”A FAST ACTING CAMERA USING LINEAR ACTUATORS AND VISUAL FEEDBACK”

Proc 1st International Conference on Visual Information Engineering

H Schmidt-Cornelius, D Young & H Buxton

Guildford, July 2003, 149-152

A FAST ACTING CAMERA USING LINEAR ACTUATORS AND VISUAL FEEDBACK

H Schmidt-Cornelius, D Young & H Buxton

University of Sussex, UK

ABSTRACT

We describe a biologically-inspired monocular camera platform for active vision systems. The design uses damped linear actuators to provide high angular velocities and angular accelerations at low cost and within a compact framework. The system lacks the mechanical precision of some high-performance heads, but we propose that use of visual feedback can compensate for loss of accuracy and repeatability. In support of this, we investigate control strategies based on visual feedback, and comment on their performance.

1. INTRODUCTION

For active vision, it is essential to have camera platforms that can emulate the fast saccades and precise tracking movements of the human eye. Very high angular accelerations and velocities are needed if image capture is to be effective. This is true even in everyday situations, such as orienting towards and then observing a person walking into a room where the camera is located.

This problem has often been addressed using highly-specified and precisely engineered mechanisms, which provide good accuracy and repeatability along with a rapid response. See, for example, Brooks et al. [1], Sharkey et al. [2]. Although such systems have remarkable performance, they tend to be expensive to build, as well as bulky and heavy – at least compared to animal eyes.

Recent work has concentrated on more lightweight platforms with improved dynamic properties. For example, Gosselin et al. [3] introduce parallel mechanisms (in which no actuator has to move another actuator) whilst Dankers and Zelinsky [4] address drive issues by replacing gearbox or harmonic drive technology with cable drives.

We have investigated an alternative approach, by designing and building a device that is inexpensive enough to be used in reasonable numbers, and light and small enough to be mounted easily on laboratory mobile robots. To achieve this we based the broad outline of our mechanical design on the human eye, using two pairs of linear actuators to aim the camera, rather than the more usual DC or stepper motors. A full account is given by Schmidt-Cornelius [5]. The use of linear actuators for this purpose is, as far as we know, novel.

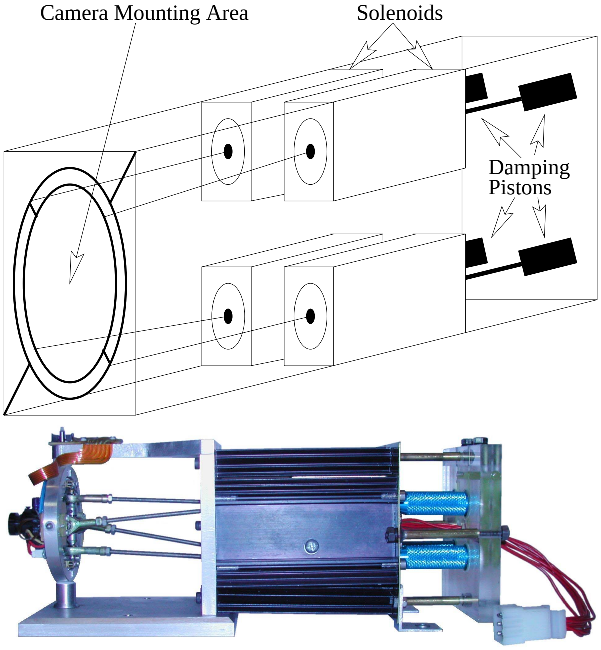

Figure 1. Schematic and photograph of the camera platform. From left to right in the photograph: gimbals with camera; actuating rods; linear actuators under heatsinks; damping pistons; pressure equalising channels in perspex block. The overall length of the assembly is 29 cm.

Such an approach naturally involves a tradeoff. It is clear that angular acceleration and velocity cannot be sacrificed without losing the value of active vision. Instead, the mechanical part of our system lacks both accuracy and repeatability. That is, the camera's axis cannot be moved to a precisely predictable direction relative to its base, even after calibration. We argue that this is relatively unimportant: visual feedback and a suitable control system can dynamically compensate for this mechanical imprecision.

It seems likely that such a strategy is also used by biological visual systems. In adopting it here, we are exploiting the rapidly decreasing cost of computational resources relative to mechanical resources to produce a cost-effective high-performance camera platform.

2. MECHANICAL DESIGN

Our Monocular Active Vision Eye (MAVE) consists of a small CCD camera on a gimballed mount positioned by damped linear actuators.

The overall layout is shown in Figure 1. The mechanical architecture is parallel: that is, each actuator operates between the camera mount and the base. (In a serial architecture, the component moved by one actuator forms the base from which the next actuator operates.) This has the advantage that the outer gimbal can be small and light, since it has only to constrain the camera motion, and not provide a mounting point for an actuator. The arrangement echoes the arrangement of the muscles of the human eye; the major difference is that control of torsional movement is via the gimbal bearings rather than through a third set of linear actuators.

The camera itself is statically balanced in its gimbals so that the actuators have only to act against its inertia and not support its weight. The outer gimbal rotates about a vertical axis to provide panning motion and the inner about a horizontal axis to provide tilting motion, as shown in Figure 2.

The actuators are inexpensive solenoids of a type often used for mechanisms such as car door locks, and can exert a peak force of about 20 N. They are mounted horizontally and couple to the gimbals through rods with low backlash bearings at each end. The lever arm of each actuator about the centre of rotation is 15 mm.

The actuators have no internal mechanical damping, which is provided instead by adjustable miniature shock absorbers as used in model cars. These are essentially cylinders containing silicone fluid which damps the motion of pistons connected to the solenoid cores. As supplied, these have a restoring spring and a membrane to contain the fluid. We removed the springs and membranes, and instead coupled the cylinders in pairs through channels in the thick perspex backplane of the assembly. The pistons in a pair move in opposite directions when a gimbal is rotated, transferring fluid from one cylinder to the other. The damping force is thus approximately linear in velocity and independent of position.

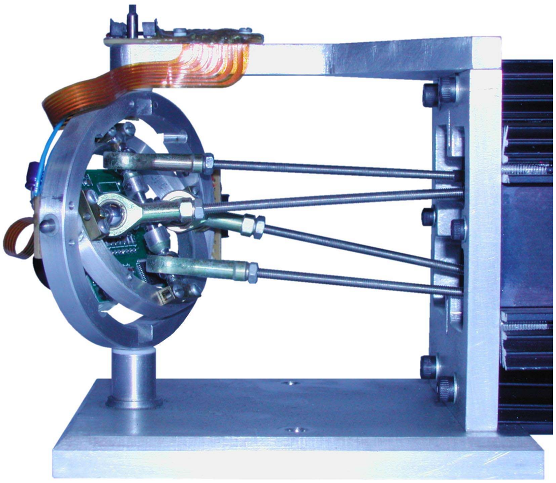

Figure 2. The gimbal assembly. The camera is mounted centrally in the gimbals and is pointing to the left, down and away from the viewer. The connecting rods and end bearings are visible. The actuators are to the right.

A more compact design could be produced, at the expense of some complexity, by mounting the dampers in the spaces between the four actuators, and coupling them independently to the gimbals. Other possibilities for damping, including magnetic and electronic damping, could be considered. We chose a mechanically straightforward and simple solution in order to facilitate investigation of the properties of the design.

Overall, these design choices gave us a device that includes a number of mechanical imperfections, notably static friction, a nonlinear response to excitation, and lack of independence between the actuators. These factors are addressed by flexible control mechanisms. On the other hand, as will be seen, the mechanical system is capable of generating very high rates of motion, an essential property since, if it is lacking, it is impossible to compensate.

3. ELECTRONIC CONTROL

The solenoids are driven from a 22 V power supply using pulse width modulation at a frequency of 400 Hz. Pulse widths range from 0.25 ms to 1.75 ms (10% - 70% duty cycle). In continuous use these values would somewhat exceed the duty ratings of the solenoids; however in practice the camera makes a series of saccadic or tracking movements which allow any given solenoid sufficient time for cooling. The 400 Hz vibration of the solenoid cores, transmitted to the damping pistons, helps overcome ”stiction” and so assists with the free movement of the camera.

Each solenoid power supply is controlled by a DC signal from a digital-to-analogue converter, which is computer controlled using a standard serial interface.

4. SOFTWARE CONTROL

Without feedback, the platform would be impossible to control, since positioning errors would rapidly accumulate. Although the gimbal axes are fitted with potentiometers for monitoring purposes, we concentrated on using visual feedback, since this allows tight coupling between the system’s goal – the accurate foveation of a visual target – and the control signals sent to its actuators.

We assumed that visual processing would provide an error signal, in image coordinates, that could be used to drive camera positioning. For most of our tests, we implemented this simply by requiring the camera to centre the image of a bright spot projected onto a dark background. For this task the image processing is trivial, and the error signal is simply the vector offset of the point of peak brightness relative to the centre of the image. Whilst most real applications require more sophisticated image analysis to identify the target feature, the result is likely to be presented in the same form to the control system.

We distinguished between two tasks for our controllers: saccadic and pursuit movements. In the former, the camera is required to make one or more fast ballistic movements to a target that may be initially a considerable distance from the image centre. In the latter, the camera is required to make a sequence of small adjustments (ideally continuous) to track a moving target whose image is normally close to the centre of the field of view. In each case, the controller must perform a mapping from the target error vector to the control signals for the actuators.

We experimented with a simple linear controller, an adaptive controller, and a controller based on a model of mammalian eye control.

We define the error vector, $\mathbf{e}$, to be the offset of the image of the target from the image centre. The activation vector $\mathbf{a}$ has one component for each pair of actuators: the magnitude gives the time for which the actuator is energised, whilst its sign specifies which one of the pair is to be powered. The gimbal axes and the axes of the camera's pixel array are approximately aligned when the camera is aimed centrally; we therefore order the components of $\mathbf{a}$ so each component controls the actuator pair that will mainly affect the corresponding component of $\mathbf{e}$.

4.1 Linear Controller

This uses pure positional feedback:

$$\mathbf{a} = s\mathbf{e} \tag{1}$$

where s is a scalar constant. The saccadic and pursuit controllers differed only in the value of s, which was roughly doubled for the pursuit controller.

The saccadic controller was activated twice for a given target position, with visual feedback between activations.

4.2 Adaptive Controllers

This is based on the linear controller, but the value of s is adapted using visual feedback. The updating rule, applied after each ballistic camera motion, is:

$$\begin{aligned} s' &= s + v|\mathbf{e}'| \quad \text{for target undershoot} \\ s' &= s - vp|\mathbf{e}'| \quad \text{for target overshoot} \end{aligned} \tag{2}$$

where $\mathbf{e}'$ is the error vector after the motion, v is a constant (typically much less than 1) determining the adaptation rate, and p is a constant (equal to 10 in our experiments) which penalises overshoot relative to undershoot in order to reduce the probability of oscillatory behaviour.

Again, the saccadic and pursuit controllers differed only in the value of a constant: v was set higher for the pursuit controller because the more uniform error magnitude allowed a higher rate of adaptation.

We also experimented with a combined controller in which the parameters are switched between saccadic and pursuit modes depending on the current magnitude of the error vector, and with more complex functions from the error vector to the activations, including polynomial functions of the error components.

4.3 Neural controller

Our final controller is based on the Dog Net controller of Burdess [6], which was inspired by a model of the superior colliculus by Ritter et al. [7]. A two-layer architecture is used to learn a mapping from target position in the image to activation times. Each layer consists of 600 units arranged in 20 concentric rings. The lower layer is known as the lattice layer, and defines the input space. The upper layer is known as the activation layer and defines the output space.

During operation, the lattice node whose weight vector $\mathbf{w}^L$ is closest to the input vector is found. This defines a centre of activation in both layers of the net. The activation vector that is output is an average of the weight vectors $\mathbf{w}^S$ of the activation nodes, weighted by a Gaussian function of their distances from the centre of activation.

During training, the lower layer forms a topology-conserving map of target locations, using Kohonen-type updating of the weight vectors $\mathbf{w}^L$. The output layer's weight vectors $\mathbf{w}^S$ also use Kohonen-type updating, but undergo reinforcement learning based on the reduction in error after each saccade. In both cases Gaussian neighbourhoods are used. The neighbourhood size and learning rates decrease according to a fixed schedule during training.

We also studied an adaptive version of the controller in which the neighbourhood sizes and learning rates depend on the current performance, in order to allow it to respond to long-term changes in the parameters of the system.

5. EVALUATION

5.1 Hardware performance

The system was reliable, running automatic tests for over 30 hours continuously without thermal damage.

We measured the angular velocity of the camera whilst driving it as quickly as possible between its physical limits. The camera reaches a peak angular velocity of about $930^\circ/\mathrm{s}$, which is comparable to published estimates of the maximum speed of the human eye. From the same observations we estimated the maximum angular acceleration to be about $16900^\circ/\mathrm{s}^2$. Both the angular velocity and angular acceleration are towards the top end of the range of published figures for existing active heads, for example [1], [2], [3] and [4].

The angular range of our system is relatively small, with about 60° of motion on each axis. This is limited by the gimbal mounts, whereas systems with rotary motors can be made to have much more free motion. We note that the human visual system operates with very fast eye movements of limited range supplemented by slower but larger head movements.

We have not yet obtained figures for the repeatability of positioning of our mechanism: this is complex since it depends on the time scale of the measurements. There is no limit in principle on the position resolution, and in practice this is more likely to be restricted by the accuracy of the visual feedback than by mechanical properties.

5.2 Controller performance

We tested the system as a whole using a precisely steerable laser to project a target onto a screen. To test the saccadic controllers we moved the target to a random position before activating the camera. The pursuit controllers were tested by moving the target along a random track generated by smoothing a random sequence of velocities.

The simple linear controller produced average errors on saccade of about 2.7°. The adaptive linear controller performed better, with typical average errors of 1.9°. However, the non-adaptive controller was affected badly by changes in the parameters of the hardware: a 20% change in the pulse width, for example, doubled the error. The adaptive controller coped well with such changes, showing only a transient increase in error. In smooth pursuit, the adaptive controller again fared better, producing average errors of about 0.8° which were stable against pulse width variation, whilst the non-adaptive controller tended to produce oscillation about the target position if the pulse width was increased.

The Dog Net was trained using a simulation of the system, and was found to converge to smooth mappings after several thousand iterations. The middle layer weight vector distribution reflected the geometry of the projection apparatus and camera. The trained network was then applied to the hardware, when it achieved average errors of about 1.9° on saccades.

We conducted an informal test of our system by generating a target position using image differencing. The camera was then able to make rapid saccades to foveate moving objects in its field of view.

6. CONCLUSIONS

The design we have presented could be made considerably lighter and more compact by developing the same design principles. Controllers that used velocity feedback and more sophisticated control strategies could improve the accuracy of saccades and pursuit.

However, we have demonstrated the feasibility of building a cheap, high-performance camera platform for active vision. Visual feedback control is necessary to compensate for mechanical imprecision, and adaptive controllers can provide greater robustness when the physical parameters of the system can change. We believe that platforms based on our design could be usefully exploited in fields such as surveillance and mobile robotics, where rapid response and compactness are at a premium.

REFERENCES

1. Brooks A, Dickins G, Zelinsky A, Kieffer J and Abdallah S, 1997. ”A high-performance camera platform for real-time active vision.” Proc 1st Int Conf on Field and Service Robotics, Canberra.

2. Sharkey PM, Murray DW, Vandevelde S, Reid ID and McLauchlan PF, 1993. ”A modular head-eye platform for real-time reactive vision.” Mechatronics, 3, 517-535.

3. Gosselin CM, St Pierre E and Gagne M, 1996. ”On the development of the agile eye.” IEEE Robotics & Automation Magazine, December 1996, 29-37.

4. Dankers A and Zelinsky A, 2003. ”A real-world vision system: mechanism, control, and vision processing.” Proc. 3rd Int Conf on Computer Vision Systems, Graz 223-235.

5. Schmidt-Cornelius H, 2002. ”Reverse Engineering an Active Eye.” DPhil dissertation, University of Sussex. Internet source: http://www.schmidt-cornelius.com/.

6. Burdess C, 2002. ”Saccadic eye movements.” Internet source: http://www.dog.net.uk/saccades/.

7. Ritter H, Martinez T and Schulten K, 1992. ”Neural Computation and Self-Organizing Maps: An Introduction.” Addison-Wesley, pp 141-161.