A BIOMIMETIC ROBOTIC HEAD USING A MODEL OF OCULAR TRACKING

Kostis P.Michmizos, Henrietta L. Galiana, IEEE Fellow
Biomedical Engineering Department, Faculty of Medicine, McGill University, Montreal, QC, Canada, H3A 2B4

ABSTRACT

This paper describes a biomimetic vision platform that tracks moving targets with self-generated pursuit and saccadic intervals. Extensions to the controller add image analysis capabilities that provide a measure of prediction and low-level target selection. A model for the bottom-up control of visual attention in primates is presented and experimentally tested in the platform. Given an input image, the system predicts which location in the image will automatically and unconsciously shift a person's attention towards it. Target selection relies on the extraction of a pair of 2D feature maps based on spatial discontinuities in the modalities of intensity and velocity (brightness and slip). Both maps are then combined into a single 2D "saliency map" which encodes the desired features for each pixel in the scene, irrespective of the particular feature which detected this location as conspicuous. A winner-take-all system then detects the highest-salience point in the map at any given time, and draws the focus of attention towards this location. That allows the selection of a target in a visual scene containing multiple distractors without the need of first recognizing the objects. The intensity of the target is also embodied into the gains of the controller altering the alertness of the anthropomorphic robot with respect to the brightness of the target. The parallel observation of multiple targets and the tracking of the most salient one enhance further the biomimetic nature of the robot allowing its controller to judge the significance of a target that suddenly comes into its visual field.

INTRODUCTION

Our biomimetic bilateral controller [1] is based on the physiological evidence that smooth pursuit -that is the slow phase movement of the eyes- and saccade-their fast phase movement- are handled by the same neural circuitry. At any instant the bilateral model is in only one phase, and the mode of operation can be altered via internal switching. During the smooth pursuit phase, the bilateral controller is composed of two unilateral controllers operating within a visual feedback loop (Figure 1).

The only visual input to the system is the visual (retinal) error, e, from each 'eye'. Internal models are used to monitor eye position: Given an accurate model M of the actual plant P, it is sufficient to feed the plant model with the same drive D as the plant. This provides an efference copy E^* of the output *E*. Then, the actual eye position *E* is compared against target position *T* to produce a visual error *e*. This error is mapped from sensory signal to motor signal *i*, suitable as input to the butterfly. This task is transparent to the controller which must drive the motor errors to zero through the feedback loop.

The left and right sides are connected via the cross-midline gain, which stems from reports of strong coupling by the commissural connections, and of interconnections between the vestibular nuclei across the midline [1]. During saccades projections across the midline are removed, so the system controls each eye independently based on its own retinal error.

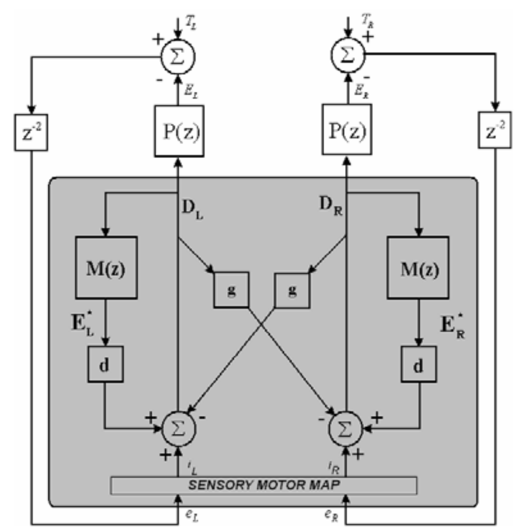

Figure 1: Smooth Pursuit Bilateral Controller. Vision provides a close loop to the controller. A delay is required to model the processing delay. Subscript L (R) stands for left (right) eye.

The implementation [6] of a previous theoretical work [4] in actual hardware resulted in a binocular robotic system with horizontal and vertical movements embodying derivative and integral signals (PDI) in the pursuit and the saccadic modes.

THE ROBOT'S HARDWARE

Our robotic platform (Figure 2) consists of a stereo vision system that is attached to a neck (the robotic head is fully described in [5]). The robot, is able to make a variety of movements in order to track targets, namely slow pursuit (slow phase), saccades (fast phase), conjugate and vergence. The platform has six degrees of freedom; each eye has an independent horizontal and vertical axis of rotation whereas the neck can produce both yaw and pitch movement. Each robotic eyeball consists of embedded color ELMO CCD MN401E cameras, which provide an NTSC signal to the frame grabber at 30 frames per second. A Pulse Width Modulation Module TL494 converts DC voltage into PWM signals in order to drive the four digital servo motors of the cameras. The robotic platform is attached to an INTEL P4 1.6 GHz operating on a Windows 2000 platform. The workstation serves as the sensory processing engine and implements the bulk of the robot's perception and attention systems.

THE PHYSIOLOGICAL MODEL OF VISUAL ATTENTION

Koch and Ullman [3] originally and Niebur [7] subsequently elaborated Caltech's hypothesis which represents one of the most tangible analyses of the cortical areas involved in the mechanism of controlling the visual attention. The authors also described a computational model that matches our implementation, namely the *saliency-based model* of visual attention.

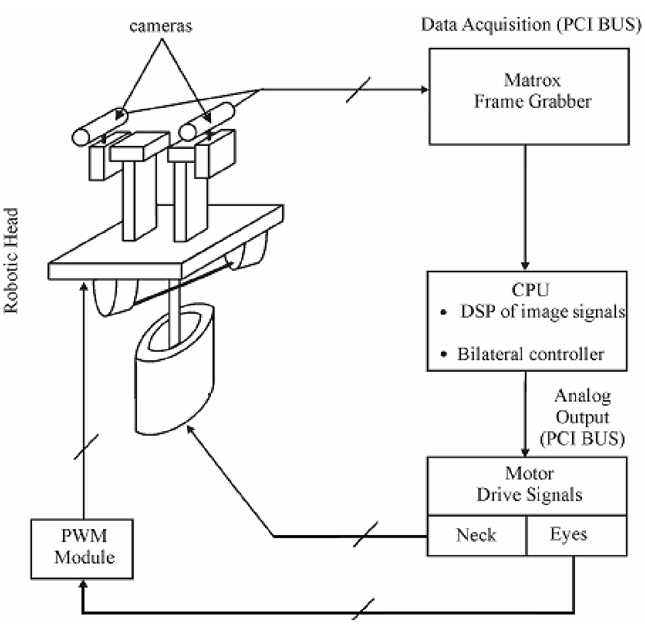

Figure 2 : Block diagram of the robotic implementation.

Based on the Itti et al. [2] implementation of the model, we developed and experimentally tested a model that estimates the extent of salient objects that solely depend on bottom-up information. Using low-level features and with negligible additional computational cost, the image region containing the attended objects is extracted from the scene.

The model is in accordance with the known anatomy and physiology of the visual system of the macaque monkey [8] and comprises two interacting stages: The first stage is a fast and parallel pre-attentive extraction of visual features across the spatial maps (for brightness and motion). The brightness feature is computed by calculating the mean intensity of each blob (an area with uniform brightness in the image) detected inside the visual field of the robot. The motion is approximated by taking the first derivative of the X and Y Center of Gravity (CoG) coordinates for each blob. Each one of the computational branches is multiplied by a gain. The second stage is an altering speed focal attention shifting mechanism that uses a Winner-Take-All mechanism to select the most conspicuous image location.

The link between the two stages is a *saliency map*, which topographically encodes for the local conspicuity in the visual scene, and controls where the focus of attention should currently be deployed. A final step of the model is required in order to detect the most salient target among other potential blobs.

When multiple blobs are found, the selection of the most salient one is based on the calculation of the value

$$C_K = \alpha \bullet Intensity_K + \beta \bullet \left(\sqrt{Speed_X^2 + Speed_Y^2} \right)_K \quad (1)$$

for each blob K.

The gains α and β can have different values depending on the feature (intensity or speed) that we want to favor. In the experiments to be described, the different spots' intensities have a small standard deviation (10 pixels, around 0.5 degrees) with a mean value of 245 in order to detect as small blobs as possible yet avoiding perturbations, under the presence of unavoidable recording noise. Hence, the speed was selected as the dominant saliency factor by using α = 0.1 and β = 0.9.

Our model does not consider the shape and extent of the attended object in determining the attended area. This conforms with the way the human eyes home in on a moving target, since commonly attention is believed to act before the visual recognition of the objects. (However, experimental evidence suggests that attention can be tied to objects, object parts, or groups of objects [8]).

The controller, with the addition of the multiple target detection feature, can handle a single target without any difference from the controller presented in [4]. However, when two or more targets are presented in the robot's visual field, a new process is introduced in sequence with the already existing computations. The output of that process is the COG coordinates of the most salient target. Hence, the process of target tracking as described previously is transparent to the controller: once the most salient target is recognized, its coordinates are given to the controller and the rest of the process remains as in [4].

ALERTNESS OF THE ROBOT

The retina projects to only about 10-20% of geniculate cells. Two layers of large cells are believed to be devoted to movement of the image in the two retinae. The remaining layers are of smaller neurons (cell bodies) and analyze the two images for color and for picture details. The remaining majority of input to the geniculate comes from other brain regions. This data apparently influences the projection to the visual cortex. Part of the afferent geniculate inflow is from the reticular system. Among other functions, this gigantic and diffuse mass of neurons governs the level of consciousness (and attention). At least a certain amount of this circuitry is a feedback loop. Hence, it can be argued that not only does the level of alertness affect what we "see" but also what we see affects our level of alertness and concentration.

For robot and human to be able to direct one another's attention to objects in the scene, the robot's visual attention control mechanism should be capable of functioning like that of a person. Therefore, we propose that visual alertness in the humanoid robot, since it is based on a model of human visual attention, also depend on the viewable object. In our biomimetic robotic implementation, this fluctuating level of alertness is implemented by altering the bandwidth: A system with a narrower bandwidth responds more sluggishly to a step input, ie. a non-expected target that is suddenly presented on its visual field. In the experiments to follow, the controller's bandwidth is a linear function of the perceived salient target's intensity, but could also include slip.

RESULTS

In the experiment presented in Figure 3, two targets (laser dots) of nearly equal mean intensity are moving in the robot's visual field. One target is moving in a sinusoidal horizontal trajectory and the other is stable 2 degrees below the first one.

The first (moving) target is not present throughout the entire experiment. For a certain period of time, we drive the target to a spot that is not visible to the robotic eyes. For these periods, only one (the stable) target is visible and the eyes fixate on it. As soon as the moving target is re-introduced to the visual field, the eyes leave the stable spot and follow the fastest (and hence the most salient) target.

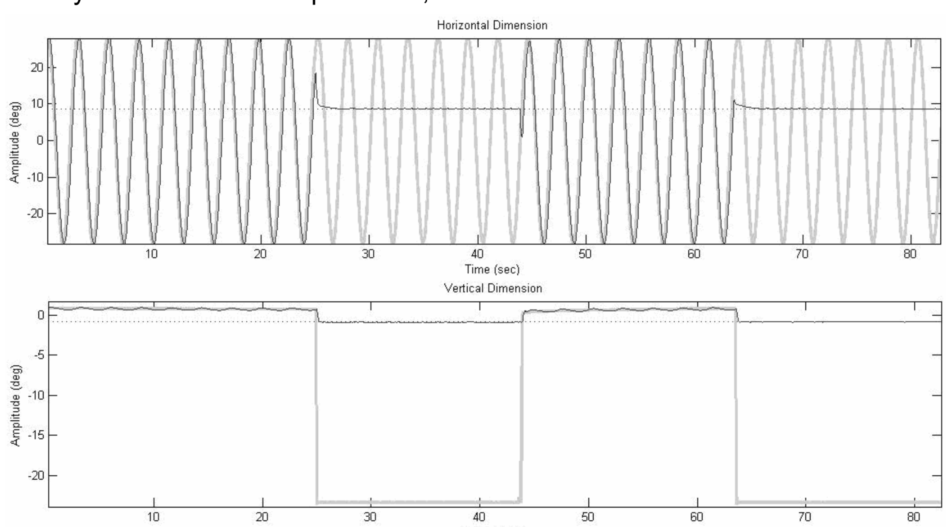

Figure 3. Individual Horizontal and Vertical Dimension of the experiment. The controller conjugate response (black thick line) follows the *B* target (gray thick line) and only on the absence of it, it fixates on *A* (dotted line).

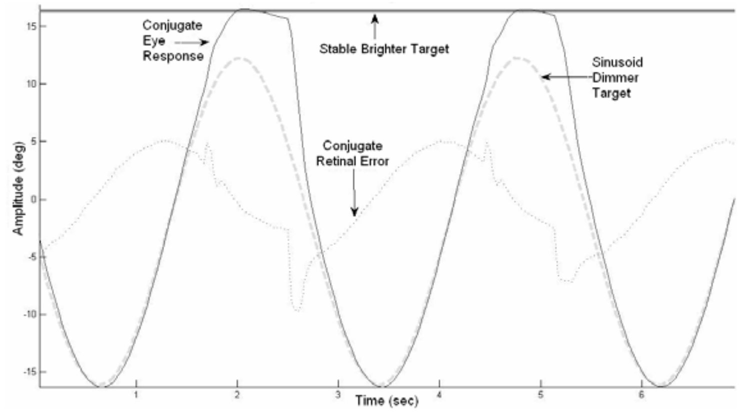

Figure 4. Experimental Response of the Horizontal Pursuit Controller under the presence of two targets with different intensities and velocities. The conjugate retinal error is drawn with a thin dotted gray line.

In the experiment shown in Figure 4, a very bright (with maximum mean intensity) light spot, *A* is used as a stable target and a laser dot *B* is used as a moving target with slightly lower mean intensity. *A* is fixed in the plane of *B*'s horizontal movement and is located nearly 5 degrees away from the right sinusoid peak.

The robotic head follows *B* accurately during the movement of the target in the left plane with respect to the origin. However, as the eye trajectory approaches the right peak of the sinusoid, its speed becomes gradually zero. The criterion *C,* described in equation (1) for *B,* becomes smaller than the corresponding *C* for *A*. Hence, the pursuit controller response is driven further from the peak of the sinusoid and lands on *A*.

The eyes stay on the stationary dot as long as $C_B \leq C_A$. As the sinusoid target moves away from the peak of its trajectory, it speeds up. When the speed reaches a certain threshold, the controller decides that the moving target is again the most salient target. The changes on the controller's decisions regarding the most salient target can also be observed by the sudden peaks introduced in the conjugate retinal error at the moment when the eyes leave one target for the pursuit of the other.

CONCLUSION

In this paper, a model for the bottom-up control of visual attention in primates was presented and experimentally tested in a humanoid robotic device. It may serve as an initial step for subsequent object detection. An implementation of techniques for object recognition could guide the robot to survive in an unknown environment. At this point, the binocular robot reacts to the visual scene simply on the basis of brightness and of moving CoGs. Two principal approaches - decision-theoretic and structural – could also be used in order for the robot to "learn" from sample patterns.

In a modification to the spotlight metaphor and the robotic alertness already described, supervised learning can be introduced in a future work to bias the relative weights of the features in the construction of the saliency map and achieve some degree of specialization towards target detection tasks.

ACKNOWLEDGEMENTS

This work was supported by the National Science and Engineering Research Council of Canada, the Québec Nature and Technology Research Fund and the Onassis Foundation.

REFERENCES

[1] H.L. Galiana and J.S. Outerbridge, "A bilateral model for central neural pathways in vestibuloocular reflex.", *J. Neurophsyiology*, 51 (2): pp. 210-241, 1984.

[2] L. Itti, C. Koch, E. Niebur, "A model of saliency-based visual attention for rapid scene analysis", *IEEE Trans. Pattern Anal. Mach. Intell.*, 20 (11), 1254–1259, 1998.

[3] Ch. Koch and S. Ullman, "Shifts in selective visual attention: Towards the underlying neural circuitry.", *Human Neurobiology*, Vol. 4, pp. 219-227, 1985.

[4] J.W. Lee, H.L. Galiana, "An Internally switched Model of Ocular Tracking with Prediction". IEEE Trans. on Neural Systems and Rehab. Engin., Vol. 13, No. 2: pp. 186 – 193, June 2005.

[5] K.P. Michmizos, "Implementation of Robotic Visual Attention motivated by Human Physiology and Behavior", *Master Thesis*, McGill University, Montreal, Canada, 2006.

[6] K.P. Michmizos and H.L. Galiana, "A biomimetic Robotic Device operating on an internally switched model of Ocular Tracking using Prediction and Memory", *submitted to IEEE Trans. On Robotics*, 2006.

[7] E. Niebur and Ch. Koch, "Control of selective visual attention: Modeling the where pathway.", Advances in Neural Information Processing Systems, Vol. 8, pp. 802-808, 1996.

[8] P. Roelfsema, V. Lamme, H. Spekreijse, "Object-based attention in the primary visual cortex of the macaque monkey", *Nature*, 395 (6700)) ,pp. 376–381, 1998.

[9] S. Treue, JHR Maunsell, "Attentional modulation of visual motion processing in cortical areas MT and MST", *Nature* 382:539 –541, 1996.