Eye-based head gestures for interaction in the car

Diako Mardanbegi
IT University of Copenhagen
Rued Langgaards Vej 7, DK-2300
Copenhagen S
+45 72 18 53 72
dima@itu.dk

Dan Witzner Hansen
IT University of Copenhagen
Rued Langgaards Vej 7, DK-2300
Copenhagen S
+45 72 18 50 88
witzner@itu.dk

ABSTRACT

In this paper we suggest using a new method for head gesture recognition in the automotive context. This method involves using only the eye tracker for measuring the head movements through the eye movements when the gaze point is fixed. It allows for identifying a wide range of head gestures that can be used as an alternative input in the multimodal interaction context. Two approaches are described for using this method for interaction with objects inside or outside the car. Some application examples are described where the discrete or continuous head movements in combination with the driver's visual attention can be used for controlling the objects inside the car.

Categories

H.5.2 [Information interfaces and presentation (e.g., HCI)]: User Interfaces – Interaction styles (e.g., commands, menus, forms, direct manipulation).

General Terms

Human Factors; Measurement.

Keywords

Gaze tracking; Head gesture; Interaction; Eye movements

1. INTRODUCTION

In the last decades, automotive user interfaces have become more complex with much new functionality. Besides controlling the vehicle and operating the primary tasks (maneuvering the car e.g. controlling the speed or checking the distance to other cars), drivers need to interact with a variety of digital devices and applications in the car when driving. However, driver's focus on driving, is still the primary task, and should have the highest priority. The other tasks should be as minimally distracting as possible for the safety reasons [11]. New interaction techniques like speech, touch, gesture recognition, and also gaze have found their way to be used for interaction with user interfaces in a multifunctional space like car. This paper proposes using *eye-based head gestures* as a potential technique for interaction with automotive user interfaces. Eye-based head gesture [13] is a technique for recognizing head gestures. It uses the driver's gaze and eye tracking data for a) distinguishing the gestures from the natural head movements, b) for measuring the head gestures, and c) for using the driver's intention in interaction with objects.

Among the new interaction methods that have so far been studied in the automotive context, techniques like speech and head gestures have the advantage of providing a way for hands-free interaction. However, speech, and head gesture recognition often require a short explicit command like pushing a button before they can be used. Therefore, they can be used in multimodal interaction systems combined with the other input modes and help to minimize the amount of time that the driver's hand is off the steering wheel.

Associated level of physical, visual, and mental workload should be considered when designing a user interface and thinking about the interaction with an automotive user interface [3]. There have been some studies that report that certain kinds of voice-activated interfaces impose inappropriately high cognitive loads and can negatively affect driving performance [5, 6]. The main reason is that we are still far from achieving high-performance automatic speech recognition (ASR) systems. There are also some tasks like controlling radio volume, opening the window just slightly, continuously zoom or scrolling the map which are not intuitive operations to perform solely via speech-based interaction. Speech input cannot also be used when the environment is too noisy. In contrast, head gesture recognition is more reliable and can be a good alternative to speech input. Even if the number of different detected gestures is relatively small, they can be used as both continuous and discrete commands. Interaction by head gestures involves less driver's cognitive load as it can use the natural human communication skills. However the head gesture recognition has been mostly concentrated on detecting head shakes and nods to communicate approval or rejection and as an intuitive alternative in any kind of yes/no decision of system-initiated questions or option dialogs.

On the other hand, much work has been done in driver fatigue detection, and a fatigue monitoring device have been studied as a tool that allow for implicit interaction between the car and the driver to improve driving safety [16]. Eye and the visual behaviors measured by a video-based eye tracker provide significant information about driver's attention [14, 15] and the state of drowsiness and vigilance [18]. A video based eye tracker can also be used for recognizing head gestures using the eye and gaze information. It is possible to detect a wide range of head gestures as well as nods and shakes, which can be used for interaction. Head gestures can also be interpreted as different interaction commands by using the other modalities like gaze and intention proving an inferred interaction.

The paper is organized as follows. Some related works are described in the next section. Then, eye-based head gesture and the interaction method are described. Some application scenarios

AutomotiveUI'12, October 17-19, Portsmouth, NH, USA.
Adjunct Proceedings.

of using the method for interaction with objects in the car are described in a subsequent section and finally we conclude in the last section.

2. RELATED WORK

Many methods for gesture recognition have been proposed and some of them are applied to the automotive environment for detecting the head and hand gestures. Among the non video-based methods, an interesting work was done by Geiger [7], in which a field of infrared distance sensors is used to locate the hand and the head of the driver and sensing the movements. Although the sensor array does not achieve the resolution of a video-based methods, but his system is evaluated to be highly robust in measuring the simple directional gestures. Here, our focus is on the video-based methods for head gesture recognition. Many video-based techniques have been proposed for tracking the user's head and mostly are based on head/face detection and tracking. For example, Althoff [1], developed a system for detecting the head nod and shake using a near infrared imaging approach for interaction in the vehicle. In general, video-based techniques use some features of the face for detecting the head position in 2-D image space [12, 19], or some of them work by fitting a 3D model to the face in each image of the video to provide estimates of the 3D pose of the face [2]. However, these methods are not usually robust enough to strong illumination changes, and usually not accurate and fast enough to be useful for interactive environments.

On the other hands, some attempts have been made to use eye image for head gesture recognition. Concentrating on head gesture recognition methods that use the eye features, Davis and Vaks [4] presented a prototype perceptual user interface for a responsive dialog-box agent. They used IBM PupilCam technology for only detecting the eye location in the image and used together with anthropometric head and face measurements to detect the location of the user's face. A Finite State Machine incorporating the natural timings of the computed head motions was employed for recognition of head gestures (nod=yes, shake=no). Kapoor and Picard [10] introduced an infrared camera synchronized with infrared LEDs to detect the position of the pupils. Recognizing the head gestures had been demonstrated by tracking the eye position over time and a HMM based pattern analyzer was used detecting the nod/shake head gesture in real-time. However, their system used complex hardware and software and had problems with people wearing glasses and with earrings. The most relevant work to this paper is conducted by Ji and Yang [8, 9]. They have proposed a camera-based real-time prototype system for monitoring driver vigilance. An infrared imaging system and the bright/dark pupil effects (similar to PupilCam) is used for detecting the pupil position. They investigated the relationships between face orientation and these pupil features and so that the 3D face (head) pose have been estimated from a set of seven pupil features: inter-pupil distance, sizes of left and right pupils, intensities of left and right pupils, and ellipse ratios of left and right pupils. They have also estimated the driver's gaze and average eye closure speed having the eye images. However, their gaze estimation was limited into nine areas: frontal, left, right, up, down, upper left, upper right, lower left and lower right. Head movements were not measured accurately and what they were interested was to detect if the driver head deviates from its nominal position/orientation for an extended time or too frequently. The same idea for detecting the limited head movement and the rough gaze estimation using the eye images (with different methods) had been also presented before in [17].

3. EYE-BASED HEAD GESTURES

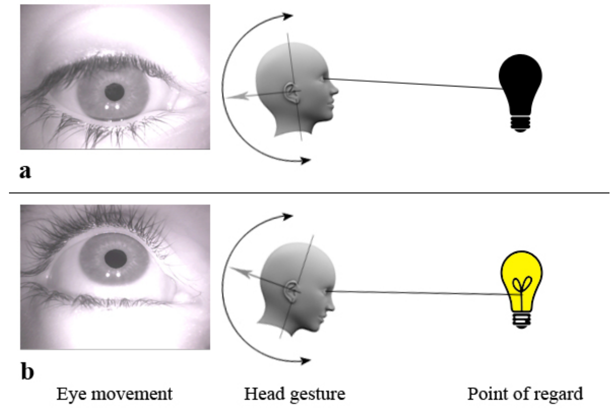

Figure 1: When the gaze point is fixed the head movements can be measured through the eye movements

Eye movements can be caused by the head movements while point of regard (PoR) is fixed or by changing the PoR when the head is fixed. When the point of regard is fixed and the head moves, the eyes move in the opposite direction and with the same speed as the head movement. These eye movements are due to the vestibulo-ocular reflexes (VOR), which are used to stabilize the image on the retina. Figure 1 illustrates a user looking at an object but in two different situations, one when the head is up (Figure 1.a) and the other when the head is down (Figure 1.b). The eye image is different in each posture even though the PoR is fixed. Since the eye trackers measure the eye movements and estimate the point of regard, they are able to measure the head movements when the PoR is fixed. In this paper, the term eye-based head gestures, denotes a predefined pattern of head movements measured through eye movements but where the PoR is fixed on a given object, and the term fixed-gaze target denotes the object that PoR is fixed on it. This method is able to measure a wide range of the head movements (including the head roll) and even though they are very small. The head roll can be detected by measuring the optic flow of the iris pattern and the yaw/pitch movements by tracking the pupil center. Figure 2 shows the basic roll, yaw and pitch movements of the head and the corresponding eye movements in the eye image.

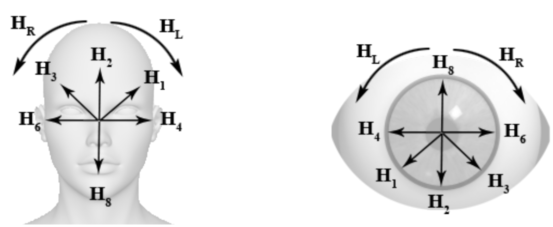

Figure 2: The basic head movements and their corresponding eye movements

This method is independent of the type of the eye tracker and where the data come from and it can be used for head gesture recognition whenever the gaze point and the eye image are available.

Head gestures together with fixed gaze point can be used as a method for gaze based interaction. A combination of fixed gaze and head gestures can be used for interaction with both the objects inside the car and also outside of the car. Two different methods are presented in this section for interaction with objects inside or outside the car. The main reason of separating these two is that fixating the gaze on the objects inside the vehicle during performing the head gesture is not acceptable, and we are interested to minimize the amount of time that the driver's visual attention is away from the forward roadway.

3.1 Interaction with the roadway objects:

For interaction with the objects on the roadway (e.g. getting information about the signs), the driver can simply keep the gaze fixed on the object and then perform a gesture. The eye tracker will recognize the gazed object even though the object and the driver may have a relative movement. When the object has a velocity less than $15°s^{-1}$ in the field of view, the eyes have a slow movement called smooth pursuit. Above this speed the smooth pursuit will be accompanied by saccades. Therefore, these eye movements need to be differentiated from the eye movements caused by the head gestures according to their range of speed. However, in this case, the head rolls can be easily detected by measuring the iris torsion, and can be used as gestures.

3.2 Interaction with the objects inside the car:

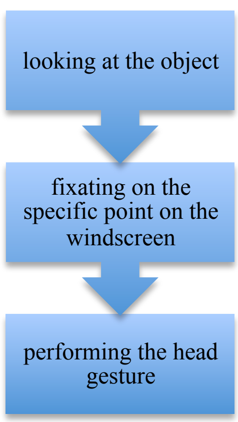

Figure 3: The main 3 steps for interacting with objects inside the vehicle

Interacting with the objects by looking at the object, fixating the gaze on the object, and then performing a head gestures can be useful for some tasks. However, when the task is more complex, this method would not be a safe approach for interaction (e.g. adjusting the side-view mirror in the car). With the method described below, we minimize the time that the gaze is away from the roadway by transferring the fixed-gaze target from a point on the object to a specified point on the windscreen. This point can be indicated by a small dot located on the windscreen in front of the driver. When the target is shown on the windscreen allows the driver to maintain attention to events happening on the road. Therefore, Interaction with the objects inside the car can be done by looking at the object, and then fixating the gaze on a specific point on the windscreen and performing the head gesture. This method uses the driver's visual attention as an implicit interaction modality, so that when the driver looks at an object in the car (e.g. the window) the eye tracker recognize that specific object and then waits for the next step. Once the user fixates on the specific point on the windscreen, the system waits for the user's head gesture for controlling the last intended object.

While performing the gesture, eye tracker measures the eye movements and tracks the gaze point. The distance between the windscreen target and the eye is basically less than 1 meter and therefore the driver's eyes converge during the gesture. The eye tracker can detect this convergence by measuring the distance between the two pupils. Therefore, the convergence of the eyes can be used as an indicator that the driver is performing a gesture.

4. APPLICATION SCENARIOS

Some example applications of using eye-based head gestures in the automotive context are described in this section.

Head gestures have a great potential to be used as an intuitive alternative in any kind of yes/no decision when a system initiated questions or option dialogs. As an example, when the mobile phone is ringing, the incoming calls can be accepted or denied by the head gestures. These simple vertical head gestures can also be used for flipping the rear-view mirror down or up.

The left and right head movements can be used for shortcut functions enabling the user to control the music player and to skip between individual cd-tracks or radio stations.

This method can also be used as a way for interacting between the driver and the head-up display (HUD), enabling the driver to do selecting and for switching between different submenus in a more intuitive way compared to standard button interactions.

Continuous vertical movements of the head can be useful for changing the volume, adjusting the air conditioning temperature, opening and closing the window, and continuously zoom or scrolling the map. In these examples, visual or audio feedback through HUD or speakers can help the driver to perform the task more efficiently. The visual feedback can be a highlight color or even displaying the image of the object. For example, when the driver wants to adjust the side-view mirrors, he/she looks at the mirror and then the eye tracker recognize the mirror as the attended object and then the system shows the real-time image of the mirror in the head-up display. Now, the driver can see the mirror image in front of the windscreen and therefore can easily adjust the mirror by the head movements.

5. Conclusion

In this paper, we suggested to use eye-based head gestures for interaction in the automobile. This method uses only the information extracted from the eye image for measuring the head movements. One of the advantages of this technique is that even very small head movements can be measured through the eye movements. Another advantage is that a video-based eye trackers can potentially be used as one multi-purpose device in the car for head gesture recognition as well as for fatigue detection, monitoring the driver's visual attention, and gaze estimation. Some example applications are described where the gaze and head gestures are used together for controlling some objects in the car. In general, whenever the head gestures are used so far in the automotive context, the new method for head gesture recognition can be applied, too.

6. REFERENCES

[1] Althoff, F., Lindl, R., and Walchshaeusl, L. Robust Multimodal Hand- and Head Gesture Recognition for controlling Automotive Infotainment Systems. In VDI-Tagung: Der Fahrer im 21. Jahrhundert, Braunschweig, Germany, November 22-23 2005.

[2] Black, M.J. and Yacoob, Y. Tracking and recognizing rigid and non-rigid facial motions using local parametric models of image motion. In ICCV95, pages 374–381, 1995.

[3] Burnett, G.E, Designing and evaluating in-car user-interfaces. In J. Lumsden (Ed.) Handbook of Research on User-Interface Design and Evaluation for Mobile Technology, Idea Group Inc. 2008.

[4] Davis, J.W., and Vaks, S. 2001. A Perceptual User Interface for Recognizing Head Gesture Acknowledgements. In Proceedings Workshop on Perceptive User Interfaces.

[5] Garay-Vega, L., A. Pradhan, G. Weinberg, B. Schmidt-Nielsen, B. Harsham, Y. Shen, G. Divekar, M. Romoser, M. Knodler, and D. Fisher. Evaluation of different speech and touch interfaces to in-vehicle music retrieval systems. Accident Analysis & Prevention, 42(3): 913–920, May 2010.

[6] Gartner, U., Konig, W., and Wittig, T. Evaluation of manual vs. speech input when using a driver information system in real traffic. In International driving symposium on human factors in driver assessment, training and vehicle design, 2001.

[7] Geiger, M. Beruhrungslose Bedienung von Infotainment-Systemen im Fahrzeug. PhD thesis, TU Mu ̈nchen, 2003.

[8] Ji, Q., Yang, X. Real-time eye, gaze, and face pose tracking for monitoring driver vigilance, in: Rea Time Imaging, pages 357-377, 2002.

[9] Ji, Q. and Bebis, G. "Visual Cues Extraction for Monitoring Driver's Vigilance." Procs. Honda Symposium, pp.48-55,1999.

[10] Kapoor, A., and Picard, R.W. 2002. A real-time head nod and shake detector. Technical Report 544, MIT Media Laboratory Affective Computing Group.

[11] Kern, D., Schmidt, A. Design space for driver-based automotive user interfaces. In Proceedings of the 1st International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI '09). ACM Press (2009), 3-10.

[12] Kjeldsen, R. Head gestures for computer control. In Proc. Second International Workshop on Recognition, Analysis and Tracking of Faces and Gestures in Real-time Systems, pages 62–67, 2001.

[13] Mardanbegi, D., Hansen, D.W., and Pederson, T. "Eye-based head gestures: Head gestures through eye movements". Proceedings of the ACM symposium on Eye tracking research & applications ETRA '12, ACM Press, California, USA, 2012.

[14] Pomarjanschi, L., Dorr, M., and Barth, E. Gaze guidance reduces the number of collisions with pedestrians in a driving simulator. ACM Transactions on Interactive Intelligent Systems, 1(2):8:1-8:14, 2012.

[15] Pomarjanschi, L., Rasche C., Dorr M., Vig E., Barth E. 2010, "Safer driving with gaze guidance" Perception 39 ECVP Abstract Supplement, page 83.

[16] Schmidt A (2000) Implicit human computer interaction through context. Pers Ubiquit Comput 4(2):191–199.

[17] Smith, P., Shah, M., and da Vitoria Lobo, N. "Monitoring. Head/Eye Motion for Driver Alertness with One Camera', The. Fifteenth IEEE ICPR. Nov. 2000.

[18] Wang, Q., Yang, J., Ren, M., Zheng, Y. "Driver Fatigue Detection: A Survey", in Proc Intelligent control and Automation, Dalion, China, pp 8587- 8591, 2006.

[19] Wren, C.R., Azarbayejani, A., Darrell, T.J., and Pentland, A.P. Pfinder: Real-time tracking of the human body. PAMI, 19(7):780–785, July 1997.