DESIGN AND PERFORMANCE OF TRISH, A BINOCULAR ROBOT HEAD WITH TORSIONAL EYE MOVEMENTS

EVANGELOS MILIOS, MICHAEL JENKIN
Department of Computer Science, York University
North York, Canada M3J 1P3
Email: eem@cs.yorku.ca, jenkin@cs.yorku.ca

JOHN TSOTSOS
Department of Computer Science, University of Toronto
Toronto, Canada M5S 1A4
Email: tsotsos@cs.toronto.edu

We present the design of a controllable stereo vision head. TRISH (The Toronto IRIS Stereo Head) is a binocular camera mount, consisting of two fixed focal length color 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 axis (torsion). Torsional movement makes it possible to minimize the vertical component of the two-dimensional search which is associated with stereo processing in verging stereo systems.

Keywords: Active vision, stereo vision, stereo ranging, robot vision.

1. INTRODUCTION

Active Vision[1] is a research paradigm inspired by the structure of biological vision systems,[2,3] and it has been shown to have benefits for a number of visual tasks.[4] 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.

Krotkov[5] reports on an agile camera system with the ability to translate on a gantry (two translational degrees of freedom (DOF)), and the capability of pan and tilt (two rotational DOF), and vergence (one DOF). Each camera has controllable aperture, zoom, and focus (three DOF for each camera), giving a total of 11 DOF. The four positional DOF are implemented by AC servomotors, while the rest are DC servomotors. Another early head is the Harvard head[6], which has pan, tilt, and vergence control of both cameras, and aperture and focus control for each camera.

International Journal of Pattern Recognition and Artificial Intelligence Vol. 7 No. 1 (1993) 51–68

Ballard[7] reports on a binocular robotic head attached to a Unimation PUMA arm. Each camera moves independently in yaw, but is yoked to the other in pitch. One of this head's goals was to perform eye movements on the time scale of the human eye (10–30 msec). Utilizing stepper motors, this head reports speed of up to 150°/second.

Abbott[8] reports on a stereo head built using off-the-shelf components. The head consists of two cameras which can be independently verged (two DOF), while the combination is mounted on a pan and tilt unit (two DOF), which in turn can be translated horizontally (one DOF). The cameras are equipped with motorized lenses with three DOF each (zoom, focus, and aperture control). This head has a total of 11 DOF. Stepper motors were used for the positional DOF, while the lenses were equipped with DC motors.

More recently, Pahlavan and Eklundh[9–12] report the construction of a 13-DOF stereo head. Aperture, zoom, and focus control of each camera lens involves six DOF, pan and tilt for each individual eye (four DOF), baseline control (one DOF), and pan and tilt for the whole head (two DOF for neck movements). This design uses stepper motors for the seven positional DOF, and DC motors for the lenses.

A stereo head has been designed by LIFIA, Institut Nationale Polytechnique de Grenoble, with controllable focus and aperture, and independent vergence control for each camera.[12,13] A head with similar specifications has been designed at the University of Surrey.[14] Both of the above heads are designed for mounting on the end-effector of a robot manipulator. Another stereo head by Aalborg University, Denmark includes independent control of vergence, zoom, focus, and aperture of each camera, and joint control of pan, tilt, and baseline of the stereo pair.[12,15]

Recently, a stereo head has been demonstrated by the National Institute of Science and Technology (former National Bureau of Standards). Its main novelties are the use of a third low-resolution camera in the center, and the use of specialized motors to achieve speeds and accelerations comparable to those of the human eye, something that cannot be achieved currently with ordinary off-the-shelf motor hardware.

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,[7] 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 the Stockholm head is a self-contained unit. A major motivation behind developing TRISH was the need to equip a mobile platform with color binocular vision with 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.

2. THE GEOMETRY OF CONVERGENT BINOCULAR STEREO

Ongoing research in stereo vision by many different researchers at many different institutions has advanced the understanding of depth from disparity to the point at which robots can be built which utilize stereo as their primary source of environmental information. One problem with stereo information is that the accuracy of the measurements is limited by the way in which the cameras are positioned. In a non-convergent stereo system (one in which the optical axes of the two eyes are parallel), the only way in which higher precision can be obtained is by extending the baseline. If the optical axes are not constrained to be parallel then the eyes can be rotated to fixate on different structures in the environment. Given a maximum allowable baseline, convergent binocular vision promises a higher accuracy of measurement, but there is an associated cost: the task of identifying interocular correspondences goes from a simple horizontal search to a search along a non-horizontal "epipolar" line of variable slope.

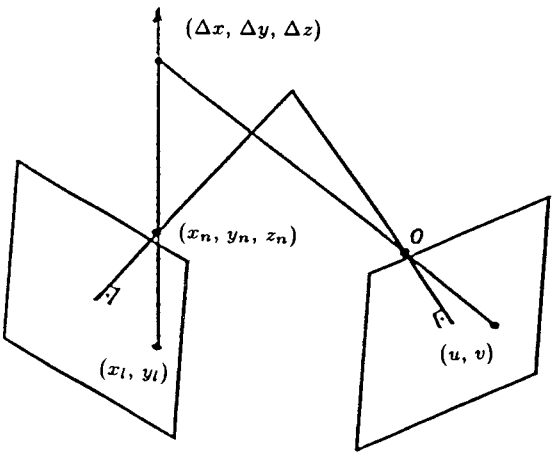


Fig. 1. Epipolar lines.

Consider the convergent binocular geometry shown in Fig. 1. A point, as viewed in the left eye is projected (via a perspective transformation) through the nodal point of the eye onto the image plane. As only the projection of this point can be measured, the true position of the point is not known, only the direction in which it lies. Now consider how the world is viewed by the right eye. Establish a 3-D

coordinate system at the nodal point of the right eye. This system is parallel to and aligned with the right eye image plane. We call the nodal point of the left eye in the 3-D coordinate system defined by the right eye (x_n, y_n, z_n). Any point which projects into the left eye lies along a line which passes through the left eye nodal point and (x_l, y_l), the location at which the point projects onto the left image plane. Let $(\Delta x, \Delta y, \Delta z)$ be a direction vector joining these two points in the right eye coordinate system. The right eye projection of this line onto the right eye image plane is given by

$$u = -f(x_n + \lambda \Delta x)/(z_n + \lambda \Delta z) \; v = -f(y_n + \lambda \Delta y)/(z_n + \lambda \Delta z) \qquad (1)$$

where λ is a free parameter. The set of points (u, v) is a line with slope given by

$$\frac{dv}{du} = \frac{z_n \Delta y - y_n \Delta z}{z_n \Delta x - x_n \Delta z}. \qquad (2)$$

dv/du defines the slope of the line which must be searched for the corresponding point in the right eye for the candidate point in the left. If the left and right eyes are raised and lowered together and if the eyes are separated by a distance d, then $(x_n, y_n, z_n) = (-d \sin \theta_r, 0, d \cos \theta_r)$, where θ_r is the rotation of the right eye in the plane containing the two eyes and is the angle between the positive z axis and the line passing through the nodal point of the two eyes. We will assume this geometry in the remainder of this paper.

If we are interested in recovering disparities over the entire visual field then knowledge of the position and slope of the epipolar line can be used to constrain the search for corresponding features. On the other hand, if we are interested in recovering disparities near the fixation point, then torsional eye movements can be exploited to drive the epipolars to be horizontal in the right image plane. Substituting the expression for (x_n, y_n, z_n) in Eq. (2) results in

$$\frac{dv}{du} = \frac{\Delta y \cos \theta_r}{\Delta x \cos \theta_r + \Delta z \sin \theta_r} \qquad (3)$$

which is independent of the eye separation d.

If (x_l, y_l) is the projection of a particular point in the left eye with direction vector (as seen by the right eye) of $(\Delta x, \Delta y, \Delta z)$, and if the left eye makes an angle θ_l with the line separating the two eyes then the direction vector in the left eye coordinate system is

$$(\Delta x, \Delta y, \Delta z) = (-x_L, -y_L, f). \qquad (4)$$

To obtain an expression of (dv/du) in terms of measurable quantities, we perform a coordinate transformation on the above expression, in order to express $(\Delta x, \Delta y, \Delta z)$ in the right eye coordinate system. The coordinate transformation is a rotation by an angle equal to $\theta_R - \theta_L$. Performing this coordinate transform

and substituting the resulting $(\Delta x, \Delta y, \Delta z)$ into Eq. (3), we obtain the following expression:

$$\frac{dv}{du} = \frac{y_l \cos\theta_r}{x_l \cos\theta_l - f \sin\theta_l}. \tag{5}$$

The maximum torsion that the system will have to cope with can be found by maximizing $\tan^{-1}(dv/du)$ for the parameters of the system. If the eyes are to verge $\pm 35°$ from straight ahead, and with a CCD array of 7×5 mm^2 and a focal length of 7.5 mm, then $|dv/du| < 0.43$, yielding a maximum slope of about 23°. The worst values are obtained when the eyes diverge and when data points are considered at the edges of the CCD array.

In addition to making epipolar lines parallel to pixel rows in the stereo correspondence problem, torsion has many other applications.

(a) In calibrating the stereo head, it can be used to correct for alignment errors in head construction. It can be used in determining the true centre of the CCD array.

(b) In stereopsis, it can be used for implementing various active sensing strategies.

Suppose that instead of doing full field stereopsis which is inherently difficult (see Ref. 16 for some of the details), we consider performing stereopsis over a limited range in both space and disparity. When the eyes converge on a particular point in space, the fixation point and the nodal points of the two eyes define a circle (in the plane of these three points) over which the horizontal disparity is zero. This is the longitudinal horopter (also known as the Vieth–Müller circle[2]), and it defines a cylinder of zero horizontal disparity in space. Now consider a fast hardware implementation of a stereopsis algorithm (such as in Ref. 17) set up to operate with images that have been roughly pre-shifted to have a zero horizontal disparity and a zero vertical disparity. By verging the two eyes, different circular regions in space are mapped to this zero disparity region and processed. Now consider the effect of torsional eye movements. If the two eyes are both aligned vertically, then the cylinder of zero horizontal disparity is perpendicular to the plane defined by the nodal points of the two eyes and the fixation point. If the two eyes are not aligned vertically, then the zero disparity region distorts to cover surfaces which are not perpendicular to this plane. Cyclodisparities can be used as a cue to local surface slant.[18]

Instead of physically rotating the cameras, it would also be possible to rotate the images after they have been digitized. One potential problem with rotating the image after digitization is that of sampling. The small rotations typically associated with torsional eye movements may give rise to difficult sampling problems which may require a large amount of effort to correct.

3. HEAD DESIGN

TRISH is a seven degree of freedom robot. 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 seven DOF, typically it will operate as though it had six.

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 available off-the-shelf. The low level controller was provided by the motor supplier.

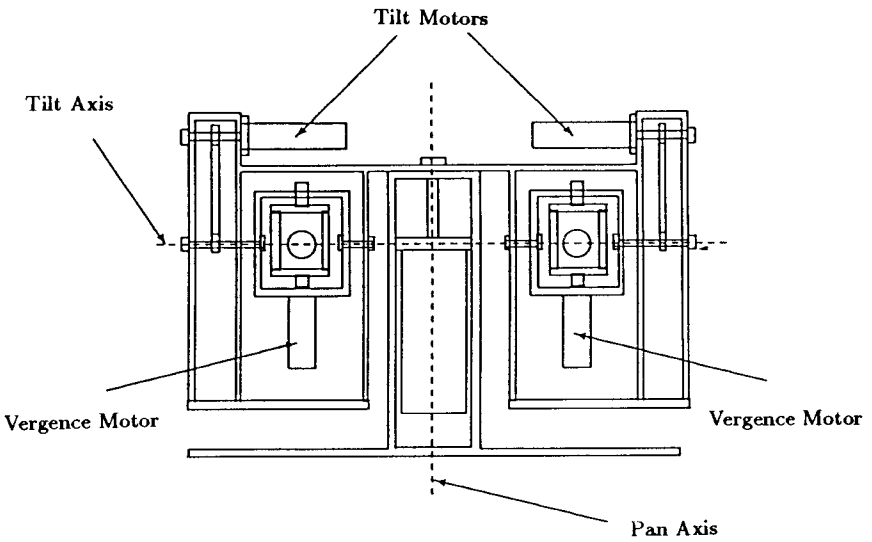


Fig. 2. Front view of the head showing pan and tilt units.

Figure 2 shows the front view of the head mounted on a flat surface. The head is roughly 62 cm across, 52 cm high and about 20 cm deep. All motors except the tilt 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. A sketch of the torsion enclosure is given in Fig. 3. The heart of the torsion enclosure is a miniature color Panasonic camera (model GP-KS102[19]),

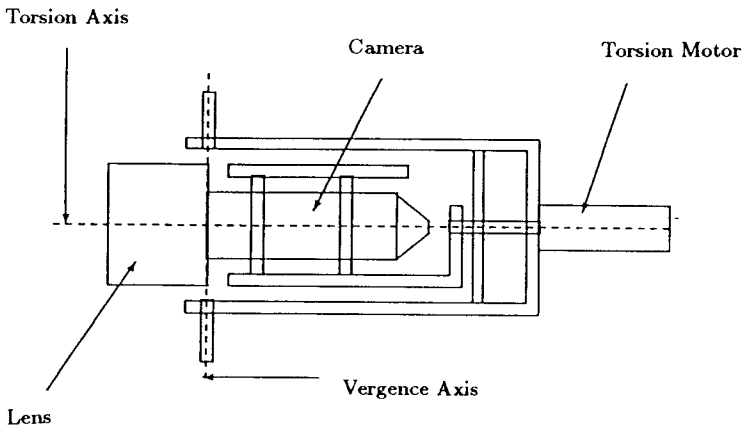


Fig. 3. Side view of torsion enclosure.

mated to a fixed focus lens with $f = 7.5$ mm (Panasonic lens GPLM3T). The miniature Panasonic camera is a tube, roughly 1.8 cm in diameter, about 5 cm 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 36 g.

Vergence assembly. The vergence motor, mounted at the bottom of the torsion assembly, rotates a vertical 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. they can remain stationary in any position without requiring the application of torque) is a basic requirement. Balancing is a potential issue for the tilt axis. However, it was not found necessary in practice to balance the tilt assembly by adding counterweights. Balancing occurs naturally in the pan, torsion, and vergence axes.

4. PERFORMANCE

In the design of TRISH, it was desirable to endow the head with highly accurate and 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. In the following, we obtain approximate measures of the characteristics of the human visual system and relate them to TRISH's design limitations.

Retinal resolution. We would like to compare the resolution of the human eye with that of our camera. We used two different informal ways of estimating the resolution of the human eye. For the first estimate, we use a count of receptor cells in a photograph of the retina. We count a total of about 50 receptor cells (rods and cones) over a retinal area of approximately 50 μm $\times$50 μm (from Fig. 2.1, p. 14 of Ref. 20). This leads to an average distance between receptor cells equal to $d = 50\ \mu\text{m}/\sqrt{50} = 7\mu\text{m}$. For the second estimate, we use the empirical fact that two objects can be resolved if the viewing angle between them and the eye is at least 1' of a degree.[21] Since two objects can be resolved if their images fall on two different receptor cells, we can use this information to compute the average distance between receptor cells as follows: Assume an average focal length of the eye of $f = 30$ mm. Then with $\sigma = 1' = \pi/(60*180)$ rad, we obtain for the average distance between receptor cells $d = \sigma f = 8.7\ \mu$m, which is comparable to the distance obtained above. Our color CCD micro camera has an array of size 7 mm $\times$ 5 mm with 682 $\times$ 492 pixels squared. This yields an average receptor cell size of 10 μm.

Baseline. The human visual system baseline is about 6 cm. This is much smaller than it will be possible to realize due to physical limitations (size of cameras, etc.) in a binocular robotic head built from existing off-the-shelf components. If the cameras are to verge then there has to be somewhere for the cameras and lenses to move to, and thus the smallest realistic baseline is about 32 cm.

Vergence. In calculating the vergence/version speed of the human eye, one can rely on a variety of experiments. Experiments with the fusion of random-dot patterns of periodically varying disparity[22] have shown that the human eye is capable of tracking temporal frequencies of up to 1 Hz for a sinusoidally varying disparity of amplitude up to 30 arc min (with less than 25° phase lag). This corresponds to a vergence speed of 30 arc min/second or 0.5°/second. Much higher speeds have been measured for version, namely of the order of 800°/second. The accuracy of human eye movements can also be measured under different conditions. Humans use

sophisticated control mechanisms to drive eye movements. For example, there is evidence that humans deliberately undershoot targets when making eye movements.[2] As a measure of accuracy, humans are capable of keeping the eyes fixated on an acquired target in the dark with an error of 2°.

The maximum vergence required for TRISH can be computed from the minimum expected distance to objects in the environment and the maximum baseline size. If b is the baseline (distance between the two cameras) and r is the distance of an object lying on the axis of symmetry of the two cameras, then the required vergence ϕ is given by $\phi = \arctan(b/2r)$. The minimum object distance for fixed focus cameras is 20 cm, 20 cm, and 30 cm for focal lengths of 4.8 mm, 7.5 mm, and 9 mm respectively. For $r = 20$ cm and $b = 30$ cm, we get $\phi = 35°$. For $r = 30$ cm and $b = 30$ cm, we get $\phi = 26.6°$. By spacing the eyes 32 cm (or more) apart, it will be possible to have sufficient space to have this range of vergence.

Torsion. Humans (with training) can obtain torsional eye movements of ±20°.[23] TRISH's design allows for a full 360° eye torsion, limited only by constraints on the video cable. Assuming a maximum change in vergence of 40°, then for TRISH to be able to complete the maximum torsional eye movement at the same rate that it can complete the maximum vergence eye movement TRISH must have a vergence speed of at least 40/70 = 57% of the vergence speed.

The epipolar line we are trying to make horizontal consists of a single row of photoreceptor elements (a horizontal rectangular strip of length 7 mm and a height 5 mm/492 = 10 μm). The maximum allowed error in torsion is the angle formed by the horizontal and the diagonal of this rectangular strip, i.e. 0.08°.

Tilt. The human visual system can perform saccadic movements at a speed of 800°/second and can raise or depress the eyes by 40°–45°.[2]

Pan. It is desirable for the head to be able to make a 360-degree turn. This is considerably in excess of the range of motions available in human vision. It is not possible to accurately relate human head motions to TRISH's limited pan movements. The human neck is a complex positioning device, while TRISH is capable only of rotational head motions.

Table 1 summarizes the ideal design specifications (human performance) and TRISH's theoretical performance 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 considerations 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 as described below.

The specific combinations used in TRISH are listed in the Appendix, and the resulting velocity and accuracy limits are given in Table 1.

Table 1.

	Human performance			TRISH design		
	range (deg)	velocity (deg/second)	accuracy (deg)	range (deg)	velocity (deg/second)	accuracy (deg)
Pan	±90			±80	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

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. Gear-boxes 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 in 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 on the tilt axis, extending away from the body of the head like ears. This resulted in a design that was an additional 20 cm 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 moment of 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 the pan motor being mounted inside the neck which protrudes into the upper part of the head.

The current status of TRISH (October 5, 1992) is the following. TRISH has been milled and assembled (Fig. 4). Figure 5 shows images of the scene shown in Fig. 4 obtained from the left eye, the right eye, and the right eye after a torsional movement has been performed on it. Figure 6 shows the overall configuration of the TRISH computing and control environment. Preliminary experiments have been performed to determine actual performance and accuracy characteristics. The results of those experiments are reported in the Appendix. The overall conclusion is that the accuracy of the head movements are better than what can be perceived with cameras of a standard resolution (e.g. 682 × 492 pixels), provided that the last movement before arriving at a given angular position is always in the same direction (so as to eliminate the gear backlash). Special care has to be taken in adjusting the controller parameters to achieve maximum speed without instabilities

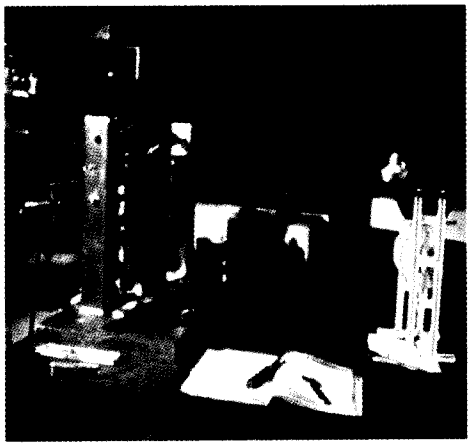


Fig. 4. A side view of TRISH in front of a toy set. The low level controller can be seen in the background.

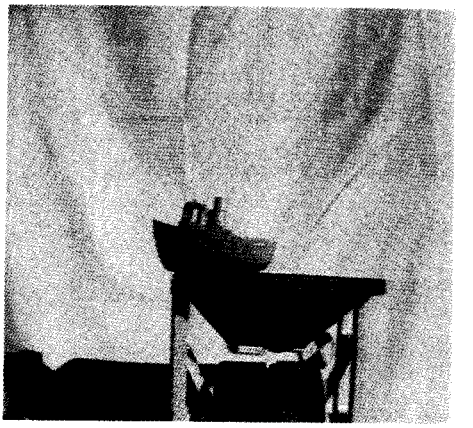
of the control system resulting in undesirable oscillations. One complicating factor here is that the inertial load on the various motors changes as the motor positions change, making an analytical study from a control theory viewpoint more difficult. Experiments are currently underway to (1) optimize the dynamic performance of the head movements, and to (2) use the torsional degrees of freedom in solving stereo matching, fixation, and visual tracking problems.

ACKNOWLEDGEMENTS

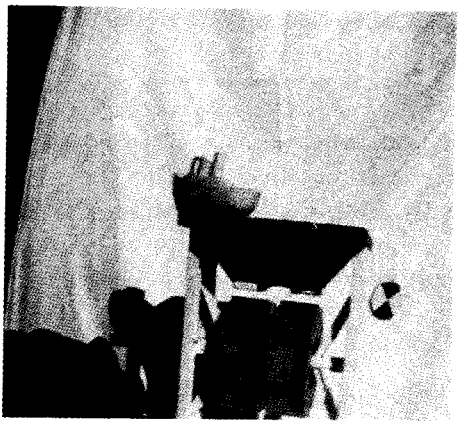
The IRIS Project received financial support from the Institute of Robotics and Intelligent Systems (IRIS), one of the Government of Canada's Networks of Centres of Excellence. J. Tsotsos is the CP-UNITEL Fellow of the Canadian Institute of Advanced Research. Financial support from NSERC Canada through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

REFERENCES

1. R. Bajcsy, "Active perception vs. passive perception", in *Proc. Third IEEE Workshop on Vision*, 1985, pp. 55–59.
2. I. Howard, *Human Visual Orientation*, John Wiley and Sons, New York, 1982.
3. D. Ballard, "Eye movements and visual cognition", in *Proc. Workshop on Spatial Reasoning and Multisensor Fusion*, St. Charles, Illinois, 1987, Morgan Kaufmann, pp. 188–200.
4. J. K. Tsotsos, "On the relative complexity of active versus passive visual search", *Int. J. Comput. Vision* 7, 2 (1992) 127–141.
5. E. Krotkov, *Active Computer Vision by Cooperative Focus and Stereo*, Springer-Verlag, New York, 1989.
6. N. Ferrier, "Harvard binocular head", in *Applications of Artificial Intelligence X: Machine Vision and Robotics, Vol. 1708, Proc. SPIE*, Orlando, FL, Apr. 1992, pp. 2–13.



(a)



(b)

(c)

Fig. 5. Images taken with TRISH: (a) left eye view, (b) right eye view, (c) right eye view after a torsional movement has been performed.

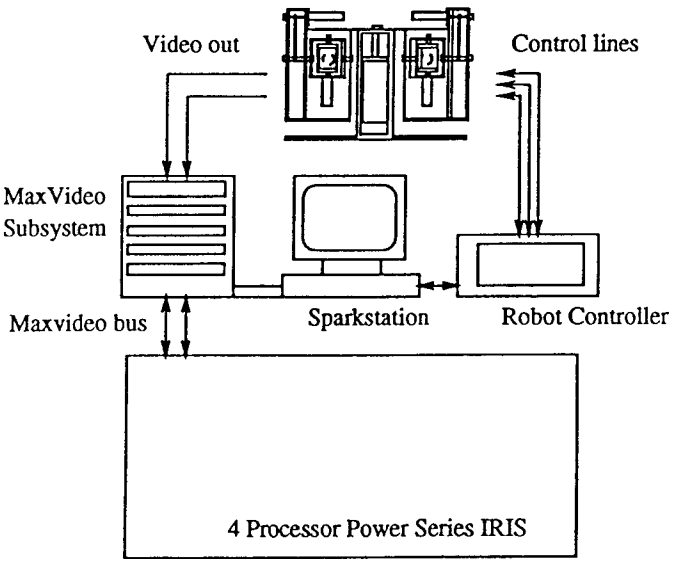


Fig. 6. The overall configuration of the TRISH computing and control environment.

7. D. Ballard and A. Ozcandarli, "Eye fixation and early vision: Kinetic depth", in *Proc. Int. Conf. on Computer Vision*, Tampa, Florida, 1988, pp. 524–531.
8. A. Abbott, Dynamic Integration of Depth Cues for Surface Reconstruction from Stereo Images, Ph.D. thesis, Electrical Engineering, Univ. of Illinois at Urbana-Champaign, 1990.
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, Oct. 1991.
10. K. Pahlavan, T. Uhlin and J.-O. Eklundh, Integrating Primary Ocular Processes, Technical report, Computational Vision and Active Perception Laboratory, Royal Institute of Technology, S-100 44 Stockholm, Sweden, Oct. 1991.
11. K. Pahlavan and J. Eklundh, "Head, eyes, and head-eye systems", in *Applications of Artificial Intelligence X: Machine Vision and Robotics, Vol. 1708, Proc. SPIE*, Orlando, FL, Apr. 1992, pp. 14–25.
12. J. Crowley, "Vision as Process", Technical report, ESPRIT Basic Research Action BR 3038/BR 7108, 1992.
13. J. Crowley, P. Bobet and M. Mesrabi, "Layered control of a binocular camera head", *Int. J. Pattern Recogn. Artif. Intell.* **7**, 1 (1993) 109–122.
14. 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, Vol. 1708, Proc. SPIE*, Orlando, FL, Apr. 1992, pp. 62–74.
15. H. Christensen, "A low-cost robot camera head", *Int. J. Pattern Recogn. Artif. Intell.* **7**, 1 (1993) 69–87.

16. M. R. M. Jenkin, Visual Stereoscopic Computation, Ph.D. thesis, Department of Computer Science, University of Toronto, 1988.
17. A. Jepson and M. R. M. Jenkin, "The fast computation of disparity from phase differences", in *Proc. CVPR 89*, San Diego, California, 1989, pp. 398–403.
18. I. Howard, "Cycloversion, cyclovergence and perceived slant", in *Spatial Vision in Humans and Robots*, eds. L. Harris and M. Jenkin, Cambridge University Press, in press.
19. Panasonic Communications and Systems Company, Industrial Camera Division. Panasonic Industrial CCD Microcameras GP-KS102, Color and GP-MS112, Black and White.
20. J. Dowling, *The Retina: An Approachable Part of the Brain*, The Belknap Press of Harvard University Press, Cambridge, Massachusetts, 1987.
21. H. Kuchling, *Taschenbuch der Physik*, Harri Deutsch, Frankfurt/Main, 1989.
22. L. Spillman and J. Werner, *Visual Perception: The Neurophysiological Foundations*, Academic Press, San Diego, CA, 1990.
23. R. Balliet and K. Nakayama, "Training of voluntary torsion", *Investigative Opthamology* **17** (1978) 303–314.
24. J. Shigley and C. Mischke, *Standard Handbook of Machine Design*, McGraw Hill Book Co., New York, 1986.
25. Winfred M. Berg Inc., *Berg Precision Mechanical Components Catalog*, East Rockaway, NY, 1992.

APPENDIX A. MECHANICAL DESIGN OF TRISH

The parts that were used in TRISH are shown in Table 2. All parts were supplied by Micro Mo Electronics Inc. of St. Petersburgh, Florida. Micro Mo also supplied a low level controller specifically adapted to each motor. Mechanical parts were purchased from BERG Precision Mechanical Components, East Rockaway, New York.

One aspect of the design that needs special attention are shaft couplings. The simplest solution and the one that leads to the simplest mechanical layout is the use of solid couplings (simply a piece of bar stock bored to receive two shafts). Solid couplings are easy to construct, and because of their rigidity, they help eliminate the movement in the transverse direction of the shaft attached to the load. The two shafts are attached to the solid coupling via set screws, and, therefore, they must be filed or ground flat where the set screws are placed to increase the coupling's torque-handling capacity. The problem with solid couplings, however, is that they do not tolerate any axial misalignment. We recommend the use of flexible couplings, instead of solid couplings.[24,25] In this case, additional support must be added for the shafts attached to the load, to prevent any transverse movement. Proper design and installation of couplings is of paramount importance, because any slippage of the shafts within the couplings will result in unreliable operation.

Another aspect of the mechanical design is balanced inertial load, i.e. having the axes of rotation pass through the center of mass of the rotated load. This leads to the neutral balance of all inertial loads, so that if zero torque is applied, the load can be at equilibrium at any angular position. In TRISH, only the load to the

tilt motors is unbalanced, which implies that part of the torque provided by these motors is used to keep the cameras in a horizontal position (they naturally tend to fall backwards so that they point at about 50° with respect to the horizontal). In our design, the axes of rotation are determined by the optics, so there is no way the load can be balanced by changing its position relative to the rotation axes.

Table 2.

Unit	Specification	
Torsion unit		
	Motor	Model 1624-E-024S
	Gearhead	Model 16/3 K185, ratio 262:1
	Encoder	Model HEM 1616, 15 parts per revolution
Vergence unit		
	Motor	Model 2230-F-024S
	Gearhead	Model 22/5, ratio 64:1
	Encoder	Model HEDS 5010, 500 parts per revolution
Tilt unit		
	Motor	Model 2842-S-024C
	Gearhead	Model 23/1, ratio 246:1
	Encoder	Model HEDS 5010, 500 parts per revolution
Pan unit		
	Motor	Model 3557-K-024C
	Gearhead	Model 38/1, ratio 246:1
	Encoder	Model HEDS 5010, 500 parts per revolution
Controller		MCX-02EYU01

APPENDIX B. COST OF TRISH

The cost of the head has been approximated as in Table 3 (in US dollars, paid between November 1991 and March 1992).

Table 3.

Milling of aluminum parts	$ 2800
Motors, encoders, gearboxes, controller	$ 8630
Cameras and lenses	$ 4700
Small mechanical parts	$ 500
Total	$16630

APPENDIX C. ACTUAL PERFORMANCE MEASURES OF TRISH

Accuracy. One full rotation (360 degrees) corresponds to the following number of quadrature encoder counts equal to 4 × the encoder parts per revolution for each

of the degrees of freedom as shown in Table 4.

Table 4.

	Encoder count per full rotation	Degrees per single encoder count
Torsion	15 720	22.5×10^{-3}
Vergence	128 000	2.8×10^{-3}
Tilt	492 000	0.725×10^{-3}
Pan	492 000	0.725×10^{-3}

The actual limitation, however, in the accuracy of the system is not lack of encoder resolution, but gear backlash (the amount by which the width of a tooth space exceeds the thickness of the engaging tooth). Typical backlash is of the order of 0.5 degrees, but it is measurable, and changes only slowly with time, as the gear teeth wear because of use. The result of backlash is that a position specified by a given encoder count (with respect to a reference or home position), actually represents a range of positions of the output shaft allowed by the gear backlash. However, if the end position is reached by the gear rotating in a constant direction (e.g. clockwise), then only one position out of the range of positions allowed is achieved. We confirmed this theory experimentally. The experiment was to place the camera in front of a calibration mark of size 5 cm and at a distance of 75 cm from it, and rotate the tilt, vergence and torsion unit "forward" and "backward" by the same number of encoder counts, so we theoretically go back to the same position. If the movement by which the motor reached its initial position was with a "backward" movement, we expect no backlash, whereas if the position was reached with a "forward" movement, we expect maximum backlash. We measure the disparity of the center of the calibration point in the two images (the center was found manually using an interactive image display program allowing cropping and zooming). In the case of torsion, we measure the difference in slope of a straight line before and after the move. Table 5 shows the result of these measurements.

The above measurement method cannot measure angle differences corresponding to image size smaller than a pixel. This minimum angle is approximately equal to 0.09°, computed both theoretically (from the size of the CCD array and the focal length of the camera), and experimentally (from the real size of the calibration mark at a known distance from the camera and the size of its image in pixels).

Table 5.

	Maximum backlash case	Zero backlash case
Torsion	0.2°	$< 0.1°$
Vergence	0.2°	$< 0.1°$
Tilt	0.76°	$< 0.1°$

We notice that in the zero backlash experiment there is no perceptible disparity between the position of the calibration mark in the two images.

Speed and Acceleration. The maximum speed that we can achieve is limited by the maximum speed allowed for each motor. Whether an axis can move between two positions in a given time interval depends not only on the maximum speed the motor can achieve, but also on the control profile followed. We operated our controller under a trapezoidal profile, according to which the motor accelerates (approximately) from zero speed to a given speed, stays at that speed for some time, and then decelerates with constant deceleration to zero speed, when it reaches the final position. The controller allows the following settings: maximum motor speed, acceleration, and gain, pole and zero of the digital filter of the feedback control. In our initial experiments, we used the default settings for the digital filter parameters (chosen for sluggish response). Maximum motor speed was set to a value corresponding to approximately 20% of the maximum speed ratings of the motors, and acceleration was set to a conservative value to ensure smooth operation. With these conservative settings, the operation of the head was smooth (mechanically) and the control system was stable and avoided oscillations about the final position. To determine the speed achieved for each motor under the same controller settings, we performed the following experiment. Each motor was rotated by the same number of encoder counts forward and backward ten times, and the total time was measured by a stopwatch. The experiment was repeated with two encoder counts, $n_1 = N$ and $n_2 = 10N$, leading to total angles a_1 and a_2 in total times t_1 and t_2 respectively. The achieved speed was calculated as $(a_2 - a_1)/(t_2 - t_1)$, leading to the results in Table 6.

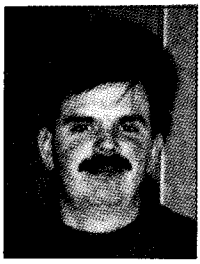
Table 6.

	a_1	a_2	t_1	t_2	Speed
Torsion	212°	42.4°	7.06 second	4.52 second	66°/second
Vergence	280°	28°	17.10 second	6.90 second	25°/second
Tilt	152°	15.2°	23.60 second	8.51 second	9°/second

It should be noted that these results are preliminary, and they do not represent the best performance possible. They only give an indication of the performance that can be easily achieved with highly conservative controller parameter settings. To achieve the best performance, careful experimentation with settings not only of the maximum speed and acceleration, but also with the digital filter parameters is necessary. They should be set to values that achieve the fastest response that avoids oscillation. The control issues of TRISH are complex and deserve treatment in a separate paper.

Received 19 June 1992.

Evangelos Milios received a diploma in electrical engineering from the National Technical University, Athens, and S.M., E.E. and Ph.D. degrees in electrical engineering and computer science from the Massachusetts Institute of Technology. While at MIT, he also worked on acoustic signal interpretation at the MIT Lincoln Laboratory. After working as a research scientist in the Department of Computer Science, University of Toronto, he joined York University, Ontario, as Associate Professor. His areas of research are shape understanding, mobile robotics, active vision, and digital signal processing.



Michael Jenkin received the B.Sc. degree from Trinity College, University of Toronto, and M.Sc and Ph.D degrees in computer science from the University of Toronto, Ontario, Canada. He is currently an Associate Professor in the Department of Computer Science, York University, Ontario. His active areas of research include stereo vision, mobile robotics, and environmental exploration.

John K. Tsotsos received an honours undergraduate degree in engineering science in 1974 from the University of Toronto. He continued at the University of Toronto to complete a Master's degree in 1976 and a Ph.D. in 1980 both in computer science. He is currently a Professor in the Department of Computer Science and Professor (status) in the Department of Medicine at the University of Toronto. Prof. Tsotsos has been a Fellow of the AI and Robotics program of the Canadian Institute for Advanced Research since 1985; he is currently the Canadian Pacific (Unitel) Fellow of the institute. He has published over 130 scientific papers, is a co-editor of a book on motion perception, and has served on numerous conference committees and editorial boards. His current research focuses on the development of a biologically plausible, computational model of visual attention.