Head or Gaze? Controlling Remote Camera for Hands-Busy Tasks in Teleoperation: A Comparison

Dingyun Zhu[1, 2] **Tom Gedeon**[2] **Ken Taylor**[1]

[1]CSIRO ICT Centre, Acton, Canberra, ACT 0200, Australia

[2]School of Computer Science, College of Engineering and Computer Science, The Australian National University, Canberra, ACT 0200, Australia

dingyun.zhu@csiro.au tom.gedeon@anu.edu.au ken.taylor@csiro.au

ABSTRACT

Head motion and eye gaze are general models of natural human interaction. Recent computer vision based head tracking and eye tracking technologies have expanded the possibilities of designing and developing more natural and intuitive user interfaces for a wide range of applications. In this work, we focus on common hands-busy situations in teleoperation activities, where operators often have to control multiple devices simultaneously by hand in order to accomplish operational tasks. This overloads an operator's hand control ability and also reduces productivity. We present an empirical user study comparing head motion and eye gaze as different input modalities for remote camera control when a user is carrying out a hands-busy task. Both objective measures and subjective measures were used for the study. According to the results, we demonstrate the advantages of using gaze for remote camera control in such hands-busy settings.

Author Keywords

Head motion, eye gaze, natural interaction, remote camera control, hands-busy tasks, teleoperation, user study.

ACM Classification Keywords

H5.2. Information interfaces and presentation (e.g., HCI): User Interfaces – evaluation/methodology, input devices and strategies, interaction styles.

INTRODUCTION

Hands-busy situations are common in teleoperation. For many teleoperation activities, operators usually have multiple devices or equipments to control via conventional hand operated interfaces, such as joysticks, switches, wheels, mice and keyboards. A typical example of such situations is controlling a robot for either navigation or manipulation as the primary task and controlling a remote camera as the major visual feedback operation to obtain situational awareness from the remote environment all by manual interfaces (Fong and Thorpe, 2001). Apart from switching hands frequently between different control interfaces, the working attention of the operator will also be distracted from the primary task under such circumstances, which results in workload increase and performance reduction.

There has been much research work on developing user interfaces and applications based on head motion and eye gaze as they are part of natural human interaction. By using such natural interaction models, users can interact with technologies, applications and systems as they are used to in everyday life (Valli, 2008). More importantly, they are able to provide effective "hands-free" control (Matsumoto et al., 2001; Tall et al., 2009) or interactions (Kang, 1998; Zhai et al., 1999). Zhu et al. (2009) introduced the use of head movements as an alternative input for the remote camera control in order to solve the hands-busy situation in teleoperation settings. From the results of a user study, they show that head tracking control is able to offer a comparable performance to traditional keyboard control.

In this work, we focus on a user study comparing head motion with eye gaze as different input modalities to control a remote camera for acquiring visual feedback from the remote site when simultaneously using both hands for another control task.

USER STUDY

Camera Control Models

In order to conduct the user study, we designed and developed two remote camera control models, one using head motion and the other using eye gaze. The head motion control only considers the distance between a user's neutral head position and the moved head position: when a user either moves (translation) or rotates their head more than a certain distance (threshold) at any direction, the camera will move along the same direction as the head. It will keep moving until the user moves their head back.

The overview of the gaze control follows a simple design concept: whatever the user looks at on the screen, it moves to the centre. The actual camera control is based on the distance between the user's current gaze position and the centre of the screen. If the gaze position is not in the central area, the camera will start moving along that direction.

OZCHI 2010, November 22-26, 2010, Brisbane, Australia.

OZCHI 2010 Proceedings ISBN: 978-1-4503-0502-0

1 http://www.seeingmachines.com/product/faceapi/
2 http://www.logitech.com/en-us/435/238/
3 http://www.seeingmachines.com/product/facelab/
4 http://www.pelco.com/products/

System Description and Apparatus

We integrated FaceAPI[1] (Ver 3.0) with a Logitech QuickCam Ultra Vision[2] webcam into our system as the head motion control. The gaze control used FaceLAB[3] (Ver 4.5) eye tracker with its Clink Tool SDK for the development. Both versions were implemented in C++, running on a standard Windows XP PC. A Pelco ES30C[4] camera was integrated to perform the remote camera control. Figure 1 shows the developed system and the experimental setting for our user study.

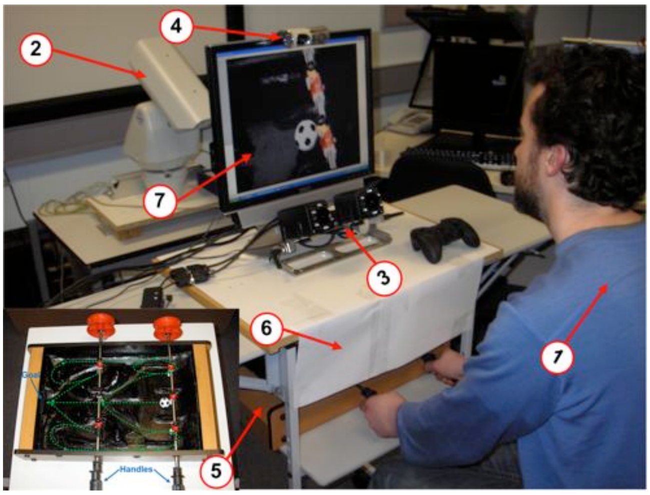

Figure 1. Experimental setting: (1) participant, (2) remote camera – Pelco ES30C, (3) FaceLAB eye tracker, webcam for head tracking - Logitech QuickCam Ultra Vision (4), redesigned foosball table (5), (6) covers for obscuring participant's vision, (7) video stream from the remote camera.

As to the hand control task, we used a physical game analogue: a redesigned foosball game with two handles. The original foosball table needs two players to control a pair of handles on each side. Therefore, a number of modifications have been made in order to change it into a single-player task (see (5) in Figure 1).

A standard 19' monitor was used as the major user screen with a resolution of 1280 × 1024 pixels, showing the real-time video stream to the participant from the remote camera. The redesigned foosball table was placed under the monitor with several covers attached on the near side to obscure the participant's direct vision to the game area. Therefore, participants had to continuously move the remote camera using either head motion control or gaze control during the experiment in order to find where the ball was and play the game using two handles simultaneously.

Participants

16 undergraduate volunteers (mostly first-year students) from a local university successfully participated in this user study, including 13 male and 3 female, ranging from 18 to 24 years of age ($M_{age} = 19.3$, $SD_{age} = 1.7$).

All participants were regular computer users (at least 2 hours a day). A few participants had previous experience playing foosball game but none of them had experience on our redesigned game as well as using an eye tracker. Several participants wore glasses and the rest had normal vision without any correction.

Experimental Design and Procedure

The experiment used a within-subject design so all the participants participated in both conditions, i.e. head motion control and gaze control. The independent variable of the study was the *camera control model* and the order effect was eliminated by switching the order of the camera control models.

Participants took part in the study individually. Prior to the experiment, a short oral introduction was given (about 5 minutes), including the system description, instructions on the head motion control and the gaze control, and how to play the redesigned foosball game. The objective of the task was to use both hands controlling the handles to score as many goals as they could during each camera control trial. All the participants were required to confirm an understanding of the introductions and the task requirements.

Participants directly started the experiment after the introduction, no pre-training was offered. An extra 3-5 minutes were spent on each participant to calibrate their eye gaze for the tracker before they started the gaze control trial. For each camera control trial, participants had 5 minutes to play the game. The video streams from the remote camera for each participant using each camera control model were recorded respectively. Additionally, the entire experimental period for each participant was recorded by another video camera for post-experiment observation.

The number of goals and kicks each participant achieved was recorded by checking against the video clips as the objective measures. After the experiment, a questionnaire with a 5-point Likert scale (rating from 1-strongly disagree to 5-strongly agree) and several open ended questions was used to collect participants' subjective feedback, including naturalness, required consciousness, distraction and time to get used to each control model.

RESULTS AND DISCUSSIONS

From a paired T-Test (2-tailed), we found a significant difference on mean scored goals ($T_{15} = -3.52$, $p < 0.01$). It showed that using gaze camera control ($M_{gaze\_goal} = 5.0$, $SD_{gaze\_goal} = 1.2$) participants scored significantly more goals than head motion control ($M_{head\_goal} = 3.68$, $SD_{head\_goal} = 1.6$). Figure 2 shows the comparison of mean goals.

The Paired T-Test (2-tailed) for mean kicks indicated similar results to the mean goals comparison. Participants made significantly ($T_{15} = -4.05$, $p < 0.01$) more kicks in the gaze control ($M_{gaze\_kick} = 27.2$, $SD_{gaze\_kick} = 4.8$) trail than using head motion control ($M_{head\_kick} = 23.5$, $SD_{head\_goal} = 4.4$). The mean kicks comparison is illustrated in Figure 3.

From the results of both goal and kick analysis, we can clearly see that the gaze control performed quantitatively better than the head motion control for a majority of the participants in the hands-busy task. More kicks and more goals were achieved. This exactly followed our expectation of this hands-busy task that more kicks would result in more goals. It also tells us that using gaze control, participants significantly obtained more opportunities to kick the ball and score more goals than head motion control in the hands-busy game.

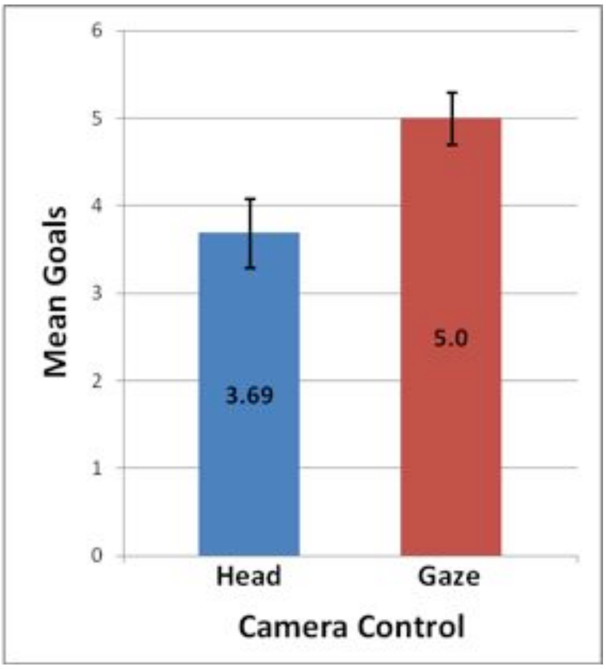

Figure 2. Mean goals for each camera control

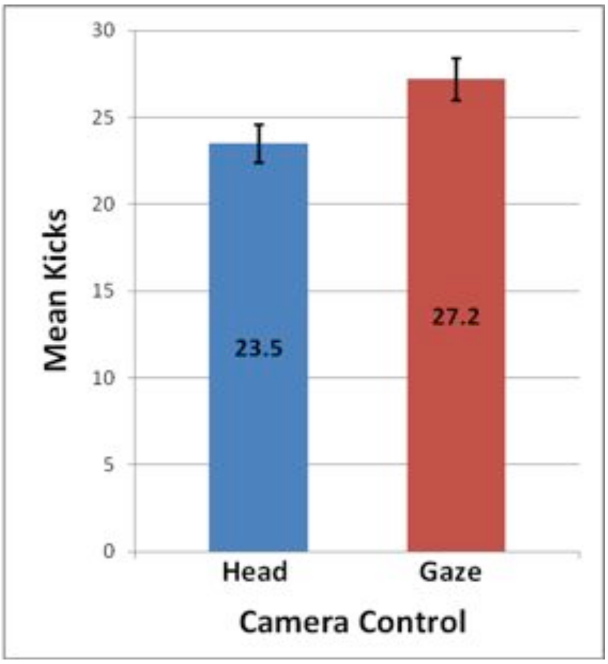

Figure 3. Mean kicks for each camera control

The mean results of the subjective measures are depicted in Figure 4. Participants rated the gaze control much better than the head motion control across all the criteria in the questionnaire, including how natural they felt, how much time they spent to get used to the control, how much attention they had to pay and how distracting when working on the hands-busy task.

Besides these questions, participants were required to give an overall preference ranking of these two camera control models in terms of their user experience gained from the experiment at the end of the questionnaire. A majority of the participants (13 / 16 = 81.2%) ranked the gaze control better and the rest (3 / 16 = 18.8%) showed their preference to the head motion control.

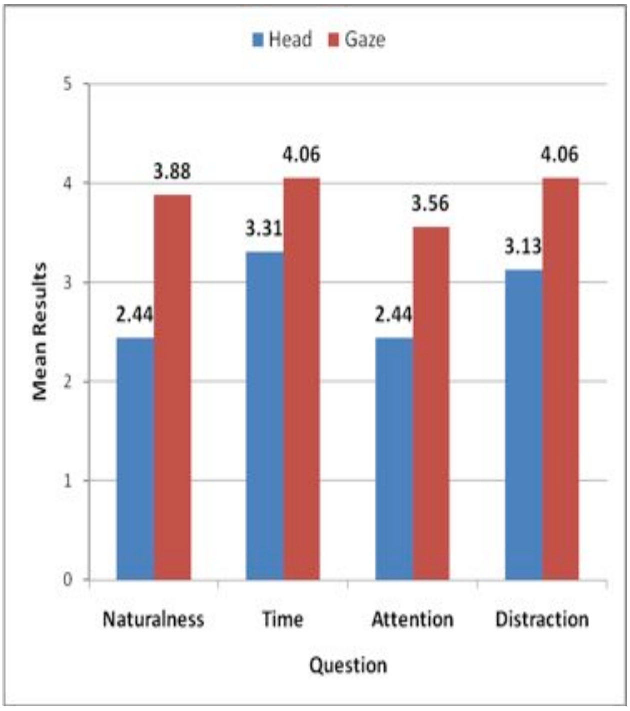

Figure 4. Summary of questionnaire results (5-point Likert scale: 1-strongly disagree, 2-disagree, 3-neither agree nor disagree, 4-agree, 5-strongly agree)

From the mean results comparisons of the subjective measures, we can see that they remain consistent with the results of the objective measures. Especially for the first two questions regarding naturalness and time to get used to the control model, most participants clearly felt that they were able to pick up the gaze control very quickly. The head motion control was reasonably fine, but it still required a bit of time to get familiar with the control mapping as a few participants stated in their questionnaires.

Furthermore, several participants commented that using gaze they did not have to pay much attention or consciousness on the camera control when performing the hand task at the same time. This means participants did not have to "deliberately think" much about the camera control in the experiment.

Since the camera motion always followed their current visual attention in the video stream, there was no particular control mapping for the participant to adapt to, which allowed them to more focus on the task doing with hands. However, they felt the head motion control was not as comfortable as the gaze control because it obviously requested much more physical movement. Also, it was slightly difficult for them to have an appropriate view of the screen content if they occasionally moved or rotated their head a bit far away from the screen.

	Head	Gaze
Advantages	1. immersive experience 2. engagement, excitement, entertainment (gaming & VR environments) 3. tracking hardware relatively cheap (can be done just by a standard webcam)	1. easy to pick up 2. not much attention or consciousness 3. less physical movement
Disadvantages	1. limitation on head motion 2. more physical movement 3. not practical for long time continuous control	1. tracking hardware still expensive 2. tracking quality could be improved

Table 1. Summary of head motion control and gaze control

Those participants who preferred the head motion control all played video games quite often. They explained that the head motion control offered more exiting and engaging experience for this game-like task regardless of performance measures. They commented that head motion might be more suitable for gaming or 3D virtual environment interaction, as users would be more engaged and entertained in these settings.

Moreover, a number of participants mentioned that although the head motion control could be a possible solution for the hands-busy situations, it might not be suitable for long time continuous control as users could be easily fatigued and annoyed. In contrast, the gaze control is more practical than the head motion control as it requires much less physical movement and attention, and it is almost fatigue free (Saito, 1992).

We further summarise the user study into advantages and disadvantages of the head motion control and the gaze control respectively and visualize them in Table 1.

CONCLUSIONS

In this paper, we present a user study comparing head motion with eye gaze for remote camera control when users are performing a hands-busy task simultaneously. From the results of both objective and subjective measures, we demonstrate the gaze control significantly outperformed the head motion control. However, participants with much game experience showed their preference using head control because of the engagement and enjoyment. Future directions can be exploring possible control modes combining both head and gaze interaction for teleoperation settings.

ACKNOWLEDGMENTS

The authors would like to express their appreciations to all the experimental participants, and also thank Chris Gunn, Matt Adcock, Leila Alem, Jock Cunningham and Eleonora Widzyk-Capehart for their great help and valuable suggestions.

This research was supported by the Transforming the Future Mine theme under CSIRO National Mineral Down Under (MDU) Research Flagship.

REFERENCES

Fong, T., and Thorpe, C. Vehicle teleoperation interfaces. *Autonomous Robots*, 11, 1 (2001), 9-18.

Kang, S. B. Hands-free navigation in VR environments by tracking the head. *Int. J. Human-Computer Studies*, (1998) 48, 247-266.

Matsumoto, Y., Ino, T. and Ogasawara, T. Development of intelligent wheelchair system with face and gaze based interface. *In Proc. ROMAN 2001*, IEEE Press (2001), 262-267.

Saito, S. Does fatigue exist in quantitative measurement of eye movements? *Ergonomics*, 35, 5/6 (1992), 71-78.

Tall, M., Alapetite, A., Agustin, J. S., Skovsgaard, H. H., Hansen, J. P., Hansen, D. W. and Mollenbach, E. Gaze-controlled driving, *In Proc. Ext Abstracts CHI 2009*, ACM Press (2009), 4387-4392.

Valli, A. The design of natural interaction. *Multimedia Tools and Applications*, 38 (2008), 295-305.

Zhai, S., Morimoto, C. and Ihde, S. Manual and gaze input cascaded (magic) pointing. *In Proc. CHI 1999*, ACM Press (1999), 246-253.

Zhu, D., Gedeon, T.D. and Taylor, K. Keyboard before head tracking depresses user success in remote camera control. *In Proc. INTERACT 2009*, Springer Press (2009), LNCS 5727, 319-331.