6	4.864	-
Tilt	429.2	57.98	16.59

Finally, Figure 6 shows the results of the pendulum pursuit experiment. The robot successfully tracked the target, but there was an evident phase shift between the robot and target. This behavior was caused by two main reasons: the inherent delay in the position control loop and the lack of feedforward techniques to predict target speed and therefore correct errors faster. When the ball decelerated to a halt, the head managed to reach it and bring the error to zero. Despite the small amount of lag, motor position history shows that the system has captured the damped oscillatory behavior of the target accurately.

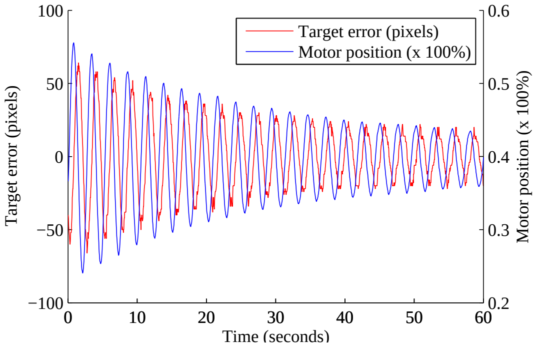


Figure 6. Target tracking: the exponentially decreasing sinusoidal profile of motor position confirms tracking of the oscillating pendulum.

6. Conclusions

A gaze control system for a robot head based on a biologically-plausible model of visual attention was presented in this work. Instead of focusing in complex eye movement and cognitive models, a simple PI control strategy and low-level saliency detection were used for real-time execution. The main advantages of this approach are that the image-based control system eliminates the need for kinematic calibration, there are no strong assumptions about the nature of the target and that controller tuning is performed in a straightforward manner.

However, there are still issues to be addressed: in highly cluttered or noisy scenes, where similarly salient objects may be competing for gaze, some higher level processing may be necessary. In addition, some lag was observed during visual tracking of moving objects, which may be compensated with some sort of predictor or speed control. Future work also includes the design of a vergence control scheme to guide de non-dominant camera to the same spot as the dominant camera, completing the proposed attentional framework.

References

[1] R. Bajcsy. Active perception. *Proceedings of the IEEE*, 76(8):966–1005, 1988.

[2] A. Borji and L. Itti. State-of-the-art in visual attention modeling. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(1):185–207, 2013.

[3] A. Borji, D. N. Sihite, and L. Itti. Quantitative analysis of human-model agreement in visual saliency modeling: A comparative study. *IEEE Transactions on Image Processing*, 22(1):1–16, 2012.

[4] C. M. Brown. The Rochester robot. Technical Report 257, University of Rochester, 1988.

[5] J. J. Clark and N. J. Ferrier. Modal control of an attentive vision system. In *Second International Conference on Computer Vision*, pages 514–523, 1988.

[6] N. J. Ferrier and J. J. Clark. The Harvard binocular head. *International Journal of Pattern Recognition and Artificial Intelligence*, 7(1):9–31, 1993.

[7] L. Itti, C. Koch, and E. Niebur. A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(11):1254–1259, 1998.

[8] D. A. Robinson. The oculomotor control system: A review. *Proceedings of the IEEE*, 56(6):1032–1049, 1968.

[9] J. Santos-Victor, F. van Trigt, and J. Sentieiro. MEDUSA - a stereo head for active vision. In *International Workshop on Inteligent Robotic Systems*, 1994.

[10] A. M. Treisman and G. Gelade. A feature-integration theory of attention. *Cognitive Psychology*, 12(1):97–136, 1980.

[11] D. Walther. *Interactions of Visual Attention and Object Recognition*. Phd thesis, California Institute of Technology, 2006.