Extending Visual Perception Through Hands

Barış Serim
Helsinki Institute for Information
Technology HIIT, Uni. of Helsinki
Dept. of Design, Aalto Uni.

Khalil Klouche
Helsinki Institute for Information
Technology HIIT, Uni. of Helsinki
Dept. of Media, Aalto Uni.

Giulio Jacucci
Helsinki Institute for Information
Technology HIIT, Uni. of Helsinki
Helsinki Institute for Information
Technology HIIT, Aalto Uni.

firstname.lastname@hiit.fi

Abstract

The concurrent action and sensing potential of our bodies exceeds what we can visually attend at a given time. The action possibilities of our bodies also provide opportunities for extending our visual perception to places that are impractical for our direct vision, namely the vision that is facilitated by head and eye movements. In this workshop paper, we describe an interactive input-feedback loop that utilizes hand movements rather than head and eye movements to visually attend to objects. The basic mechanism we employ is projecting the content around the users' hand to the location of user's visual attention. We prototyped this mechanism with eye and skeletal hand tracking and adapted it to use scenarios that involve interacting with UI widgets, while looking at other regions on the interface.

Author Keywords

Visual attention; multifocus interaction; eye tracking; manual input; hand motion tracking

ACM Classification Keywords

H.5.m [Miscellaneous]

Introduction

Human visual perception is constrained by physical factors such as the physical location and orientation of the head and the eyes. Thus, realization of many activities depend

---

Proceedings of the CHI 2017 Workshop on Amplification and Augmentation of Human Perception, May 07, 2017, Denver, CO, USA.

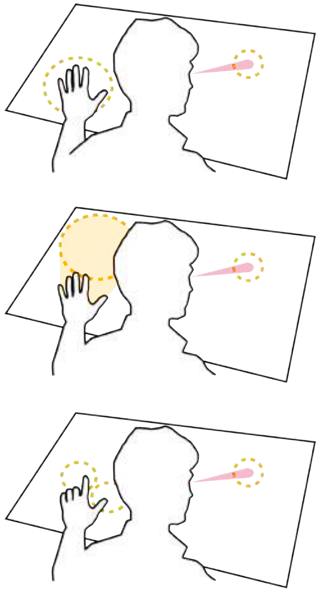

Figure 1: We project the content around the users' hand to where they are gazing at as sensed by eye tracking – in a way using hands to extend visual sense. However, the specific content can be determined by different qualities of the hand such as proximity to objects (top), palm orientation (middle) and hand posture (bottom).

on various artificial mechanisms that facilitate visual perception. Endoscopes used in surgeries, rear-view mirrors in a car or image-transmitting cameras extend our visual perception to places that are impractical for our direct vision.

Visual perception can fall short even for objects that constitute our immediate environment, in part, due to the limited nature of visual attention. At a given time, we can visually attend only to a portion of our surroundings. As visual acuity is highest in the center of the visual field, visual attention is constantly facilitated by head and eye movements that bring different regions of interest to the center of the visual field. On the other hand, the single-focus quality of visual attention stands in contrast to the concurrent sensing and action possibilities afforded by our body and limbs.

The concurrent action possibilities of human body has been the motivation for our work [6], through which we have been exploring the potential of using eye tracking and hand movements to partly overcome the limitation of visual attention. The basic mechanism we employ is projecting the content around the users' hand to where they are gazing at – in a way using hands to extend visual sense. We applied this interaction mechanism to scenarios of split visual attention.

Background

HCI researchers have long been aware that the concurrent sensing and action potential afforded by the human body surpasses what can be covered by human visual attention. Success of some input devices such as keyboard, or game controllers lies in their operability without or with little visual monitoring, enabling eyes to focus on the display. Thus, one design target in HCI has been eyes-free interaction, accomplished through physical input devices (e.g. [2]), haptic (e.g. [5]), proprioceptive [4] or audio feedback mechanisms.

On the other hand, for tasks that involve complex feedback, or where the visual exploration is the main activity, haptic and audio feedbacks might not provide sufficient bandwidth. A possible design approach in these cases is to utilize motor action, for example using hands, to facilitate visual perception more effectively, and overcome the bottleneck of limited visual attention. Various interaction techniques fall under this approach as they allow users to bring distant regions together through step-by-step interaction. This can involve opening contextual menus, declaring lenses or folding the information space (e.g. [3, 1]). While these techniques are very useful, they come at the cost of requiring an additional interaction step when compared to direct access afforded by spatial layout.

Interaction Mechanism

While earlier work relied on step-by-step interaction, tracking eye and hand movements allows us to *concurrently* extend visual perception. In the interaction mechanism we propose, *input location*, i.e. where the visual content is sourced from, is determined by hand movements, while the *feedback location*, i.e. where the content is projected to, is determined by user's visual attention. The mechanism enables quick visual overlay of different areas and decreases the need for redirecting the gaze over long distances.

The described mechanisms can potentially be realized with multiple hardware setups. For head-up displays or glasses, the feedback can be projected to a predefined location such as the center or periphery of the display. For larger touchscreens, the feedback location can be determined through eye tracking. Similarly, the input location can be determined through different technologies such as capacitive sensing or skeletal tracking. In our specific implementation (Figure 2), we combine skeletal tracking of hands using of-the-shelf equipment (Leap Motion) and tracking glasses (binocular

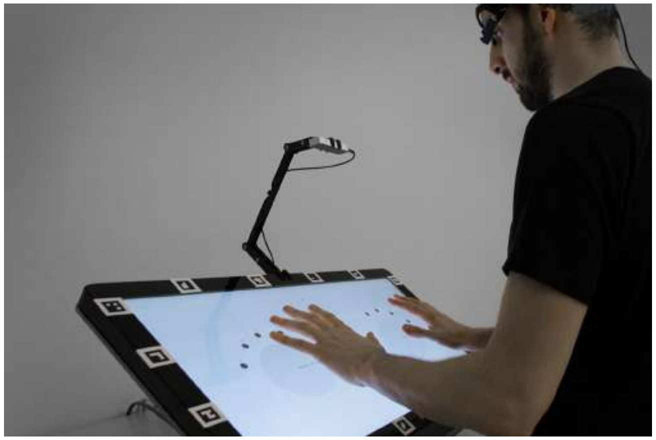

Figure 2: The hardware setup consisting of a touch screen, depth sensor (Leap Motion, attached to the upper screen edge) and eye tracking glasses.

SMI pupil) to sense the gaze location on a 27" touchscreen. The interactive mechanism also requires taking into account a number of design considerations, namely which input to show and how to provide visual feedback.

For input, the main considerations pertain to the specific qualities of the hand. *Position* of the hand in the space can be one way of determining content to be shown, by using proximity data to the content in the environment. Another way of determining the input content is through the *orientation* of the hand or the palm to the environment, similar to laser pointing. Additionally, non-positional information such as *hand posture* or *hand speed* can be used to distinguish between different content to be projected.

Where, when and how to show the feedback also requires taking into account a number of considerations. *Position* of the feedback can be determined by eye fixations; feedback can be shown near or adjacent to the focus of visual attention. However, eye fixations shift rapidly, and constant repositioning of the visual feedback can be intrusive. Thus, changes in position can be timed based hand movements or above-threshold gaze shifts. *Visibility* of the feedback is also an important dimension. Feedback that persists in the center of vision can be intrusive, requiring appropriate withdrawal of the visual feedback when it is not needed. The system can rely on hand movements to time the visibility (mapped to opacity, size or other visual variables) of the feedback.

Application case: UI widgets

Building on our earlier work [6], we applied the interaction mechanism described above to the problem of using UI widgets on a touchscreen while looking elsewhere. Touchscreens, unlike tangible controls, require visual attention for positional input. UI widgets provide a good starting point for exploration. First, when compared to images or text they have a smaller visual footprint making their overlay on the location of visual attention easier. Second, projection of their feedback can naturally precede widget manipulation through touch input. In these applications, we used skeletal tracking above a touch surface to project the visual feedback of the acquired widget near the eye-fixation location. The widget command is then confirmed by a touch action.

We prototyped two applications (Figure 3), first featuring manipulation of visual objects on a canvas and the second real-time video manipulation. Both applications enable content manipulation without having to redirect the visual attention to the toolbar. Instead visual feedback of the UI widget in the proximity of the hand can be projected near user's gaze, enabling continuous fixation on the media.

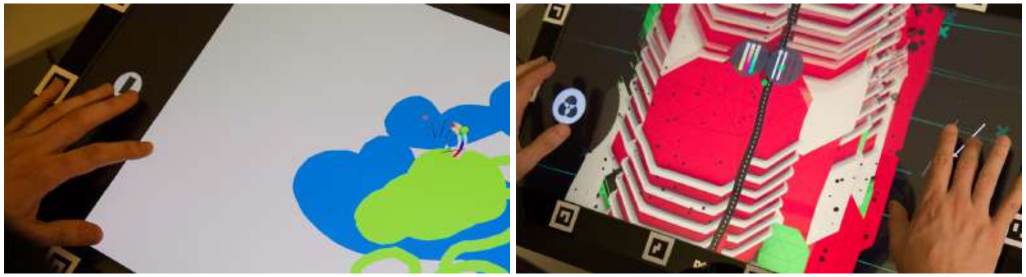

Figure 3: Left: The feedback of the hand over a color selector widget is overlaid to where the user is looking at the canvas. Right: two separate set of video filter widgets modifiers can be used concurrently when their visual feedback projected to the point of user's visual attention. In both images, visual feedback of the input widgets are overlaid on the media near the eye fixation location.

Conclusion

We presented an interaction mechanism that enables the juxtaposition of visual feedback from multiple locations, by tracking hand and eye movements. While some research in eye tracking or novel modalities target limiting motor action (for accessibility or performance gains), our research is motivated by the situations where the visual attention is the bottleneck and motor action can help overcome it. Situations where the cost of redirecting the gaze exceeds that of moving hands, such as in-vehicle interfaces or bimanual interaction make potential application cases. Other relevant situations are when direct vision is impractical due to occlusion, either by hand or other objects.

As is the case with a new sensing technique, there is a period of familiarization. Our observations from the earlier study [6] showed that the users can initially be hesitant to move their hands without visually attending to them. Another difficulty the users reported in a more recent study is the difference between the motor space and the scaled down visual feedback that is projected near user's eye fixation. Future work can work towards improving the interaction mechanism. However, we view part of the familiarization process necessary for increasing sensory skill.

Acknowledgments

This work has been partly supported by MindSee (FP7 ICT; Grant Agreement #611570).

References

[1] N. Elmqvist, Y. Riche, N. Henry-Riche, and J. D. Fekete. 2010. Mélange: Space Folding for Visual Exploration. *IEEE Transactions on Visualization and Computer Graphics* 16, 3 (May 2010), 468–483.

[2] George W. Fitzmaurice and William Buxton. 1997. An Empirical Evaluation of Graspable User Interfaces: Towards Specialized, Space-multiplexed Input. In *Proc. CHI'97*. ACM, New York, NY, USA, 43–50.

[3] Azam Khan, George Fitzmaurice, Don Almeida, Nicolas Burtnyk, and Gordon Kurtenbach. 2004. A Remote Control Interface for Large Displays. In *Proc. UIST'04*. ACM, New York, NY, USA, 127–136.

[4] Pedro Lopes, Alexandra Ion, Willi Mueller, Daniel Hoffmann, Patrik Jonell, and Patrick Baudisch. 2015. Proprioceptive Interaction. In *Proc. CHI'15*. ACM, New York, NY, USA, 939–948.

[5] Anne Roudaut, Andreas Rau, Christoph Sterz, Max Plauth, Pedro Lopes, and Patrick Baudisch. 2013. Gesture Output: Eyes-free Output Using a Force Feedback Touch Surface. In *Proc. CHI'13*. ACM, New York, NY, USA, 2547–2556.

[6] Baris Serim and Giulio Jacucci. 2016. Pointing While Looking Elsewhere: Designing for Varying Degrees of Visual Guidance During Manual Input. In *Proc. CHI'16*. ACM, New York, NY, USA, 5789–5800.