Gaze-Based Controlling a Vehicle

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

Dan Witzner Hansen
IT University of Copenhagen
Rued Langgaards Vej 7,
DK-2300 Copenhagen S
Witzner@itu.dk

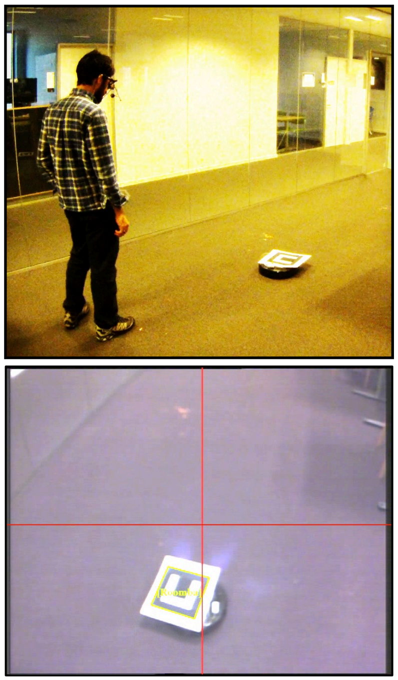

Figure 1: Controlling a Roomba vacuum cleaner by gaze in a mobile situation. The user's PoR is shown (cross hair) in the scene image of the mobile gaze tracker in the bottom image.

Abstract

Research and applications of gaze interaction has mainly been conducted on a 2 dimensional surface (usually screens) for controlling a computer or controlling the movements of a robot. Emerging wearable and mobile technologies, such as google glasses may shift how gaze is used as an interactive modality if gaze trackers are embedded into the head-mounted devices. The domain of gaze-based interactive applications increases dramatically as interaction is no longer constrained to 2D displays. This paper proposes a general framework for gaze-based controlling a non-stationary robot (vehicle) as an example of a complex gaze-based task in environment. This paper discusses the possibilities and limitations of how gaze interaction can be performed for controlling vehicles not only using a remote gaze tracker but also in general challenging situations where the user and robot are mobile and the movements may be governed by several degrees of freedom (e.g. flying). A case study is also introduced where the mobile gaze tracker is used for controlling a Roomba vacuum cleaner.

Author Keywords

Gaze-based interaction; robot; vehicle; craft; head-gestures; eye tracking; driving

ACM Classification Keywords

H.5.2. User Interfaces — Input devices and strategies.

Introduction

Gaze interaction can be generalized for usage in 3D environments where it can be used for interaction with many different types of objects present in our daily activities. This paper focuses on possibilities of using gaze trackers for controlling remote robots in 3D environment. Most approaches to gaze-based vehicle control are focused on using remote eye trackers and the point of regard on a monitor. This paper discusses how gaze interaction can be performed in more challenging situations where the eye tracker/user is mobile and where the vehicle movements may be governed by several degrees of freedom (e.g. flying). The paper categorizes different approaches for gaze-based controlling vehicles using the readily available data in eye trackers, and it discusses limitations and possibilities for the different approaches.

The rest of the paper is organized as follows. First the basic concepts in gaze controlling a vehicle are

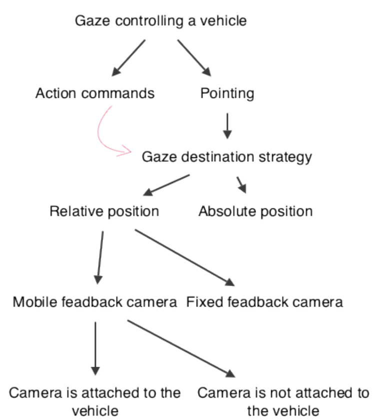

Figure 2: Gaze for controlling a vehicle.

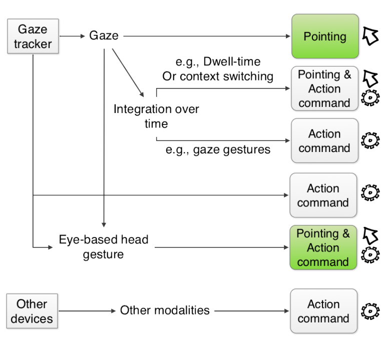

Figure 3: Gaze as a tool for pointing and for sending commands.

proposed in the following section. Then in the second section different situations and approaches are categorized based on the basic descriptions. Finally an implementation case study is presented and then we conclude the paper.

Basic Concepts and Overview

The fact that we often look into the direction of the next move when walking or driving, tells us that the gaze can somehow be used for enhancing the process of driving a car or a remote robot. Using gaze for controlling vehicles (e.g., wheelchairs and remote robots) has been studied in [2, 7, 8]. This paper studies the fundamental principles of gaze interaction with remote robots. There are many factors that influence the way that gaze can be used for controlling a vehicle (figure 2) specially in the challenging situations where: (a) mobile gaze trackers are used instead of remote gaze tracker, (b) when the user and robot can move relative to each other, (c) or when vehicle have more than two degrees of freedom (e.g. flying vehicles). In order to be able to discuss how gaze may be used in different situations, first these different factors are introduced in this section.

Controlling a Vehicle

In this paper, a vehicle is defined as a rigid body object located in 3D space with the ability of moving between two points inside its movement space. The movement space is a space of possible position pairs that the vehicle can travel between. The movement space may be one, two, or three dimensional (curve, surface, or a 3d space). In this paper, vehicles with each of these three types of movement spaces are termed as 1D, 2D or 3D vehicles.

A vehicle may have different conditions in terms of degrees of movement (translation and rotation). A 2D vehicle or a car can reach any point on the ground by having only one rotation (turning) and one translation (forward/backward), or by only having 2 translational degrees of movement (forward/backward and right/left). A vehicle may need to have more degrees of movements to be able to get any orientation in its movement space.

In this paper, controlling a vehicle is defined as below: "Sending at least one bit of information (*input information*) to a vehicle in order to start, stop, or changing the direction or velocity of the movement in at least one of the degrees of movement of the vehicle."

Consequently, when gaze is used in any form (directly or indirectly) in the process of providing the input information by the machine, we can say that vehicle is controlled by the gaze. In gaze-base controlling approaches, gaze can be used only for pointing, or for both pointing and sending commands together. This has been described more in the following.

Gaze for Interaction

Using gaze for interaction on screen-based interfaces is well known [6]. The point of regard and pupil position are locations in space. The point of regard (PoR) can at a specific time instance and due to the Midas-touch problem only be used for pointing and not be used to yield any additional commands. More information is needed to define commands e.g. to make a selection. A common way to achieve more information is to integrate the eye and gaze information over the time. Dwell-time activation is obtained when the gaze and pupil positions are fixed over time. The limitations of

Examples of two Gaze control strategies

Gaze action strategy:

- The user looks at the "left" button on the computer display for moving the vehicle to the left.
- The user looks at the vehicle for moving it or stopping it.
- The user is sitting inside a car and looks at a point shown on the windscreen and perform a head gesture for changing the direction.

Gaze destination strategy:

- The user looks at a point on the floor (space) and sends a command, and then the robot goes to that point on the floor (space). In this case, the user may be looking at the scene through a display.
- The user is sitting inside a craft and looks at a target point and the craft goes toward that point.
- The user is looking at a button called "kitchen" and then the vehicle goes to the kitchen.

dwell-time activations have already been investigated thoroughly [4]. The principles of gaze gestures [3] is based on the pupil and point of regard (in space) are both changed over time. A known limitation of gaze gestures is that they do not allow the user to keep the gaze at object - i.e. gaze is removed from its context in which the interaction is intended. Context switching [9] turns this limitation into to an advantage by defining at least two contexts and let the transition of gaze and eye positions act as the defining principle for a command. Eye-based head gesture is a novel method for enhancing gaze-based interaction through voluntary head movements [5]. The method allows the gaze position to remain fixed while the pupil position is changing over time. Eye-based head gesture is based on the fact that when the point of regard is fixed and the head moves, the eyes move in the opposite direction due to the vestibulo-ocular reflex.
Since, eye-based head gesture technique can be achieved with both remote and head-mounted gaze trackers and provides us a gaze-based interaction method for executing commands in remote and mobile situations, in this paper it has been considered as the default tool for sending action commands to a remote robot (figure 3). However, measuring the head movements through the eye movements may be challenging in situations where the gazed object moves very fast while performing the head gesture. In these challenging situations in which the eye-based head gestures are not measurable, other interaction modalities for example the head movements measured by the other devices can be used for sending the action commands and gaze is only used for pointing.

Gaze Control Strategies

The PoR is usually a point on a 2D surface [1], however gaze may also be estimated as a 3D point or as a direction. This paper assumes that whenever the gaze is estimated as a direction, it should be intersected with a surface in order to become useful for controlling a vehicle. In this paper, possible approaches of using gaze point for controlling a vehicle are classified into two strategies: 1) gaze action strategy 2) gaze destination strategy.
Gaze action strategy uses the gaze for sending an action command to the system for starting or stopping the movement of the vehicle in one of its degrees of movement. Gaze destination strategy is when the gaze is used for giving information about where the movement should be stopped (desired or destination point). In the gaze destination strategy, the user may be directly looking at a desired point in space or in an image, or he/she may look at a context that contains information about the destination position.
The gaze destination strategy in which robot goes to reach a gazed destination point, involves different approaches based on the knowledge of the system about the exact position of the vehicle, the user and the gaze point in the world coordinates system. This has been addressed in the following.

Relative or Absolute Positions

When the gaze is used for pointing the destination of the next movement of vehicle, the navigation method would be different based on the information that the system has about the locations of the gaze point and the vehicle in the world coordinates system. When the system have enough information for obtaining the absolute position of the gaze point and the vehicle in the world coordinates system, it can easily determine

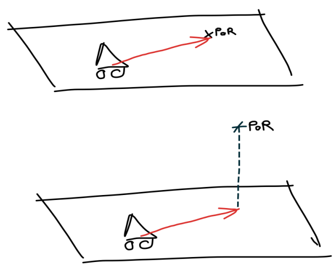

Figure 4: Minimizing the distance between the PoR and the 2D vehicle position on a plane

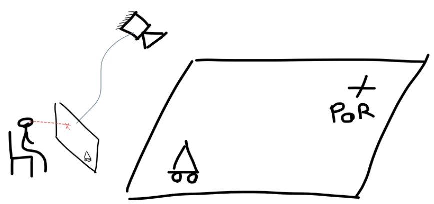

Figure 5: The user is sitting in front of a display showing the image from the feedback camera and a remote gaze tracker is used for gaze estimation.

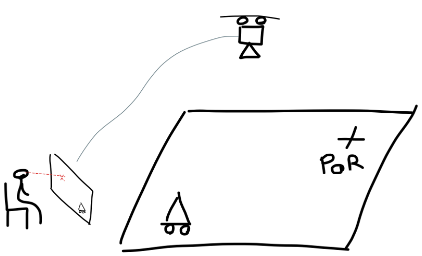

Figure 6: the feedback camera is mobile and the image is shown on a remote display

the path that the vehicle should follow in its movement space in order to reach the destination.
When the system does not have enough information about the exact position of the vehicle in the world coordinates system, but the relative positions of the vehicle and the gaze are accessible in an image, the navigation still can be done through a feedback loop. In this approach, the system measures the relative distance between the gaze point and the vehicle position in each time instance and always tries to minimize this distance.

Methods

In this section, we investigate different gaze-based methods of controlling a vehicle in different conditions.

Using Gaze action strategy

Hemin et. al [2] have followed this approach by introducing the TeleGaze interface overlaid on top of the video stream from the robot camera shown on a computer display, and used gaze for controlling a remote robot. Eye-based head gestures can be used for sending the action commands (e.g., left, right, forward, and backward) to a vehicle. It may be done by interacting with a graphical user interface or by looking at the objects in the real world. A case study of using eye-based head gestures for interacting with a mobile robot is introduced in the section in the following.

Using Gaze destination strategy

Different approaches of this strategy are categorized based of the knowledge of the system about the position of the vehicle and the PoR in the world coordinates system.

Absolute positions

This method is very straightforward when the position of the vehicle and the gaze point is known in the world coordinates system. The system infers the destination point from the gaze and then moves the vehicle toward the gazed point. In case of using mobile gaze trackers, estimating the gaze point in the world coordinates system varies based on the type of the eye tracker that is used and some times requires information about the position of the user in space. However, the destination coordinates can also be obtained indirectly from the gaze. For example when the destination is the point A, user can look at a button called "A" on the computer display, or a real object in the real world signed as "A", and the system infers the destination coordinates through the context of that object. The desired destination point may be inside or outside the movement space of the vehicle, and the system always tries to minimize the distance between the vehicle the destination point (figure 4). The way that the system navigates the vehicle is out of the scope of this paper.

Relative positions

When the absolute position of the vehicle is unknown, in some situations, the gaze tracker may still be used for navigating the vehicle. It should be noted that here we assume that the gaze is pointing to a desired position (PoR) and the system moves the vehicle toward that point. Therefore, when we want the vehicle to follow our gaze point continuously (like when the cursor follows the gaze on the screen), at each instance of time, we are actually defining a destination point for the next movement. Consequently, we only discuss one step of controlling the vehicle, where we point to one specific point in space. When the system does not have any information about the absolute position of the

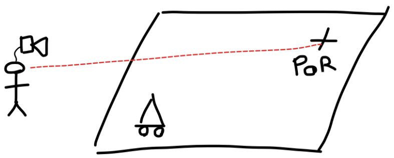

Figure 7: the feedback camera is the scene camera of a mobile gaze tracker.

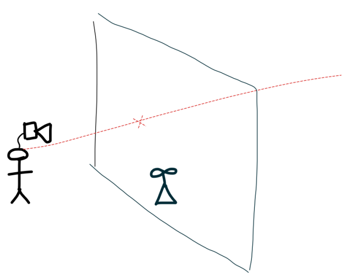

Figure 8: This figure shows the same concept of the figure 3 but with a 3D robot (craft) that be moved inside a virtual plane by the gaze. The movement in the third dimension (along the gaze direction) is controlled by looking at the craft and through continues head movements.

vehicle in the real world, but the relative positions of the destination point (gaze point) and the vehicle are known, navigation can be done through a feedback loop. This relative position may be measured by the system visually through a camera (feedback camera). The image of the camera should contain enough information about the PoR position in the image (x,y), and the posture of the vehicle (for all degrees of movement). When we only want to move the vehicle to the PoR and the final orientation is not important for us, the system only needs to be able to measure the changes of the minimum degrees of movement of the vehicle between each two frames. Two situations may happen for the feedback camera. The camera can be fixed or mobile in the world coordinate system while vehicle is moving in the feedback loop. Different approaches that can be used in these two situations are described below:

Fixed Feedback Camera

When the camera is fixed, one time sampling the gaze point in the camera image is enough and the system only needs to get feedback from the posture of the vehicle (changes in degrees of movement). If the user wants to change the destination point, he/she looks at another point and the system needs to update the PoR estimated by the gaze tracker. One example of this situation is shown in figure 5. It is obvious that when the PoR and the vehicle are along the optical axis of the camera, the system needs some extra information to move the vehicle in that direction.

Mobile Feedback Camera

In this situation, the user's gaze may move in the image when the feedback camera is moving. Therefore, the position of the destination point may be changed in each frame. Many computer vision techniques can be applied for detecting or tracking the destination point (PoR) in the image while the camera in moving. It means that the user does not need to keep gazing at the destination point while moving the vehicle, and one time sampling is enough. Two situations are shown in figure 6 and figure 7 where the feedback camera is mobile and it moves independently from the vehicle. When we have a 3D vehicle (craft) and the camera is not attached to it, the same approach can be used (figure 8). However, the craft can only move in a 2-dimensional plane unless the third degree of movement of the vehicle is activated by a command (e.g., pressing a button, or blinking, or a gesture). The eye-based head gestures can be very useful in this situation, because it allows the user to even control the third degree of movement by the continuous head movements while looking at the vehicle (craft). Furthermore, it does not require an extra device for controlling the third dimensional movement.

Figures 9 shows a situation where the feedback camera is attached to the vehicle. The image of the camera can then be transferred to a remote user outside the vehicle. The main important object that has to be considered here is that at least the axis of one of the translational movements of the vehicle should be visible in the image (either as a point or a line). In the feedback loop, the system tries to minimize the relative distance between the gaze point and the projection of the translational (forward/backward) movement in the image (adjusting the locomotion of the robot). However, the forward/backward movement cannot be controlled by the gaze point, and another modality should be used for controlling the movement in that direction. Eye-based head gesture can be used for this

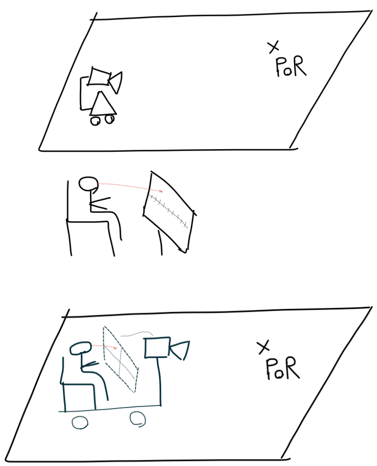

Figure 9: The feedback camera is attached to the vehicle and the image is shown on a remote desktop

Video demonstrations of this experiment can be accessed at http://youtu.be/6O2gYjRymyg.

More information about the open source Haytham gaze tracker is used for gaze tracking, eye-based head gestures, and interaction with the robot can also be accessed at http://eye.itu.dk.

purpose while the user is looking at a point. Tall et.al [8] have implemented the situation shown in figure 9. The camera may also be attached to a 3D craft in space and the gaze can be used in a same way for controlling it. However, the navigation method during the feedback loop varies based on the degrees of movement of the vehicle.

Case Study

An implementation case study has been conducted for controlling a Roomba vacuum cleaner using a mobile gaze tracker. Two strategies are tested in the experiment. The first was the gaze action strategy where the user looks at the robot ant controls the robot using the head gestures. The second, Roomba is following the user's gaze in the scene image and only the action commands (e.g., clean, turn off, and turn on) were sent using the head gestures. Figure 7 shows the mobile situation and the scene image of the mobile gaze tracker. A visual marker is used for detecting the robot in the image. This case study is an example of the situation using a mobile feedback camera in a relative position method.

Conclusion

The fundamental principles for controlling non-stationary robots have been studied. Different approaches for controlling vehicles using gaze are categorized based the knowledge of the system about the PoR and position of the vehicle in 3D space. A case study is introduced in which two of the approaches are implemented. This case study shows the potential of using only a mobile gaze tracker for controlling a remote robot in a 3D environment.

References

[1] Hansen, D.W., and Ji, Q., 2010. In the Eye of the Beholder: A Survey of Models for Eyes and Gaze. IEEE Trans. Pattern Anal. Mach. Intell, pages 478-500.

[2] Hemin, O. L., Nasser, S. & Ahmad L. (2008) Remote Control of Mobile Robots through Human Eye Gaze: The Design and Evaluation of an Interfacel, SPIE Europe Security and Defence 2008.

[3] Isokoski, P. 2000. Text Input Methods for Eye Trackers Using Off-Screen Targets. In Proceedings of the ACM symposium on Eye tracking research & applications ETRA '00. ACM Press.

[4] MacKenzie, I. S., & Zhang, X. (2008). Eye typing using word and letter prediction and a fixation algorithm. In ETRA '08: Proceedings of the 2008 symposium on Eye tracking research & applications.

[5] Mardanbegi, D., Hansen, D. W., and Pederson, T. Eye-based head gestures. In Proceedings of the Symposium on Eye Tracking Research and Applications - ETRA'12, ACM Press.

[6] Morimoto C., H., Koons, D., Amir, A., Flickner, M., and Zhai, S. 1999. Keeping an Eye for HCI. In Proceedings of the XII Brazilian Symposium on Computer Graphics and Image Processing (SIBGRAPI '99). IEEE Computer Society.

[7] Matsumoto Y., Ino, T. & Ogasawara, T. (2001). Development of Intelligent Wheelchair System with Face and Gaze Based Interface, Proceedings of 10th IEEE Int. Workshop on Robot and Human Communication.

[8] Tall, M., Alapetite, A., Agustin J.S., Skovsgaard, H., Hansen, J. P., Hansen, D.W., and Møllenbach, E. 2009. Gaze-controlled driving. In CHI '09 Extended Abstracts on Human Factors in Computing Systems (CHI EA '09).

[9] Tula, A. D., Campos, F., and Morimoto, C.H. 2012. Dynamic context switching for gaze based interaction. In Proceedings of the Symposium on Eye Tracking Research and Applications (ETRA '12), Stephen N. Spencer (Ed.).