Behaviour Based Oculomotor Control Architecture for Stereo Heads.

Alexandre BERNARDINO César SILVA José SANTOS-VICTOR Carlos PINTO-FERREIRA
Instituto de Sistemas e Robótica - Instituto Superior Técnico,
Av. Rovisco Pais, 1096 Lisboa Codex, PORTUGAL
e-mail jasv@kappa.ist.utl.pt

1 Introduction

During the last decade several stereo heads were built as tools for active vision. The mechanical system (two eye anthropomorphic form) is inspired in the physiology of many animals, hopping that millions of years of natural evolution resulted in the most adapted visual system to the environment where we live. Some researchers claim that not only the mechanical structure but also the architecture of the control system and the visual processes should be based in biological examples [6], [2]. In this paper the visual system is briefly analysed and the most relevant issues are retained to use in a possible control architecture for an active vision system. A simple enough and generic structure is attained through the definition of two basic behaviours: the gaze holding behaviour and the gaze shifting behaviour. Based on this structure, some experiences were made on the MEDUSA stereo head [8].

2 A biological process: eye movements and visual behaviours

There is an increasing interest among engineers in biological control systems, specially in the oculomotor control system, which controls the eye movements. These movements are recognised as the output of a more complex system and we must deal with them as such. The eye movements made by frontal-eyed foveal animals are the combination of three basic movements [7]:

- saccade: it is a very rapid eye movement in response to a target position error, discretely sampled one reaction time previously, and it is not controlled by visual feedback during its execution.
- smooth pursuit: it is a smooth and continuous movement in response to a target velocity error on the retina.
- vergence: smooth and conjugate eye movements in response to image disparity on both retinas.

and the nystagmus [7]:

- vestibular nystagmus: these eye movements compensate the head rotation through the stimulation of the semicircular canals, in order to stabilise the image on the retina.
- physiological nystagmus: these movements consist of drifts, high frequency tremor and microsaccades that are present during fixation.
- optokinetic nystagmus: involuntary following by the eyes of any large moving pattern.

The eye movements consist of intervals of fixation interspersed with saccades. A saccade can be a voluntary movement which occurs when the observer decides to change the direction of gaze, or an involuntary fast relocation, to keep the target position inside the fovea, during fixation. The smooth pursuit tends to stabilise the image in the retina, controlled by the visual feedback. Vergence movements are the slowest of all the kinds of eye movements and maintain the target simultaneously on the fovea of both eyes. These types of movements cooperate in several visual tasks. In the behavioural sense, these movements perform a set of visual behaviours, such as [5]:

- locating a target;
- keeping still (fixation);
- tracking targets;
- wandering (looking around).

Some characteristics of these behaviours are related as shown in Table 1. The *tracking* and the *fixation* processes differ only conceptually. In fact, these behaviours are performed by the same stabilisation system. The common features of the *locating* and *wandering* processes are also shown in Table 1.

	Locating	Fixation	Tracking	Wandering
There is an important voluntary part (depending on observer decision and non-visual perception).	x			x
The involuntary movements are dominant.		x	x	
The movement consist of intervals of short periods of fixation interspersed with saccades.	x			x
The involuntary saccades, the smooth pursuit and the vergence cooperate.	x	x	x	x

Table 1: Relation among some characteristics of the visual behaviours.

3 Control Architecture

After the analysis of some important eye movements, composing the visual systems of humans and other animals, we focus our attention on how can they be replicated and integrated in a useful control mechanism for stereo heads. We have seen that the main visual physiological movements are the voluntary saccade, the involuntary saccade, the smooth pursuit and the vergence.

3.1 Gaze holding behaviour

Basically the *tracking* and *fixation* processes make use of the vergence, smooth pursuit and involuntary saccadic movements. The corresponding neuronal mechanisms are mostly involuntary, and therefore the cognitive system intervention in these processes is minimal as the control is made in a lower level, resulting in faster reaction times to environmental changes. When implementing these processes it is strongly recommended to have fast image processing routines and if possible avoid the extraction of features dependent on the shape, size or other categorical features. The required visual information should be composed of pre-categorical cues [8] such as:

- disparity measurements obtained, for example, by correlation or cepstral filtering techniques, to control the vergence movement;
- retinal slip obtained, for example, by optical flow estimation [4], to control the pursuit movement;
- retinal position measurements in order to detect if the target is moving away from the centre of the retina and bring it back through an involuntary saccade. These measurements can be obtained, for example, with a zero disparity filter [1].

All these movements, when integrated, should result in a gaze holding behaviour able to keep gaze fixed to a point in space independently of its size or shape. As suggested in [6] and [3], the performance of such a behaviour should depend on the existence of a foveal zone in the centre of the retina.

3.2 Gaze shifting behaviour

All visual tasks require voluntary movements of the eyes in order to locate specific objects or wander around the visual field, monitoring the surrounding environment. The *wandering* and *locating* processes are composed by sequences of voluntary saccades, changing the gaze point, and stabilisation periods when the target is inspected. The difference between the two processes is that the latter needs *a priori* knowledge of the object to locate. In both, categorical cues must be extracted and interpreted by the cognitive system in order to identify the object and decide whether to keep gaze on it or to continue the *wandering* or *locating* process. The new gaze points are selected among a set of candidates. This set of points can be obtained by edge and motion detectors, and contain the interesting points where to look at in the visual field.

3.3 Integration

Having defined the two main behaviours of active visual systems (gaze holding and gaze shifting), we can represent their integration in the diagram of Fig.1.

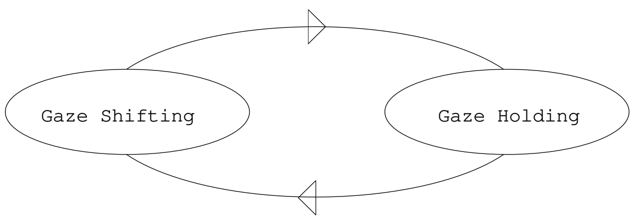

Figure 1: High level control architecture

This is a computationally attractive structure because the behaviours, as represented, are separable (do not occur simultaneously). The main coordination problem consists in managing the transitions from one to the other, as a function not only of visual information but also of internal motivational parameters.

We have now a conceptual structure that appropriately integrates all the basic ocular processes. The next sections are devoted to describe some experiences in this domain, already executed in the stereo head MEDUSA [8].

4 System setup

Medusa is a stereo head with four degrees of freedom. The head is composed of two B&W video cameras, verging independently, a common tilt unit (up and down) and a common pan unit (left and right). Medusa host system is a PC computer, equipped with a RGB colour frame grabber and two control boards. We use two buffers of the frame grabber to keep the images from the left and the right camera, captured simultaneously. The motors are controlled by two axes control boards (each board controls up to 3 axes). Two unused control channels will be used in the future for active lens control [8].

5 Experiments

5.1 Pursuit Movement

In order to demonstrate the real time pursuit capabilities of MEDUSA, a fast pursuit routine was implemented. The head follows a bright spot and therefore, the visual processing is simplified by applying a threshold to each image and calculating the centre of mass. After having determined the object position in the image planes, the required joint angles to keep the object in the centre of the retina are calculated. These angles are then introduced in a PID controller that determines the reference signals to drive the motors. Figure 2 shows the angles of each joint and the position of the bright spot in the coordinate frame of each joint, during a pursuit movement.

The cameras and pan joints are redundant (several configurations can verge on the same point). In order to deal with this issue, the pan and cameras are associated to inertial weights (I_p and I_c) and the variation of the pan angle $\Delta\theta_p$ is determined through an energy minimisation criteria:

$$J = I_c[\Delta\theta_l^2 + \Delta\theta_r^2] + I_p\Delta\theta_p^2 \tag{1}$$

where $\Delta\theta_l$, $\Delta\theta_r$ are the variations of the left and right camera angles.

Figure 2 illustrates the cameras and pan angles when I_p is zero. In the first half of the experience the target described a trajectory in the horopter - all motion is executed by the pan axis while the cameras are still. In the second part, the bright spot moves toward the cameras such that the motion is composed by almost pure vergence.

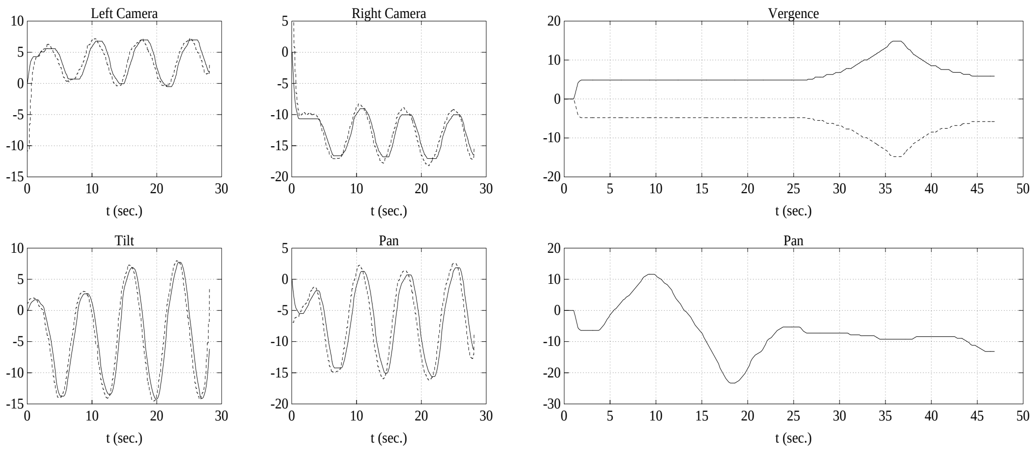

Figure 2: The four images in the left represent the response of the pursuit mechanism to a bright spot moving in circles. In the right, pure vergenge and pure pan movements illustrate decoupling between the cameras and pan joints in the case of zero pan inertial weight.

5.2 Saccadic movement

To perform a fast saccade to a point outside the fovea, a simplified calculation may not be enough to bring the point to the centre of the retina. In the smooth pursuit movement it is assumed that the object is always near to the centre of the retina and therefore an approximate calculation is sufficient. In the saccadic movement the object may be in the perifery of the retina and the head nonlinearities must be taken into account.

Therefore the kinematics of the head-eye system is determined, providing the fundamental relations to attract the target point, as efficiently as possible, to the centre of the retina. An algorithm based in the inverse kinematics was designed to specify the angular displacement for each joint as a function of the target position over the two image planes. This procedure results in a fast uncontrolled redirection of gaze but it depends on the physical parameters of the head-eye system. However, the observed sensitivity to parameter uncertainties is small and therefore a saccade movement always results in a good target approach to the centre of the foveal zone.

Figure 3 shows a sequence of saccades. The target position is defined by a fixed bright spot, somewhere outside of the fovea. We can see that after first saccade the joint positions are already near the final positions - in fact, the bright spot is already very near the centre of the fovea.

5.3 Vergence movement

By using disparity cues from the visual information we are able to control the vergence movement. There are basically two methods for doing fast disparity detection: the cepstral filtering [3] and correlation based techniques. Here we use the first approach as follows. First, we suppose that the left and right foveal images are shifted versions of each other. The junction of the left and the right foveal images results in an artificial image. Then the cepstral filter is computed as the Fourier transform of the logarithm of its Power Spectrum. By analysing the peaks of the Cepstrum (the filtering result) we identify the dominant disparities. Since the object of interest is dominant in the fovea, the disparity corresponds to highest peak of the Cepstrum. Therefore the disparity is converted to conjugate movements of cameras.

Figure 3 illustrates an experiment with a fixed object located in front of cameras (a human face in this example). We conclude that the system performs few movements until verging completely on the object.

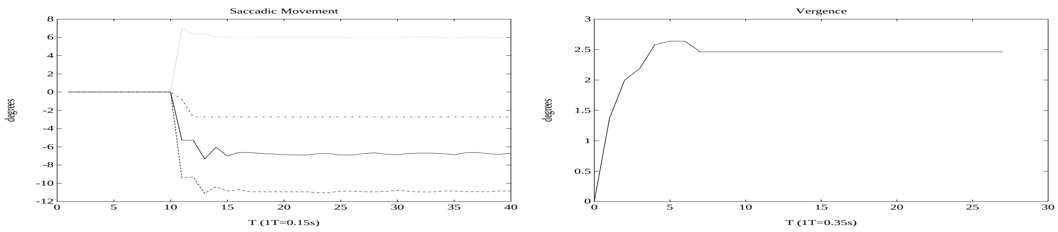

Figure 3: The left figure represents the axes movement as a serie of saccades in order to bring a fixed target to the centre of retina. The right figure represents the vergence movement of left camera for a fixed object (the right camera has a simetric movement).

6 Conclusions and future work

The analysis of some biological visual systems provides a better understanding of how several eye movements cooperate to perform visual tasks. By organising them in separable behaviours (the gaze shifting and gaze holding), it is possible to define a simple and generic control architecture for active vision. We have implemented some of the processes that constitute these behaviours with encouraging results. Future work should complete the set of required processes and focus on the problem of managing the transitions between the two global behaviours.

References

[1] C. Brown, D. Coombs, J. Soong. Real-time Smooth Pursuit Tracking, in A. Blake and A. Yuille (Eds), MIT Press, 1992.

[2] G. Cannata, E. Grosso. Explaining Biological Behaviors Through The Systems Approach, Technical Report 4-93, LIRA-Lab - DIST University of Genova, Italy.

[3] D. Coombs. Real-time gaze holding in binocular robot vision, *Ph.D. thesis*, University of Rochester, October 1991.

[4] B. K. P. Horn. Robot Vision, MIT Press, 1986.

[5] M. Land. Similarities in the Visual Behavior of Arthropods and Men, in Michael Gazzaniga and Collin Blakemore (Eds), Handbook of Psychobiology, Academic Press, 1975.

[6] K. Pahlavan. Active Robot Vision and Primary Ocular Processes, Ph.D Thesis, CVAP, Royal Institute of Technology, Stockholm, Sweden, 1993.

[7] D. Robinson. The Oculomotor Control System: A Review, Proceedings of the IEEE, Vol 56, N 6, June 1968.

[8] J. Santos-Victor, F. Trigt, J. Sentieiro. MEDUSA - A Stereo Head for Active Vision, Proceedings of the International Symposium on Inteligent Robotic Systems '94, Grenoble, July 1994.