FEASIBILITY OF USING EYE TRACKING TO INCREASE RESOLUTION FOR VISUAL TELEPRESENCE

Paul M. Sharkey¹  David W Murray²

¹Robotics and Virtual Environments Research Group
Department of Cybernetics, The University of Reading
Whiteknights, Reading RG6 6AY, UK

²Active Vision Research Group
Department of engineering Science, The University of Oxford
Parks Road, Oxford OX1 3PJ, UK

P.M.Sharkey@reading.ac.uk, dwm@robots.oxford.ac.uk
ABSTRACT

Visual telepresence seeks to extend existing teleoperative capability by supplying the operator with a 3 dimensional interactive view of the remote environment. This is achieved through the use of a stereo camera platform which, through appropriate 3D display devices, provides a distinct image to each eye of the operator, and which is slaved directly from the operator’s head and eye movements. However, the resolution within current head mounted displays remains poor, thereby reducing the operator’s visual acuity. This paper reports on the feasibility of incorporation of eye tracking to increase resolution and investigates the stability and control issues for such a system. Continuous domain and discrete simulations are presented which indicates that eye tracking provides a stable feedback loop for tracking applications, though some empirical testing (currently being initiated) of such a system will be required to overcome indicated stability problems associated with micro saccades of the human operator.

1. INTRODUCTION

Robotic sensing, perception and action are currently not of sufficient combined maturity to enable machines to perform delicate, non-routine and risky operations requiring a high degree of eye-hand, or other sensori-motor co-ordination. However, there is an industrial imperative for such operations to be carried out in hostile/remote environments. The goal of teleoperation is to fill this intelligence gap by inserting a human master operator in the sensory loop, stimulating the operator’s senses with sufficient fidelity to inspire an appropriate illusion of direct interaction with the remote environment. The sensory information transferred might include resolved force, touch, kinesthesis, proprioception, vision, and sound. Providing the operator with a clear, unobstructed view of the slave-environment interface from a natural viewpoint can greatly enhance operator performance [8]. Placing cameras within the remote environment can assist the teleoperative task but, to date, this method has been mainly limited to fixed cameras or simple pan/tilt devices at fixed locations with the workspace. Some work has been conducted on coupling head mounted displays (HMDs) to remote cameras, an early example of the is given in [5], though little research has been conducted into incorporating eye tracking to improve the very poor resolution such off-the-shelf systems offer.

It is readily apparent in visual telepresence systems incorporating stereo camera pairs and HMDs (or other stereo displays) that to fuse stereo images at different focal distances either the displayed images must be offset accordingly or the cameras must be verged. Recent trials at Robotix ’97 involving some 400 delegates confirmed this requirement. While the former approach is easier to implement, the latter will provide greater flexibility in seeking improved resolution.

A more immediate motivation for this work was the development at the Robotics Research Group (Oxford) of advanced electro-mechanical hardware and visuo-control software for studies in active vision. Such systems comprise an integrated system architecture which couples a high performance electromechanical camera platform (a number of which were designed [11], [13]) and local servo controller to a vision processing engine and high level controller to allow switching between different behaviours [12]. Here the machine vision system has control over where it points its cameras and how it uses the resulting imagery. We wish to explore how the high performance camera platforms (or robot “heads”) can be utilised as slave devices for visual telepresence, and whether and how active computer vision could be usefully exploited to assist in teleoperative tasks. We are particularly interested in investigating how such systems may be enhanced by slaving the robot cameras directly from the operator’s eye movements.

2. HARDWARE CONFIGURATION

We have developed existing work on novel camera platform designs, implementations of visuo servo-controllers and real-time vision systems. The Yorick 5–5C head platform (Fig. 1), was designed to be compact, light and portable, with a camera separation of 110mm, similar to human eye separation. Each axis is driven by a DC motor with a geared transmission using very low backlash Harmonic Drive gearbox. The low inertia on each axis allows raw accelerations of between ~20,000°s⁻² to be attained on the pan axis, ~25,000°s⁻² on the elevation axis, and 30,000°s⁻² on the two vergence axes, and maximum speeds of 600°s⁻¹ on each. For compactness, counter tensioned steel belt transmissions are used for the pan and elevation axes, and this arrangement also allows the elevation axis and the left and right vergence axis to pass through the optic centre of the respective camera.

1 Robotix 97 Exhibition, Glasgow, Scotland, 13–16 March 1997.
Figure 1. Yorick 5–5C.
The vision processing itself is performed on a network of C40 DSPs, two of which have framegrabbers. The C40s are also hosted by a PC, making accesses to and from the servo controller appear to be ones to and from the PC. Vision processes are used to implement a head tracker and autovergence of the stereo platform cameras.

3. IMAGE PRESENTATION
An alternative to autovergence is to allow individual camera movements and hence vergence to be slaved directly from the operator’s eye movements, in addition to the head movements, see Fig. 2. We have found that the use of limbus tracking using IR illumination was the most realistic: one company supplies a lightweight head-mounted system with a resolution of 2 arc minutes, a frequency response of 0–100Hz and a linearity of 3%. The implementation of eye tracking highlights the difficulty in controlling the system, maintaining stability at all times in the face of numerous delays, to ultimately ensure that the operator remains “immersed” in the remote environment.

4. EYE MOVEMENTS
There are four main types of eye movements, three conjugate, and one disjunctive. Two classifications are used by physiologists, one based on patterns of response or nystagmus (pursuit, optokinetic and vestibulo-ocular), each of which contain slow and fast phases [2].

Conjugate movements
During pursuit, the head is typically stationary while the eyes attempt to maintain fixation on a moving target. The slow-phase movements range from 1°s⁻¹ to 30°s⁻¹ and stabilise the image of the moving target or background on the retina. If the slow-phase fails, quick-phase saccadic eye movements reposition the eye.

Saccadic eye movements are also generated voluntarily during visual search. They are of the range of 1° to 40°, though head motion is often involved when the required magnitude exceeds 30°. Saccades are characterised by very high initial acceleration and final deceleration – figures appear as high as 40,000°s⁻² – and high peak velocities, up to ~500°s⁻¹. Their duration ranges from 30ms to 120ms. In response to a visual stimulus, saccadic eye movements exhibit a latency of 100ms to 300ms.

The opto-kinetic nystagmus occurs when looking at constant motion of large visual fields. Repeated periods occur comprising slow phase movements which stabilise the motion, followed by a quick phase fly-back.

The vestibulo-ocular nystagmus occurs when the head is rotate as a unit, but the eyes stabilise on a target. These are driven by signals from the semi-circular canals which have lower latency than vision, as evidenced by the higher frequencies that VOR can stabilise compared to pursuit.

Vergence Movements
Vergence belongs to eye disjunctive movement in which two eyes are move in opposite direction. This normally happens when we change the fixation from a far to a near target. It is characterised by the slower and smoother than conjugate eye movements. Its maximum velocities are on the order of 10°s⁻¹ over a range of about 15°.

Minor Movements
There are three types of eye movements in this category, all are involuntary eye motions which commonly occur during eye fixations. Flicks are very rapid saccade-like motions of less than 1°. Drifts are random, small and slow -0.1°s⁻¹ motions of the eye. Tremor is a very low amplitude (<30 arc seconds) high frequency (30–150Hz) eye vibration.
5. CAMERA CONTROL FROM THE OPERATOR’S EYE MOVEMENT

We have reviewed eye tracking methods and commercial available eye trackers [1], [3], [4], [9], [10], [15] and wish to explore the issue of the control of the robot camera axes directly from the operator’s eye movement. The vergence geometry for primate vision is given in Fig. 3.

Figure 3. Verging geometry.

The basic feedback loop for eye tracking is shown in Fig. 2, a more detailed block diagram is given in Fig. 4. The target is imaged, and these are displayed to the operator. The operator’s eye movements are measured by the eye tracker and its output used to control the vergence axes of the robot head.

Effectively, the link from head to display in Fig. 2 comprises two parallel channels, one carrying visual data sensed by the camera, the other carry angular data sensed by the head encoders. There are two possible methods of achieving movement of the display: physically, moving either the display itself or some intermediate optics; or electronically, by establishing an active area of display within a larger physical display and moving it by the DC offset on the horizontal deflectors.

While we regard the latter as the more promising, we have not discovered a display, either in the laboratory or commercially available, that has this capability. The construction and complex optics involved in building an
array of smaller LCDs makes this approach unfeasible. We are currently implementing the former in a controlled study to investigate both human interfacing and stability issues of micro saccades [14].

6. SIMULATION RESULTS

The block diagram implementations for both frequency and discrete time simulations are given in Figs 6 and 7 respectively. The equations used to develop these simulations have been omitted in favour of presenting the simulation results.

Continuous Domain Results

The simulation was carried out both in time domain and frequency domain within Matlab and Simulink. In Figs 8–10, part (a) are the system response to step input signal and part (b) are system’s Bode plots. They show the system performance as values of the camera/image delay, Δci, of 0ms, 100ms, and 200ms respectively are used (the head/display delay is fixed at Δhd = 100ms).

The first and the last figures show that with large values of Δci − Δhd of −100ms and +100ms respectively, the system tends to be underdamped and overdamped respectively. (This is as expected. From Fig. 4, we expected instability if Δci << Δhd.) When the two delays are identical, the system has good response to the step input. Even with small variation of the difference (within ± 20ms) the system can still keep its performance with nearly critical damping.

Discrete Time Domain Results

We choose to show two results which together raise questions about the feasibility of the eye tracking approach. First in Fig. 11(b) we show the output θd(t) as a result of the target making a 20° step input. The key difference from the earlier results is that there are saw-tooth oscillations of the overall display position θd(t). These are the result of sampling and holding by the imaging process during periods of fast camera motion. If we use just the slow phase tracking response, these oscillations of frequency 25Hz are low-pass filtered, and their overall effect appears minor.

However, when the saccading mechanism is installed, a saccade is triggered in response to the oscillations. Without making detailed experiments with a real display we cannot say precisely what would occur, but the observations raise the possibility that undesirable effects might arise when eye-tracking. Certainly the observation makes us more certain that method of vergence control is fraught with difficulty.

7. CONCLUSION

The review of the eye tracking methods shows that many of the methods in laboratory use such as Double Purkinje Imaging, magnetic induction, and electrodes are not applicable to telepresence. IR illuminated limbus tracking provides a good choice. But its performance cannot be guaranteed in the practice. We concluded that image based eye tracking is more practical with newly developed high rate and mini sized cameras.

Our discussion of the requirements for eye tracking highlighted that the display mechanism is inherently complicated, requiring either physical movement of the display or optics associated with the display, or electronic movement of an active area within a larger display area. The latter appears technically feasible, although no HMD devices have display areas of sufficient size to permit the large movements required.

Given error-free transfer of angular position in the forward and backward channels, our simulations indicate the system is stable during slow-phase pursuit, at least for delays of up to 100ms. Mismatches in the delay of within ± 20ms in the parts of the link between head/eye platform and the display/imaging mechanism appear to have only slight effect on the stability. The model that includes sample and hold in the vision process and introduces the possibility of the controller switching into a fast-phase saccade, exhibits less re-assuring behaviour. The oscillations in the display can trigger a second saccade. Without making repeated experiments with a real display and group of subjects, it is of course not possible to determine what discomfort this might cause an operator. The simulations merely raise the possibility that undesirable effects might arise when eyetracking.

The forward control of the vergence axes of the robot head requires the system accuracy at min arc level, and a similar quality in the display is necessary. Whereas head tracking alone allows the eyes the freedom to correct for error, head and eye tracking does not. Again without performing experiments on a real system is it futile to speculate how the operator would react to inaccuracies.

Acknowledgement. This research is funded by the UK’s Engineering and Physical Sciences Research Council, grant reference GR/J44049.

8. REFERENCES


Figure 6. Block Diagram of Simulink implementation.

Figure 7. Step response and frequency response with $\Delta_r = 0$ ms and $\Delta_d = 100$ ms.
Figure 8. Step response and frequency response with $\Delta_c = 100$ ms and $\Delta_{hd} = 100$ ms.

Figure 9. Step response and frequency response with $\Delta_c = 200$ ms and $\Delta_{hd} = 100$ ms.

Figure 10. Block diagram of Discrete simulation.
Figure 11. (a) The step move made by the target; (b) the effective angle $\theta_d$ of the target in the display with respect to the eye’s resting direction; (c) the eye output using only the slow phase controller; and (d) the resulting robot head motion.

Figure 12. (a) The step move made by the target; (b) the effective angle $\theta_d$ of the target in the display with respect to the eye’s resting direction; (c) the eye output using only a controller capable of making saccades; and (d) the resulting robot head motion.