A Control Architecture for Active Vision Systems Enabling Real Time Operation

Helder Araújo Jorge Dias Jorge Batista Paulo Peixoto

ISR - Institute of Systems and Robotics
Dep. of Electrical Engineering, FCTUC
University of Coimbra
3030 Coimbra, PORTUGAL

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 our "binocular heads". Another important issue of the operation of active vision binocular heads is their integration into more complex robotic systems. We claim that higher levels of autonomy and integration can be obtained by designing the system architecture based on the concept of purposive behavior. Finally we describe a system where the integration of an active vision stereo head in a mobile robot has been performed, with real-time performance.

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, see [?]). 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 [1]. 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 ([1], [2]). The visual system can then be considered as a set of processes working in a cooperative manner to achieve various behaviors ([6], [7]). 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 [2]. 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. Vision is considered as part of a complex system that interacts with the environment [4].

By considering vision within this framework, it can be used in real time to control a more complex robotics system such as a mobile platform or a manipulator. Also since complex representations of the world are not necessary, the visual system can interface and interact directly with all the low level subsystems enabling a high level of autonomy in terms information processing.

Next we describe the architecture we have implemented based on the application of these principles. We will also describe an application developed based on these principles.

0-8186-8058-X/97 $10.00 © 1997 IEEE

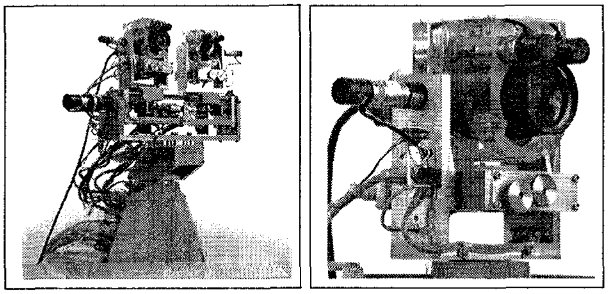

Figure 1: MDOF Active Vision Head and MDOF Eye with the MDOF Motorized Zoom lens

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 [8]. We call it the a MDOF active vision head. This head has a high number of degrees of freedom. 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.

The MDOF active vision robot head (see Fig. 1) has the following mechanical degrees of freedom:

Eyes-mechanical Each eye has three degrees of freedom (a total of six):

- elevation (tilt)
- azimuth (pan)
- cyclotorsion
- an aditional degree of freedom is included to keep the optical center at the crosspoint of the azimuth and elevation axes of the lens.

Neck-mechanical The neck has three degrees of freedom:

- tilt
- pan
- swing or lateral tilt movement

Baseline The ability of mechanicaly change the distance between the two eyes.

In addition to the common degrees of freedom for camera heads (pan, tilt and independent vergence for each of the eyes), this head includes the swing movement of the head neck, independent tilt movement for both eyes, 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.

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

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 eyes of this head are equipped with fully independent movements and azimuth, elevation and cyclotorsion. 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 equiped 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. 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 eye mechanical structure of this head and its optical center adjust-

---

[1] OCA : Optical Center Adjustment

	Precision	Range	Velocity
Zoom	$Range/90000$	$[12,5..75]mm$	$\sim 1.2 * range/s$
Aperture	$Range/50000$	$[1,2..16]$	$\sim 2.2 * range/s$
Focus	$Range/90000$	$[1..\infty]m$	$\sim 1.2 * range/s$

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

ments constraints limits the type of lenses that can be used. Only small cameras can be used, not only because of its weight but also due to its size.

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 a requirement for large acceleration, low friction, high repeatability and minimal transmission errors. These are some of the primary characteristics of systems that use motors and feedback devices mounted directly to the axes of motion. With the harmonic drive gearboxes, transmition compliance and backlash, which can cause inaccuracy and oscilations, 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.

2.1 Optical Structure

In a real world environment the range of conditions that a camera may need to image under, be it focused distance, spacial 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 controlable 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.

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 distance 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 potenciometers as feedback information. These lenses have the disadvantage of moving too slowly for real-time acommodation 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 acommodate the optical system in almost real time, 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 discretized into 90000 positions whereas the full range of iris is discretized into 50000 positions (see tab. 2). With this kind of design we hope not only to fully characterize the operation of this kind of lenses and calibrate them, but also to try them in applications where this type of speed is crucial.

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.

2.2 System Architecture

The MDOF active vision robot head is controlled by one host computer with a Pentium CPU (166Mhz) and a PCI Matrox Meteor frame-grabber PC board.

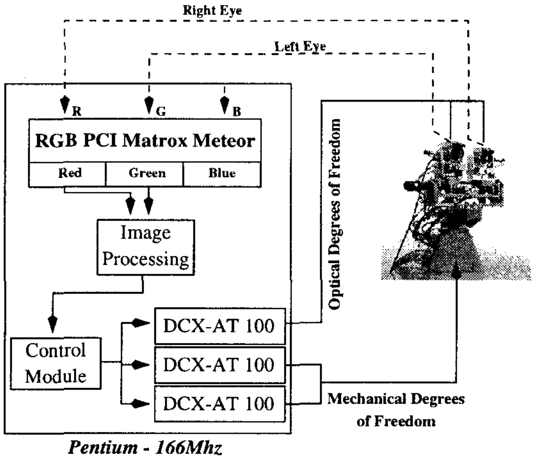

Figure 2: The MDOF system Architecture

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. Each DC servo controller module that was plugged

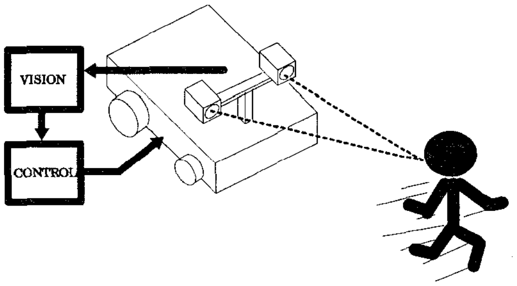

Figure 3: The information provided by the *active vision system* is used to control the mobile robot for *pursuit* a target in *real − time*.

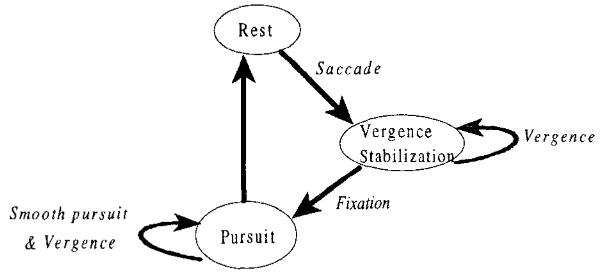

Figure 4: State diagram of the *pursuit* process.

in on the motherboard contains a trapezoidal velocity profile generator and a digital PID compensation filter.

The image acquisition is performed by an RGB PCI Matrox Meteor Board. Each one of the monochrome cameras is connected to an input of the RGB frame grabber.

The motors are controlled by fully parallel processes. The parameters of these processes can be changed on the fly.

Visual behaviors are defined by processes running on the main CPU of the Master unit.

3 Example: A Mobile Robot Pursuing a Moving Object

To demonstrate the principles above described we decided to develop a complex system integrating an active vision system on a mobile robot [10]. This system was built to perform a well defined task: pursuing moving objects. Two main problems had to be dealt with: integration and cooperation between systems. This integration [12] had two distinct aspects: the interaction and cooperation between different control systems and the use of a common feedback information provided by the vision system. The solution developed was based on the visual *gaze holding* [11] process to establish a pursuit mechanism for *targets* moving in front of the mobile robot. The system is controlled to maintain constant the distance and the orientation of the robot and the vision system. The solution for this problem deals with the interaction of different control systems using visual feedback. This problem has been addressed in different fields such as surveillance, automated guidance systems and robotics in general [13].

3.1 Pursuit of Moving Objects

To perform the pursuit of a moving *target* we use two basic control schemes: a visual *fixation* control and the trajectory control of the robot. The visual *fixation* control guarantees that the *target* is continuously tracked by the vision system, and gives information about its position to the robot control. The robot control uses that information as a feedback to maintain the distance and orientation to the *target*.

The visual *fixation* control must be one visual process that runs in the active vision system and has capabilities to define a *target*, to concentrate the vision system on the *target* and follow it. A process with these characteristics has similarities with the visual gaze-shifting mechanism in humans. The gaze-shifting mechanism generates movements in the vision system to put a new object of interest in the center of the image and hold it there. The movement used to put the object in the center is called *saccade*, it is fast and it is performed by the two eyes simultaneously. If the *target* of interest is moving relative to the world, the vision system must perform movements to hold the *target* in the image center. These movements are composed by two types of movements called *smooth pursuit* and *vergence*. These movements are the consequence of the control performed by the process that we designate as *fixation*.

The *fixation* centers and holds the orientation of the vision system on a point in the environment. The principle is described graphically by the Fig. 3 where the mobile robot with an active vision system is concentrated on a person. *Fixation* gives a useful mechanism to maintain the relative orientation and translation between the referential in the vehicle and the *target* that is followed. This results from the advantages of the *fixation* process, where the selected *target* is always in the image center (foveal region in the mammals). This avoids the segmentation of all the image to select the *target* and allows the use of relative coordinate systems which simplifies the spatial description of the *target* itself (relationship between the observer reference system and the object reference system).

The pursuit process can be described graphically by the state diagram in Fig. 4. The *pursuit* process must be initiated before starting. During the initiation a *target* is chosen, the gaze must be shifted by a *saccade* movement and the vergence must be stabilized. In our system the *target* is chosen based in visual motion stimuli. The selection corresponds to a region in the images that generates a large visual motion in the two images.

If a *target* is detected, a *saccade* movement is initialized to put the *target* in the image center, and the system changes from the state *Rest* to *Vergence Stabilization.* During the *saccade* movement no visual information is processed. In the *Vergence Stabilization* state the system adjusts its *fixation* in the *target.* This is equivalent to establish the correct correspondence between the centers of the two images, defining a *fixation* point in the *target.* Since the vergence is stabilized, the system is maintained in the *Pursuit* state.

3.2 System Control and Architecture

The main hardware components of the system are the mobile robot and the active vision system. The active vision system used in this system is different from the previously described in this paper. The head used with the mobile robot has 5 mechanical degrees of freedom. The mechanical degrees of freedom consist in the neck pan and tilt, eye pan and the baseline. Additionally, each lens has controlled focus, zoom and aperture, and it is possible to adjust the position of the optical center. The mechanical degrees of freedom are driven by stepping motors. These two basic units are interconnected by a computer designated *Master Processing Unit.* This unit controls the movements of the active vision system, communicates with the robot's onboard computer and is interconnected with two other computers designated *Slave Processing Units.* These units are responsible for processing the images provided by the active vision system. The connections between different processing units are represented in the diagram shown in Fig. 5 and a photograph of the system is presented in Fig. 6.

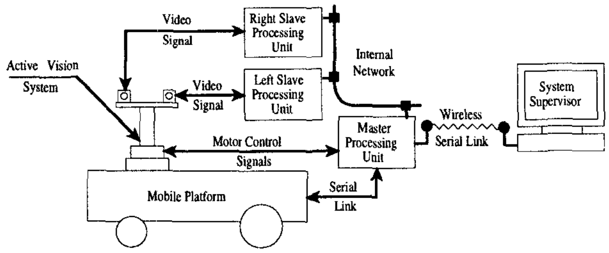

Figure 5: System Architecture.

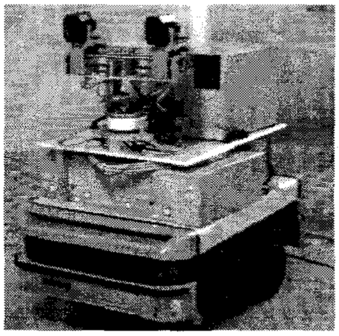

Figure 6: The active vision system and the mobile robot.

A complete implementation of the methods necessary to obtain the desired behavior of the system requires fast processing capabilities for vision processing and control. The implementation used in this work was based in control loops working in parallel and based in the visual *pursuit* process. Since the inertia of the neck and the mobile robot are greater than the vergence mechanism inertia, this type of control simulates a control system with different levels - see Fig. 7. The inmost level comprises cameras and the vergence motors of the active vision system, responsible for tracking the *target* in real-time. This sub-system controls the cameras' position to maintain the visual system fixed in the *target.* At the intermediate level there is the neck sub-system, that provides the control of the orientation of the vision system and compensates for the cameras' vergence movements. At the outmost level is the mobile robot sub-system that provides the compensation for the orientation of the active vision system and also controls the orientation and distance to the *target.*

Conceptually, the error between the actual distance and orientation of the *target* and the system is propagated from the outmost level to the inmost level. This

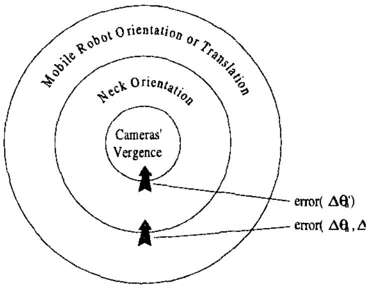

Figure 7: Graphic scheme of the principle used for the control of the system.

concept is graphically described in the Fig. 7. In each level the error is compensated for by the sub-system associated to each level. This error must be such that the maximum characteristic values of each sub-system are not exceeded. In the cases where the error exceeds these maximum values, the difference of error that can not be compensated for in that level is passed to the next in-most level. This scheme establishes a mechanism to propagate the error through the different control systems, giving more priority to the mobile robot, followed by the neck and eyes at the end.

The pursuit control loop consists basically of three stages: image acquisition, error estimation (the orientation and distance) and error correction. These steps are realized by the *Slave* and *Master Processing Units* at cycles synchronized by a general clock in the system. This clock has a $200msec$ cycle and the system's parameters are adjusted for that cycle. During this cycle the system performs different computations, depending on the state of the system. The different states of the system are illustrated by Fig. 4.

The images generated by each camera are adquired and analyzed by the *Slave Processing Units.* These units analyse the images and give the position of the *target* in each image. That position gives the necessary information to compute the system state. This state is passed to the (α-β-γ) tracker. The information provided by the *Slave Units* is delayed by one cycle of $200msec$. To avoid the lateral effects of this delay we use the prediction capabilities of the (α-β-γ) filter to estimate a value for the system state.

The *Master Processing Unit* repeatedly performs the control algorithm by using the error between the predicted system state and the desired system state. This error is passed to the different PID discrete time algorithms implemented in each subsystem. The results will be changes in the positions of the stepper motors associated with the vision system and the commands for the mobile robot to maintain the desired system state.

The movements executed by the mobile platform are based on two motors associated with each of the driving wheels (rear axle) and are essential to make the compensation for the error in the distance

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. 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.

References

[1] Aloimonos, Y.: Purposive and qualitative active vision. *Proc. Image Understanding Workshop* (1990) 816–828

[2] Aloimonos, Y., Weiss, Y., Bandopadhay, A.: Active Vision. *Intern. J. Comput. Vision* **7** (1988) 333–356

[3] McFarland, D., Bosser, T.: *Intelligent Behavior in Animals and Robots.* MIT Press (1993)

[4] Aloimonos, Y.: What I Have Learned. *CVGIP: Image Understanding* **60** No.1 (1994) 74–85

[5] Ho, Y.: Dynamics of Discrete Event Systems. *Proc. IEEE* **77** (1989) 3–7

[6] Sloman, A.: On Designing a Visual System. *J. Exper. and Theor. Artif. Intell.* **1** (1989) 289–337

[7] Horridge, G.: The evolution of visual processing and the construction of seeing systems. *Proc. Roy. Soc. London B* **230** (1987) 279–292

[8] Batista, J., Dias, J., Araújo, H., Almeida, A. The ISR Multi-Degrees-of-Freedom Active Vision Robot Head: design and calibration. *M2VIP'95–Second International Conference on Mechatronics and Machine Vision in Pratice* Hong–Kong, September, (1995)

[9] Pahlavan, K.: Active Robot Vision and Primary Ocular Processes. *PhD Thesis,* CVAP, KTH, (1993), Sweden.

[10] Dias, J., Paredes, C., Fonseca, I., Batista, J., Araújo, H., Almeida, A.: Simulating Pursuit with Machines: Experiments with Robots and Artificial Vision. *Proc. IEEE Int. Conf. on Rob. and Auto.* 472–477 May 21–27, (1995), Nagoya, Japan

[11] Brown,C.: Gaze controls with interaction and delays. *IEEE Transactions on Systems, Man and Cybernetics* **20**, May, (1990), 518–527.

[12] Almeida, A., Araújo, H., Nunes, U., Dias, J.: Multi-Sensor Integration for Mobile Robot Navigation. In *Artificial Intelligence in Industrial Decision Making, Control and Automation* Spyros Tzafestas and Henk Verbruggen (Eds.), Kluwer Academic Publishers, (1995).

[13] Burt,P., Bergen, J., Hingorani, R., Kolczynski, R., Lee, W., Leung, A., Lubin, J., Shvaytser, H.: Object tracking with a moving camera. *Proc. IEEE Workshop Visual Motion*, Irvine, 1989.