A High-Performance Camera Platform for Real-Time Active Vision

Andrew Brooks†, Glenn Dickins†, Alexander Zelinsky†, Jon Kieffer‡ and Samer Abdallah‡

† Robotic Systems Laboratory
Research School of Information Sciences and Engineering
The Australian National University
Canberra ACT 0200, Australia
email: {zoz,alex,glenn}@syseng.anu.edu.au

‡ Department of Engineering
Faculty of Engineering and Information Technology
The Australian National University
Canberra ACT 0200, Australia
email: {jon,samer}@faceng.anu.edu.au

Abstract

While active vision is a relatively new approach to computer vision, it offers impressive computational benefits for scene analysis in realistic environments. This paper describes a novel camera platform for the real-time, real-world application of active vision in robots. Requirements for performance are presented as are the figures actually achieved, along with an alternative, task-based method of specifying active visual system abilities. Details of the platform's cable drive transmission mechanism are provided as well as the advantages given by this scheme. Finally, research directions involving this platform are discussed.

1 Introduction

An important decision in the design of an autonomous robotic system is the choice of sensory mechanism(s) to be employed. The robot must be able to gather data from its environment which is both of adequate content and in a sufficiently timely fashion for it to make the decisions necessary for task-oriented behaviour. For the purposes of our research, the tasks are individual and co-operative mobile and manipulative robotics in the real world, and the time constraint is that the robots perform these tasks in real time.

We define real time as a time period commensurate to the rate of change of the environment, so appropriate sensory and processing resources must be selected with this in mind. The real world itself is an unstructured, possibly cluttered, dynamic environment which extends beyond sensor range. A robot operating in the real world must therefore be equipped with mechanisms to fixate its attention on that which is important within the time frame of its relevance, while simultaneously disregarding background irrelevancies. Attention should be directed at what can loosely be defined as "interesting items". Certain objects and locations are "interesting" in that they may be useful to the completion of goal-directed behaviour, and other agents that may be present are "interesting" because it is their interaction with the environment that causes it to be dynamic.

The sensory system must therefore be capable of shifting focus to the location of something interesting and tracking it if it is moving. In addition, as the environment is large, the robot must be able to use the means at its disposal, such as mobility, to extend its field of awareness via searching behaviour. We believe that all these requirements point strongly towards the use of vision as the primary source of sensory input. More specifically, we refer to a recently emergent computer vision methodology, active vision[Ballard 1991], processed at video frame rate (30 Hz).

This technique provides at the least saccade and smooth pursuit, as well as being capable of such behaviours as opto-kinetic reflex[1] and vestibulo-ocular reflex[2] given sufficient processing power, and so we feel that it is eminently suitable to satisfy the constraints as defined. We have therefore designed a new, modular, high-performance camera platform for real-time active vision. In this paper we describe our design and explain the reasons for our particular approaches with respect to the wide area of operations that we envision for it. We also provide performance specifications for our system and discuss intended and possible applications for it in field and experimental robotics.

2 Design Principles

The fundamentals of our design have been heavily influenced by the mechanics of the human visual system. The human eye achieves extraordinary performance through its low weight and inertia and its use of muscles for actuation. Muscles are themselves lightweight, have high

[1] Using optical flow to stabilize images and maintain fixation when experiencing unpredictable motion, for example motion due to the environment.

[2] Using proprioceptive information to hold the gaze point while the head or body is moving with respect to the target.

acceleration and do not suffer from the common problem encountered in active head design whereby each degree of freedom requires sufficiently powerful actuators to move all previous degrees of freedom including their actuators. This has previously led to large, heavy, over-engineered camera platforms[Pahlavan and Eklundh 1992] or systems with reduced degrees of freedom such as coupled vergence[Clark and Ferrier 1993]. Other research laboratories engaged in active vision have shown that it is possible to build high performance heads using this traditional incrementally dependent configuration, but by incurring corresponding disadvantages in cost, complexity or weight[Sharkey *et al.* 1993, Kuniyoshi *et al.* 1995]. We have found that by relocating the motors to a fixed base and thereby minimising the active component to a small, low-inertia 'eye', we can achieve acceptably high performance without such significant penalties. This is therefore the major guiding principle for our design, among several other important issues which have influenced our project, the complete list of which is as follows:

1. High level of performance via a cable drive system which does not require any motors to be moved about an axis. The actual performance which we consider to be acceptable is elaborated on in a later section, as is a description of the cable drive system itself, but is based on the necessity for useful real-time operation of the mechanism.
2. Light weight, both of the moving components in order to minimise inertia, and of the overall system in order to ensure its suitability for applications where weight is a factor, such as mounting on a manipulator or small mobile robot.
3. Modularity of the visual apparatus, to enable the construction of specialized seeing systems composed of single or multiple active camera platforms. It is with this concept in mind that we have modeled our design as a single 'eye' with independent pan and tilt. Although this configuration suffers from the need to ensure that the tilt movement is precisely matched between cameras in a standard vertically aligned stereo setup[Murray *et al.* 1993], this disadvantage is sufficiently correctable within the control strategy and it was felt that the difficulty was outweighed by the flexibility and reusability offered in the modular design.
4. The ability to detect and process visual information in colour. Our robots are expected to perform tasks in environments involving humans, and indeed many of those tasks may depend on interactions with the humans. It has been demonstrated that colour information is a valuable cue for detecting and interpreting the actions of humans (see for example [Appenzeller *et al.* 1997a]), so we deem it essential that our system be equipped with colour video cameras.
5. Easy reconfigurability of the active platform. It can be useful in both experimental and field situations to be able to reconfigure the visual system with a minimum of effort and disassembly. For example, some scene analysis tasks may require a wide field of view, while others may benefit from a narrow, high-resolution image stream, so we have designed our system to accept standard C-mount lenses which are easily attached by hand. Also, the single independent camera design allows quick adjustment of the imaging geometry in a multi-camera arrangement, for instance the interocular (baseline) distance, by simply moving the camera setups.
6. Low cost and standard components have been selected wherever possible to maximise affordability and ease of parts replacement, both of which are beneficial to laboratory and field operations alike. Several aspects of the mechanism have been specifically designed to avoid the use of specialized components. For example, the large reduction on the drive mechanism enables us to avoid the use of high-precision encoders on the motor shafts in order to obtain the necessary angular resolution of the camera movement, and the inclusion of all reduction within the backlash-free cable drive removes the need for expensive or difficult solutions to gearbox backlash.

3 Performance Specifications

The performance figures traditionally reported for active vision systems consist of maximum angular velocity, maximum angular acceleration, angular resolution and axis range. While the latter two are highly relevant, we consider the others to be not especially useful in that they do not detail any form of specific task competency. We therefore propose four additional specifications for an active vision system which not only involve the speed and acceleration of the axes but also express the usage intention of the system in the form of a functional requirement (Table 1).

Our platform achieves these values in order to satisfy constraints related to its desired abilities with respect to its intended visual input, 30Hz RS-170 interlaced video.

The maximum allowable full-speed saccade time was required to be 0.18 seconds, which would allow the camera to perform up to three 90-degree saccades, each preceded by four target location video frames and succeeded by one stabilization video frame, per second. The four target location frames can be used to feed motion detection algorithms which require multiple consecutive

Minimum time for 90-degree saccade in pan	0.15s
Minimum time for 90-degree saccade in tilt	0.15s
Maximum pan angle change achievable from stationary start to stationary stop within the time of one video frame	15°
Maximum tilt angle change achievable from stationary start to stationary stop within the time of one video frame	15°

Table 1: Task-oriented performance specifications of the camera platform

frames of input and/or to perform matching and error checking of the saccade destination versus the desired target. Slightly more than five video frames are captured during the saccade itself. Although the interlaced video will cause extreme motion blur in the full frames, if appropriate processing resources are available single interlaces can be used to segment scene motion from egomotion via an optical flow calculation. The extra portion of video frame is the start of the stabilization frame which may be discarded. Our system's performance exceeds this requirement, having a full-speed saccade time of 0.15 seconds.

The minimum allowable full-speed stop-to-stop angular change within one video frame was required to be 15 degrees, which equates to the ability to predictively track an object moving past the cameras at up to four metres per second at a distance of one metre, using an image processing system which is not capable of computing optical flow and hence must stop and discard every second frame due to egomotion blur before visually reacquiring the target. If the system is equipped with the capacity for segmenting scene motion from egomotion then the predictive tracking speed will of course be much higher.

Note that while our system's figures for these new specifications are the same for both pan and tilt, they need not necessarily be if increased or reduced performance for an axis was necessary to fulfil a special requirement or cope with a design constraint. We have therefore included the separate description style as a guide for other researchers who may wish to use them. The abilities of our platform expressed in the traditional performance metrics, along with our values for angular resolution and axis range, are given in Table 2.

The angular resolution has been selected with the aim of allowing the platform to perform meaningfully small camera movements with a wide variety of lenses, including those having foveal fields of view. For example, a six degree fovea occupying a full RS-170 video frame has a resolution of 0.0117 degrees per pixel in the horizontal axis and 0.0124 degrees per pixel in the vertical axis. Thus the 0.01 degree angular resolution of our apparatus allows single pixel movements in both pan and tilt for even such narrow angle or zoom lenses.

Specification	Pan	Tilt
Maximum velocity	$600^{\circ s^{-1}}$	$600^{\circ s^{-1}}$
Maximum acceleration	$72000^{\circ s^{-2}}$	$72000^{\circ s^{-2}}$
Angular resolution	0.01°	0.01°
Axis range	120°	180°

Table 2: Standard performance specifications of the camera platform

4 Drive System

For the actuation mechanism of our platform we have selected a cable drive scheme of a type that has shown much promise in robotic applications such as manipulator arms[Townsend and Salisbury 1993] and other active vision systems[Brooks and Stein 1993]. Without going into excessive detail, the cable drive principle involves the transferral of power from the motor to the final moving component via a series of cable-wrapped pinions and pulleys. The cables are fixed at both ends and rely on the wrapping friction to turn the intermediate drive train components. The cables used are 1.17mm diameter 343-core steel cables which have high levels of strength (77kg breaking tensile) and flexibility, while retaining insignificant elasticity and appropriate friction characteristics. This is sufficient to ensure that no issues of compliance or slipping arise.

The cable drive transmission confers several significant benefits upon our design:

- No backlash is contributed to the mechanism. For standard, low-cost gear mechanisms backlash is significant when compared to the desired angular resolution in this case. Cable drives obviate the need for performance compromises or expensive solutions to backlash problems.
- Cable drives impart low friction to the reduction scheme. Specifically, they are free from the friction created by the tooth meshing in a geared arrangement. For a design such as ours, where the motors have been moved away from the actuated components via a drive train, it is advantageous to avoid this in order to achieve the high accelerations required with relatively small motors.
- The low friction property of cable drives also provides advantages in that they are suitable for rapid motion without needing any special lubrication scheme. A common low-backlash alternative which was considered, the harmonic drive gearbox, proved

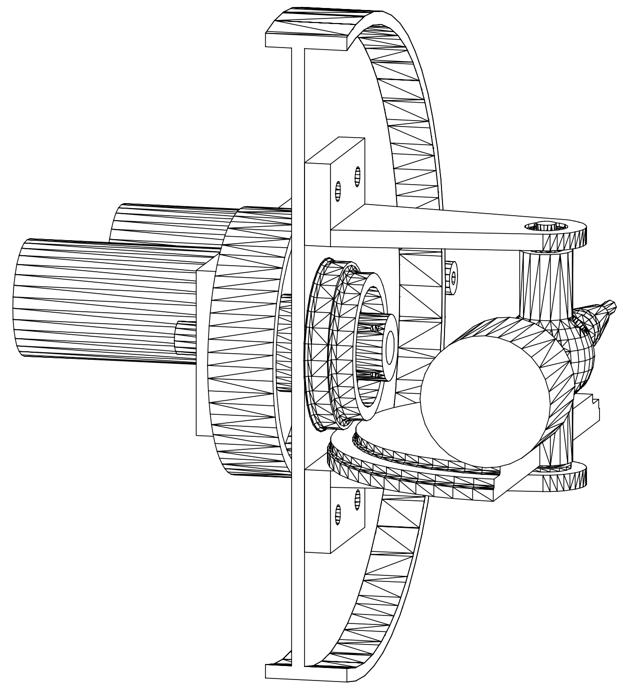

Figure 1: Front view of the camera platform showing the cable bevel gear

inadequate due to its inability to be effectively lubricated at such speeds.

- They are free from problems with compliance that may occur with mechanisms such as belt drives or certain types of gears. Again, the precision of angular resolution required for our application makes this important.
- Cable drives are a relatively low cost method of constructing an experimental system. While the cables are a specialized component, they are not particularly expensive relative to many robotic components and the machining of parts required is not especially complicated.
- Damage to the mechanism is uncommon as the most likely failure mode is cable breakage, which is the least costly component and the easiest to deal with as it need simply be replaced and retensioned. Other alternatives such as geared systems can incur unpleasant levels of downtime and expense in the event that they are stripped or otherwise incapacitated.

The drive mechanism for our system consists of two separate cable drive trains, one for each of pan and tilt, coupled at one point near the camera. The coupling is in the form of a cable bevel gear inspired by the one presented in[Townsend 1988], and this makes it possible for both motors to remain stationary during movement about both axes. The pan motor is connected to a pinion

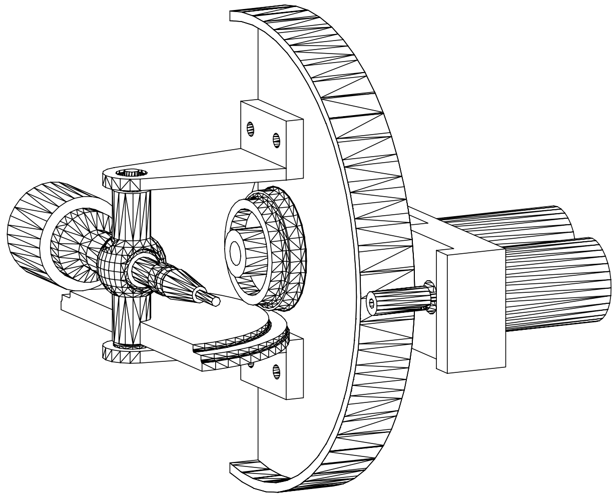

Figure 2: Rear view of the camera platform showing the pinion of the tilt drive mechanism

which winds a cable onto a larger drum to provide speed reduction. This drum is connected to the vertical wheel of the bevel cable gear, which is a complete circle. The tilt motor also drives a pinion which winds a larger drum portion, onto which are attached the immobile camera mount holders. On the camera mount itself, which is free moving, is the horizontal wheel of the bevel cable gear. This wheel is semicircular in order to provide a larger radius without adding excessive weight.

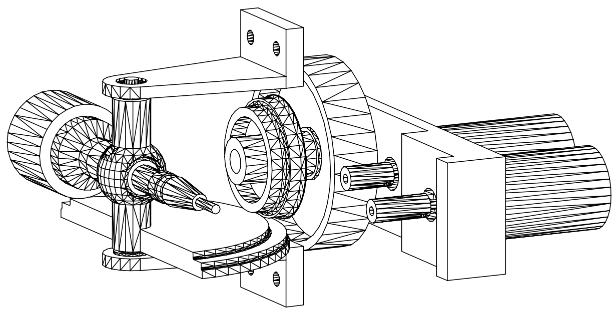

Figure 3: Cutaway rear view of the camera platform with the tilt wheel removed, showing the pinion of the pan drive mechanism

As the system is coupled here at the bevel system, pan motion can be achieved by driving the pan motor alone, but it is necessary to drive both motors when performing a tilt motion. There is no difference in performance between axis motions other than power consumption and the number of parts that move, so we have chosen pan as the single motor axis mainly because we expect most watching and tracking applications to explore more frequently in the horizontal than the vertical directions. However, if this is not the case, due to the modular nature of this system it is a simple matter to swap the axes by rotating the apparatus ninety degrees.

5 Cameras And Lenses

Our design demands a compact colour video camera of the type common in medical and industrial imaging, consisting of a small, lightweight head and an off-board video converter. We selected the Panasonic GP-KS1000 system (Figure 4), but our mechanical design is of course compatible with any of the other similar microcamera packages available. Not including lenses, the camera head of this particular system has a diameter of 17mm and a length of 42mm, weighs 20g and contains a single 1/2" CCD of 900,000 pixels. It produces full colour output as either separate RGB channels, composite, or Y/C (S-Video). We are using the RGB output of this camera for our colour image processing applications as the colour information is of considerably higher quality than that contained in a composite signal. The camera also both outputs and accepts an external sync signal, which is essential for any project such as this involving stereo applications.

Figure 4: Panasonic GP-KS1000 colour microcamera

We have fitted our camera with a C-mount adaptor to enable us to fit any lens having this standard fixture. This provides the visual system with access to many off-the-shelf varieties, rather than the handful of lenses equipped with the camera's proprietary micromount that are available from the manufacturer. We are currently using three lenses, with fields of view of 33 degrees, 63 degrees and 108 degrees in the vertical direction (the horizontal view is somewhat wider), which gives us an adequate range of choice from detailed and linear images to comprehensive but somewhat distorted views.

6 Processing Hardware

At present our image processing is performed by a MaxTD system from DataCube, consisting of one each of their MaxVideo 200 and DigiColor boards in a VME rack controlled by a MC68040 running the LynxOS real-time kernel (Figure 5). The DigiColor is used to acquire 24-bit colour through separate analog RGB colour inputs, which is then fed to the MV200 via MaxBus for such operations as filtering, motion segmentation and feature matching. Due to limitations inherent in the pipelined architecture we will be adding additional processing resources of different types in the near future.

Motor control is provided by a single Motion Engineering PCX/DSP[3] eight-axis programmable motion controller board. We have chosen to use a PCI-bus DSP system due to the fact that PCI-based architectures are rapidly becoming not only cheap and ubiquitous but capable of levels of performance approaching that suitable for real-time vision-based applications. Although a single installation of our vision platform requires only two axes of motor control, the use of an eight-axis card allows the control of up to four active eyes, or a binocular stereo head and a three degree-of-freedom neck, with the single board. We think it is advantageous to have such flexibility, especially in circumstances where a limited number of PCI slots is available, for example if PCI-based image processing hardware was also being used on a mobile robot with a self-contained computing system such as our Nomad 200[4].

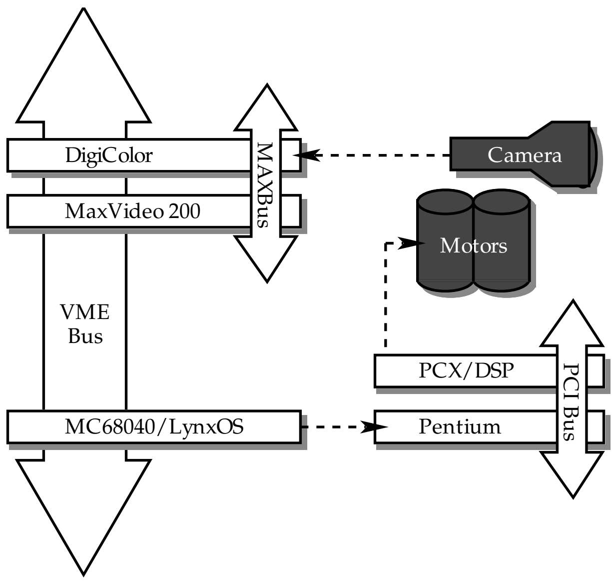

Figure 5: Processing architecture of the active vision system

[3] A trademark of Motion Engineering Inc.

[4] A trademark of Nomadic Technologies Inc.

7 Applications

Due to the modular and reconfigurable nature of this design, we envisage a range of applications for it in several distinct arrangements. Our laboratory intends to pursue and conceive projects involving active vision in the following areas:

- Single camera

  We are currently commencing an autonomous submersible project which we intend to equip with a single active visual sensor. The submarine is approximately one cubic metre in size and will carry all its navigation, movement and processing equipment on board, so a small and lightweight visual system that is able to direct attention to the environment from within an observation cupola is advantageous. We feel that a monocular active vision system is appropriate for the visual servoing tasks which this submersible will be required to perform.

- Binocular stereo

  A primary goal of our laboratory is to conduct research towards autonomous mobile robots equipped with manipulators, operating in and interacting with the real world. Inspired by biological systems, we intend to use active, binocular stereo vision as the primary sensor scheme for these robots. Stereo brings with it significant advantages for a terrestrial mobile system such as depth perception and resistance to occlusion, and the light weight and high performance of our camera platform make it ideal for use on even relatively small robots or the manipulator arms themselves. In addition, we have several ongoing projects in human-robot interaction for which we plan to capitalize on the superior tracking and distance estimation abilities of active binocular stereo.

- Multiple independent

  Passive observation or monitoring of an area has always been an application for vision sensing systems, but recent work in artificial intelligence has suggested the concept of an active space, having the ability not only to perceive the behaviour of the agents within it but also to perform functions to assist these agents in their activities (e.g. [Torrance 1995, Appenzeller *et al.* 1997b]). The fundamental visual task of tracking humans and robots within the space now requires much more detailed interpretation of their position, orientation and actions. A visual system designed for high performance detection of relevant occurrences followed by direction of attention to and tracking of that activity is desirable for providing detailed, reliable data about the requirements of the occupants. We envision multiple installations of our platform situated as independent observation stations around such an active space.

8 Conclusion

We have presented a novel pan-tilt camera unit which has high performance and is of modular design. We have explained how active vision as a sensing technique fits into the philosophy and requirements of our laboratory, and have discussed the current and potential robotic applications we have in mind for the mechanism we have constructed. We are currently developing algorithms for tracking, motion analysis, skin colour segmentation and saccade map learning for use with the hardware, and expect to have results of this intensive experimentation with the system shortly.

9 Acknowledgements

The authors would like to thank Bill Townsend of Barrett Technologies for his valuable and timely advice.

References

[Appenzeller *et al.*, 1997a] G. Appenzeller, Y. Kunii, and H. Hashimoto. A low-cost real-time stereo vision system for looking at people. In *Proc. International Symposium on Industrial Electronics*, Guimarães, Portugal, July 1997. IEEE.

[Appenzeller *et al.*, 1997b] G. Appenzeller, J.-H. Lee, and H. Hashimoto. Building topological maps by looking at people: An example of cooperation between intelligent spaces and robots. In *Proc. International Conference on Intelligent Robots and Systems*, Grenoble, France, November 1997.

[Ballard, 1991] D. H. Ballard. Animate vision. *Artificial Intelligence*, 48:57–86, 1991.

[Brooks and Stein, 1993] R. A. Brooks and L. A. Stein. Building brains for bodies. *MIT AI Laboratory AI Memo No. 1439*, 1993.

[Clark and Ferrier, 1993] J. J. Clark and N. J. Ferrier. Attentive visual servoing. In A. Blake and A. L. Yuille, editors, *Active Vision*, pages 137–154. MIT Press, 1993.

[Kuniyoshi *et al.*, 1995] Y. Kuniyoshi, N. Kita, S. Rougeaux, and T. Suehiro. Active stereo vision system with foveated wide angle lenses. In *Proc. Asian Conference on Computer Vision*, Singapore, 1995.

[Murray *et al.*, 1993] D. W. Murray, F. Du, P. F. McLauchlan, I. D. Reid, P. M. Sharkey, and M. Brady. Design of stereo heads. In A. Blake and A. L. Yuille, editors, *Active Vision*, pages 155–172. MIT Press, 1993.

[Pahlavan and Eklundh, 1992] K. Pahlavan and J.-O. Eklundh. A head-eye system – analysis and design. *Computer Vision, Graphics and Image Processing: Image Understanding*, 56(1):41–56, 1992.

[Sharkey *et al.*, 1993] P. M. Sharkey, D. W. Murray, S. Vandevelde, I. D. Reid, and P. F. McLauchlan. A modular head/eye platform for real-time reactive vision. *Mechatronics*, 3(4):517–535, 1993.

[Torrance, 1995] M. C. Torrance. Advances in human-computer interaction: The intelligent room. In *Proc. CHI'95 Research Symposium*, Denver, Colorado USA, May 1995.

[Townsend and Salisbury, 1993] W. T. Townsend and J. K. Salisbury. Mechanical design for whole-arm manipulation. *Robots and Biological Systems: Toward a New Bionics?*, pages 153–164, 1993.

[Townsend, 1988] W. T. Townsend. The effect of transmission design on force-controlled manipulator performance. PhD thesis published as AI-TR-1054, MIT Artificial Intelligence Laboratory, April 1988.