MEDUSA - A Stereo Head for Active Vision

José Santos-Victor **Franc van Trigt** **João Sentieiro**

Instituto de Sistemas e Robótica,
Instituto Superior Técnico,
Av. Rovisco Pais, 1096 Lisboa Codex, PORTUGAL
e-mail jasv@kappa.ist.utl.pt

Abstract. This paper describes the design and control of a stereo head system **Medusa**, with four mechanic degrees of freedom, (independent vergence, tilt and pan), which was designed for active vision applications. The main mechanic design specifications were met by suitably selecting the motors and carefully estimating the moments of inertia of the structure.
The control problem was split into an inner servo loop and an outer gaze control loop. The servo loop has been carefully modeled and suitable controllers were designed both for velocity and position control. Even though the visual loop is still kept to a simple level, many interesting tracking tasks can be addressed. For a single eye configuration, a test example is presented in which the tracking behaviour can be seen.

1 Introduction

Computer vision systems have traditionally been designed without paying much attention to the observer role (camera motion, stereo rig geometry, lens parameters, etc) in the perceptual process. In the last few years, however, researchers have recovered the idea of the *active observer* and *active vision* [1, 2, 3, 4], mostly inspired by studies in psychophysics, psychology and biology [6, 9].

Within this new framework, the active involvement of the observer in the perception task, by changing the general characteristics of the visual system, became an obvious advantage to improve perception. The camera/eyes of a stereo rig can verge to maintain a fixation point on an interesting object; the cameras can then foveate the object, bringing it to the image center and zooming in, and so on [5, 7, 8, 10, 11]. A key issue for all these problems is the sensorimotor coordination or, in the particular case of computer vision, visuo-motor coordination. Instead of performing exhaustive and heavy calculations for scene reconstruction, "simple" image measurements are directly used as visual feedback for motor control.

Medusa, is a stereo head with 4 degrees of freedom, designed for active vision experiments. The head is composed of two B&W video cameras, which can verge independently (hence fixating closer or further objects), a common tilt unit (up/down) and a common pan unit (left/right). To some extent, it corresponds to an anthropomorphic design and provides the main degrees of freedom available in the human ocular system and in many other living creatures [9, 12]. The inter-ocular distance, or the baseline, can be adjusted manually in order to deal with different depth ranges. Moreover, the camera fixation can be adjusted to ensure that the vergence and tilt axes intersect in the cameras optic centers.

This paper describes the design and construction of the stereo head, and addresses some early experiments and problems of active vision systems.

2 System Setup

Medusa host system is a PC computer, equipped with a DT2851 frame grabber and a DT2859 video multiplexer. The motors are controlled by two Omnitech Robotics MC3000 axes control boards (each board controls up to 3 axes). Two unused control channels will be used in the future for active lens control.

The maximum angular speed and angular acceleration were set as design specifications. To rapidly switch the attention between different targets (saccadic motion), the head should be able to move at a speed of $180^0/s$ (comparable to the human ocular system) [12] with accelerations of $1080^0/s^2$. The required motor torques to meet these requirements were determined based on the calculation of the moments of inertia of the different stereo head parts. In order to perform accurate smooth tracking of moving targets, the motion resolution should be as high as 0.01^0.

The stereo head is equipped with Harmonic Drive DC motors (with negligible backlash) coupled with encoders for vergence, tilt and pan axes. The motors are driven by Advanced Motion Controls MC3X series PWM servo amplifiers.

The baseline may be manually set between 60 and 200 mm. The camera mounting can be adjusted so that both vergence and tilt axes intersect in the camera optic centers, if desired. **Medusa** is shown in Figure 1:

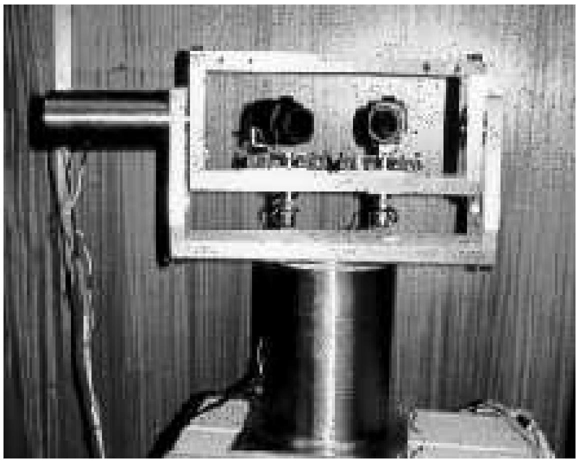

Fig. 1.: The Medusa stereo head system.

Each joint is equipped with switches for limit detection and with an inductive position sensor for fine homing during startup. Table 1 summarizes the main characteristics of the different joints of the stereo head

	Resolution (degress)	Min. speed (deg/s)	Max. speed (deg/s)	Accel. (rad/s^2)
Verg.	0.0031	1.55	180	$6\ \pi$
Tilt.	0.0018	0.9	180	$6\ \pi$
Verg.	0.0036	1.8	180	$6\ \pi$

Table 1.: Medusa main characteristics. This table shows the angular resolution, minimum and maximum speed and angular acceleration for each joint.

3 Gaze Control System

When the head is tracking an object, the following operations are performed each cycle: images are acquired; the target object position is calculated and new motor commands are generated, according to the diagram in Figure 2.

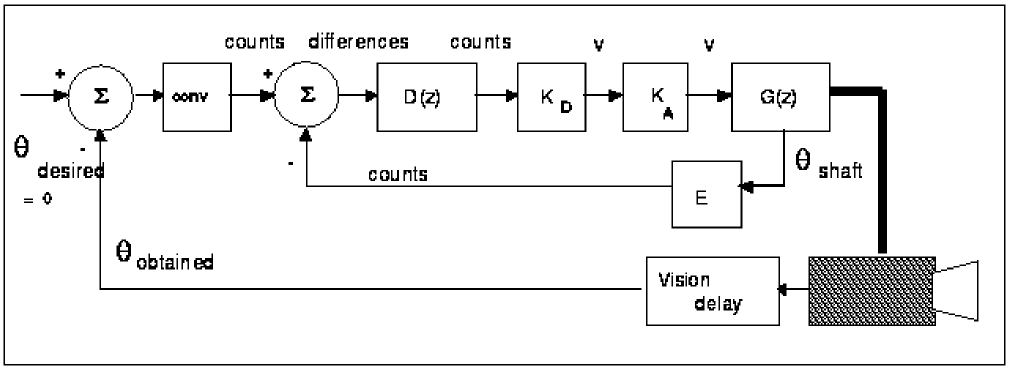

Fig. 2.: Complete loop for gaze control.

In the figure, D(z) denotes the digital servo controller, K_D is the DAC conversion factor, K_A is the amplifier gain, G(z) is the transfer function of the motor and E is the encoder function.

The control system is composed of an outer loop and an inner servo loop. The inner loop operation consists in generating new motor commands, based on the new desired joint positions and feedback from the encoders. The outer loop deals with determining new joint positions using visual information as feedback, thus providing references for the low level control system. The control of the stereo head is therefore split into a low level and a high level parts.

For the servo loop, we performed extensive modeling of the different subsystems in the control chain, and a suitable controller was designed based on control specifications, both for position and velocity control. Position control is suitable for fast movements between different targets, appropriate for saccadic motions, and the velocity control mode is suitable for smooth tracking of a moving target.

3.1 Gaze control

In the control architecture proposed for **Medusa**, the gaze control system is responsible for using the visual information to calculate new command positions/velocities, for the vergence, tilt and pan motors. These new commands are then input to the servo control loops described previously.

The main issue is then visuo-motor control, namely how to close the motor loop by using visual information. Several but interesting problems arise, such as how to decide when to pan and when to verge, since both movements will affect horizontal displacements in the image plane.

The first goal for **Medusa** is to follow a bright spot moving throughout the scene, thus allowing us to focus on the visuo-motor control problems without excessive computational burden due to the vision processing.

The images are acquired with a resolution of 512 by 512 pixels and subsampled down to 64 by 64 pixels. A thresholding operation is applied to each acquired image, separating the target from the background. At this coarser resolution, the image center of mass center is determined, yielding an estimate of the target position in the visual field.

Assuming that the vergence and tilt axes intersect in the cameras optic centers, the relationship between the vergence and tilt angles and the target position in the left and right image planes,

(x_l, y_l, x_r, y_r), results from simple geometric considerations; for vergence, we have

$$\theta_l = \arctan\left(\frac{x_l}{f}\right) \tag{1}$$

$$\theta_r = \arctan\left(\frac{x_r}{f}\right). \tag{2}$$

where f is the camera focal length. The tilt error is calculated similarly, by averaging the vertical target projections:

$$\theta_t = \arctan\left(\frac{y_l + y_r}{2f}\right). \tag{3}$$

A *saccadic reflex* can then be performed by applying these reference angles to the servo control loops, redirecting gaze as fast as possible towards the object. Alternatively, if the goal is to track the object in a smooth fashion, as in a *smooth pursuit mechanism*, velocity control should be used. The required angular velocity command, to be applied to a specific joint, may be estimated by dividing the position error by the sampling time of the overall loop (image acquisition and processing), and used to command the motors. Some experiments have been made in order to illustrate these issues.

Concerning the coordination of the different degrees of freedom, we have envisaged three main possibilities, to combine vergence and pan movements: (1) The target is followed using the vergence and tilt motors and, whenever the cameras reach an uncomfortable position, a pan saccade is performed to redirect gaze to a more comfortable position, while vergence movements compensate the neck rotation; (2) The gaze angle is kept at 90 degrees, by controlling the pan and tilt angles, while left and right vergence angles are kept symmetric, moving the fixation point closer or further from the camera; (3) A combined motion of the pan and vergence axes, based on the control effort (energy) and the comfort of the camera positions, is responsible for horizontally repositioning the target in the image center.

4 Results

An experiment has been carried out to test the overall response of the stereo head. A light source is moved at constant velocity in front of the stereo head. A single axis (camera vergence) is activated to track the moving spot. The experiment setup is shown in Figure 3. The running frequency of the

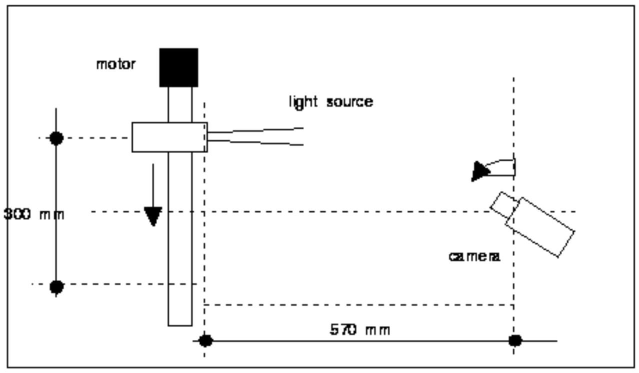

Fig. 3.: Experimental setup for the gaze control system.

global control loop is around 1.5 Hz, while the servo loop runs at a frequency of 0.5 KHz (which represents a tradeoff between speed and velocity resolution). Different tests were carried out, by trying different servo and gaze controller parameters, revealing satisfactory tracking capabilities (see Figure 4).

Figure 4, shows some of these results. Both plots show the evolution of the absolute vergence

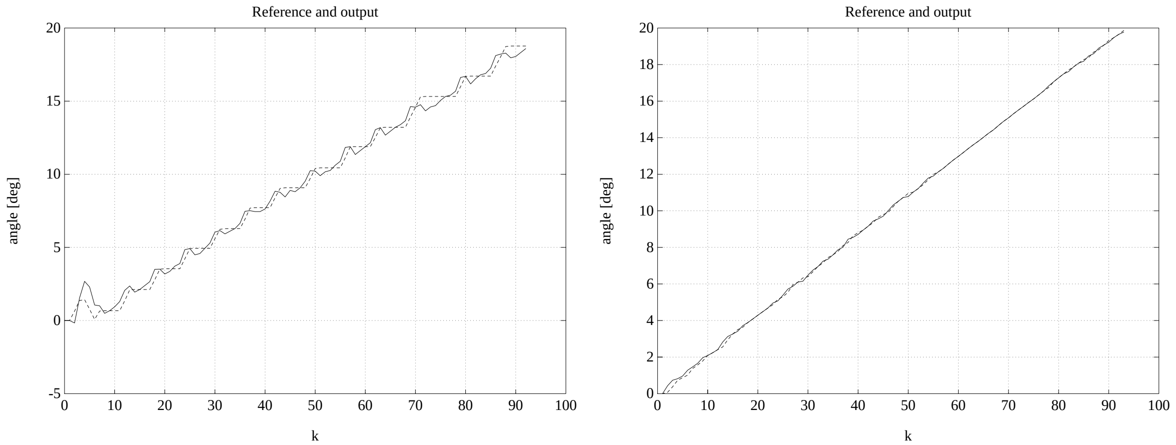

Fig. 4.: Camera (dotted) and target (solid) absolute angular positions. On the second experiment, a combined position and velocity control strategy is applied, leading to good tracking performance. The sampling period is around 0.7 s.

angle and the absolute angular position of the target, over time. The results on the left show that sometimes, even in the presence of tracking errors, the camera stops. The reason for this problem is a lower bound for the measurable velocity, due to the discrete nature of the servo loop sensor (encoder), which is in our case of about $1.5^o/s$. When the tracking error is small, the corresponding velocity command may be smaller than the velocity lower bound, and the camera will remain still, which explains the results shown. To overcome this problem, one may check whether the command velocity is less than the minimum velocity. In this case, the vergence is controlled in position mode (micro saccades), thus achieving extremely high precision. If, otherwise, the command velocity is sufficiently high, the controller will use the velocity control. This combined strategy corresponds to the right plot, showing a good tracking performance.

5 Conclusions

This paper presents **MEDUSA**, a stereo head for active vision applications. **Medusa** is equipped with 4 degrees of freedom : independent vergence for each camera and a common tilt and pan. The inter-ocular distance can be manually adjusted. The camera position can be adjusted so that the tilt and vergence axes intersect (approximately) the optical centers. We have set a number of specifications regarding the system performance, namely resolution, angular velocity and angular acceleration constraints.

The head control was split into two parts. The servo level was designed to ensure that each of the motors follows the position and velocity references, provided by the external loop. Particular care was taken in the design of these loops to ensure appropriate dynamic responses. The visual information is then linked to the motor action by the gaze control system. In this outer loop, the visual processing identifies the target position and provides references to the servo loops in order to track the target.

Finally, we have presented an example that illustrates the tracking capabilities of the system. The analysis of this experiment has revealed some control problems, that were successively dealt with. The final control configuration led to low tracking errors.

Even though the gaze control system was kept to a simple design, good tracking performance was achieved. Further improvements are now being envisaged, namely in the gaze control system to cope with the pure delay introduced by the visual processing; and addressing the problem of gaze acquisition, which may require more advanced control design, than the problem of gaze holding.

References

1. A. Blake and A. Yuille (Editors). Active Vision, MIT Press, 1992.
2. Y. Aloimonos, I.Weiss, A.Bandyopadhyay. Active Vision, *International Journal of Computer Vision*, 1(4):333-356, January 1988.
3. R. Bajcsy. Active Perception vs. Passive Perception, Proceedings of IEEE workshop on Computer Vision, Bellair, MI, October 1985, pp. 55-59.
4. D. Ballard. Animate Vision, *Artificial Intelligence*, 48:57-86, 1991.
5. D. Coombs. Real-time gaze holding in binocular robot vision, *Ph.D. thesis*, University of Rochester, October 1991.
6. J.J.Gibson. The perception of the Visual World, in Michael Gazzaniga and Collin Blakemore (Eds), Handbook of Psychobiology, Academic Press, 1950.
7. N. Ferrier. The Harvard Binocular Head, Technical Report, 91-9, Harvard Robotics Laboratory, 1991.
8. E Krotkov. Active Computer Vision by Cooperative Focus and Stereo, Springer Verlag, 1989.
9. M. Land. Similarities in the Visual Behavior of Arthropods and Men, in Michael Gazzaniga and Collin Blakemore (Eds), Handbook of Psychobiology, Academic Press, 1975.
10. D. Murray. Stereo for Gazing and Converging Cameras, Technical report OUEL 1915/92, Robotics Research Group, Dept of Engineering Science, Univ of Oxford, 1992.
11. K. Pahlavan. Active Robot Vision and Primary Ocular Processes, Ph.D Thesis, CVAP, Royal Institute of Technology, Stockholm, Sweden, 1993.
12. D. Robinson. The Oculomotor Control System: A Review, Proceedings of the IEEE, Vol 56, N 6, June 1968.

This article was processed using the TEX macro package with IRS94 style