NeuroSight: Combining Eye-Tracking and Brain-Computer Interfaces for Context-Aware Hand-Free Camera Interaction

Benedict Leung
Ontario Tech Univeristy
Canada
benedict.leung1@ontariotechu.net

Mariana Shimabukuro
Ontario Tech University
Canada
mariana.shimabukuro@ontariotechu.ca

Christopher Collins
Ontario Tech University
Canada
christopher.collins@ontariotechu.ca

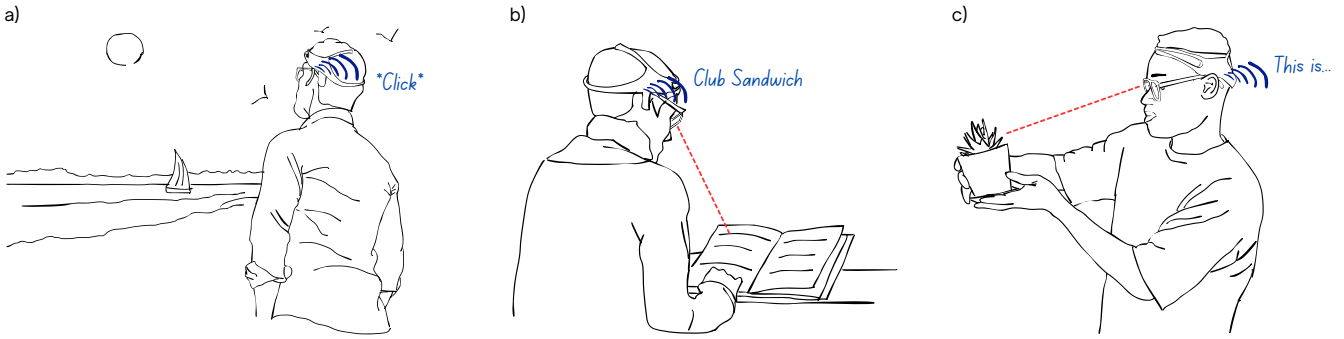

Figure 1: NeuroSight invokes a camera mode whenever a mental command is detected. Wireless earbuds are used to give audio feedback. Sample use cases of NeuroSight are a) takes a picture whenever a mental command is invoked with the brain-computer interface, b) toggles translation mode to translate a menu using the fixation point to focus on which text to translate, and c) invokes visual search where it describes the fixated object.

Abstract

Technology has blurred the boundaries of our work and private lives. Using touch-free technology can lessen the divide between technology and reality and bring us closer to the immersion we once had before. This work explores the combination of eye-tracking glasses and a brain-computer interface to enable hand-free interaction with the camera without holding or touching it. Different camera modes are difficult to implement without the use of eye-tracking. For example, visual search relies on an object, selecting a region in the scene by touching the touchscreen on your phone. Eye-tracking is used instead, and the fixation point is used to select the intended region. In addition, fixations can provide context for the mode the user wants to execute. For instance, fixations on foreign text could indicate translation mode. Ultimately, multiple touchless gestures create more fluent transitions between our life experiences and technology.

CCS Concepts

• **Human-centered computing** → *User centered design*; **Gestural input**; *Auditory feedback*.

UIST Adjunct '24, October 13–16, 2024, Pittsburgh, PA, USA

ACM ISBN 979-8-4007-0718-6/24/10
https://doi.org/10.1145/3672539.3686312

Keywords

brain-computer interface, eye-tracking, camera

ACM Reference Format:
Benedict Leung, Mariana Shimabukuro, and Christopher Collins. 2024. NeuroSight: Combining Eye-Tracking and Brain-Computer Interfaces for Context-Aware Hand-Free Camera Interaction. In *The 37th Annual ACM Symposium on User Interface Software and Technology (UIST Adjunct '24), October 13–16, 2024, Pittsburgh, PA, USA.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3672539.3686312

1 Introduction

Digital technologies such as smartphones and tablets have become essential in everyone's lives. Undoubtedly, they provide many benefits. Yet, mobile devices blur the work and home lives of adults [10, 14], and people depend on them to handle everyday tasks [7]. Mobile technology has impacted the transition between our life experiences [13], impacting our social life [1, 12]. For example, many people have been to concerts to see their favourite artists perform live on stage, a special in-person event. Unfortunately, many people would also like to record that experience on their phones, shattering the immersion of the present moment. Even a simple picture could ruin the moment, as reaching for your phone could break the immersion of the experience. Unlocking the phone and opening the camera app with notifications appearing can prolong the intended task. Smart glasses like Meta RayBan try to tackle this challenge but still require hands to operate the device. Thus, we present NeuroSight (Fig. 1), a hand-free, context-aware interface for controlling camera applications using mental commands. NeuroSight aims to make the interaction less obstructive and interruptive, creating a more immersive and fluent experience.

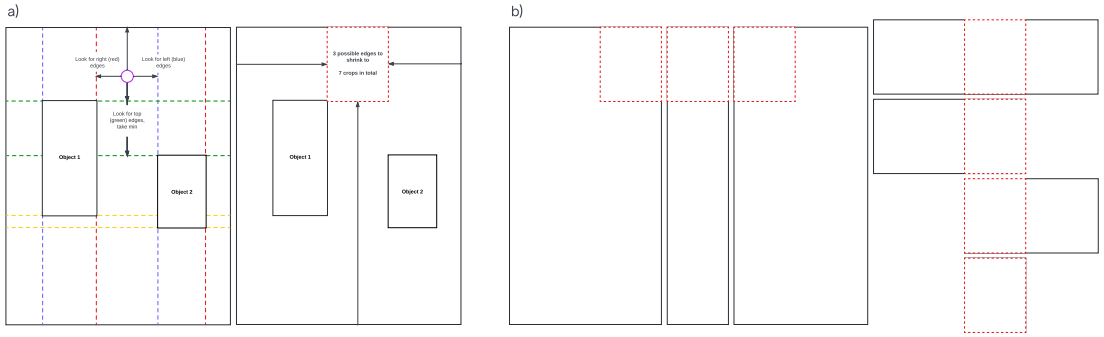

Figure 2: Steps for cropping algorithm where red, blue, green and yellow lines are right, left, top and bottom edges of the bounding boxes, respectively. The purple circle represents the fixation point. a) The crop algorithm uses a box to produce multiple crops to crop out some or all objects. b) The results of the cropping algorithm.

Touchless or hands-free interaction has been demonstrated to facilitate software interaction more conveniently and immersively, leaving the hands for more critical tasks [2, 11, 15]. Our solution eliminates the display as it can divert your attention from the current moment with the phone's display of unrelated notifications. As a result, we suggest mounting the camera in an accessible position while employing hand-free gestures to interact with the camera, along with giving feedback by audio.

2 NeuroSight

We introduce NeuroSight, a hand-free camera which uses auditory feedback for two hand-free interaction techniques: gaze and mental commands. These interaction techniques alleviate the hands and do not heavily rely on the ideal environment as voice input does. Gaze is crucial to indicate context and enable implicit task selection as it can be used to select a particular scene region, so it is possible to implement different camera modes, such as object detection, without ever needing to touch the screen. For instance, if the user has fixated on foreign text, the associated mode will be a translation (Fig. 1b). NeuroSight uses mental commands to trigger a mode. Mental commands depend on recognizing patterns in the user's brain activity and learning the difference between the user's neutral state and the desired command state [8] — the user must train and learn to trigger the desired command state. Finally, since we have eliminated the touchscreen display, NeuroSight uses audio feedback to give an interactive response back to the user. Cameras can also be used as utility tools to identify and analyze products, scan QR codes, panoramas, etc. Thus, NeuroSight offers three camera modes (Fig. 1): translation, visual search, and picture. ***Picture mode*** is the default option if translation or visual search cannot be applied according to the gaze-aware context.

***Hardware*:** Brain-computer interfaces acquire the wearer's brain signals and analyze them to execute the desired action. EMOTIV Insight was used as the brain-computer interface [3]. The EMOTIVBCI software was used to train the desired command state. To reduce false positives, the user must hold the command with power above 55 for two seconds. Pupil Core and Pupil Capture software measured gaze [9]. The glasses have two cameras: a world camera and an eye camera. The world camera acts as the camera the user will interact with. Wireless earbuds were used for audio feedback.

***Translation Mode*:** Translation may be needed multiple times in an environment with a foreign language. For NeuroSight to invoke the translation mode, foreign text must be at the fixation point. Once detected, the mode will continuously translate whenever the fixation point changes. Another mental command must be invoked to end it. Google Cloud Vision API [6], Google Cloud Translation API [5] and Google Cloud Text-To-Speech [4] were used to implement. First, it produces bounding boxes for paragraphs and words and language codes for each word. Fixated foreign text will then be translated into English and spoken back to the user.

***Visual Search*:** Visual search is designed to find information on the Internet using images. However, inputting images into the visual search engine could be difficult since noise could be present in the image, resulting in inaccurate results. For example, multiple objects may be present, or the image's background may be complex. Combining eye-tracking and object localization algorithms can crop the noise out of the image. Google Cloud Vision API [6] was used for the implementation. However, using only Google Vision is insufficient for detecting all objects in the scene. The solution presented is if no objects are detected inside the fixation point, crop out the detected objects from the scene to pass to Google Vision again. First, it finds the biggest box where if the edges protrude outwards one at a time, they will not intersect any bounding boxes and contain the fixation point (left of Fig. 2a). Finally, the image shrinks its edges to the box, one edge up to four edges at a time (right of Fig 2a). The cropping algorithm will recursively run until a bounding box is found or the image reaches minimum size. A visual search will be invoked for each bounding box found.

3 Conclusion & Future Work

We have presented NeuroSight, a hand-free camera that combines of brain-computer interface and eye-tracking glasses to create more immersive and fluent life experiences. Example use cases of NeuroSight can include translating a menu or describing the fixated object (Fig. 1). Future work should consider LLMs such as GPT-4V [16] as it complements our cropping algorithm to remove noise from an image and make queries with the image. Different hardware should also be considered to optimize comfort and setup, such as smart glasses. We hope this work can inspire new interaction techniques to be less obstructive to our life experiences.

References

[1] Elyssa M. Barrick, Alixandra Barasch, and Diana I. Tamir. 2022. The unexpected social consequences of diverting attention to our phones. *Journal of Experimental Social Psychology* 101 (2022), 104344. https://doi.org/10.1016/j.jesp.2022.104344

[2] René de la Barré, Paul Chojecki, Ulrich Leiner, Lothar Mühlbach, and Detlef Ruschin. 2009. Touchless Interaction-Novel Chances and Challenges. In *Human-Computer Interaction. Novel Interaction Methods and Techniques*, Julie A. Jacko (Ed.). Springer Berlin Heidelberg, Berlin, Heidelberg, 161–169.

[3] EMOTIV. 2024. Brain Data Measuring Hardware and software solutions. https://www.emotiv.com/

[4] Google. 2024. Cloud Text-to-Speech AI. https://cloud.google.com/text-to-speech

[5] Google. 2024. Cloud Translation AI. https://cloud.google.com/translate

[6] Google. 2024. Cloud Vision AI. https://cloud.google.com/vision

[7] Richard Harper, Tom Rodden, Yvonne Rogers, and Abigail Sellen. 2008. *Being Human: Human-Computer Interaction in the Year 2020*. https://www.microsoft.com/en-us/research/publication/being-human-human-computer-interaction-in-the-year-2020/

[8] Isuru Jayarathne, Michael Cohen, and Senaka Amarakeerthi. 2017. Survey of EEG-Based Biometric Authentication. In *2017 IEEE 8th International Conference on Awareness Science and Technology (iCAST)*. IEEE, Taichung, 324–329. https://doi.org/10.1109/ICAwST.2017.8256471

[9] Moritz Kassner, William Patera, and Andreas Bulling. 2014. Pupil: an open source platform for pervasive eye tracking and mobile gaze-based interaction. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication* (Seattle, Washington) *(UbiComp '14 Adjunct)*. Association for Computing Machinery, New York, NY, USA, 1151–1160. https://doi.org/10.1145/2638728.2641695

[10] Melissa Mazmanian, Wanda J. Orlikowski, and JoAnne Yates. 2013. The Autonomy Paradox: The Implications of Mobile Email Devices for Knowledge Professionals. *Organization Science* 24, 5 (2013), 1337–1357. https://doi.org/10.1287/orsc.1120.0806 arXiv:https://doi.org/10.1287/orsc.1120.0806

[11] Pedro Monteiro, Guilherme Gonçalves, Hugo Coelho, Miguel Melo, and Maximino Bessa. 2021. Hands-free interaction in immersive virtual reality: A systematic review. *IEEE Transactions on Visualization and Computer Graphics* 27, 5 (2021), 2702–2713. https://doi.org/10.1109/TVCG.2021.3067687

[12] Sumaiya Mushroor, Shammin Haque, and Riyadh A. Amir. 2019. The impact of smart phones and mobile devices on human health and life. *International Journal Of Community Medicine And Public Health* 7, 1 (Dec. 2019), 9–15. https://doi.org/10.18203/2394-6040.ijcmph20195825

[13] David Pauleen, John Campbell, Brian Harmer, and Ali Intezari. 2015. Making Sense of Mobile Technology: The Integration of Work and Private Life. *SAGE Open* 5 (04 2015). https://doi.org/10.1177/2158244015583859

[14] Lydia Plowman, Joanna McPake, and Christine Stephen. 2010. The Technologisation of Childhood? Young Children and Technology in the Home. *Children & Society* 24, 1 (2010), 63–74. https://doi.org/10.1111/j.1099-0860.2008.00180.x arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1099-0860.2008.00180.x

[15] Yixin Sun, Yudong Tao, Zhi Hu, Hao Fan, and Yuwei Wang. 2014. A hands-free communication solution for wearable devices. In *2014 IEEE Healthcare Innovation Conference (HIC)*. 75–78. https://doi.org/10.1109/HIC.2014.7038878

[16] Zhengyuan Yang, Linjie Li, Kevin Lin, Jianfeng Wang, Chung-Ching Lin, Zicheng Liu, and Lijuan Wang. 2023. The Dawn of LMMs: Preliminary Explorations with GPT-4V(ision). arXiv:2309.17421 [cs.CV]