Pursuit Control in a Binocular Active Vision System Using Optical Flow

Helder Araújo, Jorge Batista, Paulo Peixoto, Jorge Dias
Institute of Systems and Robotics/ Dep. of Elect. Eng.
University of Coimbra - 3000 Coimbra, Portugal
email: helder@mercurio.uc.pt

Abstract

An active vision system has to enable the implementation of reactive visual processes and of elementary visual behaviors in real time. Therefore the control architecture is extremely important. In this paper we discuss a number of issues related with the implementation of a real-time control architecture and describe the architecture we are using with camera heads. Another important issue of the operation of active vision binocular heads is their integration into more complex robotic systems. The design of the control architecture has to be suited to the integration of the system in other robotic systems. Higher levels of autonomy and integration can be obtained by designing the system architecture based on the concept of purposive behavior. At the lower levels we consider vision as a sensor and integrate it in control systems (both feed-forward and servo loops) and several visual processes are implemented in parallel, computing relevant measures for the control process. At higher levels the architecture is modeled as a state transition system. Finally we show how this architecture can be used to implement a pursuit behavior using optical flow. Simultaneously vergence control can also be performed using the same visual processes.

1. Introduction

Until a few years ago, the main goal of vision was to recover the 3D structure of the environment. According to this paradigm vision is a recovery problem being its goal the creation of an accurate 3D description of the scene (shape, location and other properties) which then would be given to other cognitive modules (such as planning or reasoning). Systems based on this approach typically used one or two static cameras (or, equivalently only considered static points of view, without the possibility of changing the viewpoint). Image acquisition, in this framework, is passive. This approach (general recovery) addresses the question of what range of mechanisms could exist in intelligent systems possessing visual capabilities. It does not address the question of how actual biological vision systems are designed as well as the question of what sort of vision systems would be desirable for particular classes of animals or robots. The "reconstructivist" approach addresses a problem which might not be directly related to the way biological or successful machine vision systems are designed [2]. Biological vision systems are designed in many different ways. They have different needs, sizes and characteristics and, in general, they do different things.

Instead of trying to find general solutions for the vision modules we can consider the problem of vision in terms of an agent that sees and acts in its environment ([2], [17]). An agent can be defined as a set of intentions (or purposes) which translate into a set of behaviors [6]. The visual system can then be considered as a set of processes working in a cooperative manner to achieve various behaviors ([16], [9]). This is a paradigm known as active/purposive vision. Within this framework we consider that the system is active because it has control over the image acquisition process and acquires images that are relevant for what it intends to do. The control over the image acquisition process enables the introduction of constraints that facilitate the extraction of information about the scene [17]. Therefore our goal when using the active vision system is not the construction of a general purpose description. The system only needs to recover partial information about the scene. The information to be extracted and its representation have to be determined from the tasks the system has to carry out (its purpose). Vision is considered as part of a complex system that interacts with the environment [3]. Since only part of the information contained in the images needs to be extracted, the visual system will operate based on a restricted set of behaviors (sets of perceptions and

1015-4651/96 $5.00 © 1996 IEEE
Proceedings of ICPR '96

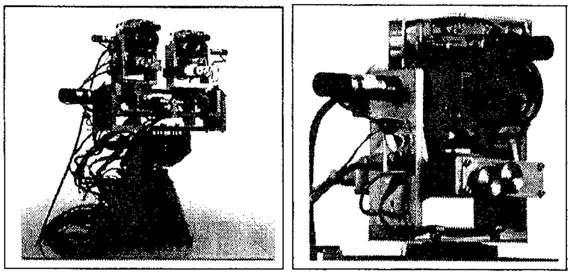

Figure 1. MDOF Active Vision Head and MDOF Eye with the MDOF motorized zoom lens.

actions).

2. A Complex Active Vision System

In order to experiment with visual behaviors and to study active vision issues (inspired by biological implementations and in particular by the human visual system) we decided to build a multi-degrees of freedom (MDOF) robot head [10]. We call it the MDOF active vision head. Other groups have built heads and demonstrated them in several applications. At the University of Rochester [4, 5] a binocular head was demonstrated and tracking was performed using vergence control by means of zero-disparity filtering. A complex head was also built at KTH [15] using stepper motors. At the University of Oxford a head was also used [7] to demonstrate the use of image motion to drive saccade and pursuit.

2.1. Mechanical Structure

The binocular head developed by ourselves has a high number of degrees of freedom. In addition to the common degrees of freedom for camera heads (neck pan, neck tilt and independent vergence for each of the eyes), this head includes the swing movement of the head neck, independent tilt for each eye,baseline control, cyclotorsion of the lenses and the ability of adjusting the optical center of the lenses. The latter is to ensure pure rotation when verging the cameras and compensate for the translation movement of the optical center when changing the focal length of the lens.

One important aspect in the design stage of these robotic systems is their performances. The analysis of some characteristics of the human active visual system can be useful for determining performance requirements for velocity and acceleration of a mechanical device that is aimed at simulating the human visual system behavior.

	Precision	Range	Velocity
Neck Pan	0.0036°	$[-110° \ldots +110°]$	$\sim 360°/s$
Neck Swing	0.0036°	$[-27.5° \ldots +27.5°]$	$\sim 360°/s$
Neck Tilt	0.0036°	$[-32° \ldots +32°]$	$\sim 360°/s$
Eye Pan	0.0036°	$[-45° \ldots +45°]$	$\sim 360°/s$
Eye Tilt	0.0031°	$[-20° \ldots +20°]$	$\sim 330°/s$
Cyclotorsion	0.0031°	$[-25° \ldots +25°]$	$\sim 330°/s$
OCA [1]	$8nm$	$[0 \ldots 80]mm$	$\sim 1mm/s$
Baseline	$20nm$	$[137 \ldots 287]mm$	$\sim 5mm/s$

Table 1. Mechanical structure characteristics of the MDOF active vision system

This design of the head inspired by biological motivation has direct consequences on the kinematics of the head. No coincident axes have been possible for all the three neck degrees of freedom. Only the pan and swing axes intersect. The tilt axis does not intersect any of these two axes and it was put 8cm ahead and 14cm above the pan and swing axes. With this particular design the eyes will have a translational component added to the pan and swing rotation movement. The eyes of this head are equipped with fully independent movements and azimuth, elevation and cyclotorsion are available. The inclusion of independent neck and eyes elevation movements was motivated by the fact that a smooth-pursuit of light loads is accomplished with much more accuracy and saccadic movements of the eye can be performed much faster than neck saccadic movements. We included the optical center adjustment due to the fact that this head is equipped with motorized zoom lenses. Pure rotation vergence movements are possible using this degree of freedom. We don't think that the adjustment of the optical center is crucial for active vision robot head, but the kinematics of the eye becomes a lot easier, in special for motorized zoom lenses. Pure rotation is also important to implement distance-independent saccade algorithms, and is essential for algorithms that assume that the relationship between motion space and motion in joint space may be learned without knowledge of the target distance. This could be extremely important for example to perform active calibration of the optical degrees of freedom. The optical center adjustment only takes place along the optical axis of the lens, since the larger variation of the center of projection occurs along this axis as a result of focus and zoom changes. A small variation on the location of the center of projection also occurs on the other two axes, but we considered that variation negligible compared with the variation that occurs along the optical axes. The dynamic performance, accuracy, and other requirements are achieved with harmonic drive DC motors. In order to simulate the performances of the human visual system there is

[1] OCA : Optical Center Adjustment

	Precision	Range	Velocity
Zoom	$Range/90000$	$[12.5 \ldots 75]mm$	$\sim 1.2 * range/s$
Aperture	$Range/50000$	$[1.2 \ldots 16]$	$\sim 2.2 * range/s$
Focus	$Range/90000$	$[1 \ldots \infty]m$	$\sim 1.2 * range/s$

Table 2. Optical structure characteristics of the MDOF active vision system

a requirement for large acceleration, low friction, high repeatability and minimal transmission errors. With the harmonic drive gear-boxes, transmission compliance and backlash, which can cause inaccuracy and oscillations, are almost eliminated. All the motors are equipped with optical encoders that provide good resolution but require initialization procedures each time the system is powered up.

2.2. Optical Structure

In a real world environment the range of conditions that a camera may need to image under, be it focused distance, spatial detail, lighting conditions or radiometric sensitivity, can often exceed the capabilities of a camera with a fixed parameters lens. To adapt the imaging conditions the camera system requires lenses whose intrinsic parameters can be changed in a precise and fast controllable manner. Motorized lenses offer greater capability and flexibility than fixed-parameter lenses. However, most active vision systems have been limited to cameras with fixed lenses because of the difficulty of modeling cameras with motorized lenses, their weight and the precision they offer. Nowadays, motorized zoom lenses become more and more important in active vision systems, e.g., for depth reconstruction, magnification, focusing, etc.. Zoom can be used to acquire images at different magnifications, e.g., simulate foveation and concentrate the view on a particular feature, focus can be used to automatically refocus on objects at different distances and compute relative depth maps, and the aperture can be used to automatically adjust the iris according to the changes in lighting conditions.

Most of the existing heads uses standard motorized lenses with potentiometers as feedback information. These lenses have the disadvantage of moving too slowly for real-time accommodation purposes (5-6 seconds to full range movement), and the accuracy for position control is not very good due to the type of information they provide as feedback. New motorized lenses have been developed to enable this head to accommodate the optical system in real time (25 images per second, with very good precision (see fig. 1). These lenses have controllable zoom, focus and iris and they use small harmonic drive DC motors with encoder feedback information. By using DC harmonic drives we are able to span the full range of zoom (12.5 mm to 75 mm) and focus (1 m to infinity) in 0.8 sec and the full range of the iris in 0.45 sec.. Also the full range of focus and zoom are subdivided into 90000 positions whereas the full range of iris is subdivided into 50000 positions (see tab. 2). With such performances, the lens is able to make continuous, small optical adjustments required by many algorithms in near real time with excellent precision. Qualitative improvements in lens performances increase the advantages of active vision techniques that rely on controlled variations of intrinsic parameters.

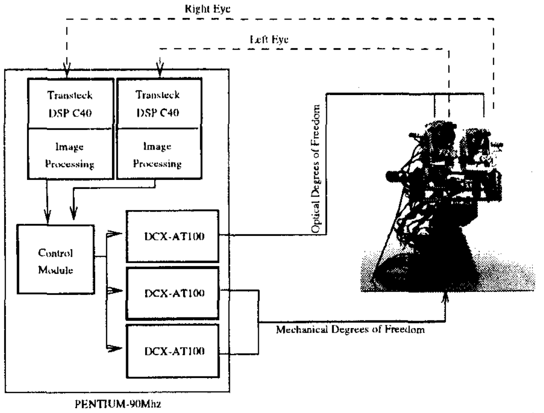

Figure 2. The MDOF system Architecture

2.3. System Architecture

The MDOF active vision robot head is controlled by one host computer with a Pentium CPU and a dual C40 Image Processing and frame-grabber PC board.

A modular multi-axis motion controller was used to control all degrees of freedom of the head. This modular system consists of a motherboard where up to six daughter-boards or modules can be connected. The motherboard is based on a 32-bit 80960 RISC CPU. On-board *Multitasking* executes up to 10 independent programs or background tasks simultaneously without interrupting motion control. Multiple boards can be built into a single system, when more than six modules are required. Three boards are used to control the 18 degrees of freedom of the robot head. Each DC servo controller module that was plugged in on the motherboard contains a trapezoidal velocity profile generator and a digital PID compensation filter. Each module is a self-contained intelligent controller.

The image acquisition and processing is performed by a dual C40 image processing board. This board also has a frame-grabber. Each one of the monochrome cameras is connected to an input of the frame grabber.

In most cases the images from the two cameras are processed in parallel by the C40s.

Most of the processes in this system run in parallel. The motors are controlled by fully parallel processes. The parameters of these processes can be changed on the fly. Typically the main CPU down-loads into the C40s the program corresponding to the elementary visual process required by the behavior to be implemented. This could be a *motion detection* module, or *visual attention* module. The processes running in the C40s directly communicate with the processes controlling the mechanical and optical degrees of freedom. Visual behaviors are defined by processes running on the main CPU of the Master unit. These processes decompose the *visual behaviors* into elementary visual processes. Elementary visual processes are implemented by the C40s. It is also possible to have different elementary visual processes implemented on both images, in parallel. In this case, each C40 processes both the left and right image. One such case is the simultaneous extraction of peripheral and foveal motion cues. The task of the main CPU is the coordination of the processes.

3. Smooth Pursuit Using Optical Flow

In order to demonstrate the architecture we have implemented a smooth pursuit process by using optical flow. The detection of motion is performed by means of image differencing. When the integral of differences is above a threshold motion is detected. The center of mass is computed and its pixel coordinates converted into the pan and tilt angles that the neck has to rotate to foveate on the origin of motion (we assume that the focal length is known). Additional pan motions by both eyes are required so that the center of mass of the detected motion is projected into the center of both left and eye images. See Fig. 3 for the image difference signaling the detection of motion.

Saccade motion is performed by means of position control of all degrees of freedom involved. Due to the latency of the saccade movement an $\alpha - -\beta$ filter is used to predict the image position of the target assuming that the target is moving with constant velocity. After fixating on the object the pursuit process is started by computing the optical flow. During the pursuit process velocity control of the degrees of freedom is used instead of position control as in the case of the saccade. See Fig. 3 for an image after the saccade. Assuming that the moving object is inside the fovea after a saccade, the smoooth pursuit process starts a Kalman filter estimator, which takes the estimated position and velocity of the target as an input. With this approach, the smooth pursuit controller generates a new prediction of the current image target velocity and this information is sent to the motion servo controller.

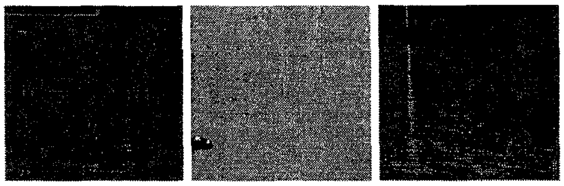

Figure 3. a)Image of the environment b) Image difference c)Image after saccade

Two different motions must be considered to exist in the scene: one caused by the motion being undertaken by the head and the other one coming from the object. Since the first is known, we only have to compute the other. For that we used the analysis of motion described by the two-component model proposed in [11].

In our case we model image formation by means of the scaled ortographic projection. Even if we model image formation as a perspective projection this is a reasonable assumption since motion will be computed near the origin of the image coordinate system (in a small area around the center x and y are close to zero). We can therefore assume that the optical flow vector is approximately constant throughout all the image, i.e.,

$$u = p_x \qquad v = p_y \tag{1}$$

To compute the optical flow vector we minimize

$$\sum_i (I_{x_i} p_x + I_{y_i} p_y + I_{t_i})^2 = 0 \tag{2}$$

Taking the partial derivatives on p_x and on p_y and making them equal to zero we obtain:

$$\begin{aligned}(\sum_i I_{x_i}^2) p_x + (\sum_i I_{y_i} I_{x_i}) p_y + (\sum_i I_{x_i} I_{t_i}) = 0 \\ (\sum_i I_{y_i} I_{x_i}) p_x + (\sum_i I_{y_i}^2) p_y + +(\sum_i I_{y_i} I_{t_i}) = 0\end{aligned} \tag{3}$$

The flow is computed on a multiresolution structure. Four different resolutions are used: $16*16, 32*32, 64*64, 128*128$. These are sub-sampled images. A first estimate is obtained at the lowest resolution level $(16*16)$, and this estimate is propagated to the next resolution level, where a new estimate is computed and so on. The optical flow computed this way is used to control the angular velocity of the motors. The sequence Fig. 4 shows images of the pursuit sequence. The position, velocity and accelerations responses of

Figure 4. Pursuit sequence

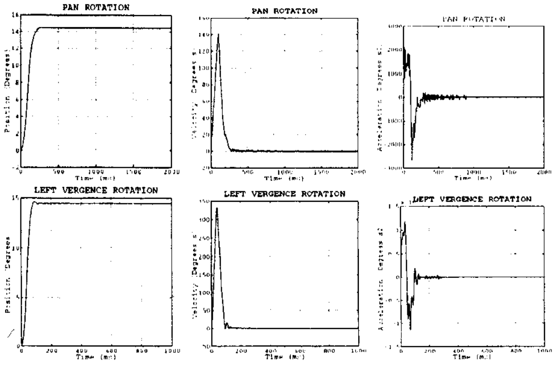

Figure 5. Initial saccade. Top row: neck-pan saccade; bottom row: eye-vergence saccade.

the neck pan during the initial saccade are displayed on Fig. 5. The position, velocity and accelerations responses of one of the eyes' vergences during the initial saccade are displayed on Fig. 5. Due to the architecture of our system we can change the control parameters on the fly so that the system can adapt itself to changes in velocity. This way the system can cope with sudden changes in velocity. We can also switch from velocity control to position control on the fly. Besides the process above described to compute flow due to motion parallel to the image plane, another process to compute flow due to translational motion along the optical axis can also be implemented, taking into account that the object is *fixated* by both cameras.

4. Conclusions

In this paper we have shown that by using the concept of *purposive behavior* it is possible to implement real-time active vision systems. The concept is essential for the design of the system architecture, if real time operation and robustness are major design goals. Another result of this approach is that computation grounded on information derived from sensation enables the achievement of autonomy. Another result of our approach is that the control architecture we have used enabled real-time operation with limited computing. On the other hand the use of parallelism enabled us the continuous processing of the image data as well as the coordination of the several actuation systems that have to work in synchrony. Parallelism is also essential to allow the visual agents to attend to the several events that are happening in the world continuously. The integration, the system architecture, the information processing modules, and the motor control processes were all designed taking into account the tasks and behavior of the systems.

References

[1] U. Nunes and al.. *Artificial Intelligence in Industrial Decision Making, Control and Automation Systems*, chapter Multi-Sensor Integration for Mobile Robot Navigation. Kluwer Academic Publishers, 1995.

[2] Y. Aloimonos. Purposive and qualitative active vision. In *Proc. Image Understanding Workshop*, 1990.

[3] Y. Aloimonos. What i have learned. *CVGIP: Image Understanding*, 60(1):74-85, 1994.

[4] C. Brown. Gaze controls with interaction and delays. *IEEE Transac. on Syst., Man and Cybern.*, 20, May 1990.

[5] D. Coombs, C. Brown. Real-time binocular smooth pursuit. *Intern. Journal of Computer Vision*, 11(2):147-165, October 1993.

[6] T. Bosser, D. McFarland. *Intelligent Behavior in Animals and Robots.* MIT Press, 1993.

[7] D. Murray and al.. Driving saccade to pursuit using image motion. *Intern. Journal of Computer Vision*, 16(3):205-228, November 1995.

[8] Y. Ho. Dynamics of discrete event systems. *Proceedings of the IEEE*, 77:3-7, 1989.

[9] G. Horridge. The evolution of visual processing and the construction of seeing systems. *Proc. Royal Soc. London B*, 230:279-292, 1987.

[10] J. Batista and al.. The ISR MDOF active vision robot head: design and calibration. In *M2VIP'95-Second Int. Conf. on Mechatronics and Machine Vision in Practice*, Hong-Kong, September 1995.

[11] J. Bergen and al.. Computing two motions from three frames. Technical report, David Sarnoff Research Center, April 1990.

[12] J. Dias and al.. Simulating pursuit with machines: Experiments with robots and artificial vision. In *Proc. IEEE Int. Conf. on Robotics and Automation*, Nagoya, Japan, May 1995.

[13] A. Meystel. *Autonomous Mobile Robots.* World Scientific, 1991.

[14] P. Burt and al.. Object tracking with a moving camera. In *Proc. IEEE Workshop on Visual Motion*, Irvine, 1989.

[15] K. Pahlavan. *Active Robot Vision and Primary Ocular Processes.* PhD thesis, CVAP, KTH, Stockholm,Sweden, 1993.

[16] A. Sloman. On designing a visual system. *Journal Exper. and Theor. Artif. Intelligence*, 1:289-337, 1989.

[17] Y. Aloimonos and al.. Active vision. *Intern. J. Computer Vision*, 7, 1988.