Natural Interaction Enhanced Remote Camera Control for Teleoperation

Dingyun Zhu
CSIRO ICT Centre,
Canberra, ACT 0200, Australia

School of Computer Science,
Australian National University.
Canberra, ACT 0200, Australia
dingyun.zhu@csiro.au

Tom Gedeon
School of Computer Science,
Australian National University.
Canberra, ACT 0200, Australia
tom.gedeon@anu.edu.au

Ken Taylor
CSIRO ICT Centre,
Canberra, ACT 0200, Australia
ken.taylor@csiro.au

CHI 2010, April 10 – 15, 2010, Atlanta, Georgia, USA
ACM 978-1-60558-930-5/10/04.

Abstract
In teleoperation, operators usually have to control multiple devices simultaneously, which requires frequent hand switches between different controllers. We designed and implemented two prototypes, one by applying head motion and the other by integrating eye gaze as intrinsic elements of teleoperation for remote camera control in a multi-control setting. We report a user study of a modeled multi-control experiment that compares the performance of head tracking control, eye tracking control and traditional joystick control. The results provide clear evidence that eye tracking control significantly outperforms joystick and head tracking control in both objective measures and subjective measures.

Keywords
Natural Interaction, head tracking, eye tracking, remote camera control, teleoperation, rock breaking.

ACM Classification Keywords
H.5.2. Information interfaces and presentation (e.g., HCI): User Interfaces – Evaluation/methodology; Input devices and strategies

General Terms
Design, Experimentation, Human Factors

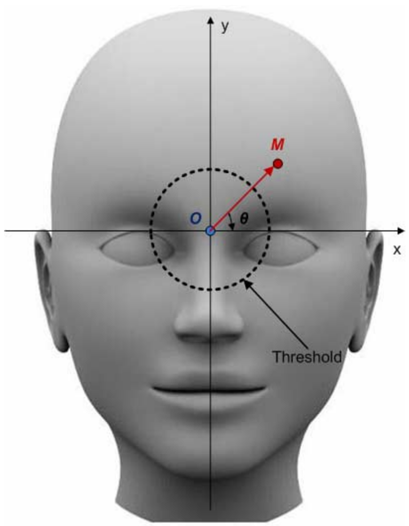

figure 2. Head tracking for remote camera control

Introduction

Teleoperation allows safe and convenient operations at a distance for hazardous or difficult working environments. Regardless of whether the remote machine is manually manipulated by a human or fully autonomous in executing its specific mission, an operator's observation, supervision and judgment still remain critical elements of the entire teleoperation activity [3].

In general, conventional control interfaces such as joysticks, switches, wheels, mouse and keyboard are predominantly used in most teleoperation activities. This leads to a typical multi-control problem as operators usually have multiple devices to control simultaneously, often more than they have hands. Figure 1 shows a real-world rock breaking instance in mining teleoperation. In this setting [1], the operator usually starts with the exploration of rocks in the bin by controlling a remote camera using a joystick interface. After the breaking spot on a target rock is identified through the camera view, the operator will switch to another two-handed joystick controller to position the tip onto the rock then fire the jack hammer to break it. This two-step process turns out to not only reduce the productivity of the task, but also increases the workload and potentially increases the number of avoidable operational mistakes. This research aims to discover whether head tracking or eye tracking enhanced camera control approach could be effective for the multi-control situation in teleoperation.

Our Design for Remote Camera Control

For both head tracking and eye tracking design, we apply the *rate control mapping* with a linear function gain to specify the moving angle and the velocity for the remote camera to carry out corresponding pan and tilt functions.

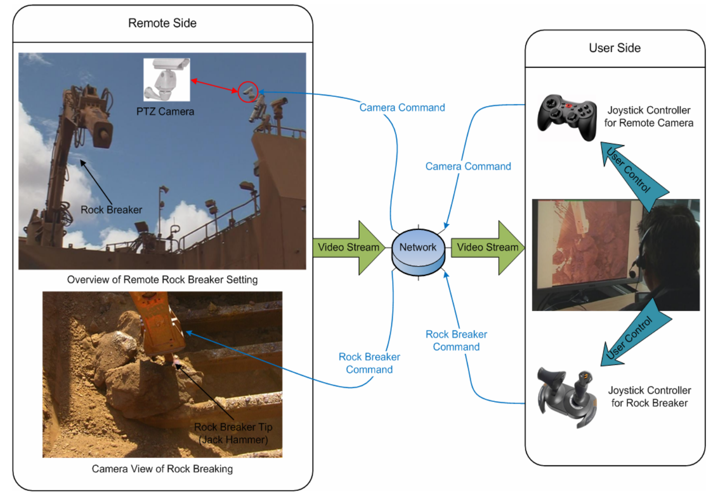

figure 1. Multi-control situation in mining teleoperation

Head Tracking for Remote Camera Control

Instead of creating a set of particular head gestures as discrete commands for camera control, our design only considers the distance between the neutral head position (the "*O*" point in figure 2) and the moved position (the "*M*" point in figure 2), the head motion can be yawing, pitching, rolling or translation: when a user moves their head more than a certain distance (threshold) either horizontally or vertically, or even diagonally, the camera will pan or tilt, or move along the same direction as the user's head move direction by conducting pan and tilt simultaneously (along the "*θ*" in Figure 2). We consider this to be appropriate for continuous control of pan and tilt functions as it allows users to have flexible head motion for the control of

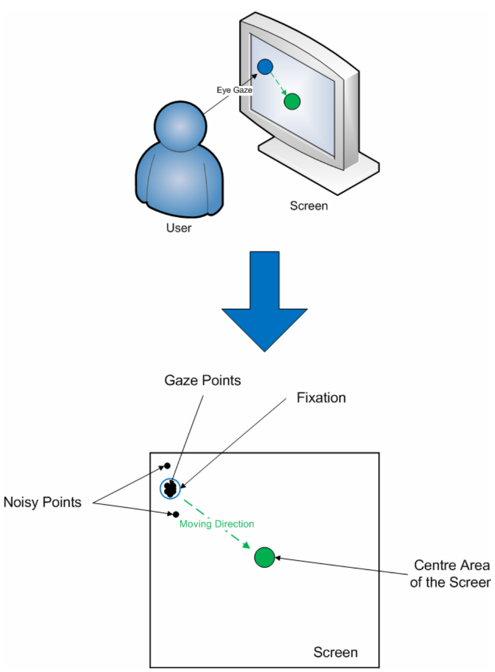

figure 3. Eye tracking for remote camera control

camera movements. The camera will keep moving the view along the direction until the user moves their head back to the neutral position (to a position smaller than the threshold). This reduces the degree of head movement required even for large amounts of camera movement.

Eye Tracking for Remote Camera Control
The eye tracking control is based on the gaze coordinates on the screen. We can break down the entire design into the following three steps (see figure 3):

1. Processing the raw gaze data to filter noisy points, recognize gaze fixation by calculating the centroid of grouped non-noisy points.
2. Calculating the distance and the angle between the current fixation position and the centre of the screen.
3. If the fixation is in the central area (distance < the radius of the central area), the camera will remain at the current position. On the other hand, if the fixation is out of the central area, the camera will start moving with the angle calculated between the current fixation and the centre of the screen.

Since human raw gaze points are quite noisy, and are not suitable for direct application [4], we used a modified version of the *Velocity-Threshold Identification* fixation detection algorithm for filtering the raw gaze points into fixations, as this method is straightforward to implement, runs very efficiently, and can easily run in real time [7].

User Study
We integrated FaceAPI 3.0 [8] with a standard Logitech webcam into our system as the head tracking control, the other eye tracking control used the FaceLAB 4.5 [9] eye tracker. Both run at about 60Hz on a standard PC using Windows XP. The system used the same model Pelco ES30C PTZ camera [5] as used in the rock breaking teleoperation on the mine site to perform the remote camera control for our study. A joystick control was also implemented by using a standard Logitech wireless gamepad as an example of the conventional control interface.

Considering the fact that we had very limited access to the real rocker equipment and the operators, we used the *functional physical modeling* [2] by which we could model the original setting by using another physical model with similar properties. We modeled the rock breaking task by using a physical game analogue: playing a re-designed table soccer game with two handles (see figure 4) and recruiting university students as experimental subjects. This design we consider to have the advantage of being more compelling and interesting for our student subjects than a more abstract industrial-like control task. In this experimental setting, participants would have to use both hands to control the handles to kick the ball and simultaneously use head motion or eye gaze or switch hands to another joystick to perform camera control, which is very similar to the multi-control situation in the real rock breaking setting.

A total of 30 undergraduate students participated in this evaluation, including 23 male and 7 female, ranging from 18 to 32 years of age with a mean of 20.3 years old and *SD* = 3.3. All of them were regular computer users with video game experience of using a joystick based interface, but none of them had previous experience with either head tracking or eye tracking

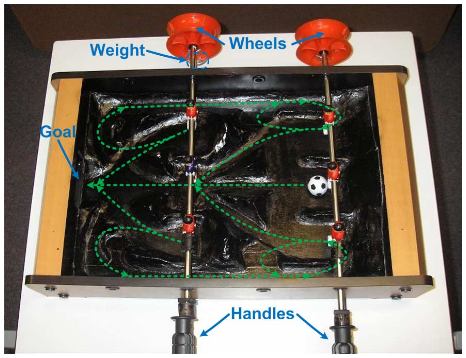

figure 4. Re-designed table soccer game as a functional physical model for the remote rock breaking task

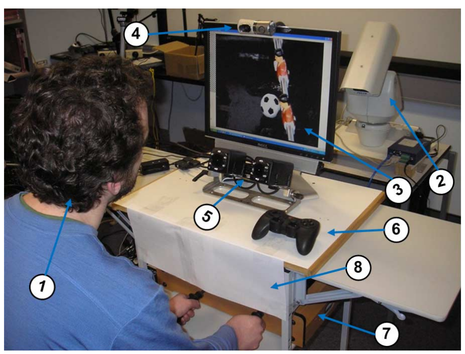

figure 5. (1) participant, (2) remote controlled camera, (3) video stream from the remote camera (zoomed view), (4) standard webcam for head tracking, (5) FaceLAB eye tracker, (6) gamepad as joystick based control interface, (7) re-designed soccer table, (8) covers for obscuring participant's direct vision.

control interfaces, and none of them had experience with our re-designed table soccer game.

The display was a standard 19" monitor with a resolution of 1280 × 1024 pixels and shows the real-time video stream to the participant from the remote camera located on the other side. The re-designed soccer table was placed under the monitor with several covers attached on the near side to obscure the participant's direct vision to the game. Figure 5 shows the experimental setting for our user study.

The experiment followed a within-subject design. Participants took part in the study individually and no pre-training period was offered before the formal test. For each control method, participants had 5 minutes to play the table soccer game. An extra 3 to 5 minutes were required for calibrating each participant's eye gaze before they started the eye tracking control trial. The order effect was counterbalanced by using a Latin square.

The video streams from the remote camera for each participant using different camera controls were recorded respectively. In addition, their entire experimental period was also recorded by another video camera for further observations. For the objective measures, the number of goals and kicks each participant achieved were recorded through checking against the video records. After the experiment, we collected participant's subjective feedback by using a questionnaire with a 5-point Likert scale, rating from 1 (strongly disagree) to 5 (strongly agree) and a short interview, in which they compared their experiences with different controls across several criteria as subjective measures, including naturalness, required consciousness, distraction and time to get used to each control.

Results

Objective Measures: Goals and Kicks

A One-Way Repeated Measures ANOVA showed highly significant differences in scored goals between the three control methods, $F(2, 58) = 9.29$, $p << 0.001$. Figure 6 shows the corresponding overall mean goals for each method. The post hoc tests revealed that participants using eye tracking control scored significantly more goals than using head tracking control ($p < 0.015$) and joystick based control ($p < 0.01$). Although the overall mean goals using head tracking ($M_{head} = 4.6$, $SD_{head} = 2.11$) is higher than the number using the joystick ($M_{joystick} = 3.87$, $SD_{joystick} = 1.78$), there was no significant difference between these two methods in mean scored goals ($p > 0.05$).

The results of One-Way Repeated Measures ANOVA of mean kicks showed very similar results to the previous mean goals analysis (see Figure 10). Highly significant differences were found in number of mean kicks between the three camera control methods, $F(2, 58) = 15.85$, $p << 0.001$. Also, the post hoc tests revealed that eye tracking control significantly outperformed both head tracking control ($p < 0.01$) and joystick based control ($p < 0.01$) in making kicks, but there was no significant difference in performance ($p > 0.05$) between using head tracking ($M_{head} = 24.23$, $SD_{head} = 4.17$) and joystick ($M_{joystick} = 23.7$, $SD_{joystick} = 4.24$). Thus, kicks correlate strongly with goals ($R(88) = 0.55$, $p < 0.01$), which relates to the rock-breaker setting in firings of the jackhammer leading to successful breaking of the rock.

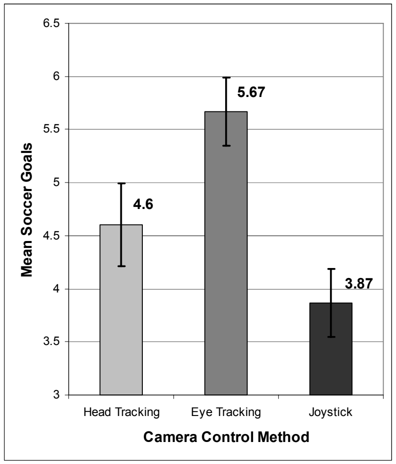

figure 6. Mean goals for each camera control method

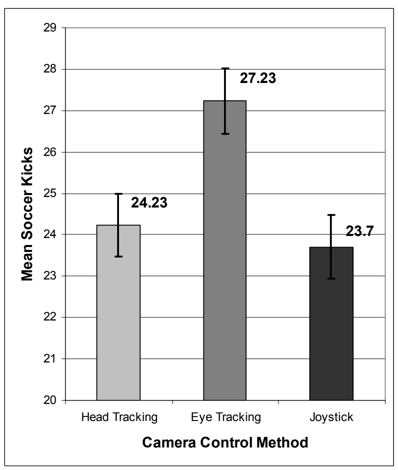

figure 7. Mean kicks for each camera control method

Subjective Measures: Questionnaire and Interview
The results of questions from our questionnaire are depicted in Figure 11 which shows the mean results of all participants. In addition, participants clearly indicated in the questionnaire that it was very annoying for them to use the joystick in the multi-control experiment as they had to frequently switch hands ($M_{annoying} = 4.1$, $SD_{annoying} = 0.92$).

All questions show significant results favoring the eye tracking control. The pair-wise comparisons between head tracking control and joystick control did not show any significant differences in questionnaire results through all the questions. From the mean comparisons, it can be seen that the first two questions regarding naturalness and time to get used to the control method show that participants ranked joystick control slightly better than the head tracking control. In contrast, the remaining two questions about consciousness and distraction show slightly better results of using head tracking control than the joystick, again these are not significant.

At the end of the questionnaire, participants were asked to state an overall ranking of the three camera control methods according to their experience gained in the multi-control experiment. The mean position for eye tracking control was 1.43, 2.17 for head tracking control, and 2.4 for joystick control. 20 out of 30 participants (66.7%) ranked eye tracking as the best method for the remote camera control in the experiment and 7 (23.3%) ranked it as their second preference. Head tracking was rated slightly better than the joystick with 6 (20.0%) participants ranking it as their first choice and 13 (43.3%) as the second choice. More than half of the participants (16/30 = 53.3%) ranked joystick control as the worst.

Discussion
Compared to head tracking and joystick control, more valid kicks and more goals were produced by using eye tracking control. It followed the general expectation that more kicks would lead to more goals. It tells us that when using eye tracking control, participants could obtain significantly more opportunities to kick the ball and score more goals as the camera motion always followed their current visual attention, which provided an effective solution for acquiring situational awareness from the remote environment through the video stream. The subjective results and feedback also remain consistent with the objective results. Eye tracking control outperformed head tracking and joystick through all the criteria we selected.

However, there were also some observations and existing issues about both prototypes. A few participants with much video game experience preferred to use head tracking control, as it provided a more interactive and immersive way for remote camera control with the experience of being more engaged in the table soccer game. In addition, some participants commented that head tracking control might not be as practical and applicable as eye tracking in a real teleoperation setting, because users would more easily be fatigued and annoyed if they had to perform camera control by continuously moving their head for a long time, and our experimental period was fairly short (5 minutes), so that this effect was not really obvious. On the other hand, participants felt eye tracking control was much better than head tracking control regarding to fatigue, as the previous study had also revealed that

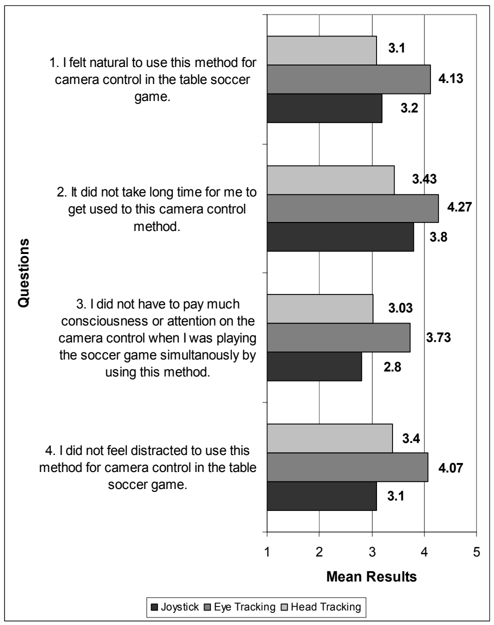

figure 8. Mean results regarding questionnaire feedbacks (scale: 1 – strongly disagree, 2 – disagree, 3 – neither agree not disagree, 4 – agree, 5 – strongly agree)

eye tracking interaction could be nearly fatigue-free [6]. Furthermore, the major issue about the eye tracking control is that the tracking process was still very sensitive and unreliable. Several participants commented that they were not able to have a completely free interaction, because when they occasionally moved their head direction along with the gaze unconsciously, the eye tracking quality was reduced or sometimes even lost if they moved their head a bit further away, which significantly affected the camera control quality. All these discovered issues will require further improvement and investigation.

Conclusion

We present two novel natural interaction enhanced remote camera control methods by using head tracking and eye tracking respectively as potential approaches for solving the common multi-control problem in teleoperation. We modeled the remote rock breaking teleoperation by using a functional physical model. Both head tracking and eye tracking controls were compared with a joystick control in a user study, including both objective measures and subjective measures.

From the results, we demonstrate that our eye tracking control is able to provide a natural and efficient control for the movement of a remote camera, and participants performed significantly better than using either head tracking control or joystick control without pre-training. We believe our eye tracking approach could be a potential replacement for conventional interfaces for remote camera control currently used in the real-world teleoperation, as it is capable of being an effective solution for the multi-control problem. Hopefully the observations described will also encourage more research on importing more natural interaction strategies into the control interface design for teleoperation and may lead to the deployment of such systems in a practical context.

Acknowledgements

The authors would like to express their appreciations to all the experimental participants, and also thank the Immersive Environments team in CSIRO ICT Centre. This research was supported by the”Future Mining” theme under CSIRO National Mineral Down Under (MDU) Research Flagship.

References

[1] Duff, E., Caris, C., Bonchis, A., Taylor, K., Gunn, C. and Adcock, M. The development of a telerobotic rock break. *In Proc. FSR 2009*, 1-10.

[2] Gedeon, T. D. and Zhu, D. Developing a natural interface for a complex task using a physical model. *In Proc. IHCI 2010*.

[3] Hughes, S. and Lewis, M. Robotic camera control for remote exploration. *In Proc. CHI 2004*, ACM Press (2004), 511-517.

[4] Jacob, R. J. K. The use of eye movements in human-computer interaction techniques: what you look at is what you get. *ACM Transactions on Information Systems*, 9, 3 (1991), 152-169.

[5] Pelco: ES30C (2009) http://www.pelco.com/products/

[6] Saito, S. Does fatigue exist in quantitative measurement of eye movements? *Ergonomics*, 35, 5/6 (1992), 607-615.

[7] Salvucci, D. D. and Goldberg, J. H. Identifying fixations and saccades in eye-tracking protocols. *In Proc. ETRA 2000*, ACM Press (2000), 71-78.

[8] Seeingmachines: FaceAPI (2009) http://www.seeingmachines.com/faceapi/

[9] Seeingmachines: FaceLAB (2009) http://www.seeingmachines.com/product/facelab/