A binocular robotic head system with torsional eye movements

Michael Jenkin*, Evangelos Milios*, John Tsotsos†, Brian Down†

*Dept. of Computer Science, York University,
North York, Canada, M3J 1P3

†Dept. of Computer Science, University of Toronto,
Toronto, Canada, M5S 1A4

Abstract

This paper presents the hardware and software designs for TRISH (The Toronto IRIS Stereo Head). TRISH is a robotically controlled binocular camera mount, consisting of two fixed focal length colour cameras with automatic gain control forming a verging stereo pair. TRISH is capable of version (rotation of the eyes about the vertical axis so as to maintain a constant disparity), vergence (rotation of each eye about the vertical axis so as to change the disparity), pan (rotation of the entire head about the vertical axis), and tilt (rotation of each eye about the horizontal axis). One novel characteristic of the design is that each camera can rotate about its own optical axes (torsion). Torsion movement makes it possible to minimize the vertical component of the two-dimensional search which is associated with stereo processing in verging stereo systems. In addition to the physical robotic system, TRISH also incorporates a real-time video processing subsystem capable of accepting and processing the images generated by the head

1 Introduction

Active Vision[2] is a research paradigm inspired by the structure of biological vision systems[7,3], and has been shown to have benefits for a number of visual tasks[12]. To implement this paradigm in practice, a particular experimental apparatus is required to provide control over the acquisition of active image data. One approach considered by a number of researchers has been the construction of robotic "heads" which provide mechanisms for modifying the geometric or optical properties of the sensors under computer control. Several research groups have built robotic heads subject to different design criteria, and as lens, motor and controller technology itself progresses, more and improved designs appear (for example, see [4,8,1,9,6,11,5]).

Each head has been built to specific design goals. Some heads have been designed to have motions similar to the human visual system[9], while other have been designed to have motion speeds similar to the human visual system[4], while others have no biological motivation in the design at all. The heads also differ in terms of their compactness. For example, the Rochester head requires a robotic arm to provide essential head motions, while Eklundh's head is a self contained unit. A major motivation behind developing TRISH was the need to equip a mobile platform with colour binocular vision with a controllable vergence. As part of the IRIS project, the University of Toronto and York University are developing a mobile platform equipped with a robotic arm, capable of interacting with a tabletop environment. The robot (known as "PLAYBOT") will maneuver around the edges of the table, and will reach onto the table using its arm to manipulate objects on the table surface. The robot's sensors are completely vision based. The robot must be able to visually acquire objects in its workspace, and to manipulate them with the arm. Each of these tasks requires a high accuracy of positional information. This information is normally unavailable from a single pair of stereo cameras with a fixed geometry, and existing heads do not offer the combination of speed, accuracy, compactness, and weight limitations required by PLAYBOT. In order to allow the binocular head to be easily integrated within PLAYBOT the head design must be self contained, with a well-specified interface to our existing hardware environment, and it must be possible to mount the head on our mobile platform.

As a robotic device, TRISH consists of three integrated but distinct subsystems.

1. The mechanical design of the head. How the various components are physically connected, and what motors and encoders are used.

2. The robotic controller. The controller operates at various levels, from the low level closed loop control of the motors that make up the head, to higher level behaviors such as fixation and tracking operations.

3. The video subsystem. Robotic heads, almost by definition, provide a large steam of data which is not directly used by the head, but is what the head was designed to obtain.

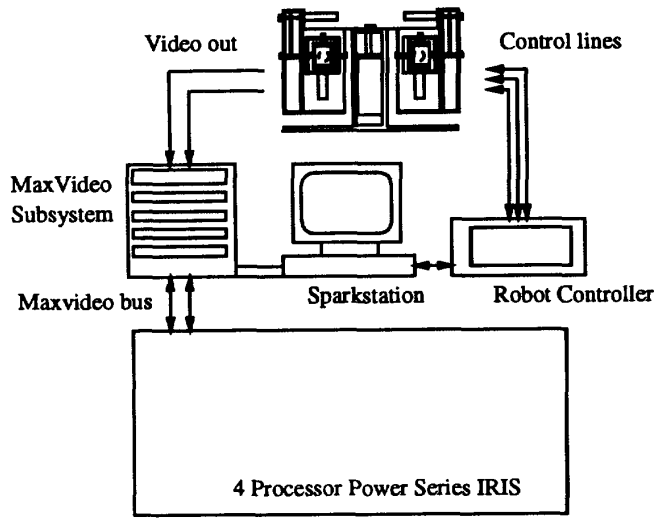

Figure 1: General structure of the TRISH system

Although each of these subsystems will be described as independent entities, the overall design of the head requires that the components work well in concert. The general structure of the TRISH system is shown in Figure 1.

2 Head Design

TRISH is a seven degree of freedom robot. Each of the two eyes are capable of torsion movements (rotation about their optical axis), and vergence/version movements (rotation about the vertical axis). The two eyes may be tilted (raised and lowered) independently, and the entire head may pan in the horizontal plane. The head is designed to roughly mimic the gross eye motions available in human binocular vision. Although the two eyes can be raised or lowered independently, in typical operation TRISH will raise and lower the eyes in concert. Thus although TRISH has 7 DOF, typically it will operate as though it had 6. A schematic of TRISH is shown in Figure 2

The head is driven by seven different DC motors, geared down to drive the various joints in the head. Each motor is equipped with a shaft encoder (optical except the torsion motors, which are equipped with a magnetic encoder due to their small size), which is used by the controller to position each shaft. Each motor/encoder pair is controlled by a single board computer, and the seven single boards are connected together and accept commands via a serial line from external processors/controllers. The mechanical parts for the head were either milled locally out of aluminum, or were constructed from off-the-shelf mechanical parts. The low level controller was provided by the motor supplier.

Figure 4 shows a picture of the front view of the head mounted on a flat surface. The head is roughly 62cm across, 52cm high and about 20cm deep. All motors except the tilt motors are "direct-drive", i.e. the motors are placed directly on the axis of rotation they control. This choice is not the most economical in terms of space, but simplifies the construction and adjustments to the hardware. Considerations in the mechanical layout include the maximum radial and axial load of the motors, and strength of the aluminum parts. Below we summarize the low-level issues we had to address in the design of the torsion, vergence, tilt and pan assemblies.

Torsion assembly The heart of the torsion enclosure is a miniature colour Panasonic camera (model GP-KS102 [10]), mated to a fixed focus lens with $f = 7.5mm$ (Panasonic lens GPLM3T). The miniature Panasonic camera is a tube, roughly 1.8cm in diameter, about 5cm long. It is secured to its enclosure via two aluminum rings, each of which holds the camera in place via three set screws, placed at 120 degree intervals. By adjusting these six screws, it is possible to adjust the orientation of the camera, and its position in the plane defined by the vergence and tilt axes. Adjusting the position of the camera along the torsion axis is achieved by making the screw holes of the L-shaped bracket elongated instead of circular. These mechanisms allow accurate positioning of the camera with respect to the drive motor. The two aluminum rings are in turn connected via an L-shaped bracket to a DC motor mounted directly behind the camera. As the motor turns, the camera is rotated about its optical (torsion) axis. The camera and lens together weigh 36g. A sketch of the torsion assembly is given in Figure 3

Vergence assembly The vergence motor, mounted at the bottom of the torsion assembly, rotates a ver-

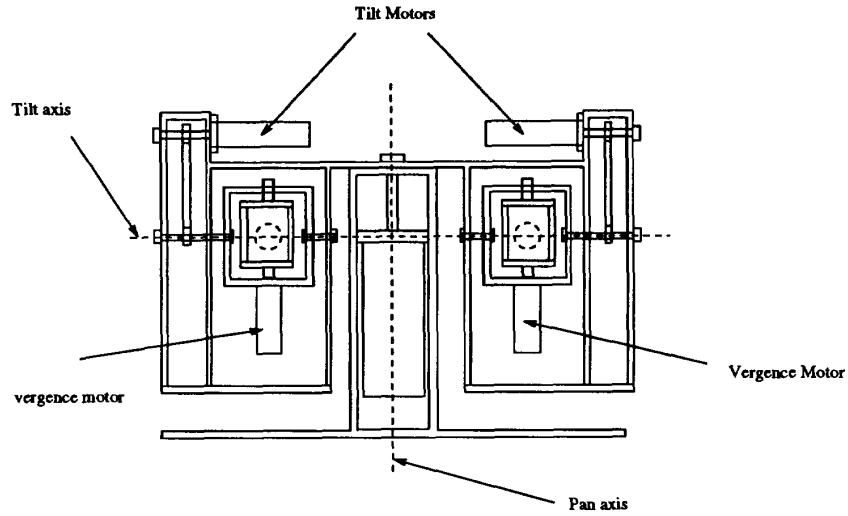

Figure 2: Front view of TRISH showing pan and tilt units

tical shaft on which is mounted the torsion enclosure. The bottom end of the rotating vertical shaft is driven by the motor, while the top end is allowed to rotate freely. A thrust bearing is used here to absorb much of the axial load on the vergence motor. An important consideration here was the depth of the rectangular vergence frame, so as to allow maximum range of vergence/version without compromising strength.

Tilt assembly Each individual eye rotates on a horizontal shaft, which is connected via a gear drive belt and pulleys to a DC motor mounted above the eye. The motors are mounted in this way to reduce the overall width of the head. A third pulley is used to adjust the tension of the belt.

Pan assembly The head can be made to rotate about its neck via the "pan" motor which is housed inside the neck. A thrust bearing at the joint between the head and the neck helps to reduce the axial load on this unit.

Balancing the various assemblies, so that they are in neutral balance (i.e. so that they can remain stationary in any position without requiring the application of torque) is a basic requirement. Balancing is an issue for the tilt axis. It may be necessary in the future to balance the tilt assembly by adding counterweights. Balancing occurs naturally in the pan, torsion, and vergence axes.

2.1 Performance

In the design of TRISH, it was desirable to endow the head with highly accurate and blindingly fast responses at a low cost. Physical limits place this goal out of reach, and it was necessary to restrict the capabilities of the robot in order to be able to build it from off the shelf components. As a design goal, we decided to attempt performance characteristics that were near the performance of the human visual system. Table 1 summarizes the ideal design specifications of our head (human performance) and TRISH's theoretical limits. Values for which no good biological values could be obtained have been left blank. As each of the components of TRISH were designed, specific motor, gearbox, and shaft encoder choices were forced to be made based on existing hardware. As a result, it was not possible to specify arbitrary velocities or accuracies for each joint. As a result, TRISH's final limiting velocities and accuracy of movement were limited to a small number of possible motor, gearbox, and shaft encoder combinations. One unfortunate limitation of the possible shaft encoders for motors under consideration for TRISH was the lack of an index pulse. An index mark or pulse would have given the controller an absolute position measure of the rotation of the drive shaft, rather than just a relative rotation. This has implications for calibrating the head.

3 Robotic Controller

The mechanical components of TRISH are driven by DC motors equipped with optical shaft encoders. Each of the motors drive appropriate shafts via gearboxes and (possibly) drive belts. Each of the seven motor/encoder pairs is driven by its own controller (two axes per controller). A commercial trapezoidal based positional controller (Micro Mo MCX-02EYU01) is used to provide the most primitive level of control of the head. This controller provides an RS232 interface to a host (either a terminal or in our case a workstation), and incorporates a primitive Basic interpreter to allow some small part of a larger system to reside on the controller. A more sophisticated controller resides on the workstation and communicates via this RS232 line to the MCX-02EYU01

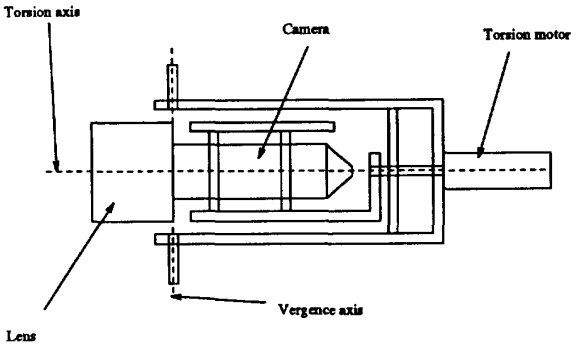

Figure 3: The torsion assembly

unit. A controller with an RS232 interface was chosen so that it would be easy to interface the controller within the general computing environment available in the lab.

Unfortunately, this link is a bottleneck in terms of communication. If no program is downloaded into the controller, then a sequence such as

```
V=2257:W=10:X=3:y#=10000:INIT:MOVE
```

would be required to move one motor to an encoder count of 10000 with a maximum velocity of 10 encoder counts per clock tick with a maximum acceleration of 3. Operating multiple axes simultaneously is quite difficult. Rather than operating this way, a partial solution to this bottleneck has been achieved by downloading a small basic program which implements some common operations directly onboard the controller. The remote workstation can then send much fewer characters to effect the same motions from the head.

The higher level controller provides an easy to use X interface for demonstration purposes, as well as a user callable library of routines to effect different motions of the head.

4 Video Subsystem

TRISH generates colour binocular video data at frame rates. If a binocular head is to be used within an active vision paradigm, then it is imperative that the video data be available to higher level processing units as quickly as is possible. TRISH utilizes two very different computing technologies in order to provide sufficient computing resources so that the video imagery obtained with the head can be used for its control. The first of these subsystems is based around a SUN Sparkstation connected to a VME chassis within which resides two MaxVideo 20 image processing boards, two MaxVideo Digicolor boards (color digitizers), and two MaxVideo FIR (convolver) boards. The analog colour video from TRISH is fed directly to the Digicolor boards where it is digitzed. From the Digicolor boards the digitized signal can be fed to the MaxVideo processing units connected within the same VME chassis. The Sparkstation provides external control of this video processing subsystem as well as operating as the host for TRISH's controller. Although the MaxVideo system provides a reasonable amount of compute processing in an image processing environment, it does not provide a powerful general purpose compute environment for less image specific operations. A 4 processor power series IRIS computer is used to provide these more general purpose computations. In order to reduce the bottleneck between the MaxVideo system and the IRIS system, the MaxVideo video bus is broken out of the Sparkstation VME card cage and is fed to two MaxVideo RoiStore boards (frame buffers) which reside within the IRIS Computer. This provides the output of the MaxVideo image subsystem to the IRIS general purpose computing system at bus rates. Feedback from the IRIS to the Sparkstation (and hence to TRISH and the MaxVideo system) is provided through standard UNIX remote procedure calls/networking operations.

5 Discussion

TRISH went through a number of different tentative designs before arriving at the design presented here. Earlier designs considered using stepper motors for motive power. Gearboxes for the stepper motor design proved too expensive, and the stepper motors were replaced with DC motors with optical shaft encoders. The change from stepper motors to DC motors mandated major modifications of the design of the controller. Some earlier designs proposed a PC or VME card based controller, rather than the stand alone controller connected via a serial connector to a remote host described here.

The size and shape of the motors and the need for a compact design forced many of the design decisions. In one design, for example, the tilt motors were also direct drive, extending away from the body of the head like ears. This resulted in a design that was an additional 20cm wide. This head was (a) too large for the intended application and (b) required a stronger motor for the pan unit due to the increased inertia of the entire head. By placing the motors above the head and driving them through belts a more compact design was obtained. A similar decision resulted in

Figure 4: View of TRISH fixating on a target. Note the low level controller in the background.

the pan motor being mounted inside the neck which protrudes into the upper part of the head.

Acknowledgments
The IRIS Project received financial support from the Institute of Robotics and Intelligent Systems (IRIS), one of the Government of Canada's Networks of Centres of Excellence. J. Tsotsos is the CP-UNITEL Fellow of the Canadian Institute of Advanced Research. Financial support from NSERC Canada through individual operating grants to each of the authors is gratefully acknowledged.

References

[1] A. Abbott. *Dynamic integration of depth cues for surface reconstruction from stereo images.* PhD thesis, Electrical Engineering, Univ. of Illinois at Urbana-Champaign, 1990.

[2] R. Bajcsy. Active perception vs. passive perception. In *Third IEEE Workshop on Vision*, pages 55–59, 1985.

[3] D. Ballard. Eye movements and visual cognition. In *Workshop on Spatial Reasoning and Multisensor Fusion*, pages 188–200. Morgan Kaufmann, 1987.

[4] D. Ballard and A. Ozcandarli. Eye fixation and early vision: kinetic depth. In *International Conference on Computer Vision*, pages 524–531, 1988.

[5] H. Christensen. Auc robot camera head. In *Applications of Artificial Intelligence X: Machine Vision and Robotics, Volume 1708, Proceedings of SPIE, The International Society for Optical Engineering, 22-24 April, Orlando, Florida*, pages 26–35, 1992.

[6] J. Crowley, P. Bobet, and M. Mesrabi. Layered control of a binocular camera head. In *Applications of Artificial Intelligence X: Machine Vision and Robotics, Volume 1708, Proceedings of SPIE, The International Society for Optical Engineering, 22-24 April, Orlando, Florida*, pages 47–61, 1992.

[7] I. Howard. *Human visual orientation.* John Wiley and Sons, New York, 1982.

[8] E. Krotkov. *Active Computer Vision by Cooperative Focus and Stereo.* Springer-Verlag, New York, 1989.

[9] K. Pahlavan and J-O. Eklundh. A head-eye system - analysis and design. Technical report, Computational Vision and Active Perception Laboratory, Royal Institute of Technology, S-100 44 Stockholm, Sweden, October 1991.

[10] Panasonic Communications and Systems Company, Industrial Camera Division. *Panasonic Industrial CCD Microcameras GP-KS102, Color and GP-MS112, Black and White.*

[11] J. Pretlove and G. Parker. Lightweight camera head for robotic-based binocular stereo vision: an integrated engineering approach. In *Applications of Artificial Intelligence X: Machine Vision and Robotics, Volume 1708, Proceedings of SPIE, The International Society for Optical Engineering, 22-24 April, Orlando, Florida*, pages 62–74, 1992.

	Human Performance			TRISH design		
	range (deg)	velocity (deg/sec)	accuracy (deg)	range (deg)	velocity (deg/sec)	accuracy (deg)
Pan	±90			±180	54	0.003
Tilt	±45	800	2	±45	54	0.003
Torsion	±20			±180	180	0.09
Vergence/Version	±45	800	2	±35	100	0.0019

Table 1: Human and TRISH design performances

[12] J. K. Tsotsos. On the relative complexity of active versus passive visual search. *International Journal of Computer Vision*, 7(2):127–141, 1992.