International Journal of Computer Vision, Vol. 12, No. 2-3, pp. 231-50, April, 1994.

TRICLOPS: A Tool for Studying Active Vision

JOHN C. FIALA, RONALD LUMIA, KAREN J. ROBERTS, and ALBERT J. WAVERING
National Institute of Standards and Technology, Robot Systems Division, Gaithersburg, MD 20899

Abstract

The design, performance, and application of The Real-time, Intelligently ControLled, Optical Positioning System (TRICLOPS) are described in this paper. TRICLOPS is a multiresolution trinocular camera-pointing system which provides a center wide-angle view camera and two higher-resolution vergence cameras. It is a direct-drive system which exhibits dynamic performance comparable to the human visual system. The mechanical design and performance of various active vision systems are discussed and compared to those of TRICLOPS. The multiprocessor control system for TRICLOPS is described. The kinematics of the device are also discussed and calibration methods are given. Finally, as an example of real-time visual control, a problem in visual tracking with TRICLOPS is examined. In this example, TRICLOPS is shown to be capable of tracking a ball moving at 3 m/s, which results in rotational velocities of the vergence cameras in excess of 6 rad/s (344 deg/s).

1. Introduction

An active or "animate" vision system is a robotic device for controlling the motion of cameras based on visual information. Active vision systems provide a testbed for studying the use of visual feedback in robotic and computer vision tasks. By using camera movements such as self-motion, tracking, focusing, and vergence in a known way, recognition and reconstruction computations can be constrained. For example, controlled alteration of viewing parameters can be used to uniquely compute shape and motion (Aloimonos & Shulman 1989). An active vision system also provides the ability to direct computational resources to important areas, therefore improving the efficiency of the visual system. Active vision systems used in this mode are sometimes referred to as "attentive" (Clark et al. 1988). These and other motivations for using actively-controlled camera systems have been thoroughly discussed in the literature (Aloimonos et al. 1987; Bajcsy 1988; Ballard 1991; Swain & Stricker 1991). In order to support active vision research at the National Institute of Standards and Technology (NIST), a new trinocular robot head, The Real-time Intelligently ControLled Optical Positioning System (TRICLOPS), was designed and built. A photo of TRICLOPS is shown in Figure 1. This paper discusses the design features and performance characteristics of TRICLOPS, as well as some experimental applications which have been implemented to date.

Active vision systems are often modeled on attributes of the human visual system since this is the most well-studied visual system. Two attributes of the human visual system, ocular motion and foveal-peripheral vision, are essential to human visual perception. Ocular motion allows movement of the eyes to direct the viewpoint of the visual field. Foveal-peripheral vision enables humans to perceive small regions in fine detail in combination with a wide field of view at coarse detail.

With respect to ocular motion, saccadic eye movements can be used to determine performance requirements for velocity and acceleration of a mechanical device. The human eye executes a saccadic motion that has an initial acceleration of up to 524 rad/s^2 (30,000 deg/s^2) (Wurtz & Goldberg 1989), which is then followed by a smaller deceleration to terminate the movement with a slight overshoot. During the trajectory, the peak velocity achieved is proportional to the magnitude of the saccade and saturates at a speed of 7-10.5 rad/s (400-600 deg/s) (Wurtz & Goldberg 1989). The amplitude of a saccade of the human eye rarely reaches above 0.26 rad (15 deg), at which point the acceleration of the motion saturates at 611 rad/s^2 (35,000 deg/s^2) (Wurtz & Goldberg 1989). Saccades of amplitude larger than 0.35 rad (20 deg) involve both eye motion and head motion, and move at even higher speeds. These and other characteristics of the human visual system are summarized in Table 1. The information in Table 1 was obtained from Carpenter (1988), Geiger & Yuille (1989), Webb Associates (1978), and Wurtz & Goldberg (1989).

The receptors in the fovea centralis of the human eye are significantly more dense than at the periphery, giving a

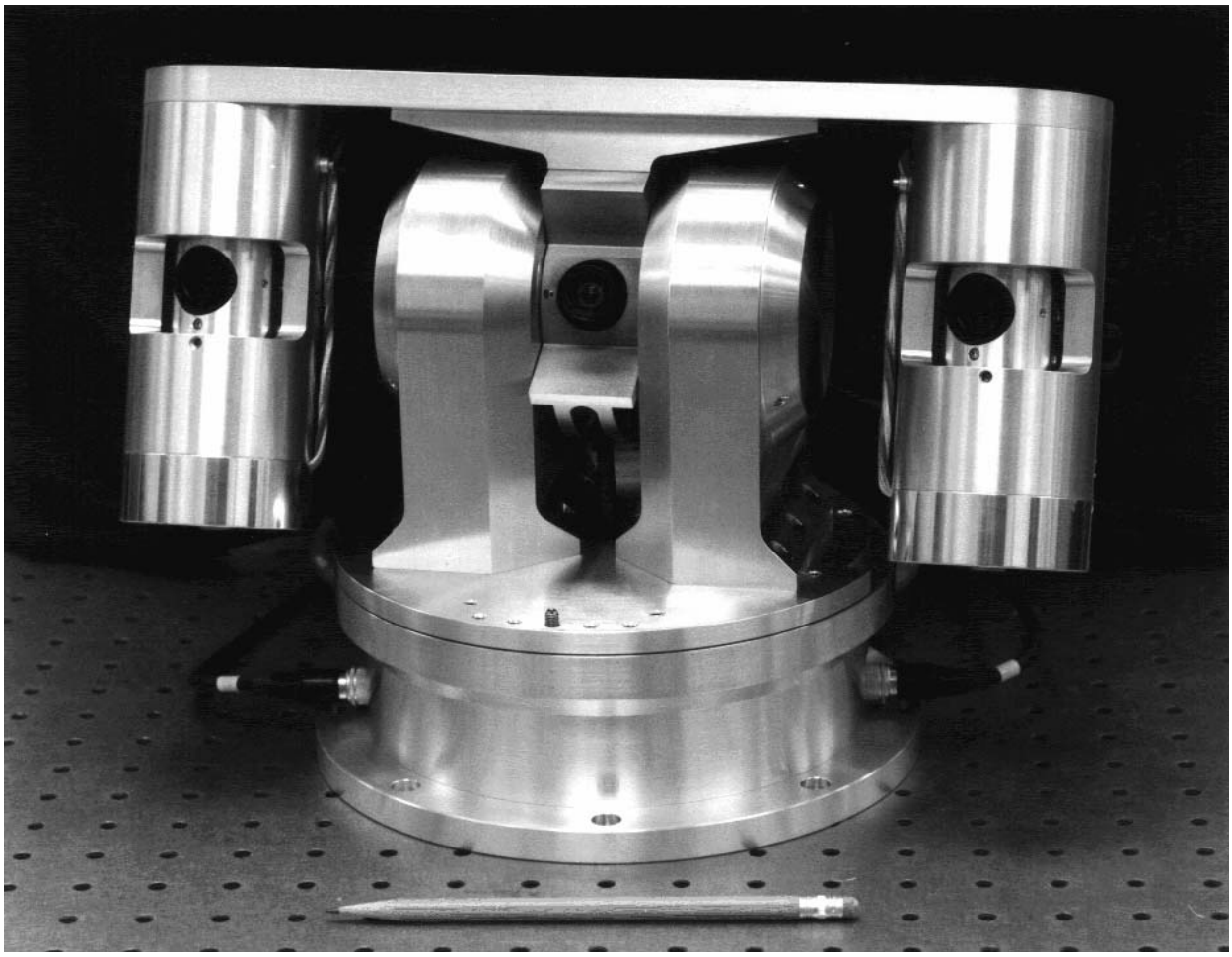

Fig. 1. TRICLOPS device.

Table 1. Human Active Visual System Characteristics.

Range of motion	
Eye pan	+/-0.78 rad (+/-45 deg)
Eye tilt	+/-0.78 rad (+/-45 deg)
Neck pan	+/-1.40 rad (+/-80 deg)
Neck tilt	1.57 rad down, 1.05 rad up (+90 deg, -60 deg)
Peak acceleration	611 rad/s^2 (35,000 deg/s^2)
Peak velocity	10.5 rad/s (600 deg/s)
Interocular distance:	~64.0 mm (2.5 in)
Foveal-peripheral resolution ratio	10:1

resolution ratio between fovea and periphery (at 0.17 rad eccentricity) of about 10:1. The area of high resolution plays a critical role in object discrimination, identification, and manipulation (Schwab & Nusbaum 1986). Although saccadic eye movements allow the fovea to be rapidly redirected, the small size of the fovea presents a problem in maintaining an object in the center of gaze. In addition, the presence of two fovea, one in each eye, creates the problem of aligning the two monocular fields for stereoscopic vision.

Much work has been done in developing active vision systems to study how these and other features of the human visual system are used to facilitate perception. One of the earliest active vision systems was built at the University of Pennsylvania (Krotkov et al. 1988). The system provides flexible image acquisition capabilities using a pair of movable cameras with computer-controlled zoom, focus, and aperture along with ten variable-intensity lamps. The positioning mechanism of the system provides two translational degrees of freedom, pan and tilt of both cameras, and coupled symmetric vergence.

At other institutions, such as the University of Illinois, researchers have assembled dynamic stereo camera systems using commercial motion control components, such as rotational and translational stages (Abbott & Ahuja 1990). Such systems can provide good repeatability and relatively high-speed motions. However, because they are constructed from general-purpose positioning units, they cannot easily be configured to meet special requirements (such as the ability to position the focal point of the camera lenses at the intersection of perpendicular motion axes).

A different approach to aiming cameras is used in the MIT Vision Machine (Poggio et al. 1988). The Vision Machine consists of two cameras rigidly mounted on a movable platform. Each of the cameras is equipped with a motorized zoom lens, allowing control of the iris, focus, and focal length. Because of the size and weight of the lenses, camera movement is achieved indirectly by pivoting a front surface mirror mounted in front of each lens. These mirrors provide two degrees of freedom in aiming the cameras. The platform rotates about two orthogonal axes to allow pan and tilt control. The MIT Vision Machine integrates several low-level vision algorithms using images from a movable two-camera eye-head system to achieve high performance in unstructured environments for recognition and navigation tasks.

Binocular active vision systems have also been built at the University of Rochester and Harvard University. The Rochester head has independent vergence axes and a coupled tilt axis, and achieves camera rotational velocities of 5.2 rad/s (300 deg/s) and positioning accuracy of 2.4 mrad (0.138 deg) (Brown 1989). The head is mounted on a six degree-of-freedom robot arm, and has been used to study gaze-holding, vergence, and kinetic depth (depth from motion during fixation) (Ballard & Ozcandarli 1988; Olson & Coombs 1991). The Harvard Head is a mobile binocular camera system with seven degrees of freedom (Clark et al. 1988). Three of these degrees of freedom control the orientation of the cameras by performing pan, tilt, and antisymmetric vergence motions. The other four degrees of freedom accomplish control of the cameras' aperture and lens focus. This camera system is mounted on a robot that provides translation in the horizontal plane and rotation about a vertical axis. The head has been used to study attentive gaze control for use in mobile robot navigation.

The head-eye system developed at the Royal Institute of Technology in Sweden (the KTH head) is a 13 degree-of-freedom device, with two mechanical degrees of freedom for each of the two eyes, two degrees of freedom in the neck, computer-controlled baseline separation, and three optical degrees of freedom for each eye (Pahlavan and Eklundh 1992). The resolution on the eye and neck axes is 126 μrad (0.0072 deg), and the maximum rotational velocity is 3.14 rad/s (180 deg/s). A particularly interesting feature of the KTH head is a mechanism which automatically compensates for shifts in the position of the lens centers caused by changes in the lens focal length and focusing distance. This keeps the lens centers near the centers of rotation, despite changes in optical parameters.

Another interesting recent development in the area of active vision hardware is the Spherical Pointing Motor (SPM) (Bederson et al. 1992). This device is a small, lightweight pan-tilt mechanism which can be configured to move a small camera. It is controlled open-loop and is capable of rotational velocities of up to 10.5 rad/s (600 deg/s). Accuracies of 2.6 mrad (0.15 deg) are achieved. The SPM shows promise for future compact, low cost active vision systems.

Although the above systems represent significant advances in the development of hardware for active vision, none of them incorporated all of the features we desired. These include human-like dynamic performance and range of motion, independent control of vergence degrees of freedom for asymmetrical vergence, ability to position vergence lens focal points at the intersection of vergence and tilt axes, multiresolution imaging capability, high positioning resolution, repeatability, and accuracy, modular design, the use of industrial-quality components, and the ability to withstand

the rigors of experimentation. In addition, a completely open controller based on the NASA/NIST Standard Reference Model (NASREM) Architecture (Albus et al. 1987) was desired to provide total freedom in modifying system parameters and algorithms. In meeting these requirements, TRICLOPS represents yet another step in the evolution of high-performance camera-pointing mechanisms.

The next section describes the details of the mechanical design of TRICLOPS. Later sections describe in detail the TRICLOPS control system, image processing system, calibration procedures, and visual kinematics. Finally, the application of TRICLOPS to a simple tracking problem is presented.

2. TRICLOPS Design

One of the primary design goals for TRICLOPS was to build a system that has very high dynamic performance—on a par with the human oculomotor system. There are several reasons for putting a large emphasis on being able to move the cameras very quickly. First, one of the primary advantages of active vision systems is to be able to use camera redirection to look at widely separated areas of interest at fairly high resolution instead of having a single sensor or array of cameras which covers the entire visual field with uniformly high resolution. It is desirable to be able to move from one area of interest to the next as fast as possible to minimize the time spent redirecting attention and to maximize the time spent acquiring useful image data. In addition, it was desired to build a device for which mechanical performance would not be a limiting factor in the development of algorithms for visual tracking and autonomous information-gathering tasks. Although the current state of image processing results in latencies which are typically on the order of hundreds of milliseconds, rapid progress is being made. We would like to be able to take full advantage of anticipated improvements in image processing hardware, which will continue to increase update rates and reduce latency.

The goal of maximum dynamic performance leads to the selection of "micro-miniature" CCD cameras, with manually-adjustable focus, aperture, and focal length, since motorized lenses tend to be quite bulky. This decision also reduces the complexity of camera calibration considerably, since optical parameters do not change while the device is in use. Also, using manually-adjustable lenses simplifies achievement of the goal of placing the focal point of the lens near the center of rotation. The focal point of motorized zoom lenses shifts a great deal as the focal length is changed, and, although it is possible to compensate for this by moving the camera and lens as the zoom is adjusted (Pahlavan and Eklundh 1992), it results in considerable additional complexity. Many useful tasks can be performed with fixed imaging parameters. However, should computer control over these functions be desired, it would be possible to mount motorized lenses to the TRICLOPS cameras. Of course, this would involve compromises in other areas, particularly in terms of dynamic performance and rotation about the focal point.

Although fixed lenses are used, it is still desirable to be able to obtain both wide-angle, low resolution images of the world (the big picture) as well as higher-resolution, narrow field of view images for detail. For this reason, TRICLOPS has been designed with an integral third center camera which has wide angle (3-4 mm focal length) lens. This color camera serves the purpose of identifying features of interest in the wide field of view. There is a significant (and obvious) advantage to using a separate camera for this purpose, as opposed to using motorized zoom lenses—the wide-angle view is always available to alert the system to new agents which may enter the scene at any time. The center wide-angle lens is used in conjunction with 15-24 mm lenses on the vergence cameras, to provide a maximum resolution ratio of 6.17:1[1]. The system is currently configured with a 4 mm center lens (1.35 rad field-of-view) and 15 mm vergence lenses (0.42 rad field-of-view). Each of the TRICLOPS cameras is mounted in a Delrin tube which has screws for fine-adjustment of camera axial position and roll.

As shown in Figure 2, TRICLOPS has four mechanical degrees of freedom. The four axes are arranged in the following kinematic configuration: 1) pan (or base rotation) about a vertical axis through the center of the base, 2) tilt about a horizontal line that intersects the base rotation axis, and 3) left and right vergence axes which intersect and are perpendicular to the tilt axis (Figure 2). The tilt and vergence axes provide eye movements, while the pan axis allows a single degree of neck rotation. This particular kinematic configuration, with intersecting pan and tilt axes, minimizes the inertia of the pan axis. An unfortunate consequence is that the motion parallax (which can be used to determine kinetic depth (Ballard and Ozcandarli 1988)) is also minimized. However, good motion parallax information can still

1. Longer focal-length lenses can be mounted on the vergence cameras if C-mount adapters are used. Again, this would be at the expense of dynamic performance and giving up rotation about the focal point.

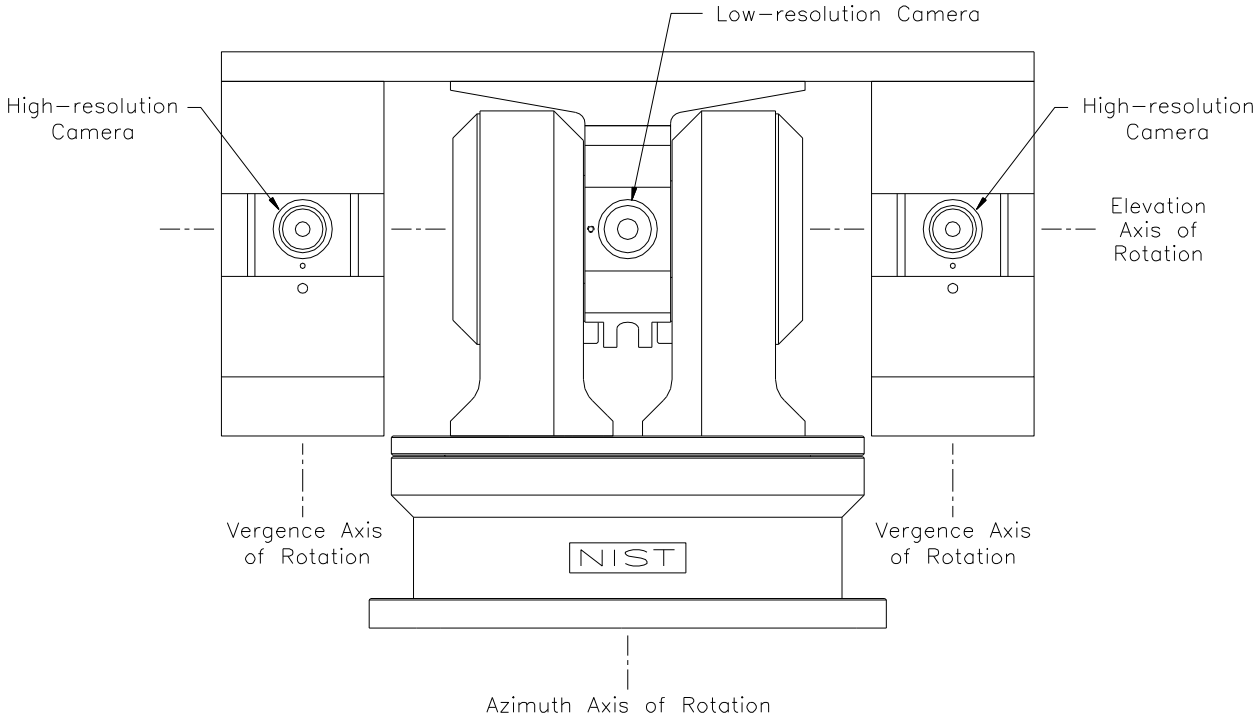

Fig. 2. TRICLOPS degrees of freedom.

be obtained by mounting TRICLOPS on a translating device such as a robot torso, mobile platform, or linear rail.

The three mounting locations for cameras are also shown; one at the intersection of the pan and tilt axes, and two more at the intersections of the tilt and vergence axes. Rotation of the cameras about the lens focal points provides many advantages. In addition to ensuring a fixed baseline separation and simplifying the imaging kinematics, rotation about the focal point enables the implementation of distance-independent saccade algorithms. Rotation about the lens focal points is also essential for algorithms which assume that the relationship between motion in image space and motion in joint space may be learned without knowledge of the target distance.

The dynamic performance, accuracy, and other requirements are achieved with a direct-drive design. In many ways, a robot head is an ideal application for a direct-drive system. There is a requirement for large acceleration, low friction, and minimal transmission errors. These are three of the primary characteristics of systems which use motors and feedback devices mounted directly to the axes of motion. Transmission compliance and backlash, which can cause inaccuracy and oscillations, are eliminated. Also, the large gravity torques which cause limitations with direct-drive robot *arms* do not exist for a *head* with the configuration shown in Figure 2; the loads on the axes are almost completely inertial. In addition, the use of frameless DC motors and brushless, frameless resolvers provides flexibility in the routing of camera and control wiring. Much of the wiring in TRICLOPS is routed through the centers of these components to minimize external wiring. This results in well-behaved cable motions during high-speed movements and reduces the likelihood of pulling out cables.

The resolvers used for position feedback provide absolute position information, which eliminates the need to perform any homing or other initialization procedures each time the device is powered up (as would be required with incremental optical encoders, for example). The resolvers are used with 16-bit resolver-to-digital (R/D) converters, which results in a position sensing resolution of 96 μrad (0.0055 deg). The R/D converters also provide an analog velocity signal.

TRICLOPS is built up of three independent, detachable modules, as shown in Figure 3. These are the pan/tilt module (which forms the base), and two vergence modules. These modules can be used separately or together as a complete

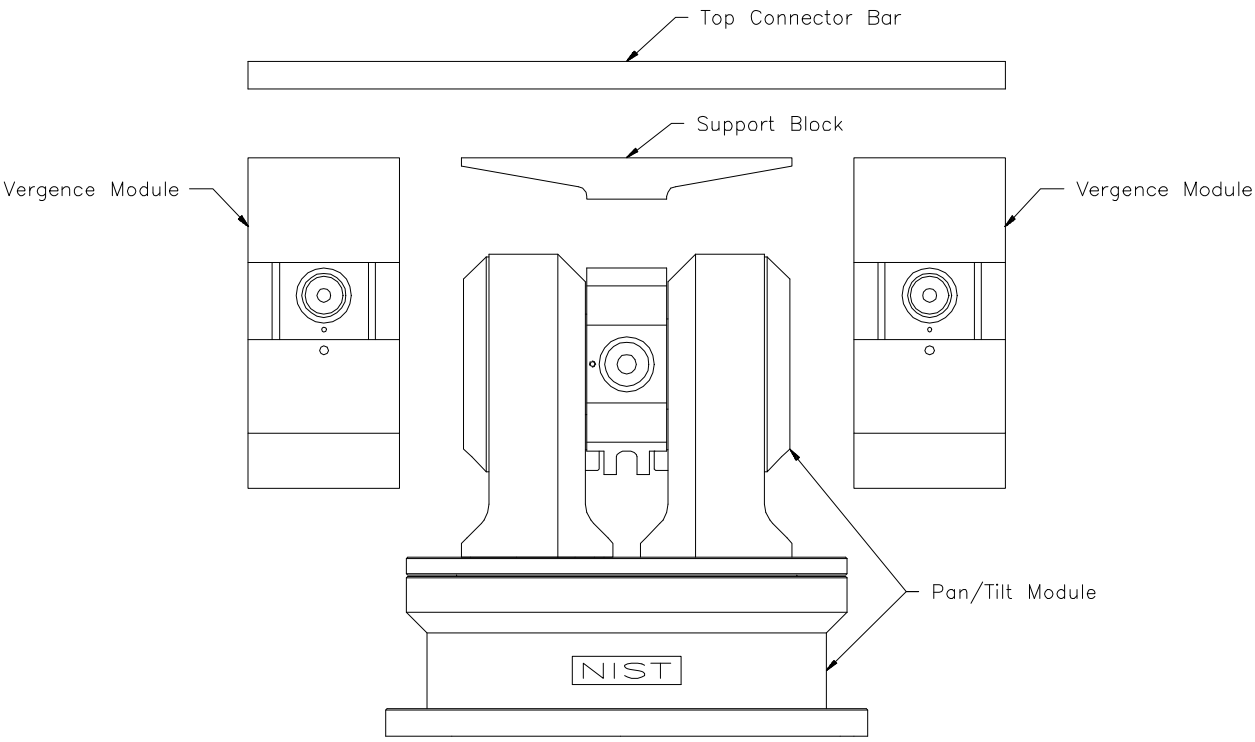

Fig. 3. TRICLOPS axis modules.

system. The vergence modules are connected to the pan/tilt module by the support block and top connector bar. The baseline separation between vergence modules is determined by the length of this bar. The current (and minimum) baseline separation is 0.2794 m.

All axes have detents which may be used with retractable ball spring plungers to provide static positioning of the shaft at known fixed locations. This is useful for determining the zero offset of the position transducer, and for other calibration procedures. Another important feature of TRICLOPS is that all axes have cushioned hard stops which prevent damage to the system when experimenting with new (and potentially unstable!) algorithms.

The performance specifications which have been achieved with TRICLOPS are summarized in Table 2. It is seen that TRICLOPS is capable of motions which are comparable to those of the human eye-head system in terms of peak velocity and acceleration, and range of motion. In fact, the vergence modules can achieve peak velocities and accelerations which substantially exceed the human capability.

3. Control System

A multiprocessing controller based on the NASA/NBS Standard Reference Model (NASREM) Architecture (Albus et al. 1987) has been constructed for controlling TRICLOPS. The NASREM Architecture is a hierarchical control system architecture which divides the sensing and control problem into discrete levels. At the lowest level are the servo control loops for the device actuators; at higher levels, groups of sensors and actuators are coordinated as equipment subsystems.

For controlling the motions of the TRICLOPS device, the architecture consists of a Primitive Level for producing coordinated motions of all axes, and a Servo Level for servoing the axes to commanded joint positions, as depicted in Figure 4. The Primitive Level is capable of accepting commands from an operator through a user interface, or from another control system. This allows the device to be used either as part of an integrated robot system or as a stand-alone device. Visual feedback from the cameras is provided by a visual processing system which is described further in the next section.

Table 2. TRICLOPS Characteristics.

Range of motion:	
Pan	+/-1.68 rad (+/-96.3 deg)
Tilt	0.48 rad down, 1.14 rad up (+27.5 deg, -65.3 deg)
Vergence	+/-0.77 rad (+/-44 deg)
Peak acceleration:	
Pan	70 rad/s^2 (4010 deg/s^2)
Tilt	320 rad/s^2 (18,300 deg/s^2)
Vergence	1100 rad/s^2 (63,000 deg/s^2)
Peak velocity:	
Pan	11.5 rad/s (660 deg/s)
Tilt	17.5 rad/s (1000 deg/s)
Vergence	32 rad/s (1830 deg/s)
Axis position sensing resolution:	96 μrad (0.0055 deg)*
Positioning repeatability:	+/-96 μrad (0.0055 deg)
Interocular distance:	279.4 mm (11.0 in)**
Vergence-center resolution ratio	6.17:1 (max) 3.2:1 (current)
Overall width:	349 mm (13.75 in)
Overall height:	248 mm (9.75 in)
Approximate weight:	17.3 kg (38 lb)

* With 16-bit resolver-to-digital converter

** Distance is determined by length of connector bar—longer distances are possible

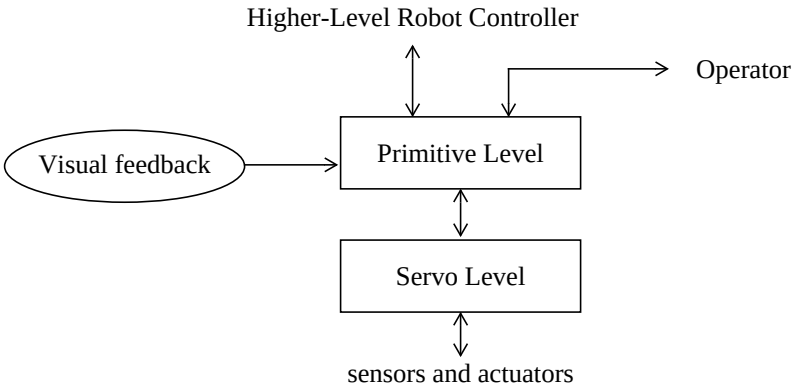

Fig. 4. Control system architecture.

Five processor boards in a VME backplane are used to implement the hierarchical controller. This hardware is depicted in Figure 5. As shown in Figure 6, the Servo Level is composed of three software *processes*, which are assigned to processors as indicated by the numbers in the upper right corners of the boxes.

As described in Fiala (1989), each process is a cyclically-executing software element that communicates with other processes in the system through common memory interfaces. Each process performs a particular part of the total control function. The Servo Planning (PL) process provides the desired signal formation for mapping the slower Primitive updates into the Servo command rate (2 kHz). This can take the form of an interpolator or a zero-order hold, for example. The Vergence Sensory Processing/Execution (SP/EX) process executes the servo loops for the vergence axes, reading the two vergence resolvers and outputting torque commands to the two vergence actuators. By completing one control cycle every 500 μs, a 50 Hz position control bandwidth is achieved for the vergence axes with a standard PD servo algorithm. The velocity signal is obtained numerically using a backward difference technique on resolver position data with filtering. Likewise, the Neck SP/EX process executes the servo loops for the pan and tilt axes at a 2 kHz control rate. A position control bandwidth of about 12-13 Hz is achieved for the tilt axis, and of about 4 Hz for the pan axis.

The SP/EX processes also each maintain a buffer of old position feedback data. These queues of old position feedback values are maintained for the purpose of matching the position feedback delay to the image acquisition and pro-

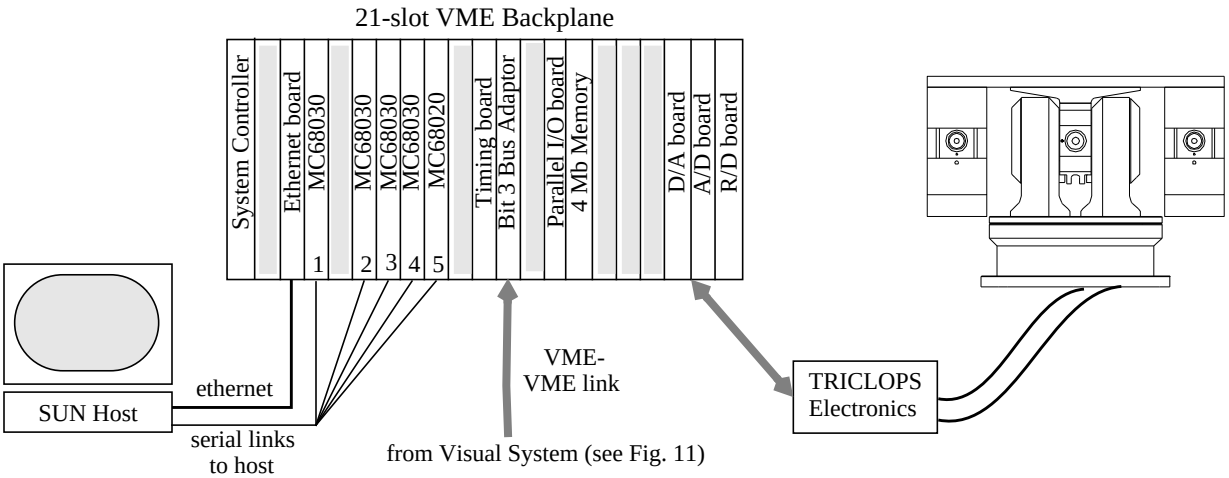

Fig. 5. TRICLOPS control hardware.

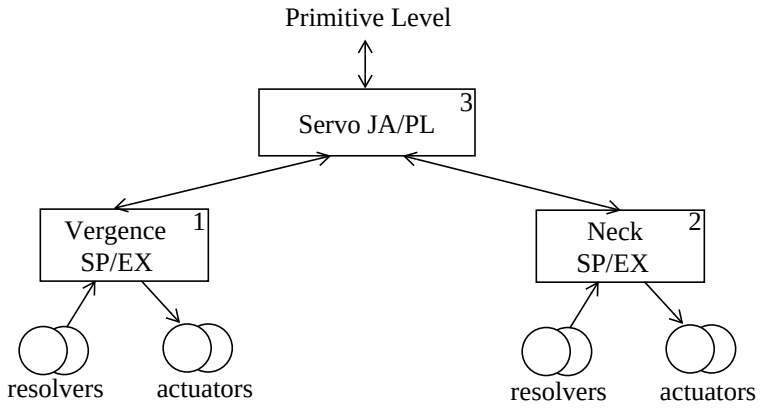

Fig. 6. Servo Level processes.

cessing delay. Position feedback values up to 0.5 s old are available. This information is used along with centroid data to compute estimates of the position of a moving object, as discussed in section 7.

The Primitive Level processes are shown in Figure 7. The Primitive Level provides trajectory generation capability for a number of different algorithms. Quintic polynomial joint space trajectories are available for executing rapid saccadic motions. For such motions, the movement of each axis is typically scaled and synchronized such that all axes complete execution at the same time (this can be changed for special circumstances). There are also sine, square, and triangle wave trajectories, primarily for test purposes. Commands to aim at desired fixation points in Cartesian space may also be entered. The above trajectory algorithms are all executed independently of visual feedback information. There are also vision-based tracking algorithms for tracking object features. One of these algorithms is discussed further in section 7. A more detailed description of the operation of the Primitive Level processes may be found in Fiala et al. (1992). In addition to the processes shown in Figure 7, there is also a data logging process which is used to record system state information for analysis and evaluation.

The joint trajectory following performance achieved by the system is demonstrated in Figure 8. This plot shows the desired and actual joint positions for quintic polynomial trajectories that move each joint at close to its maximum velocity. To obtain this data, the trajectory generation algorithm was modified to allow different motion times for each degree of freedom. The motions for all axes were performed and recorded simultaneously. For the vergence joint, the motion is from -0.75 rad to 0.75 rad (-43 deg to 43 deg) in 0.091 s. This requires a peak acceleration of over 1000 rad/s^2 (57,300 deg/s^2) and a peak velocity of more than 30 rad/s (1720 deg/s). For the tilt axis, the motion is from -1.10 rad to 0.45 rad (-63 deg to 26 deg) in 0.167 s. The peak commanded velocity and accelerations for this motion are 17.4 rad/s (997 deg/s) and 320 rad/s^2 (18,300 deg/s^2). The base pan motion of Figure 8 goes from -1.57 rad to 1.57 rad (-90 deg to 90 deg) in 0.523 s. This motion requires maximum velocities and accelerations of 11.3 rad/s (647 deg/s) and 66 rad/s^2 (3780 deg/s^2).The commanded and feedback velocities for these motions are shown in Figure 9.

As the plots show, the joint feedback positions follow the desired trajectories very closely. This degree of performance is the result of the high bandwidth of the system, as well as the use of velocity and acceleration feedforward terms which decrease following errors. These motions were performed using a PID servo algorithm. The integral term counteracts the effects of cable-twist and friction torques which would otherwise cause steady-state position errors.

4. Image Processing System

The same hierarchical control system architecture that is used for the control of TRICLOPS is also used for the visual processing of images obtained from TRICLOPS. The NASREM architecture divides the information extracted from sensors and the representation of this information into levels of increasing complexity using a sensory processing (SP) hierarchy, a world modeling (WM) hierarchy, and a task decomposition (TD) hierarchy. The sensory processing system monitors and analyzes sensory information in order to extract image features and group the features into lines, surfaces, and objects. The world model supplies current and predicted information about the environment at different resolutions. The interfaces between the sensory processing system and the world model allow updated information to

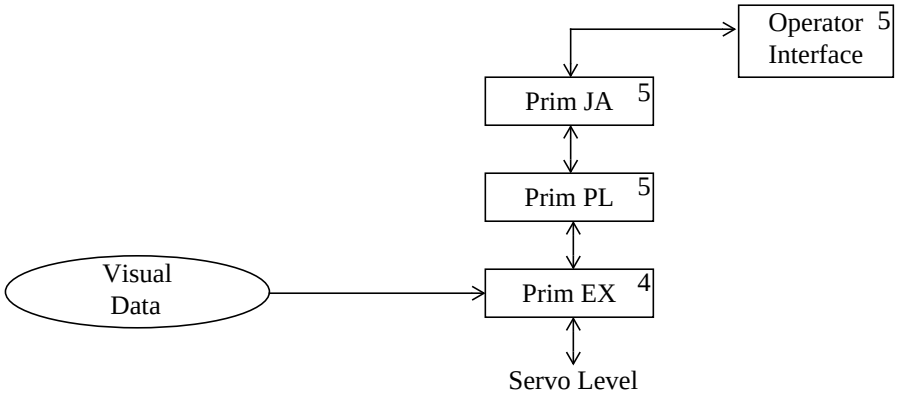

Fig. 7. Primitive Level processes.

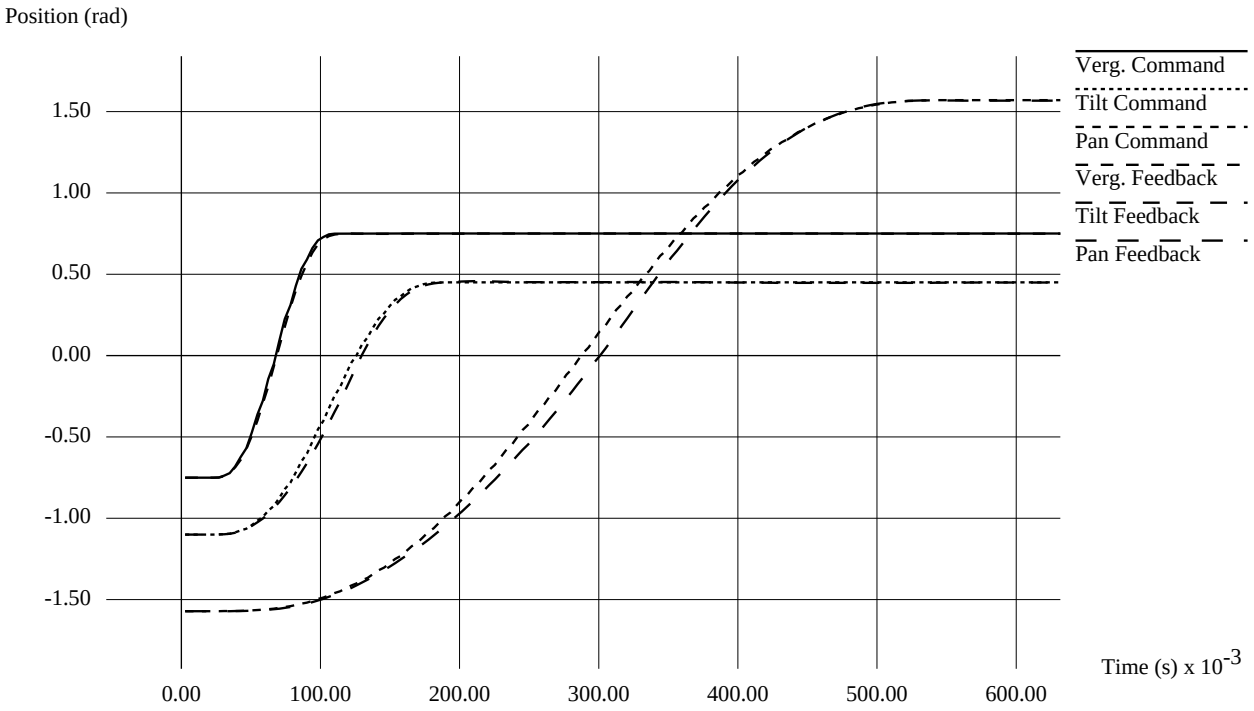

Fig. 8. High-speed trajectory following.

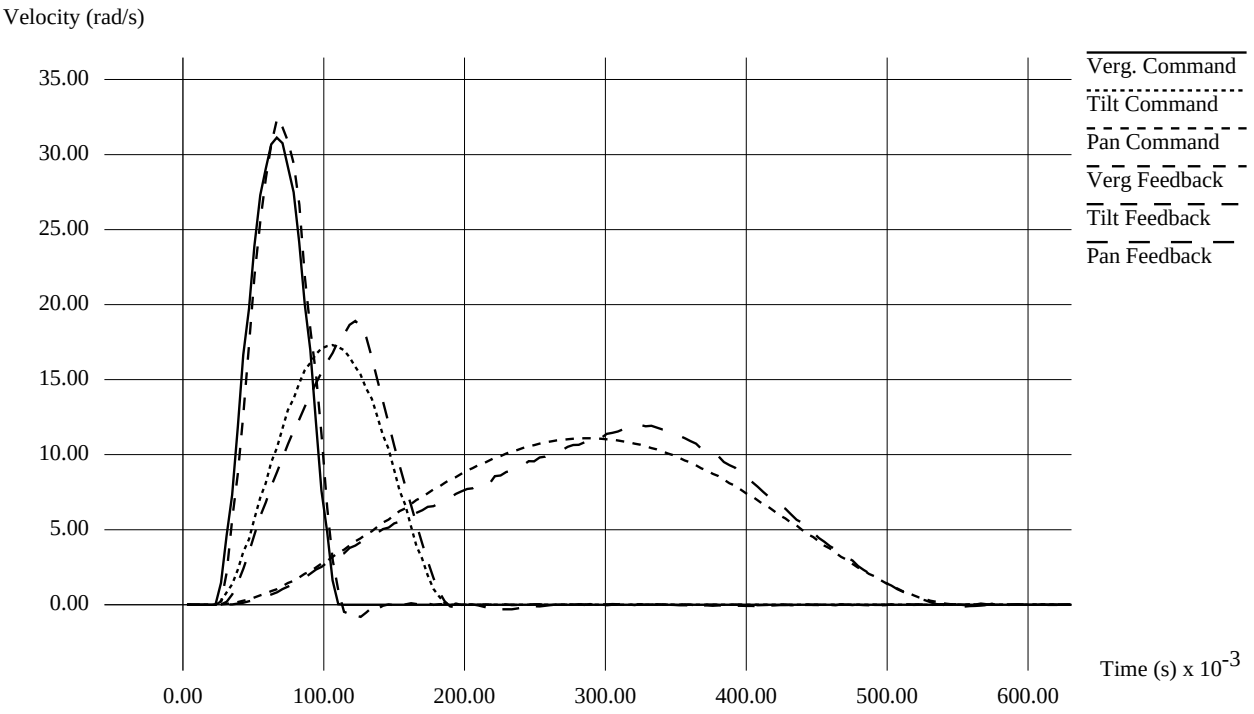

Fig. 9. Velocities during high-speed motions.

be sent to the world model and predicted information or sensory processing parameters to be sent to the sensory processing system. The task decomposition system determines which sensory processing algorithm to use and selects the appropriate algorithm parameters from the world model.

The integrated vision testbed which has been developed in our lab supplies visual feedback to the TRICLOPS control system described in the previous section. In order to provide image and three-dimensional feature information to the control system, the architecture supports visual processes which perform Level 1 SP, Level 2 SP, Level 2 WM, Level 1 TD, and Level 2 TD functions, as shown in Figure 10. In this figure, the rectangles represent the software processes, and the arrows show how the processes interact. In the most general scenario, Level 1 SP extracts iconic features from camera images, and Level 2 SP groups these iconic features. Level 2 WM computes current or predicted three-dimensional (3D) position of these features or current or predicted pose of a group of features. Level 2 TD and Level 1 TD select the necessary algorithms for the sensory processing levels to perform their functions.

As an example, consider that the Level 2 TD process requests that a centroid computation needs to be performed and that triangulation is the method to use to locate the two dimensional feature point. In this case, the Level 1 TD process would indicate that thresholding is required for Level 1 SP. The Level 1 SP process can then digitize and threshold an image from each of TRICLOPS' vergence cameras. The centroid of all of the pixels above threshold in each image can be computed at Level 2 SP. The centroid from each camera image can be used by the Level 2 WM process to triangulate the 3D position of the centroid feature. The 3D position can then be made available to the Primitive level control processes to reposition the vergence cameras.

The implementation of these software processes is based on the concept of cyclically executing modules which serve as the computational units for the NASREM architecture (Fiala 1989). After initialization, all computations are performed by cyclically executing processes that communicate via global read-write interfaces. Each unit acts as a process which reads inputs, performs computations, and then writes output. Such a process always reads and executes on the most current data; it does not wait for new data to arrive since reliable cyclic execution requires that a module be able to read or write data with minimal delay. Reading and writing involve the transfer of data between local buffers and buffers in global memory. System software has been written to prevent data corruption during these transfers.

The processes are implemented on the VME-compatible processors used in the integrated vision testbed. This hardware includes a programmable real-time image processor, the Pipelined Image Processing Engine (PIPE)[1] (Aspex, Inc. 1987) and a multiprocessor system as shown in Figure 11. Again, the numbering of the processes in Figure 10 indicates the corresponding processor in Figure 11. The Level 2 SP process that computes the image centroids runs on board #3. The Level 2 WM process that performs triangulation and the Level 2 TD processes run on board #2, and Level 1 TD runs on board #1. The only process which isn't executed on a microprocessor board is Level 1 SP which performs image thresholding and executes on the PIPE.

To provide a specific feature point for TRICLOPS to track for the visual tracking example of section 7, a subset of this implementation is used. Initially, the incoming images from each TRICLOPS vergence camera are digitized by PIPE to provide 8-bit grayscale images that are 242x256 pixels in size. The images are processed by lookup tables and arithmetic logic units that are defined in the PIPE program that performs the Level 1 SP which thresholds grayscale images. This binary information is then converted to a symbolic list by a specialized stage of PIPE. Communication with PIPE is performed on one of the remaining microprocessors. The PIPE communication process is a cyclically executing process which polls PIPE status, reads the output produced by PIPE, and writes it to the appropriate common memory locations that are shared with other processes. Direct memory accessing provides a high transfer rate of the symbolic data. The centroid of the binary data is computed by the Level 2 SP process and then this data is made available to the Primitive level control process.

5. Calibration

The visual information from this system can be used to compute object or object feature position and velocity. In order to accurately measure object or feature positions in three dimensions, we use a method involving triangulation of a 3D

1. Commercial equipment and materials are identified in this paper in order to adequately specify the experimental procedure. Such identification does not imply recommendation or endorsement by NIST, nor does it imply that the materials or equipment identified are necessarily the best for the purpose.

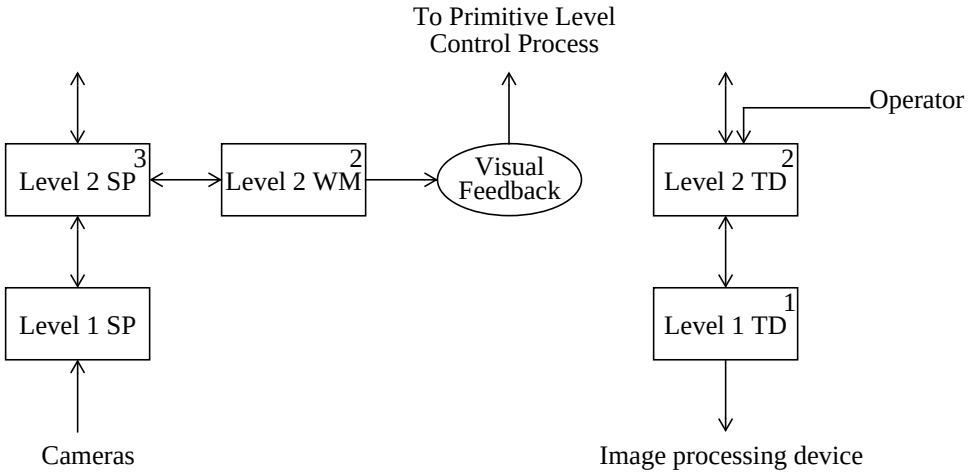

Fig. 10. Level 1 and Level 2 visual processes.

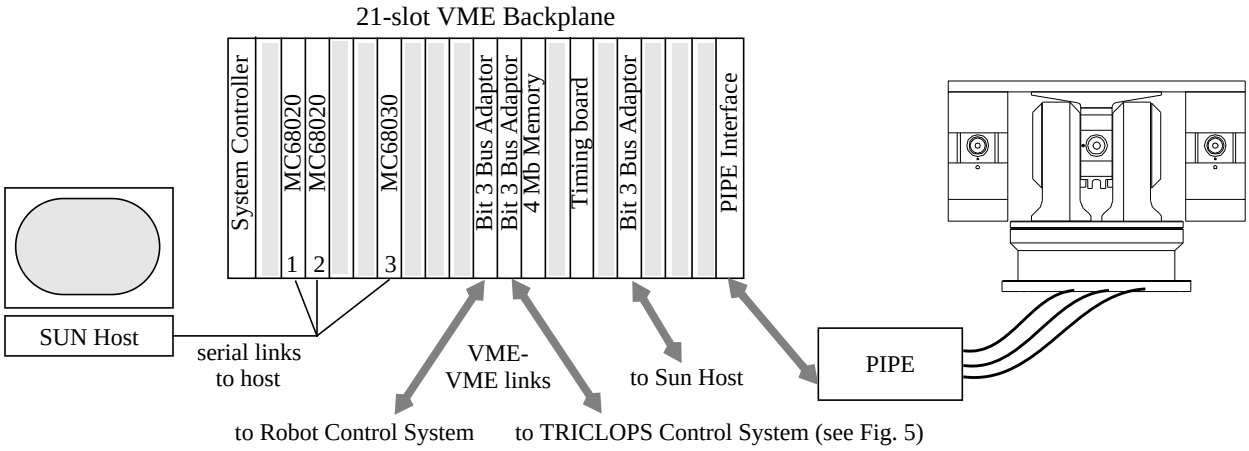

Fig. 11. TRICLOPS visual processing hardware.

feature position and its projection onto each of the two vergence cameras. In order to determine position using this method, each camera's intrinsic parameters (image center, scale factor, lens distortion coefficient, and lens focal length) and the position and orientation of each of the cameras with respect to a world coordinate system are needed. Only with this information can the positions of features on the image plane be known and the three sides of a triangle be determined so that triangulation can be performed accurately.

In order to calibrate the poses and intrinsic parameters of the cameras, the method proposed by Tsai (1987) is used. This method employs a two-stage technique that allows the camera pose to be computed separately from the intrinsic camera parameters. A non-coplanar target is used to obtain ground truth measurements of feature points. The non-coplanar target is created by moving a coplanar target a known distance on an optical bench. The camera undergoing calibration is positioned so that its image plane is parallel to the coplanar target and perpendicular to the direction of the target's motion on the optical bench. A photo of the calibration set-up is shown in Figure 12. Using the extracted feature points of the target at several depth positions, the image center is found by computing the focus of expansion. This image center is used with the Tsai method to calibrate the remaining intrinsic parameters and to determine the relative pose of the camera with respect to the target (at the initial position).

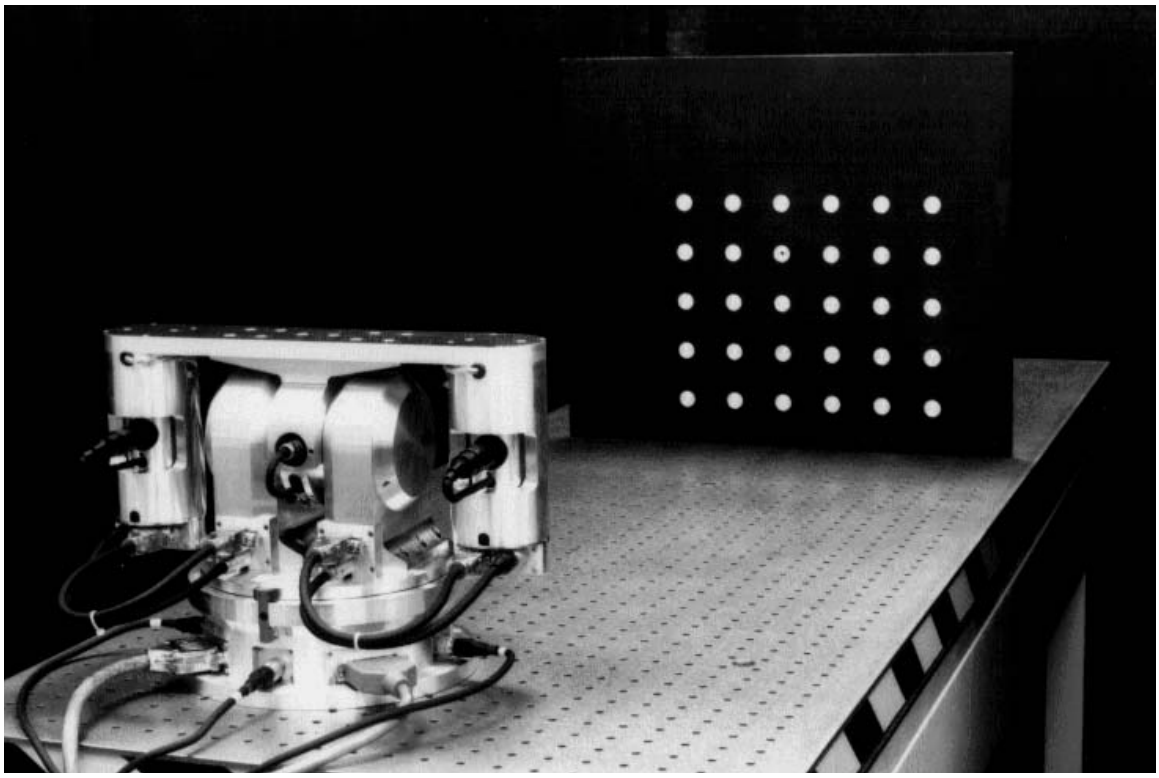

Fig. 12. TRICLOPS camera calibration set-up.

The relative pose information obtained during camera calibration is used to fine-adjust the mounting positions of the cameras (translation along and rotation about the optical axis). The axial position adjustment is performed by comparing the calibration distance from the target to the lens focal point with the measured distance from the target to the centerline of the vergence module. Using this technique, the mounting position of each camera was fine-adjusted to place the focal point of its lens within about 0.5 mm of the intersection of the vergence and tilt axes. The positions of the centroids of the target features in the camera image were used to fine-adjust the camera roll.

In addition to calibration of the camera intrinsic and extrinsic parameters, some calibration procedures were also performed on the joint axes to compensate for resolver errors and resolver offset positions. The accuracy of the position values obtained from the resolvers is determined by the size of errors in setting the resolver zero offsets, and by resolver and resolver-to-digital conversion errors. The zero offsets of the vergence resolvers are obtained using a special bar which is placed in contact across the two camera mounting tubes. This causes the vergence axes to point straight ahead and parallel to one another. The tilt axis zero offset is determined using a square between the vergence module housing and the mounting surface (an optical bench). The zero offset for the base pan axis is set by engaging the spring-loaded detent at the zero position.

Resolver and resolver-to-digital errors (3-4 mrad max combined) may be compensated for by obtaining calibration data from a resolver calibration fixture and using a third-order polynomial interpolation function. This reduces these errors by about 80% to only about 0.5 mrad (0.0286 deg) in the worst case. It is likely that even better accuracy could be obtained by using a more accurate means (such as using a polygonal mirror and autocollimator) to record calibration data.

6. Visual Kinematics

In order to use TRICLOPS to identify the position and orientation of objects relative to its own position, a fixed coordinate frame of reference is required. This world reference frame, shown in Figure 13, is located at the intersection of the pan and tilt axes (assuming ideal kinematics). The world frame is oriented such that the Z-axis coincides with the

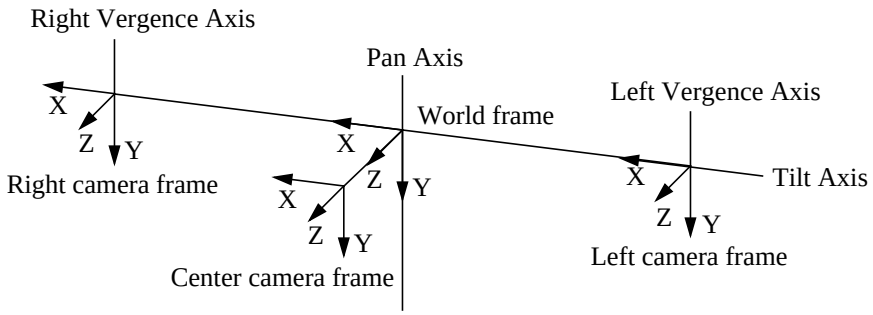

Fig. 13. TRICLOPS coordinate frames (all axes at home position).

optical axis of the center camera when the pan and tilt degrees of freedom are in the home position. The home position, also shown in Figure 13, is that in which the optical axes of the three cameras are parallel to each other and to the plane of the base plate. Positive rotation of the pan and both vergence axes is defined using the right-hand rule about the -Y axes. Positive rotation of the tilt axis is defined using the right-hand rule about the -X axis. Note that coordinate frame assignments correspond to the conventions used for cameras in computer vision literature, rather than those used in the description of manipulator arm kinematics (e.g. Denavit-Hartenberg). In Figure 14, the coordinate systems described above are shown for the case where the vergence cameras are aimed at a fixation point in space, which may be described with respect to the world frame. Here it is seen that the camera frames rotate with the joints of the head/eyes, while the world frame remains stationary.

The forward kinematics are used to compute the Cartesian (or other) coordinates of the fixation point relative to a fixed reference frame, given the joint angles. Conversely, if the coordinates of the fixation point are known, inverse kinematic equations may be solved to determine a set of joint angles which will point the cameras at the desired location. The (ideal) forward kinematic equations for TRICLOPS are quite simple and straightforward. First, the coordinates of the fixation point in the *vergence plane* (the plane formed by the positions of the two vergence nodal points

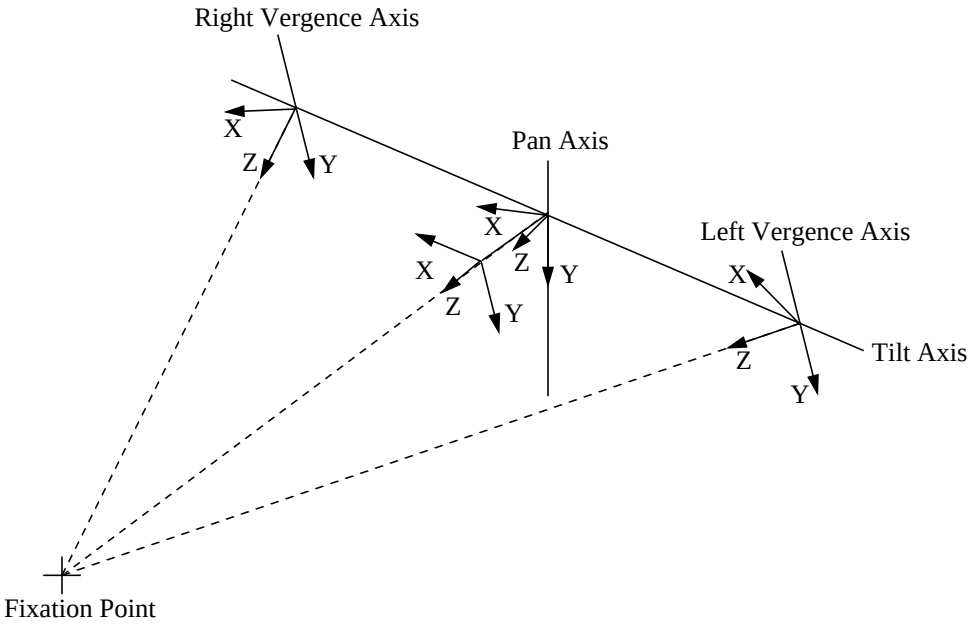

Fig. 14. TRICLOPS coordinate frames (axes not at home position).

and the fixation point) are determined using the following equations:

$$x_{fp} = \frac{-b\sin(\theta_r + \theta_l)}{2\sin(\theta_r - \theta_l)}$$

$$z_{fp} = \frac{b\cos(\theta_r)\cos(\theta_l)}{\sin(\theta_r - \theta_l)}$$

$$\text{for} \theta_r > \theta_l, \ -\frac{\pi}{2} < \theta_r, \theta_l < \frac{\pi}{2}$$

where θ_r = right vergence angle, θ_l = left vergence angle, and b = baseline separation. If $\theta_r <= \theta_l$ then the cameras are parallel or divergent, and consequently there is no frontal fixation point. The vergence plane fixation geometry is shown in Figure 15. The position in the vergence plane is then rotated by the pan and tilt rotations to get the position of the point in world coordinates.

The above relationships define where the vergence cameras are pointing in space, and, therefore, the 3D position of a feature that appears in the center of both camera images. However, in many cases we would like to know the location of a feature which appears in both vergence cameras, although not directly in the center of both images. To do this, a vector which extends out from the camera and contains the feature point on the image chip and the camera's focal point (after correcting for lens distortion) is determined for each camera. Each of these vectors is rotated by the angle of the corresponding vergence rotation. After pan and tilt rotations, the position where these two lines intersect corresponds to the world position of the imaged feature point, as shown in Figure 16.

Realistically, these two vectors will not actually intersect but rather will cross over each other in a skewed relationship. To compute the feature position using two non-intersecting vectors, the line perpendicular to the two vectors is determined. The midpoint of this line is used as the computed three-dimensional position (Chaconas et al. 1990). This method of determining range will be referred to as the triangulation method in the remainder of this paper.

Also of interest is how to determine the joint angles which will aim the vergence cameras at a particular point in

$(x_{fp},0, z_{fp})$ ┼ Fixation Point

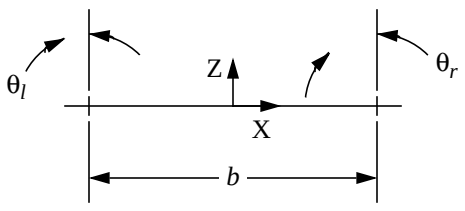

Fig. 15. Planar vergence geometry.

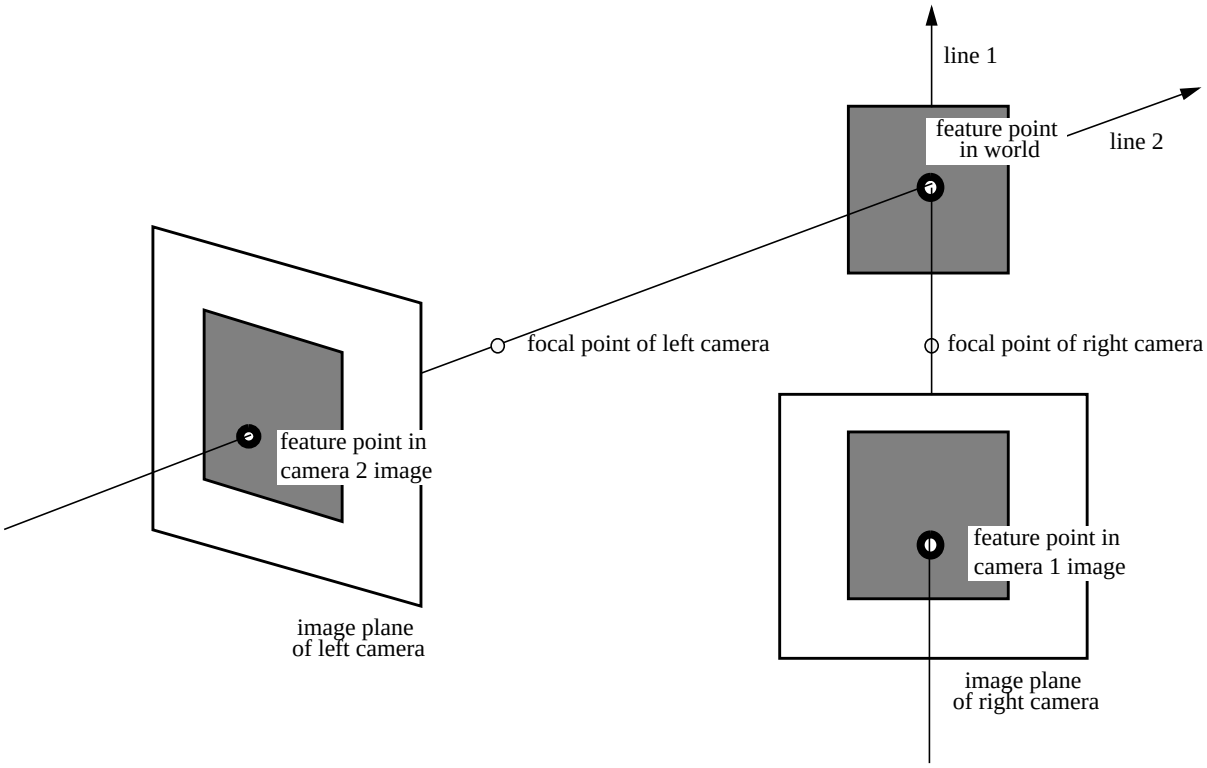

Fig. 16. Triangulation geometry.

space. Since TRICLOPS is a four degree-of-freedom device, there is one more degree of freedom than is required for this task. This redundancy is manifested in the fact that there are an infinite number of joint configurations which may be used to point to a particular Cartesian position. A convenient way to parameterize the *self-motion*, or motion of the joints which results in the same fixation point, is to specify the desired angle of the pan (base rotation) degree of freedom.

For a given fixation point, there is a range of base motion for which the vergence and tilt axes may be rotated to point the cameras at the desired location. For symmetrical vergence, the base is rotated such that the camera baseline is perpendicular to the vector from the world origin to the fixation point. Unless this results in a pan rotation close to the limits of rotation, this value is the center of the range of base self-motion. As the base rotates away from this point (while maintaining fixation), the asymmetry in the vergence angles increases until one of the vergence degrees of freedom reaches its limit. This situation is shown in Figure 17. To compute the range of base self-motion, first the angles and lengths of sides of the vergence triangle are computed for the case where the vergence degree of freedom is at its limit. Next, parametric equations are solved to find the value of θ_{pan} at the intersection of a circle of radius $b/2$ centered at the world origin (base rotation path) with a sphere of radius $|\boldsymbol{r}_{verg}|$ centered at the fixation point.

The equation of sphere centered at the vergence point is

$$(x - x_w)^2 + (y - y_w)^2 + (z - z_w)^2 = \left|\boldsymbol{r}_{verg}\right|^2.$$

where x_w, y_w, and z_w are the x, y, and z components of the vergence point vector $\boldsymbol{r}_w$. The parametric equations of the base rotation path are

$$x = \left(\frac{b}{2}\right)\cos\theta_{pan}$$

$$z = \left(\frac{b}{2}\right)\sin\theta_{pan}.$$

┼ Fixation Point

β

$\boldsymbol{r}_{verg}$

$\boldsymbol{r}_w$

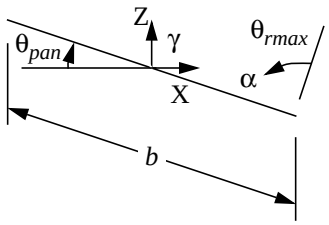

Fig. 17. Determining range of base motion.

Substituting these into the sphere equation above, and adding the constraint y = 0 gives

$$-2(\frac{b}{2})x_w\cos\theta_{pan} + 2(\frac{b}{2})z_w\sin\theta_{pan} + x_w^2 + y_w^2 + z_w^2 + (\frac{b}{2})^2 - |\boldsymbol{r}_{verg}|^2 = 0.$$

This equation is of the form $a\cos\theta_{pan} + b\sin\theta_{pan} + c = 0$, which may readily be solved for θ_{pan} using Newton's second-order method (James et al. 1977). This determines θ_{pan} to 2 mrad (0.11 deg) in usually less than 2 iterations.

Given a desired value of pan rotation within the allowable range, the remaining joint angles are determined as follows. First, the fixation point vector $\boldsymbol{r}_w$ is rotated by θ_{pan} about the world Y axis:

$$\boldsymbol{r}' = \boldsymbol{R}(\theta_{pan}, (0, 1, 0))\boldsymbol{r}_w$$

where $\boldsymbol{r}'$= the rotated fixation point vector and $\boldsymbol{R}(\theta_{pan}, (0,1,0))$ = rotation of θ_{pan} about the world Y axis. The tilt angle can then be computed using

$$\theta_{tilt} = \tan^{-1}\left(\frac{z'}{y'}\right)$$

where z' and y' = the z and y components of $\boldsymbol{r}'$.

Once θ_{tilt} is known, the coordinates of the vergence point with respect to a coordinate system aligned with the vergence plane can be obtained by

$$\boldsymbol{p}_{fp} = \boldsymbol{R}(\theta_{tilt}, (1, 0, 0))\boldsymbol{r}'.$$

where $\boldsymbol{R}(\theta_{tilt}, (1,0,0))$ = rotation of θ_{tilt} about the X axis.

At this point, we are back to the situation shown in Figure 15, and the left and right vergence angles are given by

$$\theta_r = \text{atan}\left(\frac{b/2 - x_{fp}}{z_{fp}}\right)$$

$$\theta_l = \text{atan}\left(\frac{-b/2 - x_{fp}}{z_{fp}}\right).$$

7. Tracking Experiments

One of the most basic capabilities which may be provided by an active vision system is to track an object, using visual information to direct the gaze and keep the object in the field of view of the cameras (also called gaze control, visual pursuit, stabilization). As an application example, some algorithms for performing object tracking have been implemented with the TRICLOPS system. This section will describe one of the algorithms used, and will summarize the results which have been obtained to date.

In the current tracking experiments, the emphasis has been placed on maximizing the tracking performance in terms of speed (bandwidth). As such, the visual processing has been simplified to the minimum required functionality to reduce the image processing delays as much as possible. Consequently, the image processing used for these tracking experiments consists of computing the centroid of a thresholded intensity image of a single object which presents a high-contrast image to both the left and right vergence cameras. For these experiments, only the two vergence cameras are used.

In the current approach to tracking, 3D world position-based motion modeling of the target is used with prediction to generate goal fixation points. A block diagram of the tracking control system is presented in Figure 18. New centroid data are obtained from the world model every 1/30 s. The cameras are used in non-interlaced mode, and only information from the odd fields is used. Delayed joint positions which correspond in time with the centroid data are read in immediately after reading the centroid data. The delayed positions are supplied by the Servo Neck and Vergence SP/EX processes as described in section 3. The centroid information is corrected for lens distortion, and the corrected centroids are used along with the delayed joint positions to compute the position of the object with respect to the TRICLOPS world reference frame, as discussed in the previous section.

Because of the image processing delay, which has been measured to be about 0.084 s, the computed object position corresponds to the position at t-0.084 s, where t is the current time. If this position is used directly as the goal for tracking, poor performance will result. The need for prediction to compensate for image processing delays in real-time visual servoing tasks is well known (Allen et al. 1990; Brown 1989; Robinson 1987). In contrast with the α–β–γ filters used for prediction in Allen et al. (1990) and Brown (1989), the prediction used in the current experiments is based on performing least-squares fits of low-order polynomial equations of time to a limited sequence of object position data. These object motion functions are then evaluated at a time approximately equal to the estimated processing delay. Cubic polynomials which are updated using the12 most recent object position estimates have worked well in the current

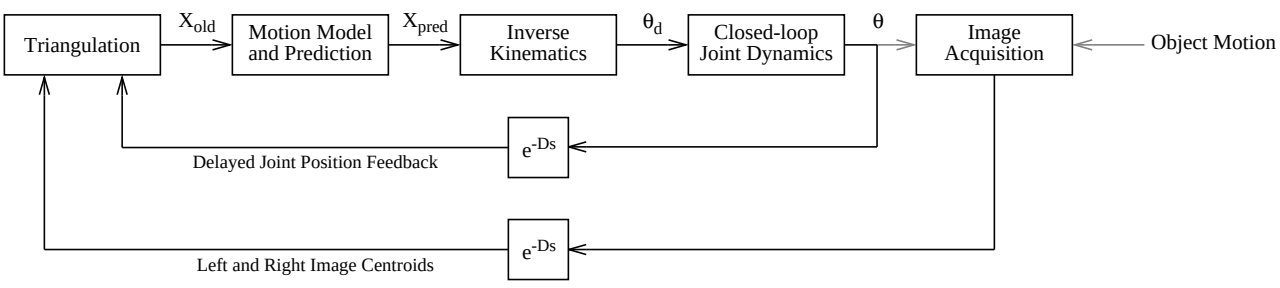

Fig. 18. Block diagram of tracking algorithm.

experiments. The object motion functions are defined such that the result of the equation when evaluated at t = 0 gives the most recent object position estimate. This allows the pseudoinverse of the functions of time matrix to be constant (and able to be computed a priori). The least squares estimate can then be updated every centroid data cycle with a single matrix multiply. More details regarding the performance of this predictive filter and comparisons with α–β and α–β–γ filters may be found in (Wavering and Lumia 1993).

After determining the predicted position of the object with respect to TRICLOPS, the desired position of the pan axis is determined. Several different means of distributing motion between the pan and vergence axes have been implemented and evaluated. These include not moving the base at all, moving the base to maintain symmetrical vergence, and moving the base some fixed percentage of the angle required for symmetrical vergence. One of the most effective methods consists of incrementing the command position by some percentage of the distance between the angle required for symmetrical vergence and the current command position. This has the effect of gradually moving the pan axis toward the direction of the object. As a result, vergence asymmetry is reduced and there is no "preferred position" of the base. The responsiveness of the base is determined by the size of the percentage parameter. A value of 0.1 is a good compromise between moving the base too sluggishly and putting too much of the tracking burden on the base.

Finally, the inverse kinematic equations are solved (with the desired base angle as an input parameter) to yield the desired joint angles. The desired joint angles are commanded to the Servo level, which executes the high-bandwidth PD servo algorithm to achieve them. New joint angle goals are computed on the same 30 Hz timing cycle as image centroids. These position command updates are interpolated by the Servo Level PL process into 0.0005 s subcommands. This interpolation greatly smooths the tracking motion, but it also incurs some additional delay which must be accounted for in the prediction of object position. This delay, combined with the image processing delay, results in a total delay of about 0.117 s. However, the predictive filter itself results in a small amount of frequency-dependent phase lead (Wavering and Lumia 1993), such that using a time of 0.090 s in the prediction equations yields better overall tracking than using 0.117 s.

Several experiments have been performed to evaluate the tracking performance attained using the above techniques with TRICLOPS. In one test, a 3.5 cm diameter white ball attached to the end of a 0.44 m rod is rotated in a horizontal circle whose center is 1.1 m in front of TRICLOPS. The set-up for this experiment is shown in Figure 20.

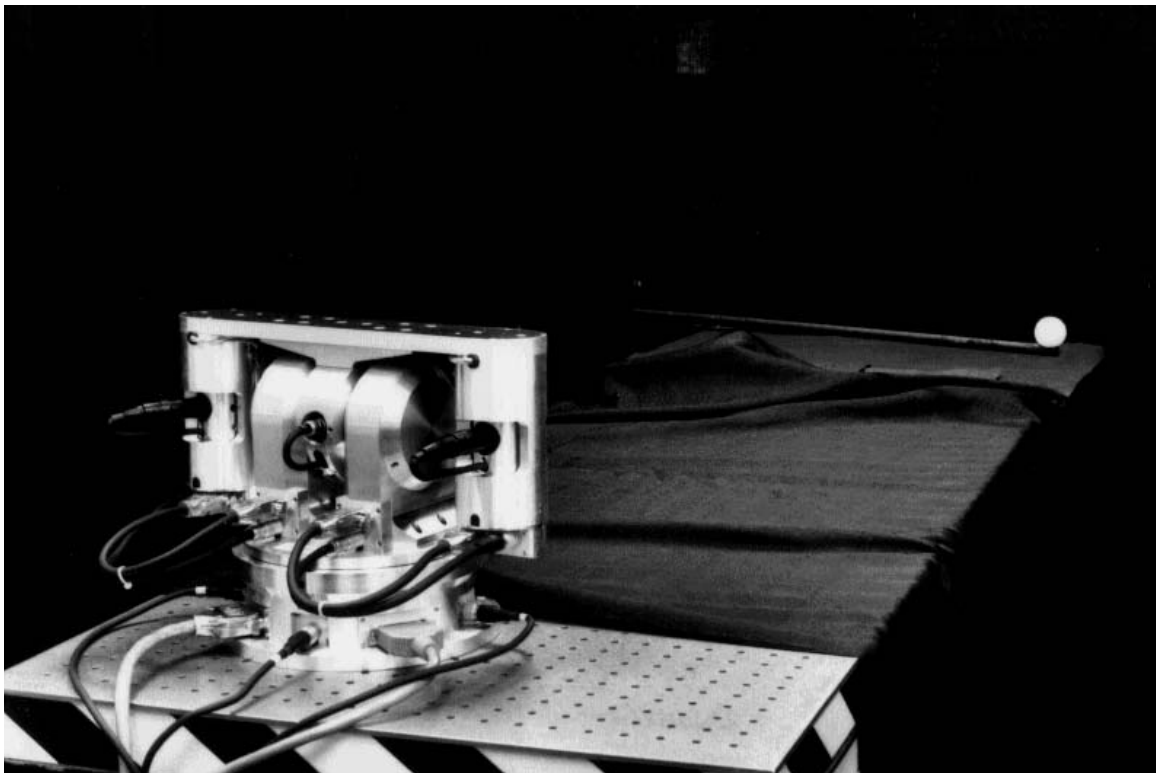

Fig. 19. Experimental set-up for circular tracking.

The tracking performance with and without prediction for a relatively slow orbital velocity of the ball (1.26 rad/s) is shown in Figure 20. Here, the tracking error is defined as the horizontal offset of the centroid from the center of the image (in pixels). The peaks in Figure 20 occur as the tracking speed of the cameras increases when the ball passes close to TRICLOPS. It is seen that when prediction is not used, the peak tracking error becomes quite large. Without prediction, the centroid is lost from the field of view if the orbital velocity is increased to above about 1.8 rad/s (100 deg/s). When 0.09 s of prediction is used, however, the tracking error is reduced by more than a factor of 2. Although the prediction increases the sensitivity to measurement noise somewhat, it enables TRICLOPS to track the ball at orbital velocities up to about 8.6 rad/s (490 deg/s) before the error causes the centroid to be lost from the image. The tracking performance with prediction for a quite fast orbital velocity of 6.9 rad/s (395 deg/s) is shown in Figure 21. For this orbital velocity, the magnitude of the velocity of the vergence axes reaches a maximum of about 6.5 rad/s (372 deg/s), and the magnitude of the peak base rotation velocity is about 1.7 rad/s (97 deg/s). The tracking error for this case is actually *smaller* than for the slower speed because the additional phase lead of the predictive filter at this frequency increases the effective prediction, such that it more closely matches the amount required for zero-latency tracking (Wavering and Lumia 1993). The predictive filter also enables TRICLOPS to track smoothly and stably when unpredictable motions of the ball at the end of a hand-held rod are performed.

The spatial accuracy of the position estimates determined by TRICLOPS during high-speed (6.9 rad/s) tracking is shown in Figure 22. The solid line in the figure indicates the actual path of the ball during a single revolution, as determined by careful manual measurement. The smooth dotted line close to the actual path is the estimated path of the ball (the input to the predictive filter), and the dashed line shows the predictive filter output. The data for both position estimates were collected while prediction was used for tracking. It is seen that the spatial error in the estimated path without prediction is quite small, with a maximum deviation of about 0.025 m at the far side of the circle. The accuracy, particularly in range, is target-dependent. The data of Figure 22 were obtained during tracking in ambient light. Accuracies on the order of a few millimeters have been obtained when a special target (for example, an internally illuminated ball) is used to produce better quality images. Figure 22 also shows that at 6.9 rad/s, the prediction adds a small amount of spatial error to the position estimate. As the orbital speed is increased, predictive overshoot causes larger deviations from the actual path of the ball.

8. Conclusions

Although it has been only about six years since the development of the first animate vision systems patterned after the human eye-head arrangement, much progress has been made. A significant recent development is the commercial availability of micro-miniature CCD cameras, which has enabled the goal of human-like dynamic performance to be achieved. TRICLOPS is a new direct-drive active vision system which takes advantage of these imaging devices to enable very large rotational velocities and accelerations. TRICLOPS is capable of peak velocities of over 32 rad/s (1830 deg/s) for the vergence degrees of freedom, 17.5 rad/s for the tilt axis (1000 deg/s), and 11.5 rad/s (660 deg/s) for pan rotations.

TRICLOPS provides concurrent multi-resolution sensing capability by locating an integral third camera between the two vergence cameras. The current lens combination uses 15 mm lenses with a 0.42 rad (24 deg) horizontal field of view for the vergence cameras, and 4 mm lenses with a horizontal field of view of 1.35 rad 77.3 deg) for the center camera. Other unique features of this system include modular construction, the use of industrial-quality components, and the incorporation of design elements which decrease the possibility of damaging the system during experimentation.

The NASREM hierarchical control system which is used with TRICLOPS provides a modular, easily-modifiable infrastructure for implementing and testing active vision algorithms. External processes can provide commands to the system at either trajectory generation or servo levels via VME common memory interface buffers. Position and velocity feedback buffers are also accessible. The PID servo loops run at a 2 kHz rate, giving the vergence axes a closed-loop bandwidth of 50 Hz. System parameters, such as servo gains, can be changed every cycle if desired. A number of different trajectory algorithms are available for high-speed saccades and for testing purposes.

The full potential of an active vision system cannot be realized without real-time image processing capability. In our lab, this capability is provided by PIPE for low-level image processing, along with a VME system for higher-level sensory processing and world modeling. The visually-based closed-loop performance of the system has been demonstrated by implementing a simple tracking application. Using triangulation and 3D motion modeling and prediction,

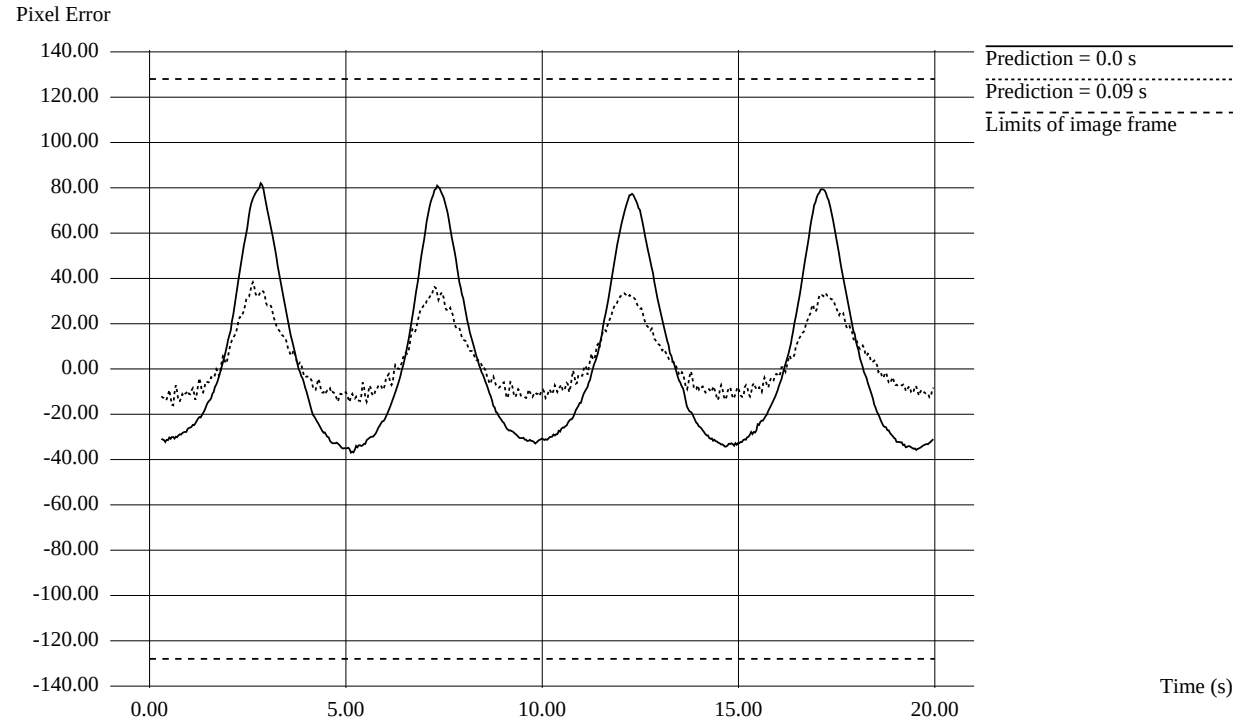

Fig. 20. Tracking error during slow target motion (orbital velocity = 1.26 rad/s).

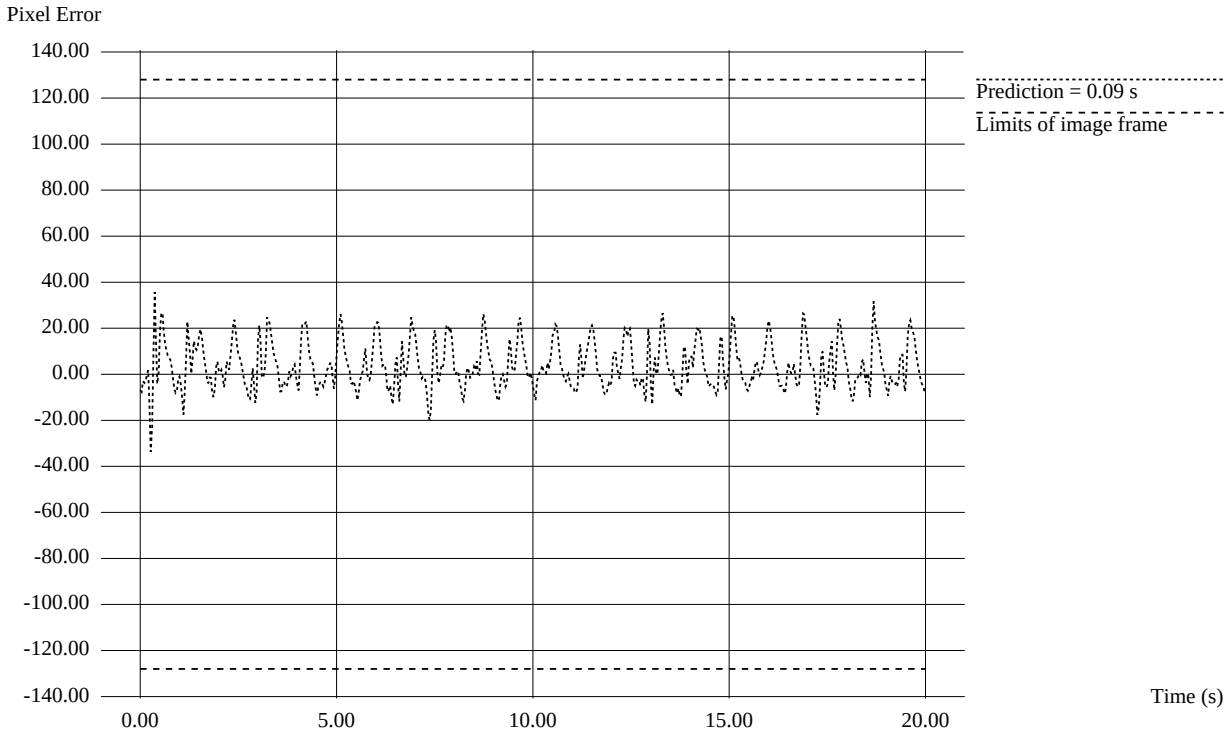

Fig. 21. Tracking error during fast target motion (orbital velocity = 6.9 rad/s).

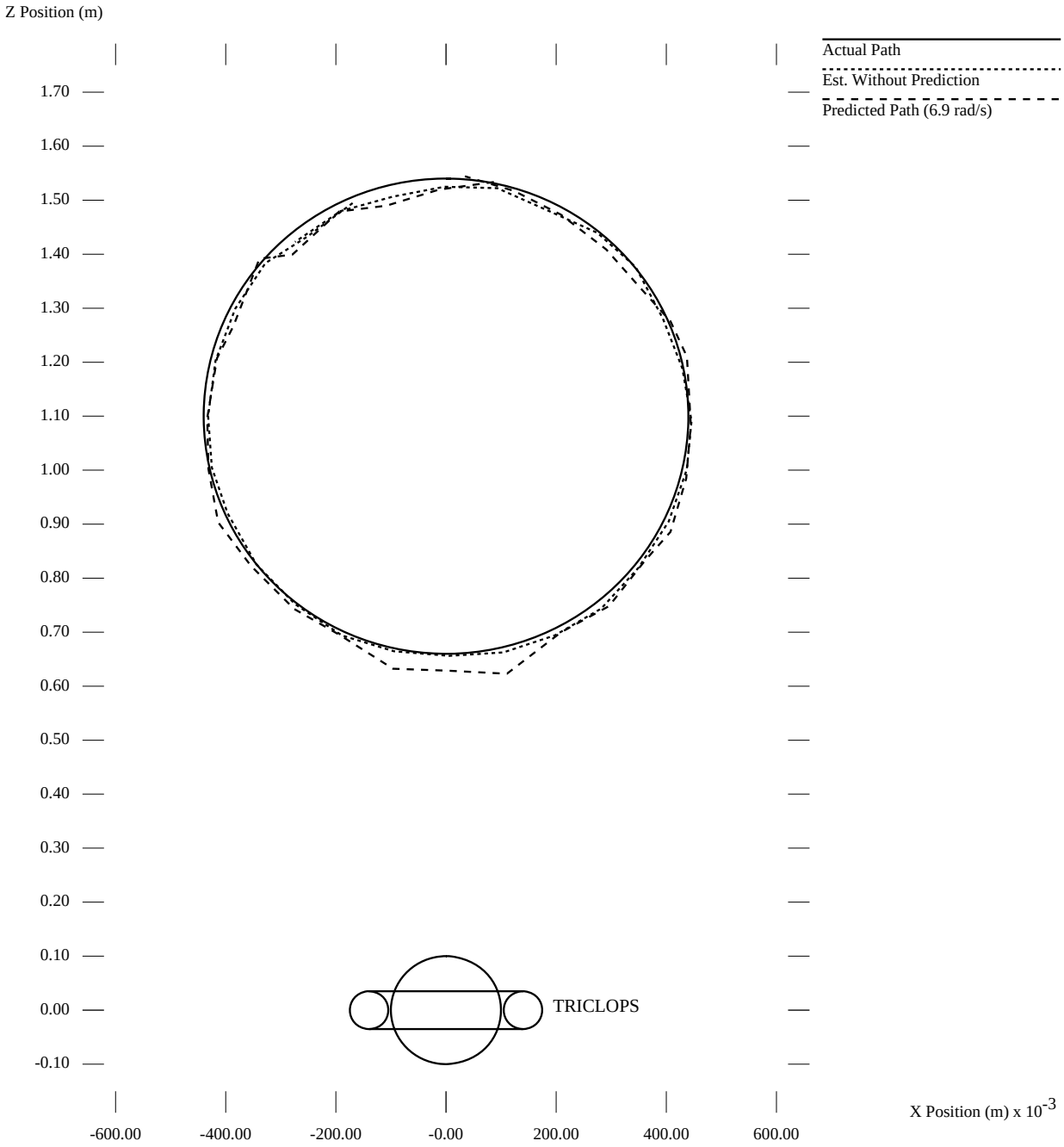

Fig. 22. Estimated ball path with and without prediction (orbital velocity = 6.9 rad/s).

TRICLOPS is able to keep a ball in the field of view of the vergence cameras for tracking velocities in excess of 6 rad/s (344 deg/s).

We have just begun to explore the potential of TRICLOPS. Future plans include using the center wide-angle view camera for the implementation of algorithms which use color and motion to autonomously direct the attention of the system. We also intend to connect the TRICLOPS control and image processing/world modeling VME systems with another existing NASREM system used to control a seven degree-of-freedom manipulator (Fiala et al. 1992). This will provide a very useful platform for studying visually-guided manipulation tasks, such as assembly. TRICLOPS will also be used in the development of active vision applications in inspection, surveillance, and mobile navigation.

9. Acknowledgments

We would like to thank our coworkers for their valuable contributions, including Tom Wheatley for building the TRICLOPS electronics, Marilyn Nashman for developing parts of the visual processing software, Adam Jacoff for aiding in the design and construction of TRICLOPS, and Henry Schneiderman for ideas and suggestions relating to prediction and curve-fitting. Our gratitude is also extended to the reviewers of this paper for their careful reading and helpful comments.

10. References

Abbott, A. L., Ahuja, N. 1990. Active surface reconstruction by integrating focus, vergence, stereo, and camera calibration. In *Proc. Third International Conference on Computer Vision*, Osaka, Japan, pp. 489-492.

Albus, J. S., McCain, H. G., Lumia, R. 1987. NASA/NBS Standard Reference Model for Telerobot Control System Architecture (NASREM). National Inst. Standards and Technol., Gaithersburg, MD, Technical Note 1235.

Allen, P., Yoshimi, B., Timcenko, A. 1990. Real-time visual servoing. In *Proc. DARPA Image Understanding Workshop*, Pittsburgh, PA.

Aloimonos, J., Weiss, I., Bandyopadhyay, A. 1987. Active vision. In *Proc. First International Conference on Computer Vision*, pp. 35-54.

Aloimonos, J., Shulman, D. 1989. *Integration of Visual Modules: An Extension of the Marr Paradigm*. Academic Press, Inc.: New York, NY.

Aspex, Inc. 1987. PIPE--An introduction to the PIPE system. New York, NY.

Bajcsy, R. 1988. Active perception. *Proceedings of the IEEE* 76(8): 966-1005.

Ballard, D., Ozcandarli, A. 1988. Eye fixation and early vision: Kinetic depth. In *Proc. 2nd International Conference on Computer Vision*, pp. 524-531.

Ballard, D. H. 1991. Animate vision. *Artificial Intelligence*, 48:57-86.

Bederson, B. B., Wallace, R. S., Schwartz, E. L. 1992. Two miniature pan-tilt devices. In *Proc. IEEE Conference on Robotics and Automation*, pp. 658-663.

Brown, C. M. 1989. Kinematic and 3D motion prediction for gaze control. In *Proc. Workshop on Interpretation of 3D Scenes*, Austin, TX, pp. 145-151.

Carpenter, R. H. S. 1988. *Movements of the Eyes*. Pion Limited: London, England.

Chaconas, K., Kelmar, L., Nashman, M. 1990. A NASREM implementation of position determination from motion. National Inst. Standards and Technol., Gaithersburg, MD, Internal Report 90-4286.

Clark, J. J., Ferrier, N. J. 1988. Modal control of an attentive vision system. In *Proc. 2nd International Conference on Computer Vision*, pp. 514-523.

Fiala, J. 1989. Note on NASREM implementation. National Inst. Standards and Technol., Gaithersburg, MD, Internal Report 89-4215.

Fiala, J., Wavering, A., Lumia, R. 1992. A manipulator control testbed: Implementation and applications. In *Proc. AAS Conf. on Guidance and Control*, Keystone, CO.

Geiger, D., Yuille, A. 1989. Stereo and eye movement. *Biological Cybernetics* 62: 117-128.

James, M. L., Smith, G. M., Wolford, J. C. 1977. *Applied Numerical Methods for Digital Computation.* Harper & Row: New York, NY.

Krotkov E., Fuma, F., Summers, J. 1988. An agile stereo camera system for flexible image acquisition. *IEEE Journal of Robotics and Automation* 4(1): 108-113.

Olson, T. J., Coombs, D. J. 1991. Real-time vergence control for binocular robots. *International Journal of Computer Vision* 7(1): 67-89.

Pahlavan, K., Eklundh, J. O. 1992. A head-eye system—Analysis and design. *CVGIP: Image Understanding* 56(1): 41-56.

Poggio, T., et al. 1988. The MIT vision machine. In *Proc. DARPA Image Understanding Workshop*, pp. 177-198.

Robinson, D. A. 1987. Why visuomotor systems don't like negative feedback and how they avoid it. In Michael Arbib and Allen Hanson, eds, *Vision, Brain, and Cooperative Computation*. MIT Press: Cambridge, MA.

Schwab, E. C., Nusbaum, H. C., eds. 1986. *Pattern Recognition by Humans and Machines, Volume 2, Visual Perception*. Academic Press: New York, NY.

Swain, M. J., Stricker, M., eds. 1991. Promising directions in active vision. Report from the NSF Active Vision Workshop, University of Chicago.

Tsai, Roger Y. 1987. A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses. *IEEE Journal of Robotics and Automation* RA-3(4): 323-344.

Wavering, A. J., Lumia, R. 1993. Predictive visual tracking. To appear in *Proc. SPIE Intelligent Robots and Machine Vision XII*, Boston, MA.

Webb Associates, eds. 1978. *Anthropometric Source Book, Volume I: Anthropometry for Designers*, NASA Reference Publication 1024.

Wurtz, R. H., M. E. Goldberg, eds. 1989. *The Neurobiology of Saccadic Eye Movements*. Elsevier: New York, NY.