Exploring Eye-Tracking-Based Detection of Visual Search for Elderly People

Michael Dietz, Daniel Schork and Elisabeth André
Human Centered Multimedia, Augsburg University, Germany
{dietz, schork, andre}@hcm-lab.de

***Abstract*—Visual search plays an important role in our daily lives and can be very frustrating whenever we cannot remember where we left objects, such as keys or wallets. This is especially true for elderly people, since they forget things more often and face this challenge very frequently. While there are several studies which analyze eye movements during visual search, none of them actually tries to detect whether a user is searching for something or not. However, this information is necessary to recognize when the user needs assistance. Therefore, we propose an eye-tracking-based multimodal approach in order to detect visual search and to support the user in that situation. Furthermore, we explore multiple strategies to inform the user of the desired object's location using a head mounted display. With the help of a prototypical implementation and evaluation of the acquired sensor data, we show that our method is feasible and capable of dealing with this challenge.**

I. Introduction

Elderly people often experience forgetfulness and even memory loss. While a declining working memory is part of the normal aging process [1], memory loss can be a sign of an illness, which is the reason why even small issues, such as a forgotten name, a forgotten way or a forgotten appointment, can lead to very unpleasant situations. Within the scope of the Glassistant project, we therefore try to recognize those situations and provide appropriate support. To this end, we create an unobtrusive virtual assistant based on smart glasses like Google Glass, which offers help as needed to ensure that elderly people can enjoy their daily lives independently. For that, the current stress level of the person is monitored with wearable sensors, which can detect when the user is confused or in a critical situation. If this is the case, then Glassistant intelligently analyzes the current context and displays additional information about the environment using augmented reality to support the person accordingly. For instance, if a user gets lost in an unfamiliar place, the system detects that and assists the person by showing navigation instructions to the desired location. Besides that, we also want to remind the user of upcoming appointments, display instructions for recipes and user manuals, provide information about surroundings and enable the option to call an emergency contact. One of the most important use cases though is the detection and support of visual search since elderly people often face this task in their daily lives. Due to the cognitive decline of their memory, they frequently forget where they left objects, such as keys or wallets, and need to find them again and again, which takes a lot of time and can be very frustrating. In this work, we therefore propose a concept, to detect when visual search is happening and to support the user by showing the location of the desired object on a head mounted display. We validate the first step of our approach with an evaluation of the sensor data from a prototypical implementation and show that the concept can be applied to deal with this challenge.

II. Background

Visual search is generally defined as the act of looking for a target object among several distractors. One of the first cognitive models to describe this process is the "Feature Integration Theory (FIT) of Attention" from Treisman and Gelade (1980), which is still the foundation for most of the current theories [2]. It proposes that the visual search task consists of two stages. In the first stage, simple features, such as color, shape, orientation and movement of an object, are perceived preattentively and unconsciously without any effort. This applies to all objects across the field of view simultaneously, which results in a very fast perception time of those attributes. However, due to its short and subliminal nature, it is very hard to detect this stage with external sensors and therefore we do not consider it in our approach. Instead, we focus on the second stage of the FIT where the previously identified features are combined in order to perceive more complex characteristics of an individual object. This requires explicit attention from the observer and can only be done in a sequential manner for each element of the visual scene, leading to a much slower processing time, but also enabling its detection [3].

As research has shown, focusing attention on an object is strongly related to certain eye movements [4], which is the reason why eye tracking is generally used as the main method to analyze the visual search process. While the field of eye tracking applications for elderly people is relatively small, a lot of research has focused on user interfaces without considering the age of the participants. Several studies in this area investigate visual search strategies by analyzing the eye movements in menus [5], [6], web pages [7], [8] and hierarchical text layouts [9]. While most of these studies only measure saccades, fixations and response times to analyze the effectiveness of the examined interface, none of them makes use of the eye tracking data to actually detect whether a user is searching. However, Credidio et al. [10] show that there are certain statistical patterns in visual search, which indicates that the detection might be possible. Bulling et al. [11] even found

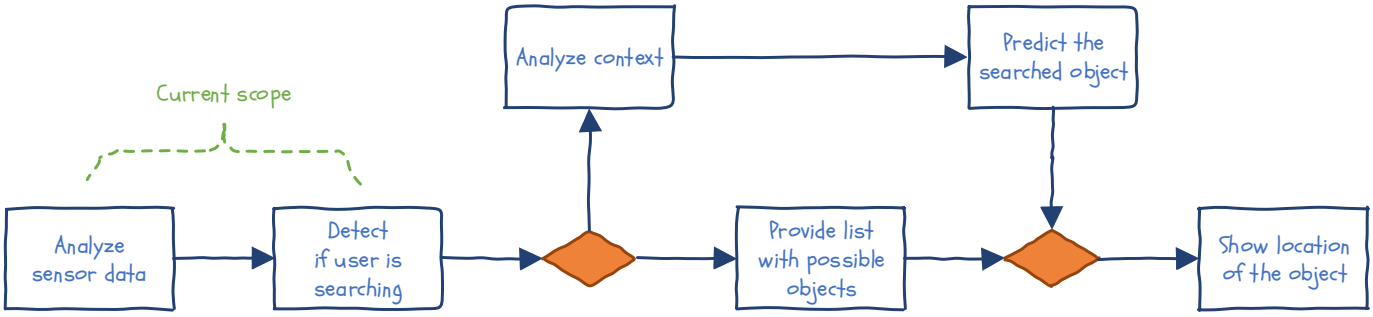

Fig. 1. Conceptual pipeline to support visual search

that eye movements can be used to recognize different office activities, supporting our assumption that this can be applied to visual search as well.

Additionally, several approaches also use pupil size [12], [13], head movements [14] or a combination of both [15] to identify the current task and the emotional state of the user. This information can be very helpful for our approach since the person might be stressed or frustrated while searching for an object and therefore we can use it as an additional indication of this process. In order to achieve the most promising results we consider all of those modalities in our approach to identify the visual search process by detecting the second stage of the FIT. Once we are able to do that, we also want to support the user with finding the target object. To this end, only a few concepts have been examined. One of them uses visual memory augmentation and tries to guide the user's attention towards unsearched places by darkening the areas he or she has already looked at [16]. This is especially helpful to prevent repeated searches at the same location and leads to a more efficient search strategy. In contrast, most other approaches try to find the objects by attaching a marker or tag to them which can be located by their system [17], [18].

III. Concept

Imagine you have an appointment soon and are about to leave the house. You look for the keys but they are not where you thought they should be. In a hurry, you start looking around, opening drawers and checking your pockets but the keys are not there. After a quick glance at your watch you get even more stressed and frustrated since you are already late. As time goes by you start to search in more unlikely locations until you finally find them, where you never assumed they could be in the first place. Many people can probably relate to this scenario and might experience similar situations from time to time. For elderly people this happens even more often and negatively affects their daily lives. Therefore, we propose the following concept to recognize this situation and to support the persons accordingly.

As shown in Figure 1, an important step of the proposed pipeline is to find out whether the user is searching for something or not. This is required to identify the point in time when the person needs assistance. Otherwise, the system could not act proactively and would require an explicit action or trigger from the user. However, this approach would lead to situations in which the system could be helpful but is not used because the person refuses to admit that he or she is in need of support. Since it is easier to accept help rather than to ask for it in the first place, we decided to always offer assistance whenever we recognize a critical situation with the option to decline if the user still does not wish any support.

In order to detect the visual search behavior we primarily analyze the eye movements of the user since they are the most promising sources of information as several studies regarding this topic have shown. To this end we use a small infrared camera, which is attached to the head mounted display and directed at the eye of the person. Combined with the scene camera of the smart glasses we are able to record the same data as with a regular head mounted eye tracker but in a less obtrusive way with the added benefit of being completely mobile. Additionally, we also consider other modalities, such as head movement, pupil size and blink frequency, which can be determined from the eye tracking sensors as well to ensure a high detection rate. Apart from the eye tracking data we also make use of the built-in sensors from the smart glasses, such as accelerometer, gyroscope, and magnetometer, to enhance the results. For the analysis of the sensor data, common machine learning techniques are applied. Basically, a binary classifier is used to determine whether the user is searching for something or not. If the results from that are positive, we analyze the current context of the user in order to predict the desired object. During this process information, such as time, weather, location, appointments and tasks, is collected from several data sources to create a probability ranking with each of the possible items.

For example, if there is an appointment in his or her calendar within the next 30 minutes it is more likely that the user is searching for objects, such as keys or a wallet, rather than a remote control for the TV. Besides that, the data from previous search tasks is compared to the current context to identify which object was searched in similar situations. If the analysis results in a high probability for a particular object then the user is asked whether he or she is looking for it and would like to have assistance with finding its location. Otherwise, a list with all previously stored objects is suggested on the head mounted display and the person is prompted if he or

Fig. 2. Comparison of sensor data during different activities

she is searching for one of them. Once an item is selected, its location is shown to the user. For that, a person needs to add items to the list of possible objects beforehand by looking at them from different angles during which the system is trained for their detection. This can either be done by a family member or the user depending his or her mental state. Upon completion, we constantly apply an object recognition algorithm to the video stream of the scene camera and try to identify the items from our list. Since this is a computationally intensive operation, we also investigate certain techniques to reduce the required application frequency. For example, the algorithm could only be used in situations where the person looks slightly downwards, which would be the case when he or she puts down an item. As soon as an object is detected, we save the video frame and assign it to the recognized item. This way we always have an image from last the point in time, when the object has appeared in the user's field of view. If the person then wants to know where an item is located, the corresponding image is shown on the head mounted display.

IV. Prototype

In order to validate that the part of our concept consisting of detecting visual search is feasible, we employed *Pupil Pro* by *Pupil Labs* [19] as a prototype. This head mounted eye tracker consists of a scene camera that captures the user's field of view and an infrared camera capturing the user's eye. Both cameras can supply up to 30 frames per second while being connected to a notebook powered by an Intel Core i7 processor. At this point, it was not necessary to use additional sensors, because the use of image processing algorithms was sufficient to supply the needed multimodal sensor data. In order to achieve real-time processing and recording, we developed a gaze tracking plugin for the Social Signal Interpretation Framework [20], which detects the pupil position on the image of the eye camera. After an initial calibration, the pupil position could be mapped to a gaze point in the field of view. The diameter of the detected pupil could also be used to calculate dilation. If a pupil could not be detected, the user's eye was considered closed. That way we were able to determine when the user was blinking. The scene camera's main purpose was to capture head movement. This was achieved by detecting features on the scene image. The features were then tracked by comparing two consecutive frames. Afterwards, outliers were removed to separate scene movement from moving objects within the scene. After taking the camera's field of view into account, the user's current head movement could be determined.

To evaluate the quality of the supplied sensor data as well as the potential to differentiate between recordings of different activities, we conducted a short evaluation. A 77 year old woman without visual or cognitive impairment wearing the prototype device was instructed to engage in four different activities. First, we hid an object and told the participant to search for it in her immediate surroundings. In order to compare this recording to other everyday activities the subject was also instructed to read text, to watch a video on a screen and to hold a conversation with another person. The evaluation revealed that the prototype supplied useful data in most areas. Gaze tracking yielded the best results. Calculation of pupil size turned out to be difficult, because the pupil appears smaller than it actually is when the user is not looking directly at the camera. A moving average filter was applied to compensate for this perspective transformation error. Detection of blinking also turned out to work in most instances. In some particular cases, the user was looking down so much that a pupil could not be detected. In these cases, blinking was registered even though the eye was not completely shut. An increased blink frequency could be observed during conversation and watching a video. Overall, the subject had a very high blink frequency, closing her eye 51 times per minute on average. This could be caused by dry eye syndrome, which has a higher prevalence in older age [21]. This result shows that age-related illnesses regarding the eyes are important to consider when developing

an eye-tracking-based solution for the elderly. Head movement could be captured on two axes using the aforementioned feature tracking method. While head nodding and shaking did not pose a problem, tilting as well as linear movement had to be ignored. Figure 2 shows that the recorded activities can be differentiated even while looking at the raw sensor data with the naked eye. Visual search, along with reading shows the highest amount of saccades per second. However, the saccade distance is much smaller when reading. Searching and conversing have a similar saccade distance, although the saccades occurred less frequent while the subject was in a conversation. While head movement during reading and watching a video was almost nonexistent, some movement could be observed during conversation. As expected, the user had to look around quite a lot during the visual search task. Pupil dilation remained almost constant when looking in a certain direction (reading, watching and conversing). A change in size was noticeable when turning towards a darker or brighter area and was mostly seen during the searching task. One could also argue that the dilation shows the user's distress as part of his or her affective processing.

The data shows that a differentiation between searching and any other task can be achieved by calculating eye-tracking features like path length of the gaze point, average saccade distance and fixation count, as well as blink count, average time of closed eye during blinking and change of pupil size. Features for other modalities could include average head movement on both axes and standard deviation of head movement. The features could then be used to train a binary classifier like a Bayesian network or a support vector machine, which can be applied to detect visual search in real time. Once the system identifies that the users are unable to find an object on their own, it can offer assistance by providing the location of the object as proposed in Chapter III. While we still use a notebook to process the data in our current prototype, we intend to utilize a much smaller and more portable device in future implementations. Though, as the study has shown, our applied algorithms are computationally intensive and might need to be adjusted in order to work on mobile devices.

V. Conclusion

In this work, we presented a concept to detect and support the visual search process for elderly people. Following that, we began to explore the feasibility of our approach by creating an initial prototype for the first step of the conceptual pipeline. Through the evaluation of the collected data, we showed that eye tracking can be used to detect when a user is searching for a particular object. In our future work, we will use those findings and conduct a study with several participants in order to identify the most suitable features to detect this process with machine learning techniques. Besides that, we intend to move on to more advanced hardware capable of making our sensor data even more reliable. For instance, the combination of smart glasses with an infrared eye camera could take advantage of the built-in sensors to replace our current camera based solutions in order to achieve more accurate results.

Acknowledgment

This work was partially funded by the German Federal Ministry of Education and Research (BMBF) under project Glassistant (16SV7267K).

References

[1] T. A. Salthouse and R. L. Babcock, "Decomposing adult age differences in working memory." *Developmental Psychology*, vol. 27, no. 5, p. 763, 1991.

[2] A. M. Treisman and G. Gelade, "A feature-integration theory of attention," *Cognitive Psychology*, vol. 12, no. 1, pp. 97 – 136, 1980.

[3] A. M. Treisman and S. Sato, "Conjunction search revisited," in *Journal of Experimental Psychology: Human Perception and Performance*, 1990, pp. 459–478.

[4] J. M. Findlay and I. D. Gilchrist, "Eye guidance and visual search," in *Eye Guidance in Reading and Scene Perception*. Elsevier, Oxford, 1998.

[5] M. D. Byrne, J. R. Anderson, S. Douglass, and M. Matessa, "Eye tracking the visual search of click-down menus," in *Proc. CHI*. New York, NY, USA: ACM, 1999, pp. 402–409.

[6] S. K. Card, "Visual search of computer command menus," in *International Symposium on Attention and Performance*, 1982, pp. 97–108.

[7] L. Granka, M. Feusner, and L. Lorigo, "Eye monitoring in online search," in *Passive Eye Monitoring*, ser. Signals and Commmunication Technologies. Springer, 2008, pp. 347–372.

[8] W. Lu, M. Li, S. Lu, Y. Song, J. Yin, and N. Zhong, "Visual search strategy and information processing mode: An eye-tracking study on web pages under information overload," in *Information and Automation*, 2011, vol. 86, pp. 153–159.

[9] A. Hornof, "Cognitive strategies for the visual search of hierarchical computer displays," *Human-Computer Interaction*, vol. 19, no. 3, pp. 183–223, 2004.

[10] H. F. Credidio, E. N. Teixeira, Reis, Saulo D S, A. A. Moreira, and J. S. Andrade, "Statistical patterns of visual search for hidden objects," *Scientific reports*, vol. 2, p. 920, 2012.

[11] A. Bulling, J. A. Ward, H. Gellersen, and G. Tröster, "Eye movement analysis for activity recognition," in *Proc. UbiComp*. New York, NY, USA: ACM, 2009, pp. 41–50.

[12] S. Alghowinem, M. AlShehri, R. Goecke, and M. Wagner, "Exploring eye activity as an indication of emotional states using an eye-tracking sensor," in *Intelligent Systems for Science and Information*. Springer, 2014, pp. 261–276.

[13] J. Beatty, "Task-evoked pupillary responses, processing load, and the structure of processing resources." *Psychological Bulletin*, vol. 91, no. 2, p. 276, 1982.

[14] S. Ishimaru, K. Kunze, K. Kise, J. Weppner, A. Dengel, P. Lukowicz, and A. Bulling, "In the blink of an eye: Combining head motion and eye blink frequency for activity recognition with google glass," in *Proc. Augmented Human*. New York, NY, USA: ACM, 2014, pp. 15:1–15:4.

[15] E. Ohn-Bar, S. Martin, A. Tawari, and M. Trivedi, "Head, eye, and hand patterns for driver activity recognition," in *Proc. ICPR*. IEEE, 2014, pp. 660–665.

[16] D. Roy, Y. Ghitza, J. Bartelma, and C. Kehoe, "Visual memory augmentation: using eye gaze as an attention filter," in *Proc. ISWC*. IEEE, 2004, pp. 128–131.

[17] J. A. Kientz, S. N. Patel, A. Z. Tyebkhan, B. Gane, J. Wiley, and G. D. Abowd, "Where's my stuff?: Design and evaluation of a mobile system for locating lost items for the visually impaired," in *Proc. ASSETS*. New York, NY, USA: ACM, 2006, pp. 103–110.

[18] T. Nakada, H. Kanai, and S. Kunifuji, "A support system for finding lost objects using spotlight," in *Proc. MobileHCI*. New York, NY, USA: ACM, 2005, pp. 321–322.

[19] M. Kassner, W. Patera, and A. Bulling, "Pupil: An open source platform for pervasive eye tracking and mobile gaze-based interaction," in *Proc. UbiComp*. New York, NY, USA: ACM, 2014, pp. 1151–1160.

[20] J. Wagner, F. Lingenfelser, T. Baur, I. Damian, F. Kistler, and E. André, "The social signal interpretation (ssi) framework: multimodal signal processing and recognition in real-time," in *Proc. ACMMM*. New York, NY, USA: ACM, 2013, pp. 831–834.

[21] D. A. Schaumberg, D. A. Sullivan, J. E. Buring, and M. R. Dana, "Prevalence of dry eye syndrome among us women," *American Journal of Ophthalmology*, vol. 136, no. 2, pp. 318–326, 2003.