

Design of an anthropomorphic robot head

K Undbekken, D Wilson, P Mowforth, S Solbakken and B Eiklid
The Turing Institute
Glasgow, UK *

Abstract

This paper describes the design of a robot head featuring cameras and microphones in a system geared to active perception. The principal motivation behind the design is that the head should exhibit degrees of freedom, physical dimensions and sensor location equivalent to that of a human head. The overall goal for the project is to develop sensory reflexes for the head with a speed and operation similar to that found in biological systems. Implementation of the vergence reflex is described and some performance results are presented.

1 Introduction

A robot head system supporting active perception [1, 2] offers significant scope for application in areas such as teleoperation and security and surveillance in addition to providing a valuable basic research tool. Such a system combines sensors with actuators and control strategies to carry out a number of useful reflex behaviours. The overall performance of the robot head and the range of reflexes it can support is critically dependent on the design of the underlying hardware platform.

2 Robot Head Design

Figure 1 shows the first prototype of RICHARD. To support a full range of reflexes, the head features six motorised degrees of freedom with azimuth and elevation control for both the neck and each of the two eyes.

A pair of miniature, remote-head CCD cameras form the eyes of the robot head. Each camera is mounted on an angle bracket which can be rotated independently in azimuth and elevation by two separate motors fitted respectively to the base and side of the bracket. The first prototype makes use of stepper motors which supply an accuracy of 0.03° per step and allow the cameras to be rotated at a speed of approximately 60° per second. The cameras are mounted with a baseline separation of approximately 6cm.

Azimuth control of the neck is provided by a motor fitted at the base of the head platform which drives the head across a 4:1 gearing system with an accuracy of 1.8° per step. The elevation control motor is mounted vertically

*This work has been supported by the Turing Institute and the DTI IED project *Active Stereo Probe*

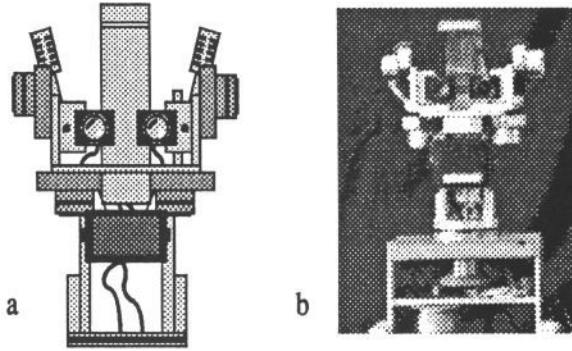


Figure 1: Schematic and photograph of the first prototype, RICHARD 1st.

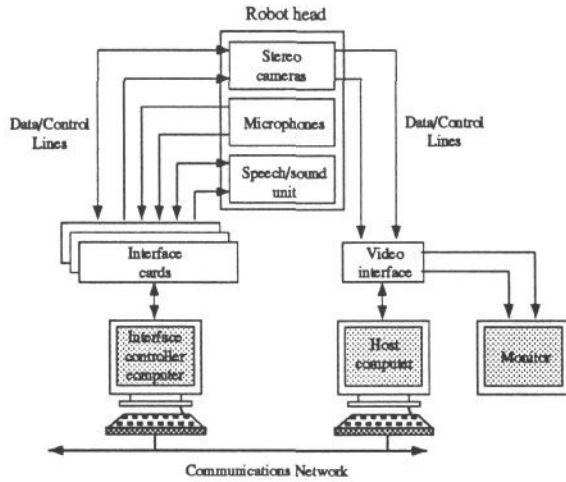


Figure 2: Computational architecture for the robot head.

within the neck structure and rotates the head about a central horizontal pivot across a drive-belt. This motor supplies an accuracy of 7.5° per step at a speed of approximately 90° per second.

An essential part of the anthropomorphic design is that the head is not restricted to visual sensory input. A pair of miniature microphones are fitted in fixed positions, one at either side of the central head pivot, to permit the capture of sound data. A speech unit which can handle both text-to-speech and synthesised tone generation is fitted below the gantry housing the cameras.

3 Computational Architecture

The robot head operates within an integrated computational architecture. The principal components of this architecture are identified in figure 2.

Interface electronics for the head (motors, speech unit and microphones) are inserted in an interface controller computer together with the necessary software drivers. This effectively runs as a device driver for the robot head

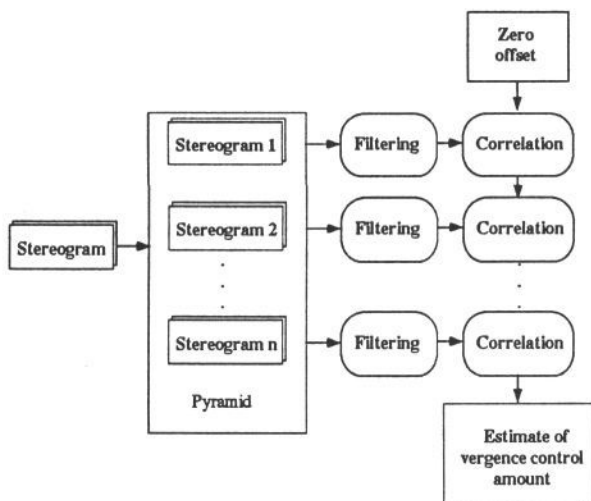


Figure 3: Software architecture of the vergence control algorithm.

and permits algorithms to be developed for the reflex behaviours which are independent of the underlying hardware.

The cameras are connected to a framegrabber unit which is installed in the host computer so that stereo images of the scene viewed by the head can be captured and displayed to the user via a monitor.

The system which is currently in use with RICHARD 1 employs a PC as the interface controller computer and a SUN SparcStation as the host. These computers are connected across an RS-232 link. This architecture is particularly suited to experimentation with the robot head and combines flexibility with considerable software processing support. An alternative solution which is more suited to end-product applications would implement the host and interface controller functions in a single computer using dedicated hardware or software systems.

Routines to implement each of the reflex behaviours execute under the control of a software workbench on the host computer. The processing sequences for the various reflexes typically involve three stages. Firstly, some sensory data (either video or sound) is acquired. Secondly, this data is processed in some intelligent manner. Finally, commands are generated to reposition the eye and neck motors. The host computer performs each of these operations and sends commands to drive the robot head across a communications link to the interface controller.

4 Vergence

Our current research is aimed at developing a range of useful sensory reflex behaviours. Vergence has already been implemented and demonstrated with the current prototype head.

Vergence control [3] is necessary to bring two images captured from two locations in space into approximate correspondence. This may be used to derive information about range as well as facilitating the task of stereo solutions.

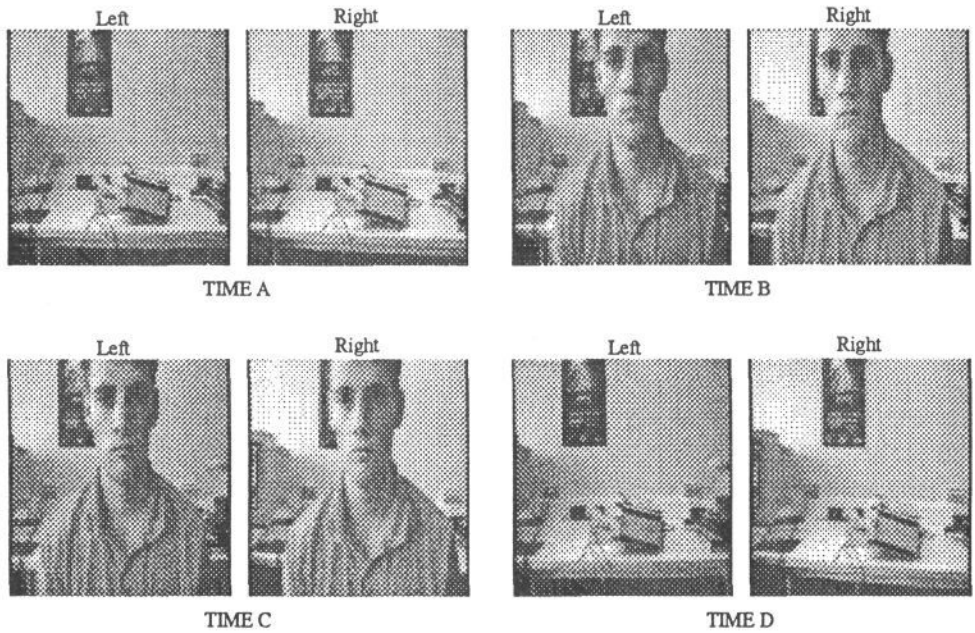


Figure 4: Selected examples of stereo images captured during vergence experiment in figure 5: (a) initially verged on background, (b) subject jumps into foreground, (c) verged on subject, (d) subject moves away.

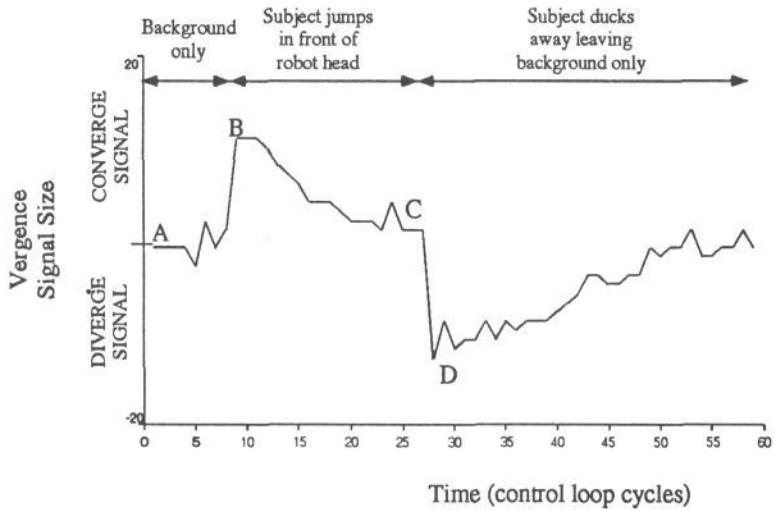


Figure 5: Real data for vergence control signal for example in figure 4 showing times A,B,C and D. Three vergence signal steps approximately correspond to 1 pixel of the original 128 x 128 images.

Stereo matching algorithms contain fusional limits for the maximum allowable disparity; the vergence mechanism simply attempts to maximise the amount of the image pair which can be fused.

4.1 Algorithm Design

To meet the requirement for fast execution, our vergence algorithm makes use of a foveated pyramid architecture, see figure 4. After pyramid formation, the stereo pair is blurred at each pyramid level using a small sigma gaussian function. This is necessary in order to achieve left-right matches at each level. Next, correlation is carried out between the image pair in the region of a small fixed-size window which can be moved from left to right. This process continues as a search operation through the pyramid and produces a vergence control estimate. The control loop for the reflex effectively moves the cameras so as to null this estimate.

This algorithm is based on the *Multiple Scale Signal Matcher* algorithm developed for stereo and motion correspondance [4]. On the SUN Sparcstation, vergence signals are calculated within the control loop in approximately 100ms. Further details of this algorithm and results for various test images are given in [5].

4.2 Performance

The performance of the vergence reflex is illustrated by the sequence of stereo image pairs in figure 4 and the associated graph in figure 5. This sequence shows the robot head initially verged on a point in the background of a typical lab scene, figure 4(a). Next, a person moves into the foreground, figure 4(b) and the cameras converge accordingly, figure 4(c). Figure 4(d) shows the cameras still verged on the foreground point after the subject has moved away. Finally, the cameras return to the original vergence point in the background. These five points are identified on the graph in figure 5 which plots the vergence control signal against time in control cycles.

References

- [1] J.Y. Aloimonos and A. Bandyopadhyay, Active Vision, Proceedings of the First International Conference on Computer Vision, pages 35-54, IEEE Computer Society Press, 1987.
- [2] R. Bajcsy, Active Perception vs Passive Perception, Proceedings of the Third Workshop on Computer Vision, pages 55-59, IEEE Press, 1985.
- [3] R.H.S. Carpenter, Movements of the Eyes, Pion London, 1977.
- [4] Z. Jin and P. Mowforth, A Discrete Approach to Signal Matching, Research Memo TIRM-89-036, The Turing Institute, Glasgow, 1989
- [5] P. Mowforth, P. Siebert, Z. Jin and C. Urquhart, A Head called Richard, Proceedings of the British Machine Vision Conference, pages 361-366, BMVA, 1990.