Eye Tracking and Gaze Interface Design for Pervasive Displays

Yanxia Zhang

B.Sc., Huazhong University of Science and Technology

M.Sc., Universiteit van Amsterdam

Submitted to Lancaster University for the degree of

Doctor of Philosophy

September 2015.

Eye Tracking and Gaze Interface Design for Pervasive Displays

Yanxia Zhang

Submitted for the degree of Doctor of Philosophy
September 2015

Abstract

Eye tracking for pervasive displays in everyday computing is an emerging area in research. There is an increasing number of pervasive displays in our surroundings, such as large displays in public spaces, digital boards in offices and smart televisions at home. Gaze is an attractive input modality for these displays, as people naturally look at objects of interest and use their eyes to seek information. Existing research has applied eye tracking in a variety of fields, but tends to be in constrained environments for lab applications.

This thesis investigates how to enable robust gaze sensing in pervasive contexts and how eye tracking can be applied for pervasive displays that we encounter in our daily life. To answer these questions, we identify the technical and design challenges posed by using gaze for pervasive displays.

Firstly, in out-of-lab environments, interactions are usually spontaneous where users and systems are unaware of each other beforehand. This poses the technical problem that gaze sensing should not need prior user training and should be robust in unconstrained environments. We develop novel vision-based systems that require only off-the-shelf RGB cameras to address this issue.

Secondly, in pervasive contexts, users are usually unaware of gaze interactivity of pervasive displays and the technical restrictions of gaze sensing systems. However, there is little knowledge about how to enable people to use gaze interactive systems in daily life. Thus, we design novel interfaces that allow novice users to interact with contents on pervasive displays, and we study the usage of our systems through field deployments. We demonstrate that people can walk up to a gaze interactive system and start to use it immediately without human assistance.

Lastly, pervasive displays could also support multiuser co-located collaborations. We explore the use of gaze for collaborative tasks. Our results show that sharing gaze information on shared displays can ease communications and improve collaboration.

Although we demonstrate benefits of using gaze for pervasive displays, open challenges remain in enabling gaze interaction in everyday computing and require further investigations. Our research provides a foundation for the rapidly growing field of eye tracking for pervasive displays.

Declaration

I declare that this thesis and the work in it are my own, and the result of my own original research. No part of this thesis has been submitted elsewhere for any other degree or qualification. This work was carried out under the supervision of Professor Hans Gellersen and Dr Andreas Bulling at Lancaster University.

Author: **Yanxia Zhang**

Acknowledgements

I would like to express my sincere gratitude to my supervisors Hans Gellersen and Andreas Bulling for their wisdom, guidance and inspiration. I want to thank my friends, colleagues in InfoLab21 and other collaborators. It was a wonderful time. I dedicate this thesis to my beloved father, mother and sister. I am extremely fortunate to grow up with so much support and love. I am very grateful to my lovely husband Ming Ki for his understanding and encouragement throughout this journey. Thank you for making my life fun and joyful.

List of Publications

- Y. Zhang, M. K. Chong, J. Müller, A. Bulling, and H. Gellersen. Eye Tracking for Public Displays in the Wild. In Personal and Ubiquitous Computing (PUC), August 2015, Volume 19, Issue 5-6, pp 967-981, 2015.

- Y. Zhang, A. Bulling, and H. Gellersen. Calibration-free Remote Eye Tracking based on Eye Symmetry. *To Appear* in book chapter of Context-Aware Systems: Methods and Applications.

- Y. Zhang, J. Müller, M. K. Chong, A. Bulling, and H. Gellersen. GazeHorizon: Enabling Passers-by to Interact with Public Displays by Gaze. In Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp′14), pages 559–563, 2014.

- Y. Zhang, A. Bulling, and H. Gellersen. Pupil-canthi-ratio: A calibration-free method for tracking horizontal gaze direction. In Proceedings of the 12th International Conference on Advanced Visual Interfaces (AVI′14), pages 129–132, 2014.

- Y. Zhang, A. Bulling, and H. Gellersen. Sideways: A gaze interface for spontaneous interaction with situated displays. In Proceedings of the 31st SIGCHI International Conference on Human Factors in Computing Systems (CHI′13), pages 851–860, 2013.

- Y. Zhang, A. Bulling, and H. Gellersen. Towards pervasive eye tracking using low-level image features. In Proceedings of the 7th International Symposium on Eye Tracking Research and Applications (ETRA′12), pages 261–264, 2012.

- Y. Zhang, A. Bulling, and H. Gellersen. Discrimination of gaze directions using low-level eye image features. In Proceedings of the 1st International Workshop on Pervasive Eye Tracking and Mobile Eye-Based Interaction (PETMEI 2011), pages 9–13, 2011.

Contents

List of Figures

LIST OF TABLES

1

Introduction

With the emergence of ubiquitous computing, it is foreseeable that computing devices will become invisible and input sensing will shift from the foreground to the background in everyday life [35]. The environments around us (such as homes, offices and public spaces) will be equipped with smart sensing capabilities that understand human behaviours and can serve human needs unobtrusively, without deliberate user actions.

As sensing devices integrate into the background, user input moves away from traditional keys, mice and styluses, where human input is explicit, unambiguous and fully attentive while controlling information and command flow [131], to naturally occurring activities or actions (e.g., gesture, voice or location) [13, 23]. Ubiquitous sensing systems should understand the interaction intended by users, and the interface should respond accordingly to satisfy the users' actions. To achieve such systems, an open challenge is to sense human intended actions, to interpret their behaviours, and to design appropriate reactions to serve human needs.

While researchers have extensively studied body movements and verbal behaviour (e.g., gestures, speech), gaze is also an efficient input for interaction [28, 196, 88, 58, 79, 155, 175, 48]. Gaze provides rich context information. The use of gaze (i.e. what and where users see, how their eyes move) plays important roles in human daily activities [103]. Our gaze is a good indicator of human interests and attention, and the motion of our eyes is tightly coupled with human cognitive and perceptual processes [191, 90, 76]. Gaze has also been considered as an attractive hands-free input modality as we naturally look at objects of interests [28, 196, 88, 58, 79]. Our eyes can move very fast with minimal fatigue and we can express our gaze from a distance. By integrating gaze sensing capability in everyday computing, it can enhance the way we interact with the environment. Our surroundings can then accommodate human attention and interests, adjust to our visual perception, and adapt to the cognitive states of users.

Despite the advantages of eye tracking, it has largely been employed in constrained environments for lab research [86]. With the advances in sensing technologies, eye tracking technologies are becoming affordable, compact and flexible to set up [74]. It is foreseeable that gaze sensing capabilities will be integrated in our daily life in the near future. Recently, eye tracking applications are expanding from diagnostic tools [132] and desktop user interface [28, 81, 156, 87, 196, 22] to new applications in the real-world contexts, such as life logging [84], activity recognition [33, 55] and attentive user interfaces [155, 175, 48]. These applications show that gaze can be useful in our daily life. However, there is only limited research and understanding of how gaze is used in everyday computing.

Currently, the computing paradigm is shifting from desktop computing to ubiquitous computing [13, 23]. The goal of this thesis is to explore the use of eye tracking in daily life applications where people can seamlessly interact with their surroundings by gaze. We aim to contribute new knowledge to the design of gaze interactive systems in the new paradigm. In our daily life, we encounter many displays around us (e.g. in museums, shopping malls and offices) which provide information or services [123, 43]. This work particularly explores the use of gaze information for these pervasive displays. As we naturally use our eyes to look around and seek information, we believe eye tracking can enhance people's interaction with these displays. For example, people look for information or interact with media contents in public spaces (e.g., shopping malls, train stations) or at home (e.g., living room). We envision that pervasive displays can sense human gaze, so that people can simply walk up to a display and use their eyes to interact with it spontaneously.

1.1 Eye Tracking for Pervasive Displays

The objective of this thesis is to inform the design of eye-based systems[1] for pervasive display applications. The scenarios considered in this thesis are public spaces, offices and home environments in contrast to constrained lab settings. To achieve this goal, we explore both the sensing techniques and interface design aspects for applying eye tracking in everyday environments.

Eye tracking systems and devices exist either as research prototypes [74] or commercial products (e.g., Tobii, LC Technologies, SMI, Ergoneers). They are

[1] Systems that take users' gaze and eye movements as input.

available in two forms: *wearable* and *remote*. Wearable eye trackers require users to attach electrodes on their skin around their eye areas (e.g., [33]) or wear a headset (e.g., [84]). Remote eye trackers are non-intrusive standalone devices that require users to face tracking sensors from a distance.

For pervasive display applications, users encounter these displays spontaneously. It is desirable to sense users' gaze non-intrusively, without anything attached to the user. In this thesis, we are interested in sensing users' gaze remotely using sensors that are commonly available, are flexible to setup and can accommodate diversity users in real-world environments. We are also interested in eye tracking systems that are deployable in large scale so that many people are able to use them in realistic settings (e.g., public spaces). Existing eye tracking systems are usually optimised for accurate and high-speed gaze estimation, but require additional illumination and a tedious calibration for each individual [120, 74, 38]. In addition, users are required to remain in front of the device and accuracy is only possible when users maintain their positions in a confined tracking area. For example, the state-of-the-art supports the range of approximately 35cm in any direction [1].

Besides using specialised hardware for high-fidelity eye tracking, researchers in computer vision have been investigating alternative approaches using off-the-shelf cameras [167, 25, 190, 165, 60]. However, much of these works focus on off-line algorithm optimisation for existing datasets (i.e., a prior training process) but neglect the developments of robust and unobtrusive real-time continuous eye tracking. No existing solutions satisfy the needs of eye tracking in real-world conditions. Our goal is to develop new robust systems that support continuous gaze sensing, with the potential for wide deployment, and require no prior user setup.

To address this, we explore computer vision and machine learning techniques for eye tracking.

The application of eye tracking for pervasive displays also raises new challenges for human-computer interaction (HCI) research. The second goal of this thesis is to design pervasive display interfaces that respond to human gaze behaviour and react to our needs appropriately. One of the difficulties is to move away from constrained lab settings where users perform well instructed tasks [13]. In pervasive contexts, interactions occur in spontaneous, short phases where systems have no prior information of their users. At the same time, users are often unaware of systems' interactivity, as human assistance is not always available.

Using gaze for interaction has a long history in HCI research, but much of the previous works focus on the WIMP (windows, icons, menus, pointer) graphical user interface [196, 152, 86, 57, 101, 100]. Their gaze interface designs are optimised for user performance that use gaze as a pointer for desktop input, such as typing [104] and selection [196, 152]. There is little knowledge of how to design gaze interfaces for everyday computing situations and to accommodate opportunistic behaviour. To enable user interactions in pervasive contexts, we explore the design of suitable gaze-based user interfaces and interaction techniques.

1.2 Research Statements

To address the research challenges set above, this work integrates knowledge from the research domains of eye tracking, computer vision, machine learning, ubiquitous computing and human-computer interaction. This thesis aims to explore eye tracking for pervasive display applications in out-of-lab environments, and intends to address the following two research questions:

1. **How to enable robust gaze sensing for everyday usage?**

   Existing eye tracking solutions focus on accurate gaze sensing using specialised hardware. These systems often require the use of additional illumination and an explicit calibration procedure. In pervasive contexts, the users and the environments are unconstrained, such as uncontrolled lighting conditions and variable user positions. For large scale deployment and long term usage, sensing systems should provide robust real-time performance for diverse users and out-of-lab environments. Towards these goals, this thesis targets real-time performance and robustness for pervasive display applications in daily life settings. We aim to answer which technologies, methods and systems can be used for gaze sensing when no prior user information is known, and do not require additional illumination, while remaining flexible to set up and deployable in the wild.

2. **How to design gaze interfaces for everyday pervasive displays?**

   The previous question addresses technical challenges, which leads to new forms of gaze tracking systems that are suitable for pervasive display applications. The second issue addressed in this thesis is the design of gaze interactions and interfaces. Pervasive displays could be used by single users to acquire information (e.g., large screens like a television at home or in public environments) and could also be shared among multiple users for collaboration (e.g., digital boards in office environments). This thesis considers how to accommodate opportunistic behaviour in everyday computing situations. Our objective is to enable spontaneous interaction, such that users can walk up and use their eyes to interact with digital contents without any human assistance. In addition, we aim to find out how people interact with gaze-based interfaces in uncontrolled settings.

1.3 Methodology

This thesis addresses technical and design challenges of applying eye tracking for pervasive displays (illustrated in Figure 1.1).

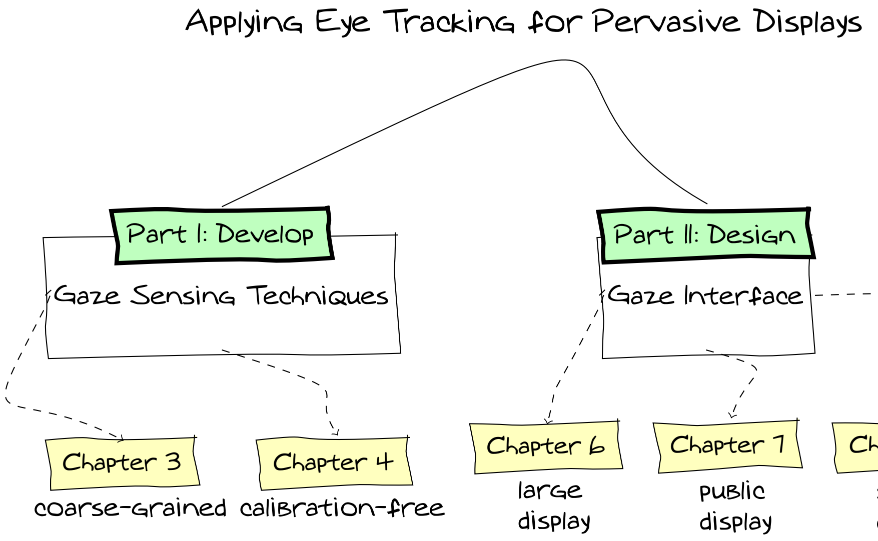


Figure 1.1: Overview of the thesis structure

To enable robust real-time eye tracking under uncontrolled conditions, we explore gaze sensing techniques with off-the-shelf web cameras, as they are widely available. We employ computer vision and machine learning techniques to develop novel gaze estimation methods (described in Chapter 3). The adopted approach is data driven. However, datasets for gaze estimation recorded with normal web cameras are limited [2]. To address this, we collect gaze datasets which cover di-

[2] Some datasets have also become available while we are conducting this research. These datasets include: UT Multiview [165], Eyediap [61], Columbia gaze data set [154] and MPI-IGaze [197].

verse users in naturalistic environments. We use machine learning techniques to develop gaze estimation methods that use the collected data for prior training. We then optimise our methods based on the datasets to achieve robustness across individuals. In everyday environments, systems sometimes have no prior information about the users. To address the challenge of prior training for individual users, we explore eye movement symmetry using facial features detected from the user's video images for person-independent eye tracking (described in Chapter 4). We achieve calibration-free horizontal gaze detection which require only a single RGB web camera. We further conduct lab experiments to evaluate the gaze detection accuracy and robustness of our methods. These techniques provide the foundation for developing gaze-interactive applications for pervasive display.

To address the design challenges of enabling gaze interaction in pervasive contexts, we need user interfaces that support spontaneous interaction, where users can simply walk up to a gaze-interactive display and start using it immediately. We therefore design new interaction techniques and interfaces that employ the proposed person-independent gaze sensing system. One restriction of the system is that it only supports the detection of coarse gaze directions. To exploit this, we propose gaze interactions that are based on display regions (described in Chapter 6). To understand user performance and subjective experience, we conduct lab studies to understand how people use our new gaze interfaces design.

Gaze interactive systems are usually evaluated in conditions where users are fully informed of the system's functionality and in controlled usability laboratories. To integrate eye tracking into pervasive displays, we aim to find out how to inform novice users to use a gaze-based interactive system without expert assistance. To address this, we deploy a novel gaze interface based on the proposed system in a public setting (described in Chapter 7). We conduct a series of field studies to gather information from passers-by to understand what guidance people need to comprehend the operation of the system. Finally, we integrate this guidance into the user interface and deploy the system in the wild. We observe users' behaviours in naturalistic environments. The knowledge gained from this experience provides a foundation for deploying eye-based technology for public displays.

Pervasive displays also support multi-user collaboration. In everyday life, there are many scenarios that people need to find information together, e.g., look for a parking lot on the map. This raises the open question of how we can apply gaze to improve co-located collaboration on a shared screen. To explore multi-user gaze interfaces, we select a collaborative search task where users look for information together (described in Chapter 8). We conduct lab studies to understand how sharing gaze information on a shared screen can improve users' performance, and we also find out people's subjective experience for gaze-assisted co-located collaboration.

1.4 Contributions

This dissertation makes the following contributions:

- Two novel person-independent vision systems that support low-fidelity gaze tracking in unconstrained environments. The systems are lightweight. They employ efficient computational algorithms that can process gaze estimation in real time. In addition, they need only a single monocular RGB webcam, and they are robust across diverse users. One of the systems enables remote calibration-free gaze sensing in dynamic real world conditions that are deployable in public spaces.

- The design of multiple novel gaze-based interaction techniques for pervasive displays. These techniques support spontaneous interactions where people can walk up to a display and control displayed information using only their eyes. The techniques enable intuitive, fast and hands-free interactions that are suitable for numerous scenarios (e.g., retail, exhibition and workplaces).
- The results and understanding from studies of multiple novel gaze interfaces for pervasive displays. These provide foundation for the rapid growing field of applying eye tracking in ubiquitous computing.
  - Our lab evaluations provide quantitative measurements on user performance and subjective feedback. We identify the challenges and limitations posed by these new form of gaze interfaces. These results are valuable for designers in early application development, as we cover a broad range of design aspects that the designers could refer to, such as system parameters, interface options, feedback mechanisms and user preferences.
  - We develop and evaluate the first walk-up-and-use gaze interface through three iterative field studies with over 190 users. Our in the wild deployment let us gain insights on users' natural behaviours and derive design considerations for applying our systems in public environments.
  - Our studies of using gaze for collaborative search show that visualising gaze on a shared display can enhance co-located collaboration. Our

results inform the challenges and considerations of designing multi-user gaze interfaces.

1.5 Structure of The Thesis

This thesis consists of two parts (illustrated in Figure 1.1): Part I focuses on the technical aspects of developing real-time gaze sensing systems; and Part II focuses on the design aspects of novel interaction techniques and interfaces.

Chapter 2 summarises state-of-the-art eye tracking techniques. The chapter covers existing commercial systems and methods proposed in research literature.

Chapter 3 describes the technical details of a novel vision-based algorithm that tracks coarse gaze directions using computer vision and machine learning techniques. This chapter is a revised version of two publications [199, 200].

Chapter 4 describes the technical details of a remote calibration-free system (called PCR) that is based on eye movement symmetry to track coarse gaze directions using vision techniques. This chapter is a revised version of [201, 202, 198].

Chapter 5 reviews existing gaze interfaces and interaction techniques.

Chapter 6 describes the design and evaluation of a novel interface (called Sideways) that allows spontaneous gaze interaction with large display. This chapter is a revised version of [201].

Chapter 7 describes the design and deployment of a novel interface (called GazeHorizon) that enables passers-by to interact with a public display using gaze. This chapter is a revised version of [204, 203].

Chapter 8 describes the exploration of multiple users sharing gaze on a display for co-located search collaboration.

Part I

Gaze Sensing Techniques

2

Background

Eye tracking is the process of tracking the motion of our eyes or the directions of our gaze in a visual scene (i.e. where a person is looking). In psychology and cognitive science, eye tracking has been demonstrated to be a useful method for studying human behaviours in daily activities [56, 191, 90], for example, reading and binocular vision training. Driven by lab research [80, 56, 191, 194, 90], the primary goals of eye tracking development are to achieve high speed and high accuracy in analysing eye movement patterns and what people are looking at. These goals are challenging because our eyes are never completely stable and, move rapidly [36, 149] and they have a very high visual acuity only at the fovea, which covers only a small area of the retina.

In recent years, tremendous engineering efforts have aimed to develop less intrusive eye trackers that allow free body movements [120, 38, 164, 59, 165, 60, 197]. In this chapter, we first introduce the basics of how the human visual system works.

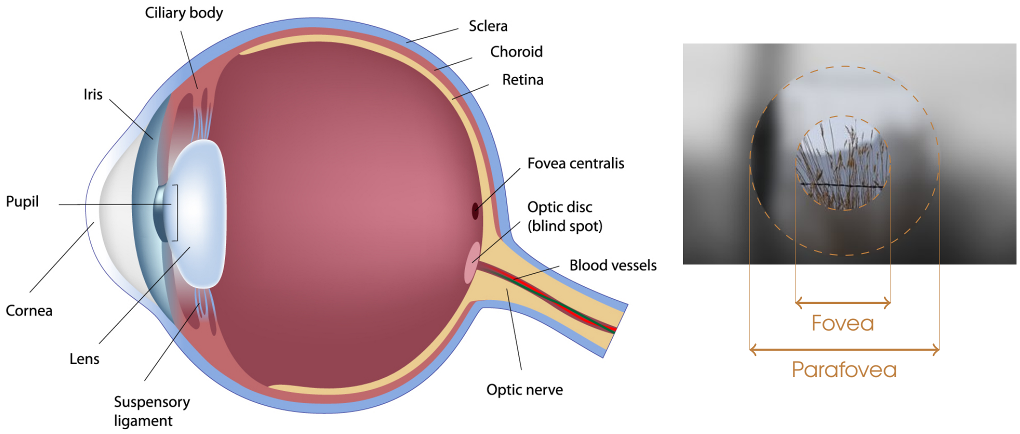


Figure 2.1: The human eye anatomy (Source: (2)) and fovea vision

Then, we review existing eye tracking devices and explain state-of-the-art systems. We then focus on the technical aspects of video-based eye tracking methods that are most relevant to this thesis. In particular, we describe works that have the same goal as our research, which aim to achieve robust eye tracking in unconstrained environments. At the end of this chapter, we identify open challenges of sensing gaze in pervasive contexts.

2.1 Human Vision

Our eyes are the perceptive organs of human vision. The retina is the light-sensitive inner layer at the back of the eye. It acts like an image sensor and converts optical images into signals (see Figure 2.1). The optic nerve then transmits these signals to the visual cortex that controls our sense of sight. Light focused by the cornea (which acts like a camera lens) reaches the retina. By automatically adjusting the size of the pupil, the iris of the eye controls the amount of light that can reach the back of the eye. The ciliary body of the eye controls the contraction of the ciliary muscle to adjust the lens. This helps the eye to automatically focus on near or far objects through a process called accommodation.

Human eyes have a fovea vision feature due to different light sensitivity at the retina as illustrated in the right picture of Figure 2.1. The fovea is the area at the retina with the eye's sharpest vision that is most sensitive to colour perception. The fovea area is about 1.5 mm wide (5° of visual angle) and provides the greatest acuity [3]. Foveola is located in the centre of fovea with a diameter of around 0.2 mm (1° of visual angle) [3]. The area on the retina further away from the fovea - peripheral vision - has lower resolution, and is weak at distinguishing colour and shape but is good at detecting motion.

2.2 Overview of Eye Tracking Systems

The developments of eye tracking devices have progressed tremendously within the last decade. There are multiple mechanisms to track eye movements and gaze positions for different applications, such as clinical and lab research, assistive technologies, etc. These methods can be categorised into the following three types: *eye attached* methods which apply special contact lenses to the eyeballs [144, 191, 92, 39], *Electrooculogram* (EOG) based techniques that measure the electric potential of the skin around the eyes [32] and *video-based* techniques that use optical sensors to measure eye movements and gaze without attaching anything on the users [74]. The first two types are usually used by specialists for clinic applications and laboratory research. Video-based eye trackers are widely used in many applications [74]. The majority of modern eye trackers have sampling rates ranging from 25 - 2000 Hz [16].

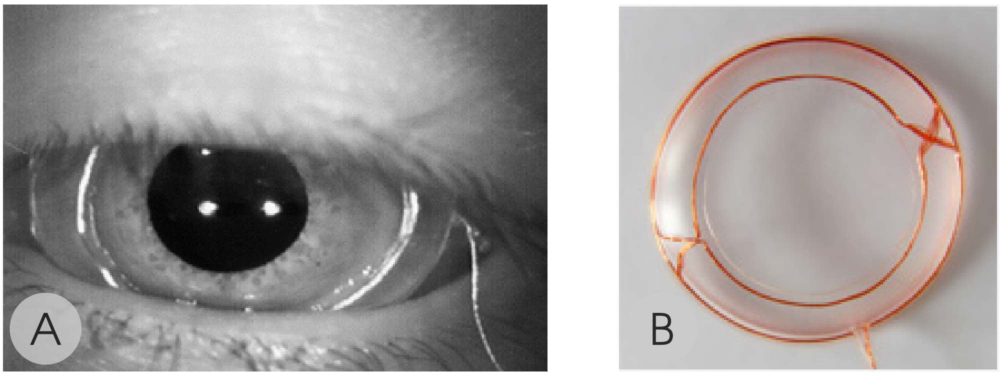


Figure 2.2: Search coil (A) A user is wearing a special contact lens for eye tracking (Source (4)) (B) A contact lens connected with a search coil (Source (5))

Eye Attached Eye Tracking

In this category, eye tracking is achieved by attaching a contact lens to the eye ball. The contact lens can be connected to an external device, for example small mirrors [191] or magnetic search coils [144, 92] (see Figure 2.2). The eye trackers with magnetic search coils require users to stay in a magnetic field. The eye movements can be recorded at high precision from the voltage generated in the coil on the lens. Although these methods provide high temporal and spatial accuracy, the devices are intrusive and eye tracking can only last for a short period of time before corneal swelling occurs. Recently, commercial companies are beginning to invest in eye tracking contact lenses without external connectors (e.g., Sony). For example, magnetic sensors can be placed on a video game console to track the location and polarisation of the magnetized contact lenses worn by the users [39].

EOG Eye Tracking

EOG techniques measure the resting potential of the retina in the eye. The basic principle is based on the fact that the human eye ball is polarised, with the front of the eye being positive and the back being negative (see Figure 2.3 (A)). Eye movements can cause changes in the electric fields. These changes generate EOG signals that can be recorded using electrodes. By recording this signal, the eye movements and gaze directions can be determined. The EOG signals are usually acquired with two pairs of electrodes, one pair for horizontal direction signals and the other pair for vertical direction (see Figure 2.3 (B)).

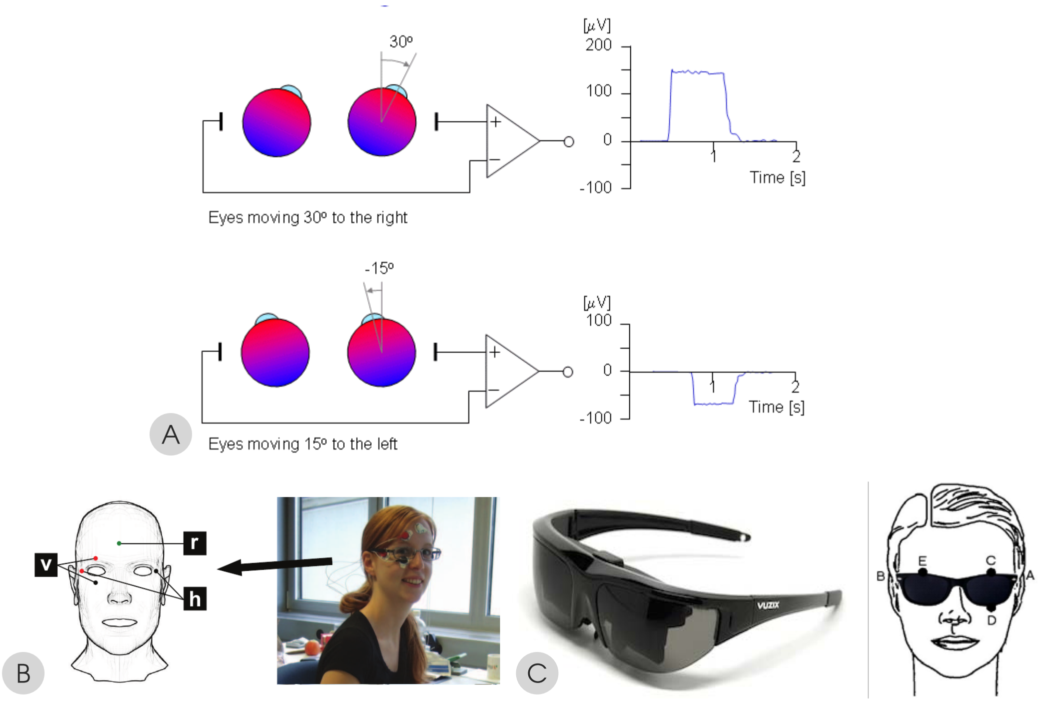


Figure 2.3: (A) EOG signals change during eye movements (Source [114]). (B) The placement of electrodes to record EOG signals (Source [32]). (C) An eyewear system integrates electrodes on to a commercially available smart glasses to capture the EOG signals (Source [21]).

EOG systems are primarily adopted for clinical applications because they provide high frequency eye tracking. Recently, some research prototypes [21] (see Figure 2.3 (C)) and some commercial products (e.g., JINS) have integrated EOG eye tracking into normal daywear glasses.

Video-based Eye Tracking

This type of eye tracker uses video cameras to track eye positions and analyse the video images to extract eye features, such as pupils and corneal reflections [74]. Most of these systems require additional illumination (e.g., infrared (IR) light) to make pupils easy to detect and to track.

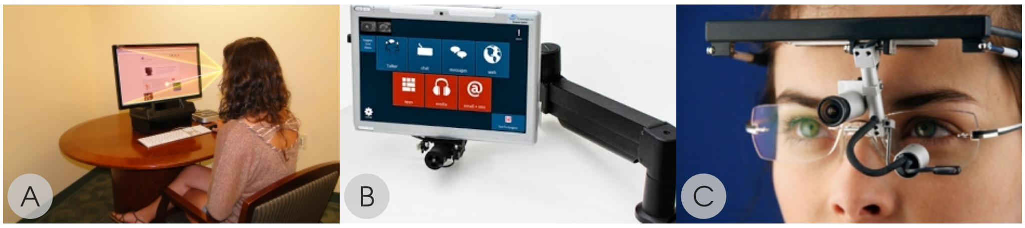


Figure 2.4: Commercial video-based eye trackers (source (6))

Both research and commercial eye trackers come in different forms: wearable and remote (see Figure 2.4). With remote eye trackers, camera sensors and light sources are placed at a distance from the user (see Figure 2.4 (A,B)). To track users' gaze positions on a screen (e.g., desktop monitors or tablets), they are required to directly face the eye tracking sensors [206, 72, 209, 166, 190, 118]. The sensors are usually fixed to a monitor and users need to maintain their position within the tracking range in front of the sensor (see [41] for comparison of tracking range in existing systems). With wearable eye trackers, users wear a headset and can move freely in 3D space (see Figure 2.4 (C)). Video cameras are attached to the headset to track eyes. These are combined with a scene camera to determine where people are looking at in the scene [84, 169]. In both types of video-based eye trackers, a calibration procedure is required before a new session starts.

2.3 Video-based Gaze Estimation Methods

The previous section gives an overview of the existing eye tracking systems. We now describe the general approaches of gaze estimation to achieve video-based eye tracking systems. Although this thesis mainly focuses on eye tracking using off-the-shelf components, we provide a general overview of existing methods that require and do not require specialised hardware.

Gaze estimation systems are capable of exclusively recording the foveal vision. As illustrated in Figure 2.5, the gaze direction - line of sight (LoS) - is defined as the direction connecting the fovea to a fixation point in the outside world. The line of sight is the approximation of the visual axis of the eye. The pupillary axis is the line perpendicular to the cornea that intersects the pupil centre, and is an approximation to the optical axis. Angle κ is the angle between the pupillary axis and visual axis [105].

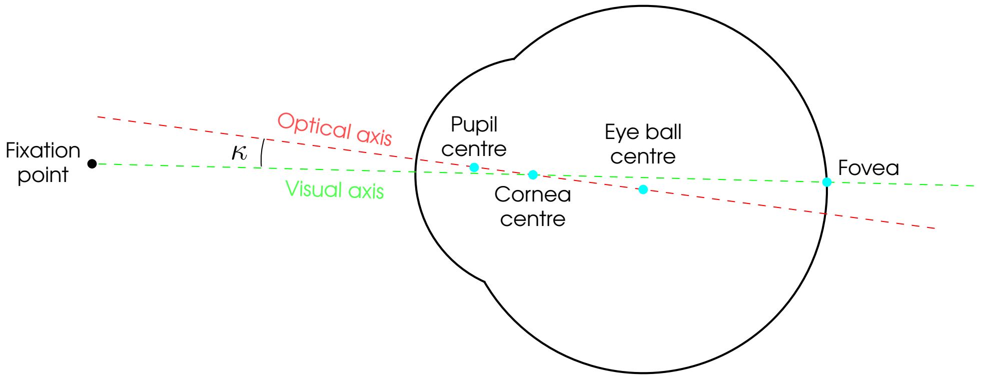


Figure 2.5: Gaze model based on eye anatomy.

Gaze estimation is used to determine where people are looking in the scene. It consists of two processes: gaze tracking (the direction of the gaze[1]) and eye

[1] The visual axis of the eye

movements tracking (the motion of the eye). There are two main types of video-based gaze estimation methods: *feature-based* and *appearance-based* [74]. The majority of existing eye tracking systems are feature-based and offer high accuracy and fine-grained detection of eye gaze. Appearance-based methods stem from the vision community and are still at early stage. These methods aim to achieve eye tracking using low-cost off-the-shelf components, e.g., web cameras without active illuminations.

2.3.1 Feature-based Methods

Feature-based approaches use explicit geometric features of the eye to estimate gaze direction and are based on local features and patterns of eye images, such as limbus contour, pupil centre, eye corners, and corneal reflections [74].

IR-based

The majority of existing IR-based feature gaze estimation methods are based on the anatomy and reflectivity of the human eye. When infrared light shines into the human's eye, reflections occur on the lens and cornea. The light that passes into the eye through the pupil is reflected off the back of the retina. As illustrated in Figure 2.6, pupil detections use the red-eye effect [194], i.e., the difference in reflection between the cornea and the pupil, to extract the pupil centre from infrared illuminated eye images.

An established principle in feature-based methods is to use infrared light for corneal reflections (glint), and gaze estimates are derived from techniques such as Pupil Centre Corneal Reflection (PCCR) techniques [121, 193, 25, 207, 120, 68,

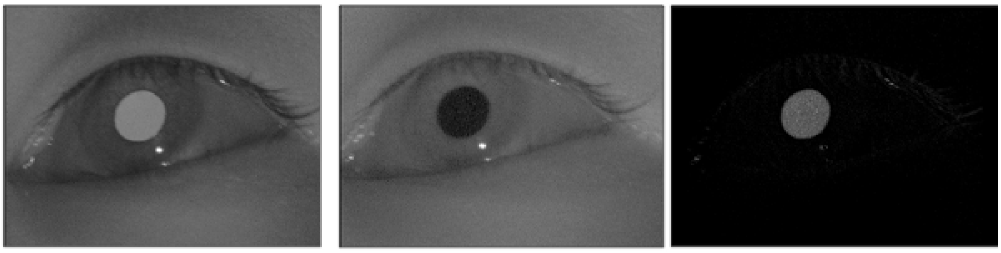


Figure 2.6: The bright and dark pupil response under IR illumination (Source (7))

209]. Corneal reflection serves as a static reference point for the moving pupil, because the position of the glint remains constant when the eye moves. Gaze direction is subsequently estimated by measuring the moving pupil centre of the eye and the corneal reflection (see Figure 2.7).

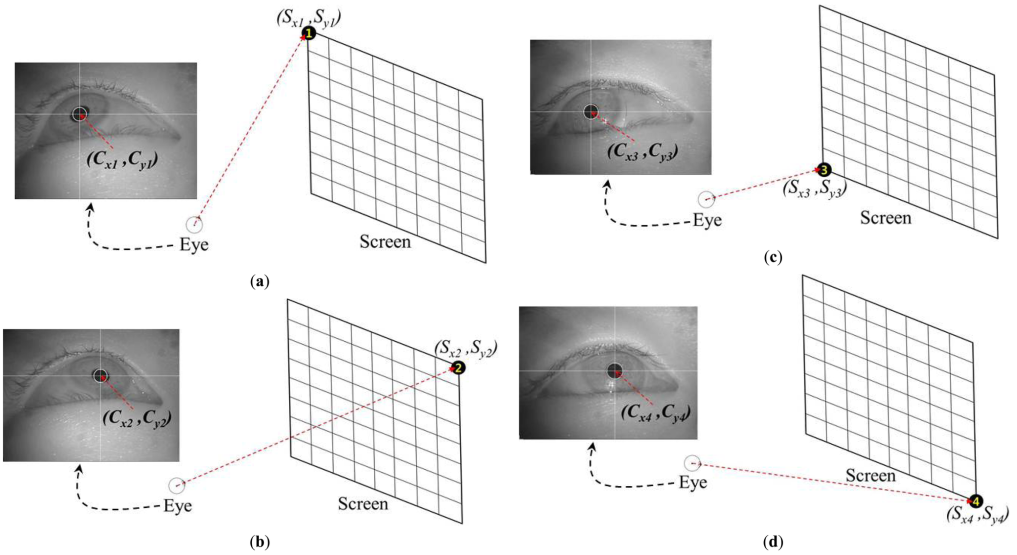


Figure 2.7: The Pupil Centre Corneal Reflection (PCCR) techniques (Source (107))

As the curvature at the edge of the cornea varies, the performance of the glint-based methods degrade when the gaze tracking range increases. For example, when people look at extreme angles the glint can be missed [102]. Some methods employing stereo approaches show potential to alleviate this problem [95, 102]. The approaches first reconstruct the position and the orientation of the pupil in a 3D-space. Gaze is then determined by the direction which connects the pupil centre and the eye centre.

The robustness of IR-based feature gaze estimation methods primarily depends on the intensity and the size of pupil response in the infrared images. However, various factors affect the robustness, e.g., ambient light, people's cognitive state, and individual differences [128]. One of the main effects is that the size of our eye pupil varies depending on the amount of light that goes through it [8]. In bright conditions the pupil gets smaller to let less light in and vice versa in dark conditions. These limitations make it hard to apply infrared-based methods for outdoor environments or uncontrolled conditions.

Many existing commercial video-based eye trackers require additional IR illumination and adopt glint-based methods. These systems can achieve high accuracy gaze tracking in controlled settings when users remain within the optimal tracking range. For example, some products report an accuracy of $< 0.5°$ with a highest temporal resolution of 300 Hz at a working distance between 55-75 cm[2].

Nautral Light

Previous works have investigated the use of feature-based gaze estimation methods with low resolution face images under natural lighting conditions (e.g. [206, 72, 190, 151]). While these approaches eliminate the use of IR illumination, gaze estimation becomes more challenging. Firstly, eye features (e.g. pupil centres) are less detectable in normal video eye images compared to the bright or dark pupil response in infrared images under active IR illumination. Secondly, the appearance of eye features varies in scale and with different users, and so detecting certain features reliably requires immense prior training data [190, 169, 60]. Con-

[2] `http://www.tobii.com/en/eye-tracking-research/global/`

sequently, the gain in accuracy of the approach is achieved at a significant cost of computational speed and flexibility.

2.3.2 Appearance-based Methods

Another common gaze estimation approach is to use computer vision techniques to analyse video eye images captured using normal RGB cameras [20, 167, 186, 130, 164, 200, 109, 59, 165, 110, 60, 197]. These methods are appearance-based and use mapping functions to estimate gaze directly from cropped eye image contents, without explicitly extracting features. As illustrated in Figure 2.8, the appearance of eyes from captured images is different when people look at different positions on the screen. The general principle of appearance-based methods is to map this high dimensional image input data to 2D or 3D gaze point.

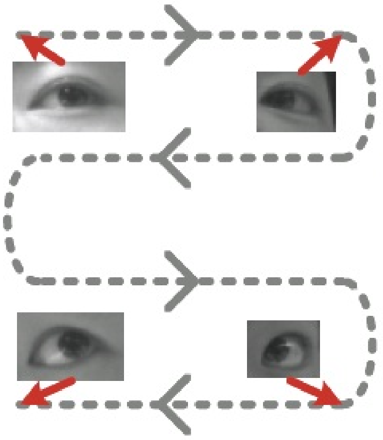


Figure 2.8: The appearance of eyes from captured images is different when people look at different screen positions. Appearance-based methods estimate gaze directly from cropped eye image contents, without explicitly extracting features (Source [109])

Calibration of the cameras is typically not required as gaze mapping is learned directly from raw image data [74]. Early appearance-based approaches are largely based on neural network methods in which the intensities of eye images are used to map directly to screen coordinates [20, 163, 188]. However, gaze estimation under natural lighting conditions is challenging due to low contrast images, which can contain multiple specular and diffuse components. In addition, the appearance of the eyes can change easily with different head poses. Other shortcomings of these approaches are that they require a large training data set and they are computationally intensive for high resolution images as increasing amounts of hidden nodes are required in the network.

To overcome the limitations posed by large amounts of training data, Williams [186] proposed a sparse, semi-supervised Gaussian process regression method to map input images to the gaze coordinates with partially labeled training samples. Basilio *et al.* [130] developed a calibration-free eye gaze detection system which is also based on gaussian processes. While they investigated techniques that estimate gaze from low resolution images without IR illumination, in real-world settings their approaches face considerable challenges from influences of head and body movements.

Recent efforts focus on improving robustness of head movements by utilising rich datasets that are either synthesised [165] or collected from a large amount of users under natural head movements [59, 197]. These works integrate the model of 3D head poses in the training of their approaches to improve robustness in unconstrained settings.

Compared to feature-based methods, appearance-based methods are less accurate (with gaze accuracy around 2° to 4°), are usually not robust to head movement, and many of them are processed offline rather execute in real-time. The advantage is that appearance-based methods require only a standard camera as sensor without the need of any special-purpose hardware. Hence, it is more flexible, easier to setup and less sensitive to influences from the environment.

2.3.3 Gaze Mapping and Calibration Procedure

In order to know where the user is looking on a computer screen, gaze estimation usually requires a calibration procedure to obtain gaze mapping. Gaze mapping is the process of deducting the function of descriptors from the eye images to inferred gaze. The majority of existing eye tracking methods require an explicit procedure for calibrating system parameters to fit individual users. During the calibration process, users are asked to follow or fixate on screen stimuli (usually nine points on the screen). The eye tracker then collects data to generate the underlying gaze mapping model for individual users. Various models have been proposed for obtaining an optimal mapping that achieves both accuracy and robustness under large head movements. The underlying gaze mapping methods can be classified as *2-D mapping-based* and *3-D model-based.*

Mapping-based approaches model the mapping relationship from eye features to gaze points. Polynomial interpolation is one of the most commonly used in gaze estimation [120, 37]. Corner-iris vectors and pupil-glint vectors are commonly employed in interpolation methods. A detailed description of polynomial interpolation methods based on different feature inputs can be found in Cerrolaza *et al.*'s paper [37]. Although polynomial interpolation method is easy, fast and simple to implement, it is usually not robust to large head movements. Other methods adopt a non-linear regression model to establish direct mapping from either raw pixels or image features to gaze points, i.e., Neural Networks [20, 208, 200], Gaussian Regression [186, 130, 164], Adaptive Linear Regression [110], Random Regression

Forest [165] and Multimodal Convolutional Neural Networks [197].

Model-based mappings model the 3D physical structures of the human eye geometrically to calculate 3D gaze direction [209, 74]. The gaze direction is expressed as a function of the 3D configuration of the system and the user, based on geometry and eye physiology. Therefore, a calibration process is required to obtain user-dependent parameters, e.g., cornea curvature, the angular deviation of the visual axis from the optical axis (see Figure 2.5). Model-based mapping methods can accommodate natural head movements, but they generally require complex system setup. It is essential to calibrate the eye tracking hardware, e.g., the camera parameters or the position between the screen, camera and light sources.

2.4 Eye Tracking in Unconstrained Environments

The objective of this work is to enable robust gaze sensing under uncontrolled conditions when no prior user information is known. We aim to develop systems that do not require additional illumination, are flexible to set up and are deployable.

Gaze estimation in unconstrained environments has been a challenging problem in eye tracking research. Previous works investigate different aspects of overcoming the difficulty of tracking gaze in the wild. The main objectives of these approaches are eliminating the requirements of IR illumination to improve robustness under different lighting conditions, using off-the-shelf components to make eye tracking easy to set up and widely available, easing or removing the calibration procedure to improve usability and increasing tracking range and robustness to allow free user movements.

Hansen and Pece [72] propose an active counter tracker which is based on image statistics. Their method requires only off-the-shelf components and are robust to light changes and camera defocusing, however, to estimate gaze, a calibration with small sets of points is required before each use. As identified in previous work, the calibration process to obtain gaze mapping takes time and hinders eye tracking applications in daily life settings [120]. To overcome this challenge, various research works propose the use of machine learning to achieve person-independence [130, 154]. These methods learn a generic person-independent model from a large number of eye images collected from different users. Other works propose the use of a visual saliency map to achieve calibration free gaze estimation [164, 38]. The calibration procedure is implicitly integrated from a person watching a video clip. The idea is based on treating saliency maps of the video frames as the probability of gaze distributions, but their setup requires a chin rest [164] or active illumination [38].

Other works attempt to explore eye tracking in mobile contexts [118, 46]. The EyePhone system [118] tracks users' eye movements using a camera mounted on the front of the phone. It tracks users' gaze position on the mobile phone display and allows users to activate an application via eye blinks. To estimate where a user is looking, the system requires calibration prior to each use. In other mobile contexts, prior research proposes wearable eye tracking systems for daily life logging [84, 168] and monitoring high-level contextual cues [33].

Another type of system is specifically designed to detect people's attention towards an object (direct gaze) [174, 154, 49]. Some of these systems use an infrared (IR) tag and are commercially available (e.g., eyebox2 [9]). The eye contact with the objects embedded with the IR tag can be estimated by determining whether the reflection of the tag on the user's eye appears central to the pupil. Direct gaze detection can also be achieved using machine learning methods with a normal web camera, such as cameras integrated on mobile tablets [154]. However, one restriction of this approach is that each object needs to be fit with a tracking sensor.

Recently, researchers attempted to bring eye tracking for large displays, e.g., smart TV in the living room [77, 106], digital boards in shopping malls [126, 153, 146]. These works show promising results of gaze tracking in less controlled environments; however, the majority still require specialised hardware and additional illumination and user calibration [146, 77, 106]. Other systems use only an off-the-shelf video camera and infer users' gaze from their head poses [126, 153].

2.5 Open Questions and Opportunities

Eye trackers have a long history of application in lab research. For most lab applications, high accuracy and sampling rate are required for eye movements analysis. As a result, work in eye tracking generally aims to achieve best possible accuracy (e.g., [207, 186]) with hardware optimized for the task and careful calibration to the individual user. These systems require specialized hardware including high resolution video cameras and infrared illumination. Despite considerable advances in tracking accuracy and speed, most video-based eye trackers are still stationary and restrict free movements of the user's head and body. The trade-off of high accuracy and high sampling rate results in constraining natural human movements which prevent eye tracking systems from being widely adopted for daily usage.

To use eye tracking in daily life applications, existing IR-based eye tracking methods face considerable challenges in unconstrained environments. For example, robustness is largely influenced by ambient light sources and sunlight in outdoor environments. Secondly, the hardware and the calibration procedure are tedious to setup. Daily life settings require eye trackers that are easy to setup and adaptable for new users. The aim of this work is to design eye tracking systems that supports walk up and use applications. The techniques should be robust across different users. In addition, it should be lightweight and able to run in real-time.

With the wide adoption of video cameras, computer vision and machine learning have played an important role in developing vision systems for daily life applications. We have seen tremendous improvements in robustness and efficiency of vision systems that run in real time. This progress enables ample applications that employ vision-based systems (e.g., body movement tracking, face recognition). To address the challenges of enabling robust gaze sensing in pervasive contexts, we apply the knowledge in computer vision and machine learning to advance eye tracking techniques in unconstrained environments.

3

Exploring Appearance-based Method

Computer vision and machine learning have made great progress in recent years in terms of accuracy and real-time performance. This holds promise to investigate how to apply the knowledge in these fields to advance eye tracking techniques. Work in eye tracking generally aims to achieve best possible accuracy with hardware optimised for the task. One of the research problems investigated in this work is to achieve eye gaze estimation with pervasive equipment, such as web cameras we might find mounted on large displays. The motivation is to use equipment that we expect to find in our environments for eye gaze interaction. As cameras in our everyday environments are not optimised for gaze tracking, the challenge is to advance methods that are capable of estimating gaze using eye images captured with real-world constraints.

In this chapter, we introduce a novel gaze estimation technique, which is adaptable for person-independent applications. This technique is data driven and employs computer vision and machine learning techniques. We first describe the details of our approach and the data collection process. Finally, we report an evaluation of the robustness of the proposed method with respect to a diverse user population and discuss the application of our approach in pervasive contexts.

3.1 Motivation

In computer vision, researchers have investigated a large variety of image features for applications such as object detection and tracking or image segmentation. Common features describe intensity and colour of image pixels or are based on filter responses. These feature types have been widely used in computer vision but haven't been extensively explored in eye tracking research.

The proposed approach for processing eye images is *appearance-based.* Appearance-based approaches work under normal illumination and directly infer gaze from video images. Previous works have shown the potential of this approach for estimating gaze from low resolution images, under laboratory conditions [20, 188, 186, 166]. These works are based on raw pixel images - by representing images as input vectors with raw pixel values for machine learning. Williams *et al.* combine the use of steerable filters on eye images and raw pixel data to achieve better accuracy for regression-based gaze estimation [186].

As described in Section 2.3.2, image intensity provides powerful features that are widely used in appearance-based gaze estimation methods (see [20, 163, 188, 186] for examples). Similar to intensities, the colour distribution in the eye region is different for different gaze directions. While RGB histograms are often used to represent image colour information, the RGB colour system depends on image characteristics. Gevers *et al.* [64] found that while converting to a colour invariant system such as normalized *rgb* the colour model is less sensitive to illumination and object pose. A survey by Hansen *et al.* [74] showed that applying different filters on images will result in enhancing particular image characteristics while suppressing others. Daugman *et al.* [47] showed that a set of Gabor filters with different frequencies and orientations can be used for iris recognition. Williams *et al.* [186] applied steerable filters to the eye images for gaze estimation.

This chapter introduces a method based on extraction of low-level features from images of the eye. Through image transformation into feature space, we are encoding information such as texture and edges in a feature vector that represents an image more compactly (i.e. with reduced dimensions) than a raw pixel image. The feature vector serves as input for a two-layer regression neural network that produces gaze estimates. The neural network requires a priori training to learn the relationship between image features and gaze direction.

3.2 Gaze estimation using Image Features

Typical appearance-based approaches use the entire eye image for gaze estimation. Among them, approaches that use raw pixel values can only represent the raw appearance information (colour/intensity) of the overall image. However, in unconstrained environments, the raw colour and intensity pixels vary largely under different light conditions and head poses. These values also differ across users due to the diversity of eye appearance. By transforming the image into another feature space allows us to encode information such as texture, edges, etc., and potentially reduce the data dimension.

Feature	Description
Colour (C)	f_C extracts the red-green (RG) and blue-yellow (BY) colour.
Intensities (I)	f_I extracts the grey scale intensities.
Orientation (O)	f_O is obtained by convolving the intensity image with a set of Gabor filters in four orientations $\{0°,45°,90°,135°\}$.
Haar (H)	f_H represents Haar features using two rectangular patterns which extracts local borders.
Spatiogram (S)	f_S encodes RGB colour histogram and their spatial distribution.

Table 3.1: The five types of low-level image features used in this work.

3.2.1 Feature Extraction and Selection

In this approach, five different types of features, namely *colour*, *intensities*, *orientation*, *haar-like* features as well as *spatiogram* (see Table 3.1) were adopted. These features were chosen because of their low computational complexity as well as their prevalent application in computer vision.

For calculating colour f_C, intensities f_I, and orientation f_O features, the system used the method from Walther and Koch [182]. The input image I was processed for low-level features at multiple scales, and centre-surround differences were computed. Three individual feature vectors $f_{K\in\{C,I,O\}} = \{f^{\{i\}}\}_{i=1}^{1200}$ were extracted to represent I in colour, intensities and orientation feature space respectively.

Haar-like features have a low computational cost and provide local edge information of an image [177]. They consist of a set of simple rectangular features.

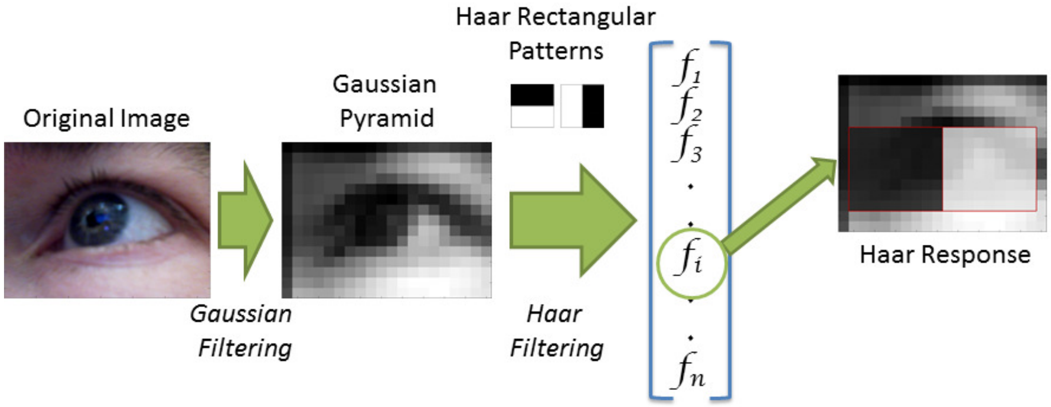


Figure 3.1: Extraction of haar-like features. The image was first normalised and then sub-sampled into a Gaussian pyramid by convolution with a 3x3 Gaussian smoothing filter and decimation by a factor of two. Two rectangular patterns were used which detect where the border lies between a dark region and a light region, horizontally and vertically. Each value in the resulting Haar feature vector f_H is a response of the rectangular patterns at a certain position and scale in the image.

Rectangles can be placed at any position and scale within the original image. The sum of the pixels which lie within the white rectangles is subtracted from the sum of pixels in the dark rectangles (see Figure 3.1). Each feature type can indicate the existence of edges or changes in texture.

While human look at different directions, the colour distributions and shape configuration between the pupil and the sclera of the user's eye appeared in images from the camera change. To represent pixel values and spatial distribution of colours in an image simultaneously, we employed the spatio-histogram (spatiogram) [26]. It expresses local colour patches over entire image. This allows us to encode object information about the texture and shape, as well as the spatial relationships between the pixels, such as the average locations of different colour patches. Given an input RGB image I, let $H^{\{k\}}$ represent the total number of pixels in the k_{th} bin of an ordinary colour histogram. We define the mean of the x,y coordinates of all pixels in the k_{th} bin as $C_x^{\{k\}}$ and $C_y^{\{k\}}$. The spatiogram of the image is a vector $f_S = \{H^{\{k\}}, C_x^{\{k\}}, C_y^{\{k\}}\}_{k=(0,0,0)}^{(24,24,24)}$ where the quantisation level was fixed to $M = 25$ for each colour channel in this work.

Five feature vectors $f_{K\in\{C,I,O,H,S\}}$ were computed from each image I in the database. To yield a fast and efficient gaze estimation, a feature selection procedure was followed instead of directly using the high-dimensional raw vector as input to the neural network. By selectively choosing the essential elements in the feature vector, we can reduce the input dimension and minimise redundancy, while maximising features' relevance. We employed mRMR (minimum Redundancy Maximum Relevance [134]) feature selection on the original feature vector extracted from I, thus reducing the high dimensional image data I into a low dimensional feature vector $z_{K\in\{C,I,O,H,S\}} = \{f^{\{i\}}\}_{i=1}^{m}$ where m $=$ 50.

3.2.2 Gaze Estimation Using Regression Neural Network

Gaze estimation was performed by mapping extracted feature vector $z = \{f^{\{i\}}\}_{i=1}^{m}$ from image I to output gaze location $g = (x, y)$ of the user's. Using raw eye image pixels as the input to a 3 layer feed-forward Artificial Neural Network (ANN) for gaze estimation has been proposed in previous work [20, 188]. In this study, the dimension-reduced feature vector $z = \{f^{\{i\}}\}_{i=1}^{m}$ extracted from raw image was supplied as the input to the Regression Neural Network (RNN). The input vector was first normalised to ensure all input features were in the same data range. We adopted a feed-forward neural network model using a 2-layer perceptron with linear output unit activation function to learn the gaze mapping function $g(z)$. This is a two-layer network where the first layer has $tanh()$ unit and the second layer is linear. The RNN was trained on a set of labelled eye image/gaze coordinates pairs by minimising a sum-of-squares error function using the scaled conjugate gradient

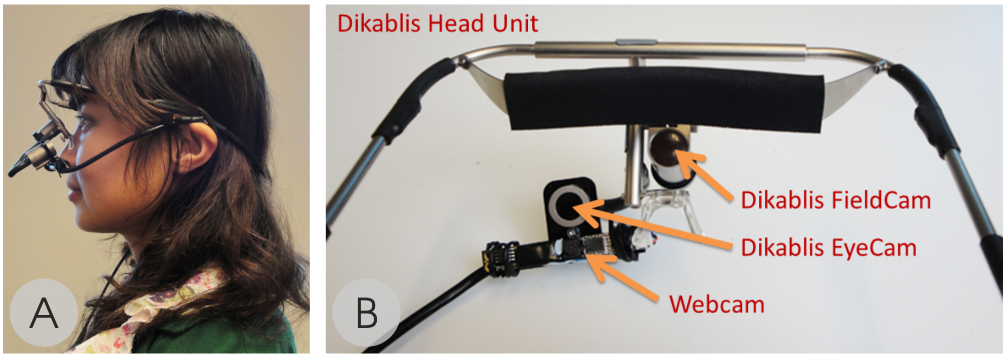


Figure 3.2: (A) Participant wearing the Dikablis eye tracker. (B) The additional webcam is mounted on the head unit to record close-up left eye images.

optimizer. The number of hidden units was decided by averaging the input and output units size.

3.3 Experiment

We conducted an experiment to collect a data set of naturalistic eye images and to evaluate the performance and accuracy of our gaze estimation technique. 17 participants (five female, 12 male), aged between 18 and 40 years (mean=26.9±6.8) took part in the study. We took particular care to include participants of different ethnicities, with different eye lashes and pupil colours. Specifically, we had nine participants with dark (i.e. brown or black) and eight with bright eyes (i.e. green or blue)). None of the participants wore glasses, but two wore contact lenses during the experiment.

We used a standard webcam (Microdia Sonix USB 2.0) with a resolution of 640x480 pixels and a frame rate of 30Hz. In addition, we used a Dikablis eye tracker from Ergoneers GmbH for collecting gaze data (see Figure 8.2). The webcam was mounted to the eye tracker on a plastic frame attached to the head unit. The camera recorded images of the participant's left eye.

The experiment took place in a real office environment with fluorescent illumination. Participants were seated about 60cm away from a 23-inch LCD monitor with visual angles of 43° horizontal and 27.6° vertical (see Figure 3.3). Free movements of the head and the upper body were allowed at any time but we encouraged the participants to move as little as possible during the experiment.

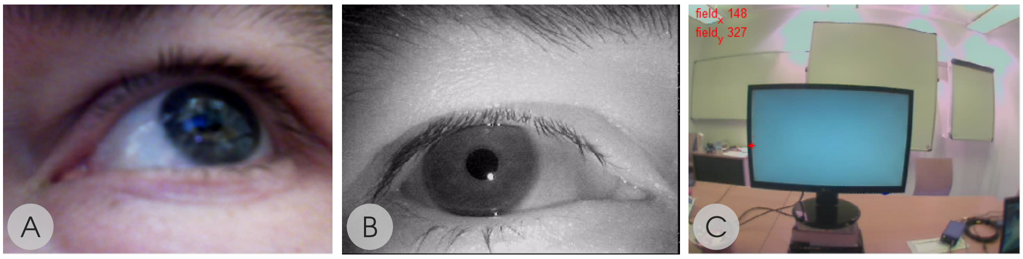


Figure 3.3: Examples of recorded data. (A) Eye image from the webcam. (B) Eye image from the Dikablis eye camera. (C) Scene image from the Dikablis field camera. The point of gaze obtained from the Diskablis eye tracker was plotted in red on the scene image.

The visual stimulus consisted of a red dot with a radius of 20 pixels (i.e. 0.5° of visual angle) shown in front of a light grey background. The system guided each participant through a sequence of 13 different predefined gaze locations. Participants were asked to fixate on the red dot at each location for five seconds (see Figure 3.4). After five seconds, the stimulus was shown at the next location. Thereafter, the participant was asked to follow a moving stimulus along several predefined paths. For each path, the stimulus moved horizontally, vertically and diagonally at constant speed. The entire procedure was performed three times. While the data was recorded, the system labeled the recorded images according to the stimulus's gaze point on the screen (see Figure 3.3 (C)).

We define the gaze direction by two rotation angles: Θ_h and Θ_v, for horizontal and vertical directions, respectively. The origin, $(\Theta_h, \Theta_v) = (0, 0)$, is the eye position when the gaze direction is perpendicular to the screen surface. Each

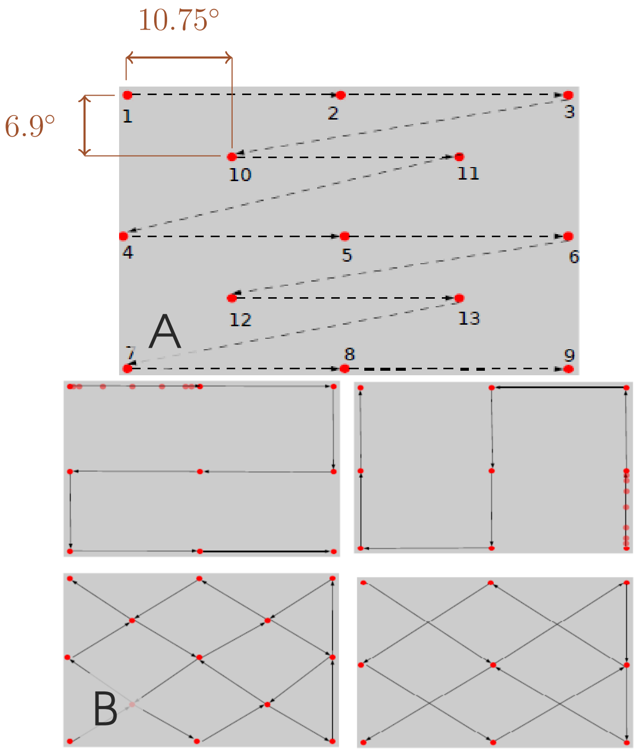


Figure 3.4: Experimental stimulus. (A) A red point is displayed on the screen. It stays at each location for 5 seconds and moves to the next location. (B) The point moves horizontally, vertically and diagonally at constant speed.

direction of gaze $(\Theta_h,\ \Theta_v)$ corresponds to only one gaze point $g = (x, y)$ on the screen. d denotes the distance of the users from the monitor. Given an estimation $g' = (x', y')$, the angular error was calculated by:

$$(\Delta\Theta_h, \Delta\Theta_v) = (|tan^{-1}(\frac{x-x'}{d})|, |tan^{-1}(\frac{y-y'}{d})|) \tag{3.1}$$

3.4 Evaluation

Around 1900 images/gaze coordinates pairs were collected for each participant. We first evaluated our system using a person-dependent evaluation scheme. For each participant 70% of the images were randomly selected for training (the “training set”); the remaining 30% (i.e. the test set) were used for gaze estimation on the

	Feature	**Horizontal [°]**	**Vertical [°]**
Individual	Colour (C)	4.21 ± 0.56	2.41 ± 0.69
	Intensities (I)	4.03 ± 0.65	1.89 ± 0.62
	Orientation (O)	4.02 ± 0.65	2.04 ± 0.57
	Haar (H)	3.73 ± 0.56	1.52 ± 0.49
	Spatiogram (S)	5.48 ± 0.51	1.97 ± 0.53
Combined	All	3.44 ± 0.48	1.37 ± 0.40

Table 3.2: Person-dependent mean and standard deviation of the gaze estimation angular error in horizontal and vertical direction averaged over the 17 participants for different feature types.

	Feature	**Horizontal [°]**	**Vertical [°]**
Individual	Colour (C)	11.29 ± 3.10	9.11 ± 1.59
	Intensities (I)	10.59 ± 2.71	8.26 ± 1.44
	Orientation (O)	10.59 ± 2.47	8.95 ± 1.03
	Haar (H)	15.37 ± 3.86	11.21 ± 1.77
	Spatiogram (S)	15.66 ± 1.82	9.17 ± 2.39
Combined	All	13.89 ± 3.84	8.63 ± 2.58

Table 3.3: Person-independent mean and standard deviation of the gaze estimation angular error in horizontal and vertical direction averaged over the 17 participants for different feature types.

same participant. The random splits of training and testing data were conducted 5 times for each participant.

Table 3.2 shows the average results for the different image features (cf. Table 3.1). The errors were calculated under the assumption that the participant looked at the centre of the stimulus on the screen. Using person-dependent evaluation the system achieved an average gaze estimation angular error of 3.44° horizontally and 1.37° vertically (1° corresponds to about $1.1cm$ on the screen plane). Figure 3.5 illustrates the angular errors for each participant using the five types of image features individually and all combined. Figure 3.6 shows the error distribution of each instance in both horizontal and vertical directions.

To test the algorithm's robustness across different people we further performed a leave-one-person-out cross-validation. In this scheme, gaze data of 16 partic-

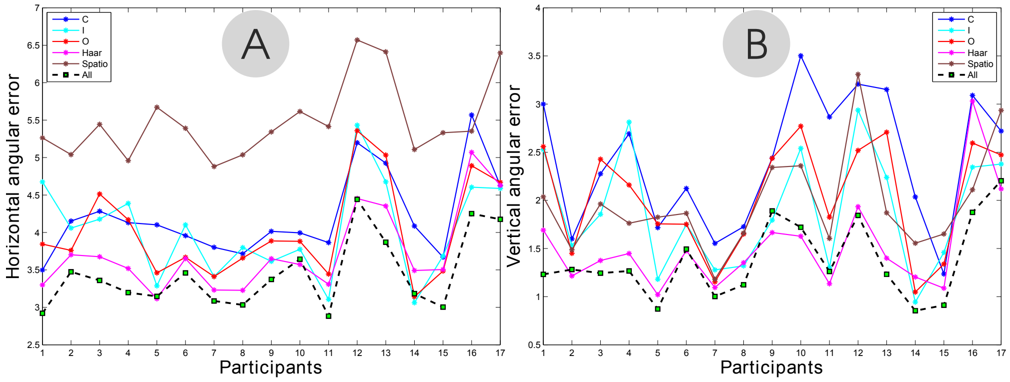


Figure 3.5: Gaze estimation angular errors of 17 participants in horizontal (A) and vertical (B) directions. The sold lines represent angular errors using the five types of features individually, while the dashed lines illustrate the errors when using all the features combined.

ipants was used for training the regression neural network and the data of the remaining person was used for testing. This was performed repeatedly over all 17 participants. The resulting gaze estimation performance averaged across all iterations and participants are summarised in Table 3.3.

Our data set includes outliers (e.g. image blurring and blinks), noises (e.g. reflection from bright objects, such as the monitor), as well as human errors (e.g. a participant failed to follow the stimulus). Despite these challenges our results show that estimating gaze with a small set of low-level image features from webcam images is feasible. A survey by Hansen and Ji identified that the accuracy of existing systems using web camera without any additional illumination is 2-4 ° [74]. Our system achieved similar accuracy with a user-dependent setup. Furthermore, the angular error was influenced by squinting (occlusions by eyelids) and frequent blinking. In our post-study analysis, we observed that participant 12 blinked frequently and participant 16 squinted his eyes (see Figure 3.5). A blinking detection method can be developed to increase the robustness of the system. The stimulus spreads less in vertical direction (10.75 ° horizontally, 6.9 ° vertically) which result in less error vertically. This suggests that our system can be improved by using more spatially closed training points.

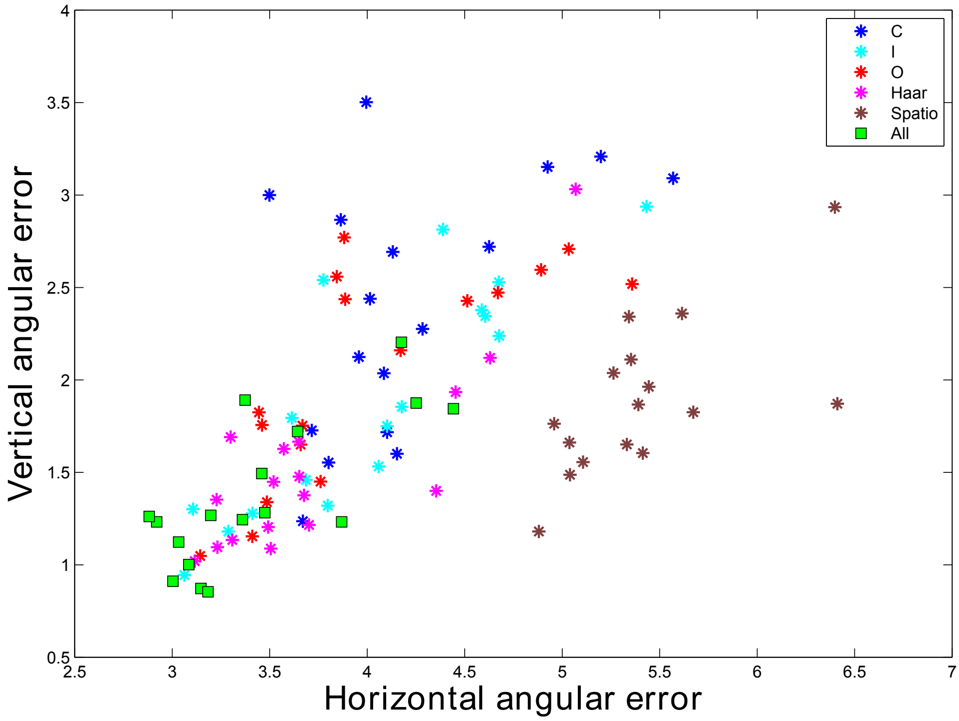


Figure 3.6: The error distribution of each instance in both horizontal and vertical directions.

Figure 3.6 shows that although several individual features achieve better accuracy, overall the best performance is achieved by combining all features. Among the individual features, colour, intensities, and orientation perform similarly, while spatiogram performs worst in horizontal direction. Haar features achieve better performance in person-dependent than in person-independent evaluation. This suggests Haar features are sensitive to eye appearance variance.

3.5 Discussion

This chapter presents a novel appearance-based technique that uses a small set of low-level image features and regression neural network for gaze estimation. This technique has the benefits of no calibration, non-intrusiveness and adaptability to new users. Results from a 17-participant user study show that the technique is robust across users with diverse characteristics and achieve decent performance for discrete gaze estimation.

Our dataset covers a large variance of eye appearance from people with different ages, gender and races. Consequently, person-independent gaze estimation shows higher angular errors than person-dependent. Although eye appearance differs across people, by using machine learning, our method learned common features which are effective in estimating gaze. Without re-calibration our method is able to provide sufficient accuracy to distinguish different areas of the monitor.

The results on different visual feature analysis show that it is difficult to select a single best feature set for different users. We plan to investigate other features that are potentially robust to different image scales (varying distance from eye to camera) and lighting variance, as well as methods to optimally select features for different applications. In addition, further improvements include adopting advanced gaze regression methods.

The optimisation of the method rely on the collected datasets. Although we include a variety of users' in the data collection, it is not clear how the method perform with unknown eye appearance. Especially in pervasive contexts, people might wear different glasses and their eyes might be occluded by eye lashes. This could potentially pose challenges to apply the system in scenarios where no prior training data from users can be obtained.

Another aspect that we didn't consider is the condition where users move their heads freely. Our data is collected using camera attached on a wearable headset.

The camera moves along with the headset as users turn their heads. This could potentially influence the performance of the method on a remote setup where camera sensors are fit on the screen and away from the user. Also, in mobile settings, other challenge are changes in the geometry relationship between the actual plane of visual gaze. To address this issue, future work can include prior geometric information in the learning process.

This method has the limitation that it does not provide fine grained gaze position estimations. The person-dependent eye tracking achieves an accuracy of 3.44 degrees, which is about one tenth of the screen size in a desktop setting. For person-independent eye tracking, the accuracy is about one third of the screen size. Although the accuracy is not comparable to existing commercial eye trackers, this method could be useful for applications that do not require selections of small targets, such as scrolling content.

3.6 Conclusion

This work shows that by applying knowledge from the field of computer vision and machine learning it is possible to design eye tracking applications using off-the-shelf web-cameras. The method has the flexibility of easy setup and do not need dedicated hardware and infrared lights. These initial results are promising and open up interesting applications in pervasive eye tracking. Example usage of the method can be in home or public environments. Accurate eye tracking might not be necessary while users might require a system that is easy to setup.

4

Calibration-free Remote Eye Tracking

The objective of this work is to enable robust remote gaze sensing, so people can interact with pervasive displays spontaneously without any preparation. Our goal is to have a gaze sensing system that is widely deployable, non-intrusive, and can accommodate unknown users. This chapter addresses the problem of sensing gaze using off-the-shelf components that require no user calibration.

Many existing eye tracking systems aim to offer fine-grained detection of eye movements but are primarily designed for lab research. These systems employ specialised hardware that requires additional illumination (e.g. infrared) and constrains users' head movement for optimal performance [86]. However, they face considerable challenges in pervasive contexts, as it is impractical to control external settings, such as lighting, users' position and their body movements. In addition, these systems require an explicit calibration procedure for individual users before each use. The calibration process hinders eye tracking applications in daily life settings, as it is often cumbersome and unnatural [38, 120].

Calibration-free eye tracking using off-the-shelf cameras remains a challenge. Previous works proposed to integrate the calibration procedures into normal activities (e.g., watch videos) instead of being a separate procedure [164, 38]. These methods employ visual saliency from monocular images for auto-calibration, but they require users to watch videos before each use, which is inherently an implicit calibration. Alternative solution is to use machine learning techniques to construct a generic person-independent model [130, 154]. As we have investigated in the previous chapter, these approaches are data driven and use supervised learning. Their performance relies heavily on sufficient labelled training data that can represent the general population. However, obtaining such datasets is highly challenging and costly.

This chapter introduces a calibration-free solution that requires no specialised hardware, works across different users, and is suitable for deployment in uncontrolled environments. We present a geometric feature-based approach to address the challenge of requiring no prior training of individual user. To achieve this, we exploit the characteristics of geometric eye features similarity across different users. We define *Pupil-Canthi-Ratio* (PCR), a novel feature for tracking horizontal eye movements based on the symmetry of our eyes. Specially, our method uses the inner eye corner as a reference point, and it calculates a displacement vector from the moving pupil centre. PCR is calculated as a ratio of the displacement vectors of both eyes (see Figure 4.1). The changes of PCR describe the degree

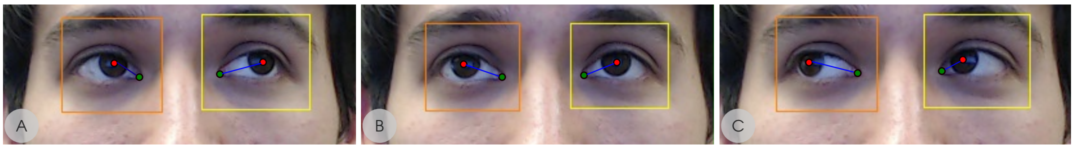


Figure 4.1: Eye symmetry: our method exploits the symmetry of eye movements. (B) With 'looking straight ahead' as a default state, the pupil centres of both eyes will be similarly distant from the respective inner eye corners. (A) and (C) When looking away from the centre, one eye will move further away from the inner eye corner while the other eye will move closer to it.

of users' eye movements towards left or right, and we leverage this characteristic for detecting horizontal gaze directions person-independently. The PCR feature is lightweight and can be obtained in real time by extracting eye features from users' video images. To evaluate performance, we compare three regression methods for gaze estimation using the PCR feature to compute its gaze mapping. To further understand the robustness of the PCR features, we test its performance against different eye images from diverse users, noises from feature detection, and variations in image scales.

4.1 Gaze Tracking using Eye Movement Symmetry

In this chapter, we present a remote calibration-free gaze tracking system using a single web camera. The method is based on the symmetry of eye movements (illustrated in Figure 4.1). We consider both left and right eyes at the same time. Similar to other feature-based methods, an analytical gaze estimation algorithm employs at least one moving point on the eye ball and one stationary facial reference point to compute a gaze direction. We choose the inner eye corners as the stationary reference points because it is the most stable feature in a face and relatively insensitive to facial expressions [206], and the iris centre as the moving point on the eye ball.

The system requires only a single off-the-shelf RGB video camera. The system takes video frames as its input and analyses them in real time using a sequence of image processing steps.

4.1.1 Eye Features Extraction from Colour Images

In the following paragraphs, we describe the image processing pipelines to obtain the pupil centre and inner eye corner features.

Eye Region Extraction

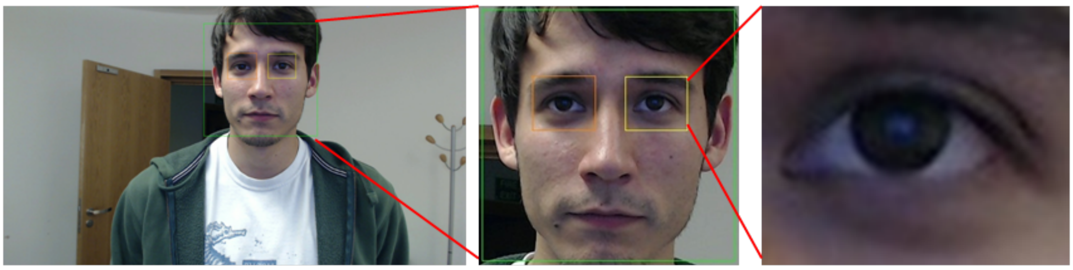


Figure 4.2: Extraction of eye regions: we apply the camshift algorithm for face tracking. The eye detectors are only run at the top third of the face region.

The images from the video camera are subject to variable lighting conditions, such as shadows, bright lights, low contrast, motion blur or noise. To tackle these problems, we first apply a bilateral smoothing filter to smooth continuous image regions while preserving and enhancing the contrast at sharp image intensity gradients. After the image is pre-processed, we use a Viola-Jones face detector provided by the OpenCV library[1] to detect the user's face and eyes in real time [177] (see Figure 4.2). The face detector identifies a rectangular area of the largest face in the scene. In order to improve performance and reduce the computational cost, we only search for eyes in the top half region of the face.

[1] http://opencv.org

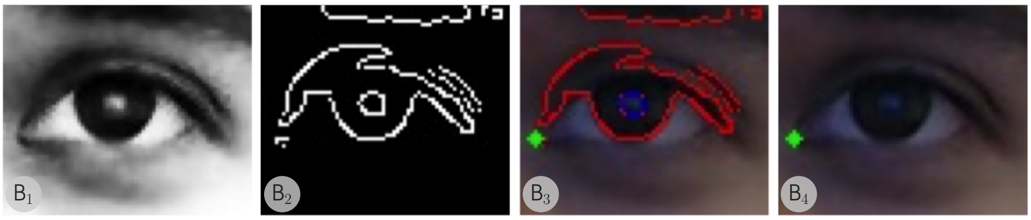


Figure 4.3: Detection of the inner eye corners: our system first applies a Harris corner detection to each eye image patch separately and further applies a Canny edge detector. The (B_3)&(B_4) eye corner is obtained by considering only those candidate points that lie on the detected eyelid edges.

We use the $haarcascade\_frontalface\_alt2.xml$ haar cascade for detecting the presence of frontal faces with a minimum size of 150x150 pixels. Face tracking is performed on consecutive video frames using the camshift algorithm. To reduce computational costs and improve real-time performance, the system assumes that the eyes are located in the upper third region of the face and only searches there. A histogram equalisation is conducted separately to the left and right half of the face region to minimise the influence of uneven lighting condition.

We compare different haar cascades eye detectors for extracting the eye regions. All cascade sets achieve good results with forward looking open eyes. However, some of them fail in eye detection at extreme gaze directions. The combination of $haarcascade\_mcs\_lefteye.xml$ for the left eye and $haarcascade\_mcs\_righteye.xml$ for the right eye provides the best accuracy and optimal processing speed. Our system finally extracts two image patches that represent the output of this first processing step, one for each eye.

Detection of the Inner Eye Corners

At the core of our gaze estimation approach is a method to measure and track the horizontal distance between the inner eye corners, where the upper and lower eyelids meet (the so-called eye canthi). Our system first applies a Harris corner detection to each eye image patch separately (see Figure 4.3). The method detects the locations of interesting windows which produce large variations when moved in any direction in an image. Locations which are above a threshold are marked as corners. Besides eye corners, the output could also include other feature points of local maximum or minimum intensity, line endings, or points on a curve where the curvature is locally maximal in the image. As this may result in several candidate eye corner points, the system further applies a canny edge detector (Figure 4.3 (B_2)). A canthus region should exhibit strong edges where the upper and lower eyelid edges meet. Accurate canthi locations can then be obtained by considering only those candidate points that lie on the detected eyelid edges (Figure 4.3 (B_3)). All other candidate points are discarded.

Localisation of Eye Centres

For eye centre detection we exploit the semi-circular structure of the iris and pupil as described in Valenti and Gevers' method [172]. To reduce negative effects from shadow cast and screen reflections our system obtains colour edges from cropped RGB eye images using a Gaussian colour model [63] (Figure 4.4 (C_1)). To detect eye centres, our system then computes an isophote curvature map to calculate the isophote radius on the edges (Figure 4.4 (C_2)). The isophotes of an image are curves connecting points of equal intensity. The eye centre is finally detected as the location with maximum isophote centre votes (Figure 4.4 (C_3)).

In summary, the output of the image processing are the eye corners and eye enters (See Figure 4.5).

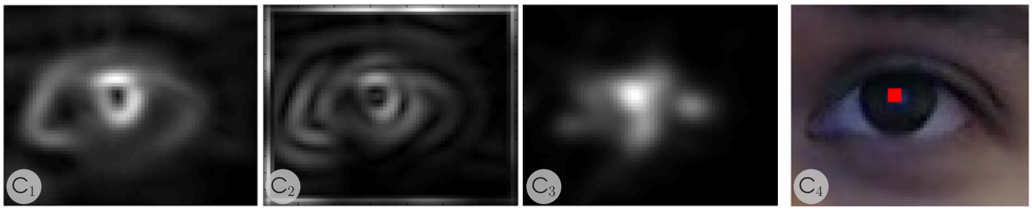


Figure 4.4: Localisation of eye centres: we compute the isophote centre votes on the cropped eye image. The eye centre is detected as the location with the maximum votes. We only take account of locations where the eye centre is likely to occur, hence locations around the edges of the cropped eye image (around 5 pixels) are not considered.

4.1.2 Definition of Pupil-Canthi-Ratio (PCR)

The horizontal distances between the inner eye corners and pupil centres from the right and left eye d_R and d_L are calculated as (see Figure 4.6):

$$\begin{aligned} d_R &= |c_R - p_R| \\ d_L &= |c_L - p_L| \end{aligned} \tag{4.1}$$

where c_R and c_L denote the x coordinates of inner eye corners from the right and left eye images and p_R and p_L denote the x coordinates of pupil centres from the right and left eyes. We define the PCR feature as the distance ratio between d_L and d_R as follows:

$$r = \begin{cases} -(\frac{d_L}{d_R} - 1), & \text{if } d_L - d_R > 0; \\ \frac{d_R}{d_L} - 1, & \text{if } d_L - d_R < 0; \\ 0 \end{cases} \tag{4.2}$$

A negative r indicates a gaze direction to the left while a positive r indicates a gaze direction to the right.

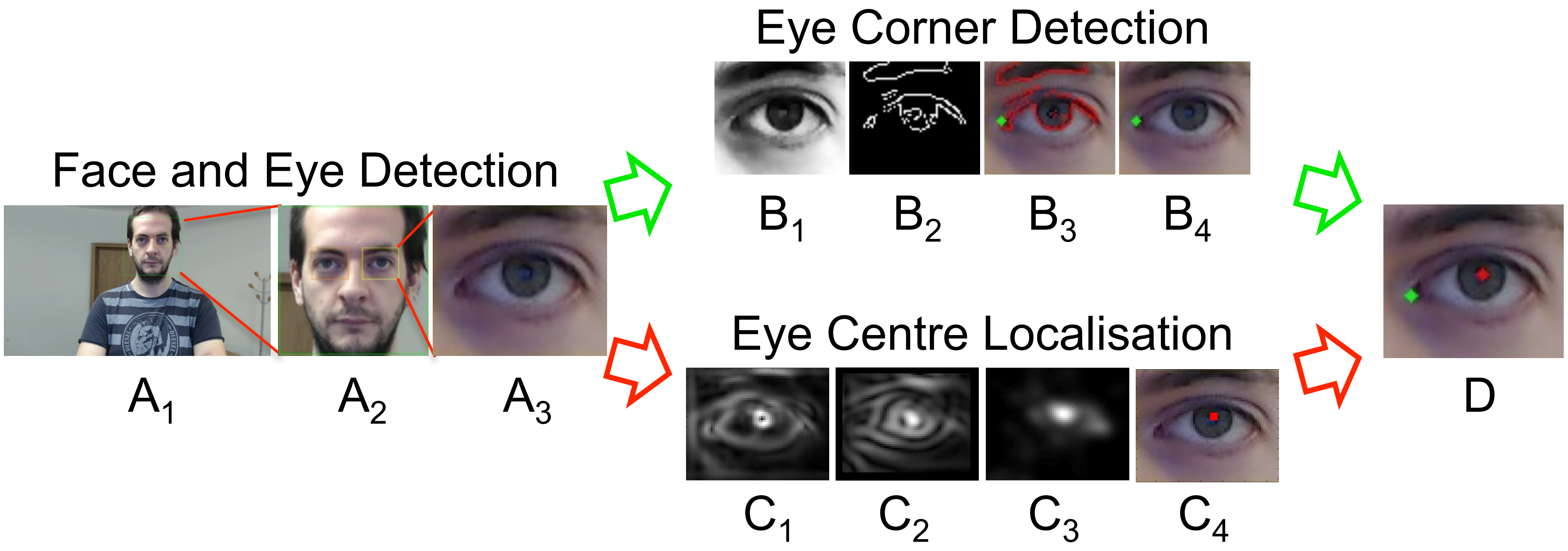


Figure 4.5: This diagram illustrates the image processing of the PCR system (201)

4.1.3 Horizontal Gaze Estimation using PCR

In the last phase, we describe how to use the PCR feature for continuous gaze estimation. Gaze mapping functions determine the gaze direction or coordinates on the screen based on interpolation or approximation from input eye images or features. We denote θ as a visual angle variable, which corresponds to the horizontal gaze point on the display that a user is looking at, and r as the observation value of the PCR feature (see Figure 4.7).

As discussed in Section 2.3.3, there is no generic model that can estimate gaze from obtained eye feature information. Various gaze mapping functions have been adopted in eye tracking research to achieve both accuracy and robustness of head-movements. To gain a deeper understanding of the general applicability of the PCR features, we consider most commonly employed gaze mapping functions. Specifically, to investigate the underlying mapping relationship between θ and PCR along the continuous horizontal space, we consider the following three commonly used gaze mapping methods: Gaussian Process Regression [186, 130, 164], Polynomial Regression [120], and Neural Networks [20, 208, 200].

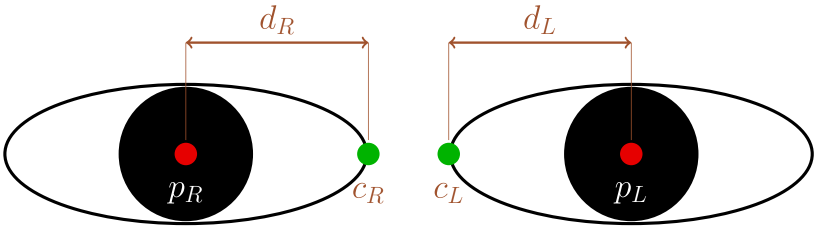


Figure 4.6: The distances of the pupil centres from the inner eye corners are measured to calculate a ratio that captures horizontal gaze. When the user moves their eyes to the left, the pupil-corner distance increases for the left eye, and decreases for the right eye. The ratio will increase the further the user moves their eyes.

Gaussian Process Regression

We assume a noisy observation model $\theta = f(r) + N(0, \sigma_n^2)$ with independent noise function. $f(r)$ is assumed to be a zero-mean gaussian process with a squared exponential covariance function $k(r, r') = \sigma_f^2 \exp(\frac{-|r-r'|^2}{2l^2})$ [143]. We use the GPML toolbox.[2] Given a set of labeled training samples $\{(r_i, \theta_i)|i = 1, ..., M\}$, the goal is to infer gaze visual angle θ^* for unseen PCR r^* calculated from an input test eye image. The parameters $\{l, \sigma_f, \sigma_n\}$ are learned during the training process by maximising the marginal likelihood. With this assumption, the best estimate for θ^* is:

$$\theta^* = k(r^*)^\top (K + \sigma_n^2 I)^{-1} \theta \tag{4.3}$$

where $K_{ij} = k(r_i, r_j)$, $k_i(r^*) = k(r_i, r^*)$ and θ represents the vector of observations $\{\theta_i|i = 1, M\}$ from the training samples.

Polynomial Regression

We assume the mapping from PCR to gaze direction has a second order or cubic polynomial parametric form. With this assumption, the estimate for θ^* is:

$$\theta^* = \begin{pmatrix} a_1 & a_2 & a_3 & a_4 \end{pmatrix} \begin{pmatrix} 1 \\ r \\ r^2 \\ r^3 \end{pmatrix} \tag{4.4}$$

[2] `http://www.gaussianprocess.org/gpml/code/`

The parameters of the polynomial function (a_1, a_2, a_3, a_4) ($a_4 = 0$ for a second order polynomial function) are trained during the training process by minimising the sum of the squares of the residuals.

Neural Networks

Similar to [200], we adopt a feed-forward neural network model using a 2-layer perceptron with the following specific parameters: 5 hidden neuros, the coefficient of weight decay prior (0.01), and maximum iterations of 100.

As a final step, we convert gaze direction in visual angles to display coordinates. If the user is at a distance d from the display (with a width of W mm and a horizontal resolution of HR pixels), the screen coordinates can be approximated by intersecting the gaze direction characterised by the visual angle θ and the screen plane as:

$$p_\theta = \frac{HR \times d \times \tan\theta}{W} \tag{4.5}$$

For consistency with the visual angle representation, we denote the screen coordinate from left to right as $(-\frac{W}{2}, \frac{W}{2})$ in millimeters.

4.2 Experiments and Results

To evaluate the performance of the proposed method, we implement the system on a computer with 2.7 GHz quad-core Intel Core i7 processor with 16 GB of RAM. The data collection application is implemented in Visual Studio C++ in Windows 8, using the image processing procedures described in the previous section. Our recorded data is analysed offline in the MATLAB environment.

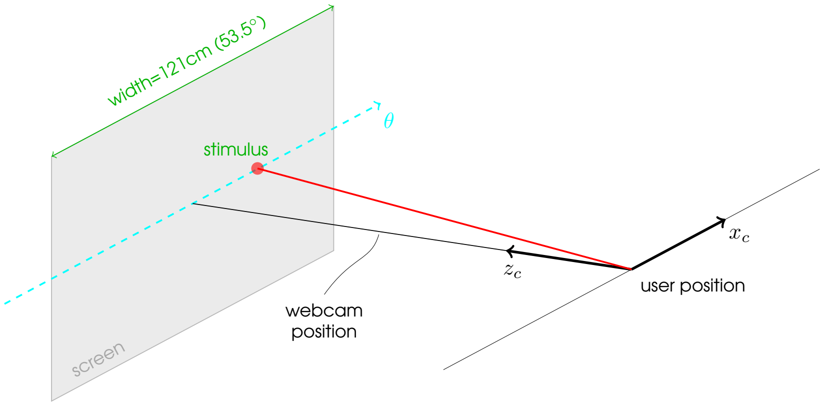


Figure 4.7: The participants stand at a distance of 1.2m in front of the display. They are asked to look at the visual stimulus shown on the display which appears at one of the eleven locations in a randomised order. The eleven locations are horizontally distributed across the display (12cm apart).

4.2.1 Data Collection

In this study, we use a 55 inch (121cm×68.5cm) display, with a resolution of 1920x1080 pixels, mounted on a wall at the height of 120cm (lower bezel) above ground (see Figure 4.7). To record eye images, we use a Logitech HD Pro Webcam C920 with a resolution of 1280x720 pixels and a frame rate of 30Hz. The camera is mounted on a tripod and positioned at a distance of 50cm in front of the display.

The study is conducted in an office environment under normal lighting conditions. We recruit 12 participants (seven female), aged between 19 and 33 years. None of them wear glasses during the study. The participants stand at a distance of 1.2m in front of the display (hence, the visual angles of the display are 53.5° horizontal and 31.9° vertical). The captured eye image resolution is 80×70 pixels, which varies slightly across the participants.

The participants' task is to look at eleven visual stimuli shown on the display one at a time, in a randomised order. Each stimulus consists of a red circle with a diameter of 40 pixels (i.e. 1° of visual angle) shown on a light grey background. The centre of the red circle is marked with a small black dot with a diameter of 5 pixels.

The eleven locations are horizontally distributed across the display (12cm apart), which corresponds to the horizontal visual angles of 26.6°, 21.8°, 16.7°, 11.3°, 5.7°, 0°, -5.7°, -11.3°, -16.7°, -21.8°, and -26.6°. The stimulus location changes every five seconds. To minimise errors, images collected during the first and the last second are discarded for each stimulus location caused by the participants moving their eyes to the next location.

4.2.2 Evaluation over Multi-subject

We test the performance of using PCR for gaze estimation with three different regression methods as described in the previous sections. To evaluate the accuracy of our proposed method, we adopt the leave-one-out cross validation method over the data of the 12 subjects. We first use training data of 11 subjects to learn the parameters of the mapping function, and then test all the frames from the remaining subject. We vary different training data sizes (3 iterations of random sampling) and achieve a mean accuracy of 3.9°. Table 4.1 summarises the gaze estimation error. The accuracy of the gaze point on screen is computed at a 1.2m user-display distance. Table 4.1 shows that the proposed method converges very

		Training data frame size (max. 4000 frames)									
		100 (2.5%)		200 (5%)		300 (7.5%)		1000 (25%)		2000 (50%)	
		(deg.)	(mm)	(deg.)	(mm)	(deg.)	(mm)	(deg.)	(mm)	(deg.)	(mm)
Gaussian Regression	*M*	**4.00**	**90.41**	**3.96**	**89.77**	**3.94**	**89.38**	**3.9**	**88.51**	**3.89**	**88.29**
	SD	0.37	8.22	0.33	7.17	0.37	8.16	0.31	6.58	0.31	6.55
Neural Networks	*M*	**4.00**	**90.54**	**3.93**	**90.64**	**3.87**	**88.33**	**3.88**	**88.40**	**3.86**	**88.10**
	SD	0.35	8.62	0.40	7.95	0.36	7.86	0.38	8.47	0.40	9.03
Cubic Polynomial	*M*	**4.16**	**92.97**	**4.22**	**94.20**	**4.13**	**92.53**	**4.16**	**92.82**	**4.14**	**92.41**
	SD	0.49	10.60	0.50	10.87	0.50	10.95	0.52	11.25	0.51	11.15
Second-order Polynomial	*M*	**5.55**	**121.37**	**5.65**	**123.39**	**5.39**	**117.86**	**5.45**	**119.10**	**5.45**	**118.96**
	SD	0.86	19.06	0.90	19.89	0.85	18.66	0.82	18.22	0.82	18.10

Table 4.1: Summary of gaze estimation errors of 12 subjects, standing at a distance of 1.2m away from the display, with different fractions of training data for learning. We used the leave-one-subject-out cross validation to test three different regression methods.

fast and requires few training samples for learning the model. The training is only performed once for obtaining a set of parameters for the model. Once the model is trained, only a small amount of sample points and pre-trained parameters are required to be saved to make real-time predictions.

4.2.3 Computation Time Comparison

We further compute the training time required by the three regression methods. For the gaussian process, the computation time is normalised over different numbers of minimisation iterations at 15, 20, 30 and 100. For neural networks, the latency is normalised over 100 echoes. Table 4.2 shows the computational expense of the gaussian process and neural networks increases drastically with the training data sizes. In contrast, the average computational time for training is 1.18ms (SD = 0.13ms) for the second order and 1.25ms (SD= 0.52ms) for the third order polynomial function. The gaussian regression, neural networks and cubic polynomial regression methods using 7.5% training data were chosen for subsequent experiments considering both the accuracy and computation time.

	Latency (ms)				
Training data frame size	100	200	300	1000	2000
(max. 4000)	*(2.5%)*	*(5%)*	*(7.5%)*	*(25%)*	*(50%)*
Gaussian Regression	4.9	8.1	14.6	308.6	1304.6
Neural Networks	1.3	2.5	2.5	3.0	3.8
Cubic Polynomial	1.1	1.2	1.2	1.2	1.6
Second-order Polynomial	1.2	1.1	1.1	1.2	1.3

Table 4.2: Summary of computation time of three regression methods with different fraction of training data for learning.

4.2.4 Different Image Resolutions

The potential influences of the method performance come from: 1) the user standing at different distances to the camera; 2) camera parameters such as sensor resolution. This experiment evaluates the robustness over different image scales. We simulate the effect of image resolution variation by downsampling and enlarging the original pictures without any smoothing. We use the model trained at its original scale and tested with features extracted from different scaled images. Table 4.3 shows the gaze estimation errors at different image resolutions. In these experiments, the original eye pictures of the collected data has an average size of 75x67 (SD 9.0x9.5) pixels. We found that the feature detection algorithms could not detect eye centre (e.g., detection is outside of the iris area) when the resolution is reduced to 40% of the original size, which is approximately 30x27 pixels.

	Average Accuracy (degrees)				
Image resolution (percentage)	–	-60%	-40%	-20%	+20%
Gaussian Regression	3.94	4.50	4.21	4.02	3.99
Neural Networks	3.87	4.59	4.22	4.14	4.01
Cubic Polynomial	4.13	4.61	4.15	4.03	4.11

Table 4.3: Gaze estimation error in effect of different image resolutions. The "–" column represents original image resolution.

4.2.5 Detection Noise Sensitivity

In order to characterise the robustness of the method, we investigate the average accuracy in the effect of noise in the extracted image features. We add random Gaussian noise with zero mean and a fixed standard deviation to the x coordinates of the extracted eye feature from only one eye (left or right) respectively. The gaze estimation error is calculated on the data with added noises. The process is

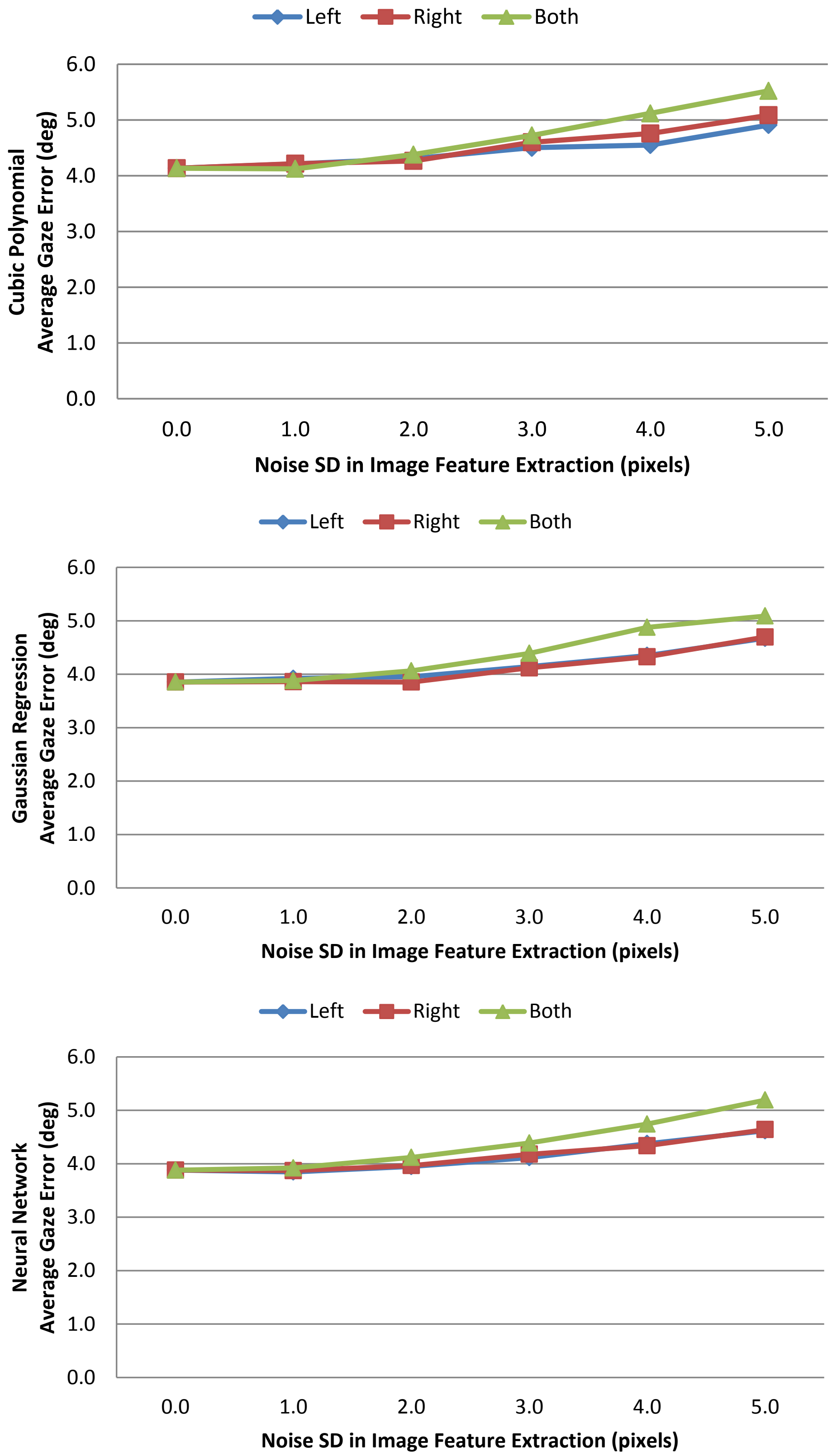


Figure 4.8: System accuracy in effect of noises in the eye corner and pupil centre detection in the video images.

repeated with increasing values of the standard deviation. The same procedure is repeated to extracted features from both left and right eyes. The system accuracy results are illustrated in Figure 4.8.

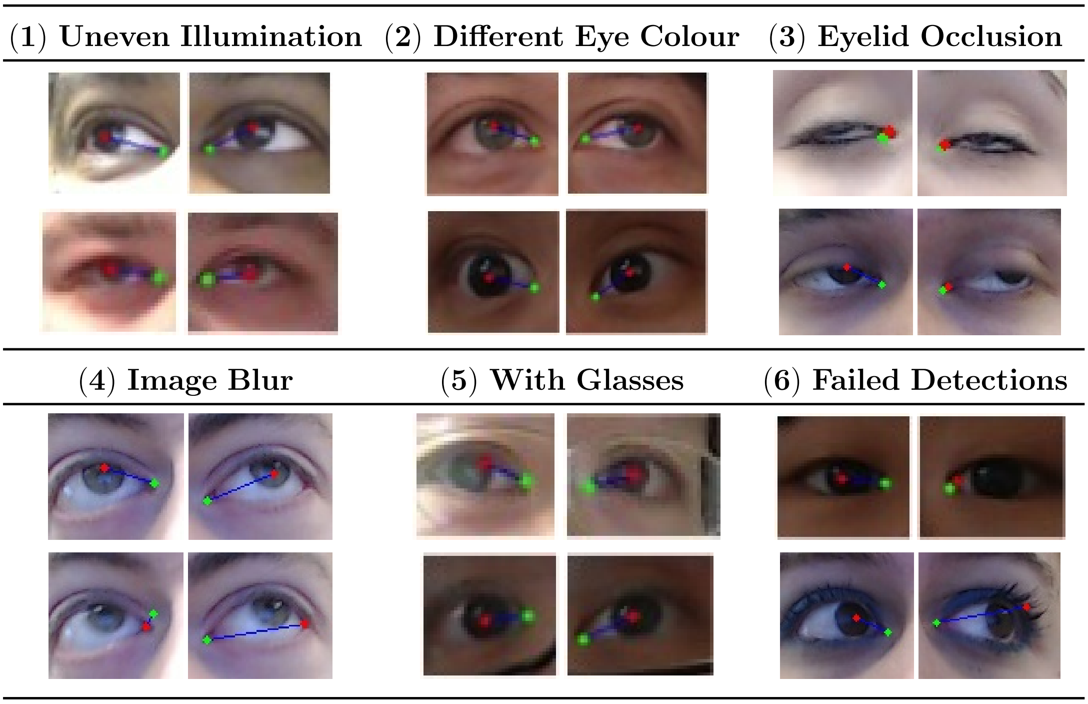


Table 4.4: The table illustrates different examples. The system can process images under various conditions as illustrated above. However, performance decreases in dark lighting conditions and detection error rate increases if the user is wearing glasses.

4.2.6 Uncontrolled Conditions

To understand the robustness of the system, we conducted an out-of-the-lab study, where we deployed our system in an open public area. In addition, by putting the system in the field, the study evaluated robustness with uncontrolled conditions, such as lighting, users, etc. Over a two-day period we invited a total of 30 participants (8 female, 22 male). Of these participants, twenty-two had dark eye colours (i.e. brown or black) and the other eight had bright eye colours (i.e. grey, green or blue). Seven participants wore glasses, but five of them removed their glasses during the study. Our system captured the participant's face at an average resolution of 245×245 pixels (SD=40). The purpose of this study is to collect data in a realistic scenario to understand potential challenges under uncontrolled conditions. Table 4.4 illustrates different example eye images recorded during the field study.

Our study illustrates that the eye centre is sometimes falsely detected when the image is blurry (see Table 4.4 (4)) or with thick eye lashes (see Table 4.4 (6)). The system response also decreases with thick eye glasses (see Table 4.4 (5)). Another influence factor is head movements. We observed that in uncontrolled conditions some people will turn their head while moving their eyes to the side (see Figure 4.9).

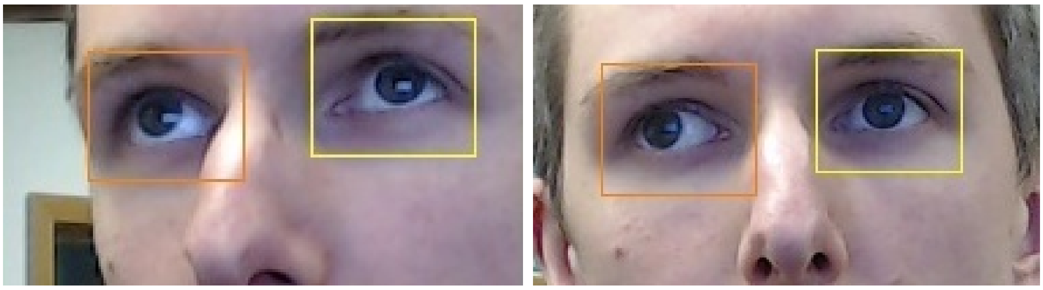


Figure 4.9: Influence of rotation parallel and perpendicular to the image plane

4.3 Discussion

We present a simple and efficient solution for estimating gaze from web camera images without calibration. Our goal is to develop a non-intrusive, calibration free system with minimum hardware cost. Our results show that the method is robust and computationally efficient. Hansen and Ji [74] provide a summary of different existing gaze estimation approaches, and identify that systems using a single webcam without additional illumination have an accuracy of between 2° to 4° with individual user calibration [72, 186]. Our method achieves a similar accuracy of 3.9°. The accuracy is comparable to current state-of-art methods, but has the advantage of being able to work in real time without prior calibration with each individual user. This is promising for application scenarios where rigid setup and user calibration are less desirable.

We compare the system performance of three commonly used gaze mapping methods for learning the generic model. The results suggest a non-linear mapping relationship from the PCR feature to gaze direction. We show that a second order polynomial is not sufficient to approximate the underline mapping. Overall, the gaussian process and neural networks achieve better accuracy than 4° which is approximately 0.2° to 0.3° lower than the cubic polynomial. Over different training data sizes (see Table 4.1), similar accuracy can be achieved even using a very small amount of training samples (e.g., 100 frames consisting of 100 pairs (r, θ)). Regarding the computational efficiency, the polynomial regression outperforms the other two methods with a minimum latency of 1.2ms. The process of learning the mapping function presented in Table 4.2 is required only once to train the model. Once the training process is done offline, the model could be applied for real-time detections. Cubic polynomial is a suitable approximation of the underlying mapping function from PCR to gaze if short training time is required. The computation time of neural networks and gaussian regression increases drastically with more training date, but they provide better accuracy.

To evaluate the robustness, different factors influencing the gaze tracking accuracy are investigated by determining the effect of feature detection noise and image scale variance. As illustrated in Figure 4.8, the system accuracy declines linearly as the error of noise in the eye corner and pupil centre detection in the video images gets larger than 2 pixels. With an error of 4 pixels, the gaze direction tracking error increases from approximately 4° to 5° in the most optimal cases.

Zhu and Yang point out that an error of 1 pixel in iris detection will generate a gaze direction tracking error of about 3° when using a method based on relative positions of the iris and any facial features that serve as a stationary reference point [206]. Our results suggest that the PCR feature is less sensitive to the noises in the feature detection (a gain of about 1° with an error of 4 pixels). This could be due to the PCR feature being a ratio measure. For example, if the iris detection has an error of 1 pixel this can bring in errors of $a_2\Delta r + 2a_3\Delta r + a_3\Delta r^2$ in Equation 4.4 for a second order polynomial. Using the x coordinates of the iris centre to eye corner vectors as input feature introduces $\Delta r = 1$ in the error function. In contrast, using the PCR feature will introduce $\Delta r = 1/d_L$ or $\Delta r = 1/d_R$ which has a normalising factor.

In order to understand the observed error, the effects of different image resolution are studied through simulations. As shown in Table 4.3, although the generic model was trained at its original scale, it is applicable on different image scales and can work with very low resolution eye images. An average gaze error of 4° at its original scale of 75x67 pixels is increased to 4.6° at 30x27 pixels . A reason for this could be that PCR is a relative ratio measure. Also, our data set covers variance in image scales from multiple users. During the data collection, although the 12 users are standing at the same distance there are variations from 90x85 to 60x50 pixels in the sizes of captured eye images.

As PCR is based on the horizontal symmetry of eye movements, which requires both eyes to be on the same horizontal level, the performance of PCR can be influenced by head tilting. For a very few cases, the height of two eyes are different by 20 pixels. Similar to existing gaze estimation methods, PCR works best when a user is facing forward towards the camera. When the user turns their head, this results in ambiguity of the PCR features. Due to the 2D representation of the feature points, a change in d_L and d_R could come from eye movements or head turning as shown in Figure 4.9. Also, the eye corner feature might be occluded at extreme head turning positions. Future work could focus on combining head pose and gaze direction to extend PCR so that head movement and eye movement are seamlessly accommodated.

We notice that our gaze tracking system has consistent performance under uneven illumination (see Table 4.4 (1)) and with users having different eye colours (see Table 4.4 (2)). However, the current eye detection algorithms cannot handle blinks or eyelid closure. One of the main sources of the tracking failure is from cases when large part of the eyes are occluded by eyelids (see Table 4.4 (3)). Future work would include integrating a blink detection to improve performance.

In this work, we did not calibrate the Logitech Webcam C920. The manufacturer already provided the webcam with an internal default calibration (i.e., minimise distortion), which uses the RightLight Technology [10]. However, with different hardware, camera calibrations might be required to achieve similar results as in this work. We would assume limited influence without an external calibration, as our approach only predicts 2D gaze coordinates on a planar surface. However, an external calibration for extrinsic parameters is necessary if 3D gaze information is computed. This process can be done through basic geometric equations and only needs to be done once, which is available in the OpenCV library [11].

4.4 Conclusion

In this chapter, we presented a lightweight method for person-independent tracking of horizontal gaze directions. Our method requires no prior adaption to individual users, and it tracks people's eye gaze in real time by analysing eye images captured using an RGB camera. The system detects gaze directions by examining the distance symmetry between iris centres and eye corners from both eyes. With "looking straight ahead" as a default state, we define PCR as a novel measure of how far a user looks towards the left or to the right. Our data analysis shows that using PCR to estimate continuous gaze achieved an average accuracy of 3.9 degrees, with minimum gaze estimation latency of 1.2ms. Our empirical study further shows that PCR is robust against different eye images from diverse users and noise in the detection of eye corner and pupil centre. The proposed method is computationally efficient and works with low resolution eye images. It requires only a single off-the-shelf camera, which makes it potentially ready for wide scale deployment.

The advantages of being lightweight and calibration-free make our method suitable for out-of-lab applications, where control setup and individual user calibration are not feasible. We believe that the proposed method has a great potential in handsfree applications, such as gaze controlled interfaces when hands are busy (e.g. in an operating room) or when hygiene is required (e.g. in a hospital). Our system is computationally efficient, which makes it potentially ready for wide scale deployment. In the future, we envision such vision-based methods can be installed in public places (e.g., museums).

Our results demonstrate the possibility of enabling robust gaze tracking that is suitable for pervasive display applications. While this approach overcomes some challenges of robust eye tracking in uncontrolled environments, it should be noted that the method can only provide coarse horizontal gaze estimation and has certain limitations, e.g., minimum required image resolution. These are important factors that should be considered in the design of gaze interface for pervasive displays.

Part II

Gaze Interface Design

5

Background

The first part of this thesis addresses the technical challenges of enabling robust gaze sensing in uncontrolled environments. The focus of the second part is to answer the question of how to design gaze interfaces for everyday pervasive displays. Although eye tracking has a long history in HCI, prior works on gaze interfaces and interactions are primarily designed and evaluated in constrained lab environments for desktop computing. As the computing paradigm shifts from desktop computing to ubiquitous computing, there is little understanding of how to design gaze interfaces to support spontaneous interaction and to accommodate opportunistic behaviour in pervasive contexts.

In this chapter, we first provide an overview of eye tracking research in HCI. Then we review state-of-the art gaze interfaces and interaction techniques. Our focus is to analyse the design challenges of eye-based systems for interactive displays and existing solutions. Finally, we highlight the unaddressed issues and open challenges in gaze interface design for pervasive displays.

5.1 Early Adoption of Eye Tracking in HCI

Eye tracking technologies have a long history in HCI and contribute considerably to our understanding of human perception and cognition. The use of eye tracking started in the 1880s. At that time researchers attached lenses to people's eyes to observe where they were looking and which words they paused on while reading text [80]. The results revealed people's eye movement patterns and reading process. For example, eye tracking reveals that readers' eyes make a series of short pauses (i.e. *saccades*) at different places. Since then, numerous studies have been carried out to understand why people's gaze stops at certain places, eye movement durations, and strategies during reading. Their results demonstrate early evidence that eye movements provide insights into understanding human behaviours in performing well structured tasks.

In 1947, Fitts et al. conducted studies that used eye tracking for usability testing in a human-machine interface [56]. They analysed the eye movements of pilots to examine how they used cockpit controls and instruments when landing a plane (see Figure 5.1). They found out that the dwell time on each instrument reflected the difficulty of interpreting the instrument and individuals with different experiences had different scan patterns. Their research demonstrated that eye tracking can be an effective tool for understanding the usage of visual interface design. However, at that time, data collection and analysis required tremendous manual effort and were done offline due to lack of computerised tools. With the advances in eye tracking devices, researchers can employ real time eye tracking to study the usability of web pages and evaluate advertisement quality [45, 139, 129, 34]. Nowadays, eye tracking measures (e.g., number of fixations, fixation durations) are commonly applied in different domains to evaluate product designs [139, 129] and to develop training programs [178].

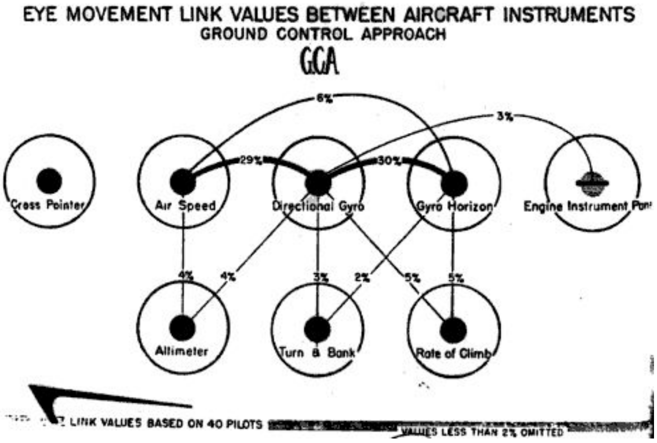


Figure 5.1: Fitts et al. analysed the eye movements of pilots while using cockpit controls and instruments (Source (56)). Their results provided understanding of pilot's scanning habits and performance which led them to identify shortcomings in the cockpit display design.

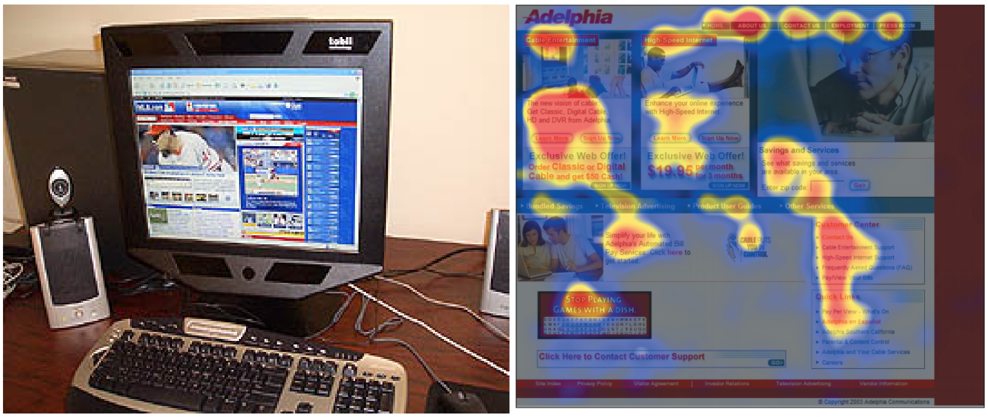


Figure 5.2: Eye tracking has been commonly used in user research to evaluate the effectiveness of visual design (Source (129)). For example, a gaze heatmap illustrates contents that are being looked at the most (coloured area).

Other information can be observed from eye movements such as underlying perceptual and cognitive processes. In 1967, Yarbus conducted an influential study to explore the link between eye movements and people's cognitive tasks in an image viewing experiment [191]. He found out that eye movements are highly task dependent and are linked to our cognitive goals. The arguments were further strengthened in later investigations when people perform different activities such as reading [90], real-world scene perception [76] and complex control actions in daily life [103]. Their research provides foundations that eye movements can be applied to infer context information, such as people's interests, attention and mental states.

With the advent of real-time eye tracking in the 1980s, researchers in HCI started to investigate the use of eye movements as computer input [27, 81, 156, 87, 196]. For instance, people can use their eyes to enter texts on a screen [58, 184, 113] and to perform a point and click action [87]. Over the last two decades, eye trackers have become less intrusive, easy to use, and affordable. The use of eye tracking for interactive applications has progressed rapidly towards the general public in areas such as desktop user interfaces [86], mobile devices [125, 118] and public display interaction [54, 146]. In the following section, we describe different gaze interfaces in detail.

5.2 Eye Tracking for Interactive Applications

Gaze input has been widely applied in assistive technologies with special-purpose visual interfaces. The idea is to replace manual inputs with eye tracking for performing computer-based tasks such as typing [184, 75, 111, 112] and drawing [79] (see Figure 5.3). Likewise, using gaze as an input has been explored in a variety of user interfaces, but has mainly focused on using gaze in the "windows, icons, menus, pointer" (WIMP) paradigm for desktop computers.

There are two ways to utilise eyes for interactive applications: using gaze *ex-*

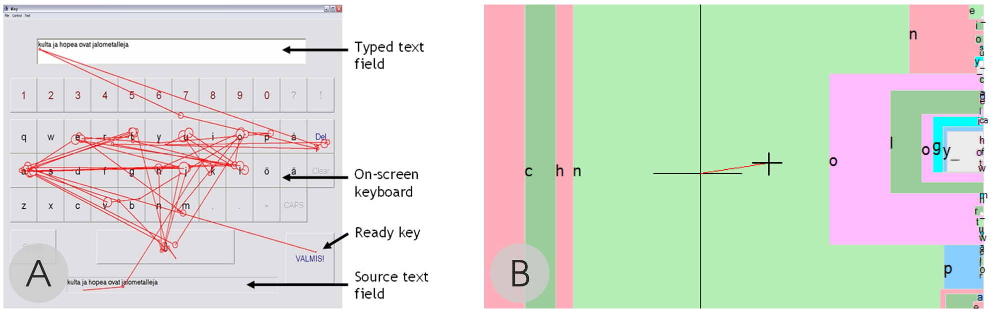


Figure 5.3: Gaze interfaces for text entry: (A) A commonly used gaze interface for text entry lets users enter text by looking at elements on a onscreen keyboard (113). (B) Dasher lets users enter text by looking at directions to zoom in, and each direction is corresponding to a piece of text. (184).

plicitly by directly applying momentary gaze positions or directions as a one-to-one stylus input; using gaze *implicitly* by interpreting gaze behaviours to infer context information. In this section, we first discuss the characteristics of gaze input and review eye-based graphical user interfaces in desktop settings. Then we cover recent progress on gaze interfaces in non-desktop contexts, such as mobile, wearable interfaces, large interactive surfaces, and interactions with physical objects. Lastly, we review existing works on using gaze in multi-user interfaces, which provides a background on prior works on gaze in shared display.

5.2.1 Characteristics of Gaze Input

While eye tracking has been used primarily for behavioural studies in lab research, prior research works have investigated different designs of applying eyes as an input for interactive systems. Researchers and designers consider eye gaze as an attractive hands-free modality for several reasons:

1. Eye movement patterns (e.g., eye fixations) have been considered as a good pointer. Our eyes can move very fast with minimum efforts [88, 196]. When changing the point of fixation, our eyes make ballistic movements (e.g., sac-

cades). These movements can be very fast with velocities up to 900 degrees per second. Using gaze has the potential to make interactions faster and reduce physical fatigue compared to other modalities such as hands. In addition, people always look at what they are working on in well-structured tasks. Our eyes are proactive and move prior to manual input while working in combination with other modalities [87, 103]. Combining gaze interaction with other input modalities requires little additional effort and enables natural interaction [152].

2. People naturally look at objects of interest [27, 28]. The amount of time looking at an object reflects different levels of interestedness towards objects in the scene. Aggregating gaze data in a time interval can infer users' interests. These implicit cues can be used to tailor contents that are most relevant to users' interests in interactive systems [156, 96, 15].

3. Eye gaze indicates people's visual attention and what they focus on [195, 22, 175]. The point-of-regard corresponding to the centre of our fovea vision has the highest acuity, while the vision becomes less sharp in our peripheral vision (see Figure 2.1). This suggests that a user can only attend to a relatively small part of a visual scene at a time and has a limited *perceptual span* within the fovea [90]. Hence, gaze can provide the context of users' attention. For example, this has been exploited in gaze contingent displays to address hardware issues that demand high rendering power and display resolution (e.g., to accelerate graphics computation) [22, 67].

Although using gaze as input is appealing, multiple challenges exist. A classic challenge is the *Midas Touch Problem* [88]. We use our eyes to obtain visual information, but our eyes as a perceptual organ cannot issue explicit actions (e.g., like clicking with a mouse) without causing ambiguity. Thus, gaze-only interfaces are susceptible to accidentally triggering.

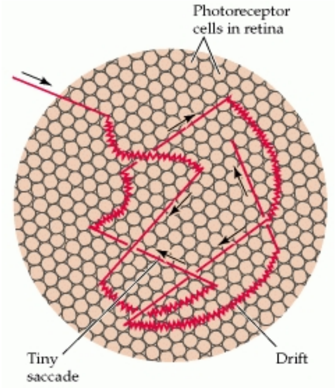


Figure 5.4: Our eyes are not completely stable. They make small and involuntary movements during fixations including micro saccades and drifts (Source [12]).

Another challenge associated with gaze pointing is that gaze only works with large targets [196, 100], because eye movements are noisy and jittery. Due to physiological properties of our eyes, they never stay completely stable (see Figure 5.4). When we fixate at a target, our eyes make small movements such as drifts and microsaccades [191]. Besides the physiological properties, the tracking accuacy is also limited by current technologies. The state-of-art video eye tracking systems can only provide an accuracy of 0.5 degree of visual angles, which is about a thumbnail size when the arm is fully extended [74].

Triggering explicit actions by gaze is challenging, as eye movements can be involuntary and often subconscious. It is also undesirable to overload the visual perceptual channel with motor tasks [196]. In gaze interfaces, balancing the use of gaze input with other input modalities still remains an open challenge.

The main goal in previous HCI research of different gaze interface designs is to address a central problem: *how to employ gaze input use by the general public.*

While the majority of previous interfaces and interactions are designed to be used by a user sitting in front of a monitor, researchers are beginning to explore the use of gaze in other contexts such as mobile devices, interactive surfaces (e.g., tabletops, large displays) and physical objects. In the following sections, we review the existing gaze interfaces. We do not intend to cover all existing works but rather focus on representative interfaces that overcome some of the above interface design challenges.

5.2.2 Gaze for Desktop Interfaces

Prior research largely focuses on using gaze explicitly and aim to integrate gaze into desktop GUI to improve pointing, selection, window switching and navigation. Other works use gaze implicitly, e.g., to adapt content in interactive graphics.

Pointing and Selection

Early work on eye gaze interaction demonstrated selection on menu grids [81] and gaze-based techniques within the WIMP paradigm for desktop computers [87]. The underlying interaction model is to treat gaze as a pointing (target acquisition) device, for example as an alternative to a mouse to move a cursor in conventional interfaces [152, 104]. The primary focus of these works is to integrate gaze into applications in traditional graphic user interfaces (GUI).

To find out the effectiveness of gaze input, Sibert and Jacob conducted studies to compare mouse versus eye gaze for object selections in desktops computers [152]. In their studies, the onscreen position of where people were looking at (i.e. *fixation*) indicated object of interests (i.e. target acquisition) and a *dwell time* threshold was used to confirm a selection. Although their results demonstrated that gaze was faster than mouse in object selection tasks, normal users are still more efficient with traditional interfaces (e.g., mouse or keyboard) [52].

Researchers proposed a variety of designs to optimise the speed and accuracy to improve user experience of gaze interfaces for selection tasks. In the gaze-only interfaces, *dwell time* and *explicit eye movements* (e.g., blinks, gaze gestures) are common techniques to overcome the Midas Touch Problem.

Dwell time is based on fixations [104]. However, the threshold of dwell time is difficult to adjust. If the interval is too short, it risks unintentional commands, e.g., a fixation during casual observations can activate a false selection. If the interval is too long, the interaction becomes slow, and prolonged fixations induce eye fatigue and cause user frustration [89, 73]. It is unnatural for humans to keep staring at an object once it is found [87].

Gaze gesture is based on saccades and lets users issue commands by performing a sequence of predefined eye motion patterns [51, 82]. Gaze gestures overcome the eye tracking accuracy issue as it uses relative eye movements [51]. However, the gaze gesture design requires prior training of a predefined gesture set, sometimes including complex patterns, so that the motions are not confused with unintentional eye movements. Another challenge is that users have higher cognitive load in gesture based than in dwell time based techniques [53].

In traditional HCI research, gaze-only solutions are less appealing. In particular in desktop GUI, target sizes can be small (e.g., less than half of the fovea size) and targets can be cluttered together. Using eyes to trigger explicit actions is considered to be unnatural [88], as Zhai *et al.* suggested that it is undesirable to overload the visual channels with motor controls [196]. Therefore, researchers explored using gaze as an augmented input and in combination with other inputs (e.g., keyboard, mouse, touch), for example, to speed up pointer input [196, 100, 52], to assist head pointing [180]. In these interfaces, other modalities are treated as a clutch for gaze input or for fine motor control. The benefits of combing gaze and manual input reduced fatigue due to less manual work and also made the interaction faster [196].

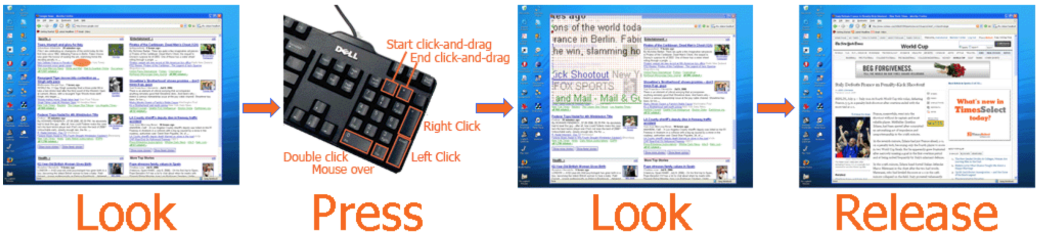


Figure 5.5: EyePoint lets users perform accurate selection with a look-press-look-release action (Source [100]). Users first look at a desired target and then press and hold a hotkey. Users' focus point is magnified with overlaid visual anchors (red grid dots) to ease hyperlink selection in the text box. Finally, users look at a target in magnified view and release the hotkey.

To improve the accuracy of gaze pointing, existing solutions increase the target size, expand target selection in motor space [119], or use magnified views such as a distorted fish-eye lens [18] and a zoom lens [100]. EyePoint, for example, presents a practical gaze selection in windows GUI [100] (see Figure 5.5). The design aims to overcome the Midas Touch Problem by using keyboard to confirm a selection. To select small targets, the user's focus point is zoomed in with an overlaid visual anchor (e.g. a grid of dots). However, multi-modal coordination is challenging due to synchronisation and can cause errors [99].

Window Switching

Eye tracking has been proposed as an alternative channel for sensing attention. In the EyeWindows system, users can select and switch between windows using eye gaze [57]. The technique allows parallel manual inputs (e.g., keyboard) in focus selection tasks, which accommodates continuous shifts in user attention. For example, the user can select a focused window by looking at its title by pressing a space bar key or by eye fixation. The focused window is restored to full dimension while the other windows are shrunk, for example, using an elastic windowing algorithm (see Figure 5.6). A similar concept has been exploited to switch input devices between multiple computers [48] and change visualisations in multiple displays [49].

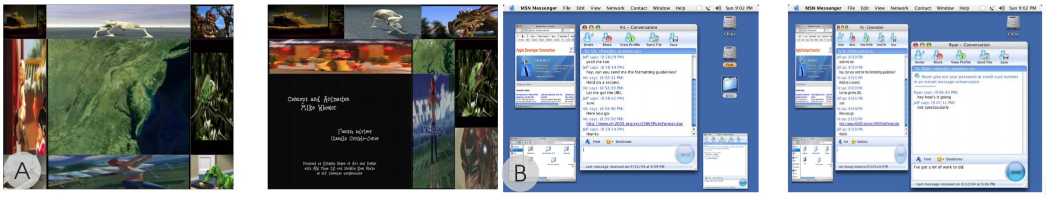


Figure 5.6: Two example applications of the EyeWindows: users select a focus window by looking at a window while pressing a space bar on the keyboard. The window focussed on is enlarged while the other windows are shrunk(Source (57))

Navigation: Pan and Zoom

As we move our eyes to navigate in the real world and information space, researchers also explored gaze input for navigation interfaces such as pan and zoom, including gaze typing [73], navigating in large scale images (e.g., maps) [14, 159], controlling camera's viewpoint in teleoperation [205], and enhancing text scrolling [101].

For example, users can perform pan actions by looking at different areas overlaid on the border of the display (see Figure 5.7 (A)). The zooming interactions

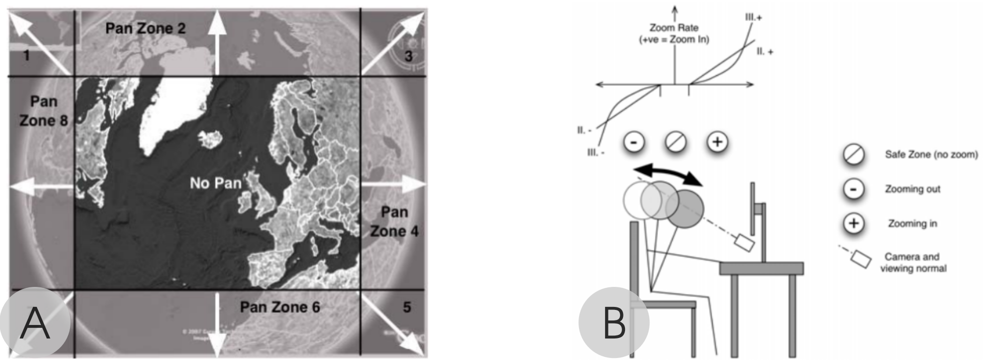


Figure 5.7: Pan and zoom (Source [14]): (A) The image zooms in when users stare in the central region. When they look at the pan region, the zooming is also accompanied by panning towards the screen centre. (B) The zooming rate is controlled by head-screen distance which is combined with gaze to control pan and zoom directions.

adopt a "fly-where-I-look" approach. Gaze-only pan and zoom interfaces have a "one-way zoom" problem so users cannot go back to the previous zoom level [73]. This issue can be overcome by using another input such as moving the head forward and backward to control the zoom actions (see Figure 5.7 (B)) [14].

Implicit Use of Gaze Input

Gaze provides context information about users' current interests and focus. By analysing fixation patterns, displayed items and areas can provide customised realtime services according to users' interests, such as revealing additional details [156, 141] and displaying relevant ads for digital advertising [15]. Instead of providing direct real-time gaze cursor feedback, the interfaces implicitly adjust content that are presented to the users. Another system uses a camera to monitor users' blink rate to infer eye fatigue caused by long term use of computer monitors. The system presents eye-blink stimulus (e.g., screen blurring, screen flashing) when users have not blinked for a while [46].

Another implicit use of gaze is to make a graphical display gaze contingent (see

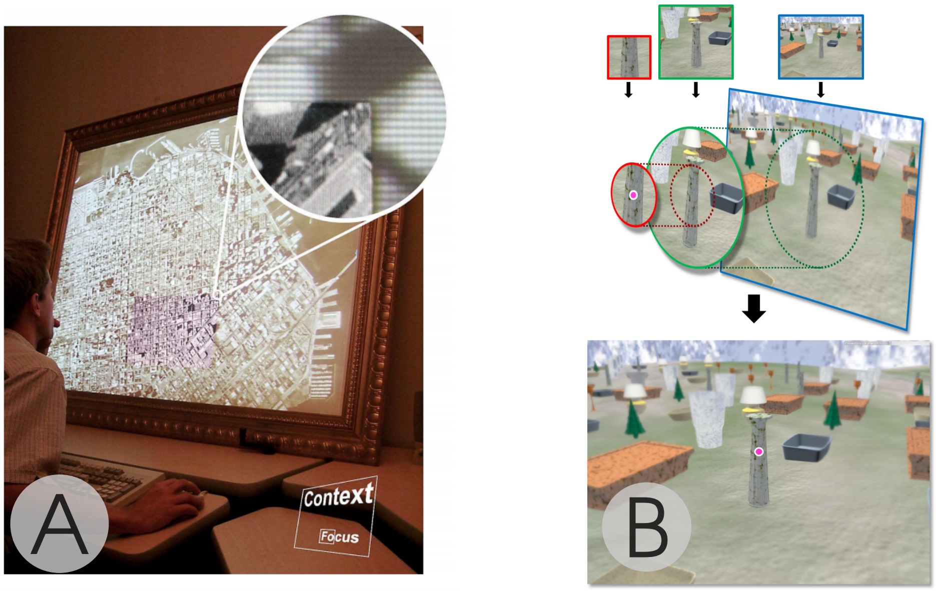


Figure 5.8: Gaze contingent display in 2D (A: Source (22)) and 3D (B: Source (67)) graphic displays.

Figure 5.8). By exploiting the falloff of acuity in the visual periphery, displays can be rendered dynamically with the sharpest resolution at where people are looking [27, 28, 22, 67]. In contrast to rendering a display at our fovea vision, Bailey *et al.* developed the *subtle gaze direction* technique that draws users' involuntary gaze by modulating the image area in their peripheral vision [19]. Other interfaces employ gaze data as a previous focus marker in attention-demanding contexts (i.e. driving) to ease attention switching [93] or to set prior focus regions to assist image cropping in desktop tasks [147].

While the majority of the gaze-based interaction techniques focused on 2D user interfaces, there are some works that has considered using gaze input for 3D interfaces [94, 73, 192, 140, 158]. Some research works have applied similar techniques as 2D interfaces [192, 140, 158], while other works require 3D gaze to enable interaction with content at different depths [94].

5.2.3 Gaze for Mobile and Wearable Interfaces

Besides desktop interfaces, researchers investigated the use of gaze input for small hand-held (e.g., mobile phones) [50, 118, 125] and wearable devices [84]. Some researchers proposed to use dwell-time and gaze gestures to issue a selection command (e.g., initiate a phone call from contact lists) or to trigger a scroll on mobile phones [50, 91]. For instance, four gesture patterns are assigned to trigger four commands: scroll up, scroll down, select and cancel. Users can scroll through the contact list and select a contact to initiate a call by using these gestures. To scroll up, users need to look across the top edge of the device and move the selection one position upwards in the list. Other work used eye movements (i.e. wink) to activate applications that users look at on the mobile phone [118] (see Figure 5.9 (A)). In addition to the gaze-only approaches, Nagamatsu *et al.* proposed a gaze-and-touch interface which supports a map browser application on mobile devices [125] (see Figure 5.9 (B)).

5.2.4 Gaze for Large Surfaces

With the recent advances in large interactive surfaces, researchers started to explore gaze input for remote interaction, such as in settings where a display is unreachable or cannot be manipulated directly [160, 170, 176]. A plethora of research focused on the combination of gaze and touch inputs. The interactions are based on the concept of using gaze for indicating an object of interest on a distant screen and touch gestures for precise positioning and manipulation. For example, gaze input is combined with touch on a mobile device for target selection and manipulation [160, 161] and content transfer across devices [170, 171]. Some research

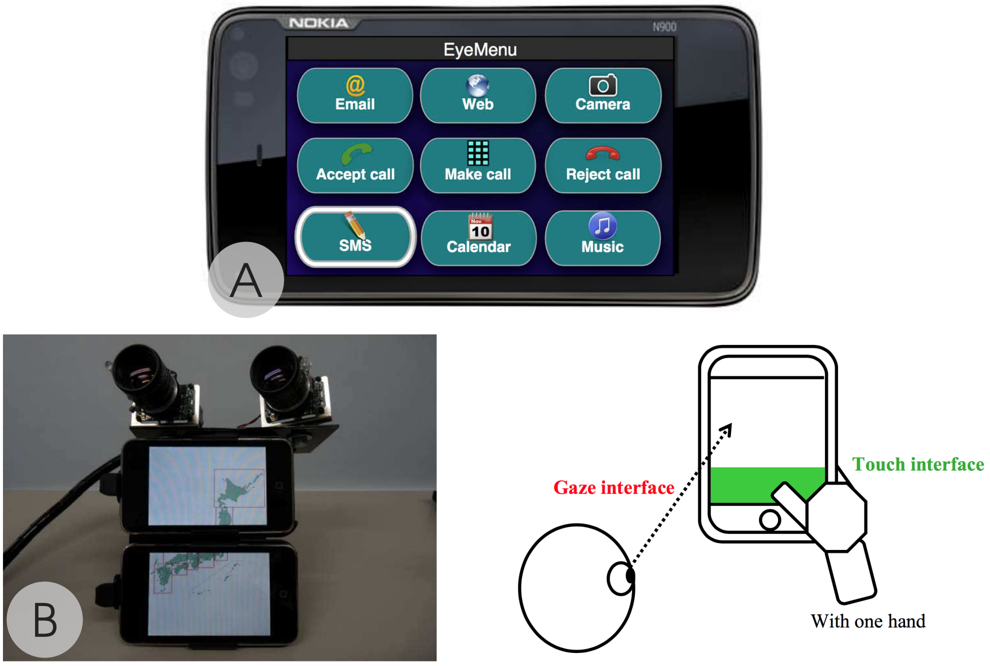


Figure 5.9: (A) EyePhone: users can launch an app by looking at an icon and wink (Source (118)). (B) MobiGaze supports gaze and touch interaction on mobile devices (Source (125)). For example, users' gaze indicates region of interests, and they can touch the area near the thumb to perform touch actions (e.g., swipe to zoom).

also explored the combination of gaze and hand gestures to manipulate 3D media contents on a smart TV [192]. In addition to remote interaction, researchers also investigated novel gaze enhanced interaction techniques on touch surfaces such as tabletops [78, 189, 136].

5.2.5 Gaze for Human-Object Interaction

A person's eye movements and fixations correlate strongly with their attention and focus on their surroundings. For example, we tend to look at a person we are talking to. Based on this concept, researchers introduced eye contact sensors to make displays or devices attention-aware [175, 154] (see Figure 5.10 (A)). This has been realised by detecting gaze direction towards infrared tags placed on devices and objects [155, 174] or cameras embedded in the devices [154]. Prior research

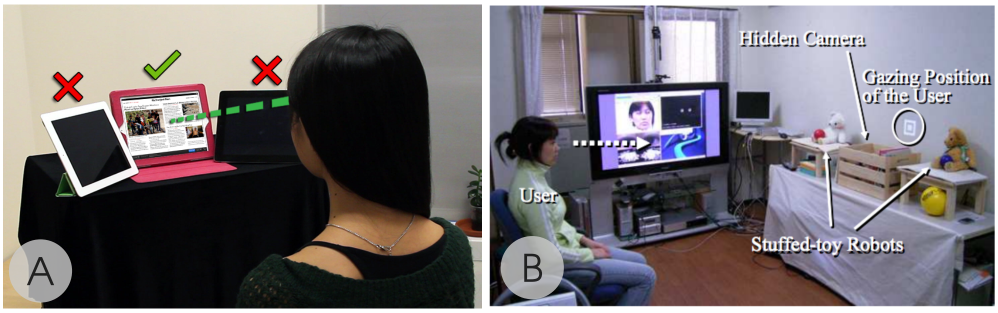


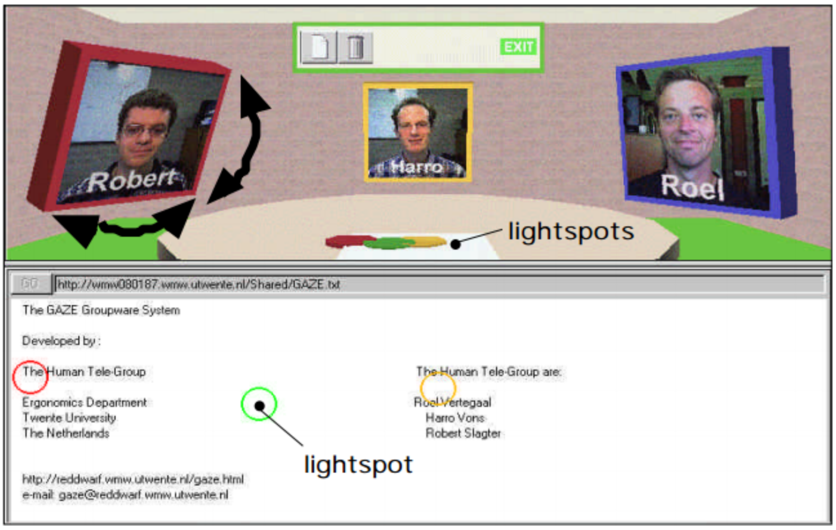
Figure 5.10: Gaze interactions with physical objects: (A) The GazeLock system allows users to unlock a screen when the tablet detect their direct gaze towards it. (Source (154)) (B) GazeCoppet enables a stuffed-toy robot to react to a user's eye contact by voices and gestures and joint attention by facing the direction towards the same object that the users look at. (Source (190))

also considered gaze in daily life applications. For example, the "Gazecoppet" system (see Figure 5.10 (B)) is designed to enable a robot to react to users' gaze for enticing communications with people who have lost their desire to communicate (e.g., those with dementia or trauma patients) [190].

5.2.6 Gaze for Multi-user Interfaces

Eye Contact for Video Conference

Gaze has been shown as an important cue for face-to-face communication [44, 31]. One of the major challenges in remote communication systems is to enable gaze awareness, because gaze cues can get easily lost in video conferences when users move freely in spaces. A plethora of research in HCI has investigated how gaze cues, mainly eye contact and mutual gaze, affect communication in video conferencing systems [150] and in immersive virtual environments [162]. One example of such systems, the GAZE Groupware, conveys gaze in multiparty communication and cooperative work, such as in meetings [173] (see Figure 5.11). Their work suggest that eye contact and gaze cues can help regulate conversation flow, provide feedback for understanding, and improve deixis in remote video conferences systems.



http://wmw080187.wmw.utwente.nl/Shared/GAZE.txt

The GAZE Groupware System

Developed by :

The Human Tele-Group

Ergonomics Department
Twente University
The Netherlands

The Human Tele-Group are:

Roel Vertegaal
Harro Vons
Robert Slagter

http://reddwarf.wmw.utwente.nl/gaze.html
e-mail: gaze@reddwarf.wmw.utwente.nl

Figure 5.11: The GAZE Groupware system is designed to support gaze awareness in multiparty video conferences (Source (173)). Personas rotate to where users look to provide eye contact cues. Each user's gaze is also displayed on shared documents using colour coded cursors.

Gaze for Remote Collaboration

In collaborative work systems, the use of gaze has been investigated in remote setups. Similar to using gaze in remote communication systems, the *Clearboard* system enables gaze awareness between remote collaborators by using the metaphor of a transparent glass window [85]. Users are virtually located opposite each other to work on a shared board and can look through the transparent board to see what their partner is looking at. Although mutual gaze and the perception of eye contact can enhance the perception of co-presence, it seems to be far less important than the view of a group's shared work space on collaborative activities [62].

Some studies investigated the role of shared gaze in collaborative systems. The motivation comes from allowing remote collaborators to share their gaze over each other's screen space (i.e., seeing a collaborator's visual focus of attention).

Previous research has pointed out that gaze plays a role as a "conversational resource"' during spatial reference [97]. Gaze has been proposed to assist verbal collaboration in remote setups, due to the verbal communication problems like misunderstandings and noise. In a tourist planning application, Qvarfordt and Zhai applied gaze in a dialogue system [142]. They discovered that a remote assistant that is following remote users' gaze patterns while conversing with them can detect the users' interest. In a remote collaborative visual search task, Brennan *et al.* demonstrated that sharing gaze is more efficient than speech for the rapid communication of spatial information [29]. Similar results were found in [127] where shared gaze was shown to be more efficient than speech during collaborative tasks that require rapid communication of spatial information. Shared gaze has also been found useful to detect misunderstanding to overcome the lack of deixis at a distance [40].

5.3 Open Questions and Opportunities

The second part of this thesis aims to enable gaze interactions for pervasive displays. While previous research designed prototype interfaces for gaze-based interactions, their designs were optimised for their selected controlled environments. To advance gaze interfaces in pervasive contexts, we need appropriate interfaces that support the new paradigm of gaze interactions in ubiquitous computing. However, to enable gaze in pervasive contexts poses several challenges that have not been addressed in prior research.

Firstly, several issues are induced by the current eye tracking technologies. Previous research focused on gaze pointing, where gaze is used as an alternative or complement to other pointing devices [87, 196]. However, gaze pointing requires careful adaptation to individual users who need to complete a calibration procedure prior to interaction, which hampers spontaneous use [120]. In Part I, we developed the PCR method which is particularly designed for gaze tracking on pervasive displays without any prior calibration. Instead of tracking fine grained gaze positions, the system tracks rough gaze regions on a display. This has led us to develop novel interfaces for this new form of eye tracking system to support spontaneous interaction.

Secondly, users engage with pervasive displays spontaneously, driven by opportunity or emerging information needs. Users and displays come together as strangers, and yet interactions have to work without preparation, as they are usually unplanned. Therefore, when a system is deployed in pervasive contexts, it is impractical to assign an expert to provide assistance for novice users. This presents a design challenge, as novice users are unaware of the interactivity of the interface. Hence, another open question is how to design a gaze interface for pervasive displays that users are able to use without any prior awareness of how the system works.

Thirdly, pervasive displays can be shared by more than one person for collaborative tasks, such as people looking for information together. However, existing eye tracking hardware can only support gaze estimation of a single user. While existing research has mainly designed gaze interfaces for single-user operations or remote-user collaborations, designing multi-user interaction on a shared interface remains an open challenge. This raises the open question of how gaze can be useful for co-located collaboration that we aim to explore in this thesis.

6

Techniques for Spontaneous Gaze Interaction

In the previous chapter, we summarised existing gaze interfaces. We realised that current designs cannot support spontaneous interaction in pervasive contexts. as they usually require specialised hardware and calibrated gaze position. To address this issue, we exploit the PCR method introduced in Chapter 4. The PCR method is particularly designed for gaze tracking on pervasive displays without any prior calibration, however it can only track rough gaze regions and cannot estimate fine grained gaze positions. To overcome this challenge, this chapter aims to answer how to design interfaces for the PCR system to support spontaneous interaction.

Public display applications vary in their need for input. There are many scenarios in which gaze control of low complexity can improve interaction with information displays. Arrival/departure screens in airports, mounted overhead and

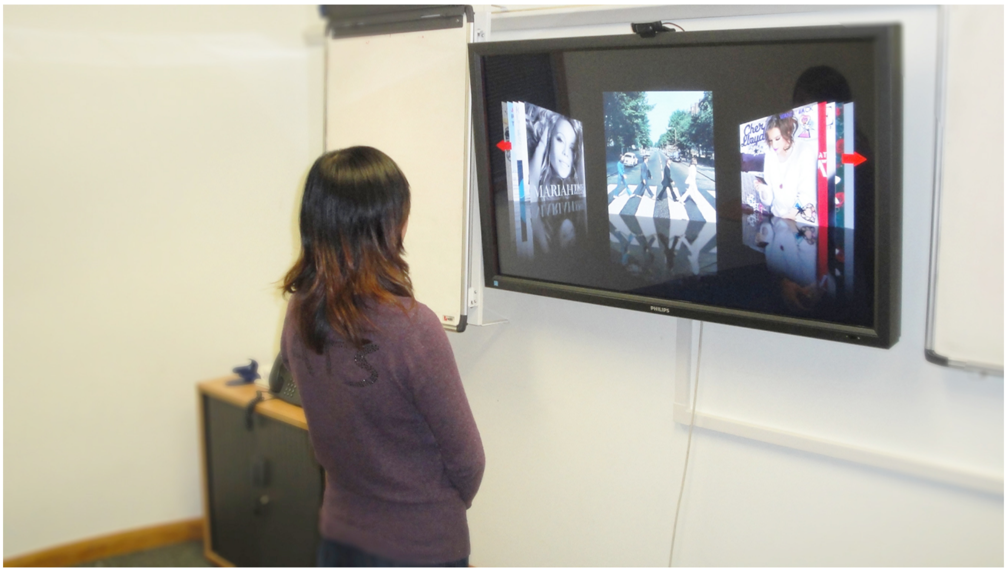


Figure 6.1: SideWays enables spontaneous gaze interaction with displays: users can just walk up to a display and immediately interact with it using their eyes, without any prior calibration or training. In this example, the user controls a cover flow with looks to the left and right.

out of reach, display content that is split over a number of pages which could be navigated by gaze. Situated map displays could be made gaze-responsive to provide more detail on areas of interest. Shop displays could employ gaze to let window-shoppers scroll through offers of interest. These scenarios have in common that they describe spontaneous interactions, where users walk up to a display that they may never have seen before and should be able to interact without further ado.

We envision *walk-up gaze interaction*, where eye tracking will be pervasively embedded in everyday interactive systems. Without specialised devices and calibration, people can spontaneously walk up to a system and start interacting with it immediately, while at the same time the system implicitly tracks its users' gaze in the background. Based on this information, the system could customise realtime services according to the users' interests, such as digital advertising.

In this chapter, we present a novel eye gaze interface, designed for users to be able to just walk up to a display and casually assume control, using their eyes only. Our system, *SideWays*, requires only a single off-the-shelf camera and distinguishes three gaze directions (to the left, right and straight ahead) as input states. The interface and interaction is developed based on the PCR system described in Chapter 4. Using the PCR system, these input states are detected in a spontaneous, robust and person-independent manner. There is no training or calibration involved, and no adaptation to individual users. Any user can step up to the display, and the system will be able to respond immediately to their attention (see Figure 6.1).

We propose three techniques for Sideways: users can use their eyes to select an object, scroll contents and control a slider on a large display. We describe a study in which we evaluated SideWays with 14 participants on three interactive tasks (selection, scrolling, and slider control). The selection task served to characterise the system in terms of correct detection of input depending on time window of observation of the eyes, while scrolling and slider control assessed usability of the interface and interaction model for control tasks. We analyse the participants' behaviours in performing the different tasks. This lets us gain insights on users' gaze control strategies, and derive design considerations for application of our system.

6.1 Design of SideWays

SideWays targets pervasive settings and adopt a deliberately simpler gaze model to facilitate spontaneous and calibration-free interaction, between users and displays that have never seen each other before. We argue that public display settings require a different approach to eye tracking and gaze input, less fixated on high fidelity but optimised for instantaneous usability by any user without prior configuration, calibration or training steps.

6.1.1 Interaction Model

The interaction model we have designed for SideWays assumes a central region of interest on the display to which users align their gaze (e.g. guided by visual design of the interface), and adjacent areas to the left and right that users can select with “sideways glances”. A look straight ahead is interpreted as attention to the centrally displayed content. In this state, the interface is kept stable, and the eyes do not trigger any action. In contrast, glances to the left or to the right are associated with input actions. In terms of application logic, these glances are like pressing (and holding) a button, but the user experience may be more subtle and fluid with interface designs that have such actions appear natural and implicit.

Prior research have considered eye contact detection in pervasive settings [175, 154]. In SideWays, we likewise detect gaze for attention to pervasive displays but in addition enable users to provide input with looks to the left or right from their centre of attention. Beyond eye contact, Vidal *et al.* showed that content displayed in motion can be gaze-selected without calibration by exploiting smooth pursuit eye movement, however relying on higher fidelity tracking of gaze with specialist hardware [176].

A variety of projects have used head orientation towards large displays in presumed approximation of what people look at [122, 126, 153]. However, Mubin et al. found in an “interactive shop window” study that only few users aligned their heads with their gaze [122]. Other work has focused on low-cost extension of public displays for gaze pointing however still requires a calibration phase prior to interaction [146]. EyeGuide [54] explored the use of a wearable eye tracking for interaction with pervasive displays. In contrast, our focus is on enabling interaction with public display without any instrumentation of the user. Magee *et al.* reported a vision-based system that is similar to ours as it detects rapid eye movements to the left and right as command input [112]. However their system was specifically designed for a user with severe cerebral palsy and was primed to detect occurrences of a left/right movement, while we continually classify eye gaze direction.

6.1.2 Apply PCR for Discrete Gaze Tracking

We adopt the PCR method to robustly detect three horizontal gaze directions in a person-indepdent manner. Using the image processing method described in Chapter 4, we obtain the pupil centre and the inner canthi in each video frame. Figure 6.2 illustrates how we use these to derive gaze direction. Consider first that we look straight ahead. In this case the pupil centres of both our eyes will be similarly distant from the respective inner eye corners. If we look to the left, then the distance of our left pupil from its inner eye corner increases, while the distance of the right pupil from its inner corner decreases. Conversely, a look to the right means that the left pupil moves closer to its inner eye corner, while the right pupil moves further away. Consequently, to determine different gaze directions, we calculate the ratio r of the eye-centre P_{cx} to inner canthi C_{ix} distances of both

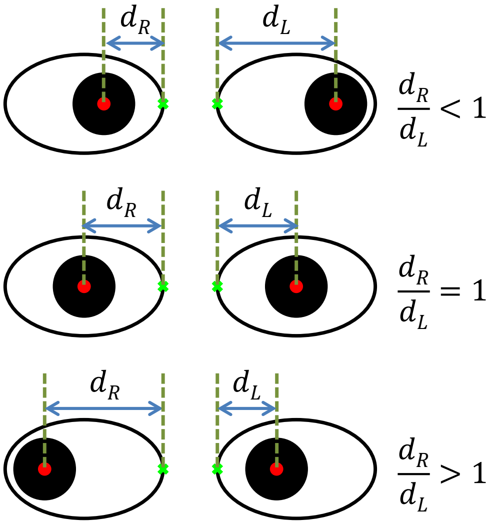


Figure 6.2: Three gaze directions are determined by the distance ratio of eye-centre P_{cx} to inner canthi C_{ix} for both eyes.

eyes as

$$r = |\frac{C_{iR} - P_{cR}}{C_{iL} - P_{cL}}| \tag{6.1}$$

where C_{iR} and C_{iL} are the x coordinates of inner eye corners from the right and left eye images, P_{cR} and P_{cL} are the x coordinates of pupil centres from the right and left eyes.

For each video frame, we calculate the inner canthi to pupil centre distances for both eyes. At time t ($t = 1, 2.., T$), we denote the measurement of our observed gaze direction for frame f_t as O_t. A general threshold T_r is set for classifying frame f_t as three gaze directions which are L(left), C(centre), R(right) according to the following rules:

$$O_t = \begin{cases} R(right), & \text{if } T_r < |r| < T_{r_{MAX}}; \\ L(left), & \text{if } 1/T_{r_{MAX}} < |r| < 1/T_r; \\ C(centre), & \text{otherwise.} \end{cases} \tag{6.2}$$

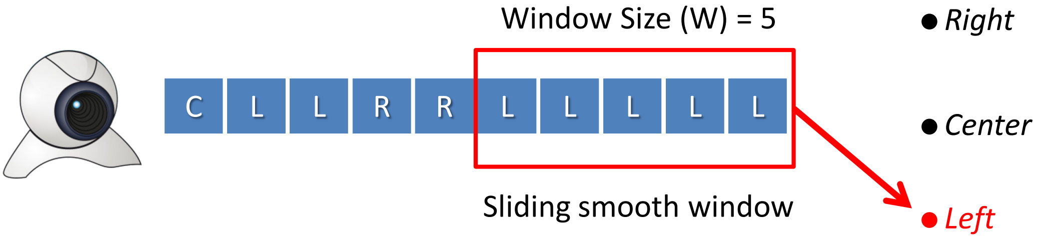


Figure 6.3: The activation of a gaze direction is based on a smoothing window. As the camera captures images in real-time, a stream of gaze directions fill the frames in the sliding window. An activation is triggered when the same gaze direction is detected consecutively.

where $T_r = 1.3$ is optimised for the user study (derived from the preliminary study) and $T_{r_{MAX}}$ is a constant upper bound.

Activation Using a Smoothing Window

To smooth the decision made for each video frame, we adopt a sliding window approach (as shown in Figure 6.3). We perform a smoothing window with a size of W over the observations of W image frames. We consider all observations within time span $[t, t+W-1]$. The input to the decision system is a set of measurements $O_{1:W} = \{O_t | 1 \leq t \leq W\}$ where $O_t \in \{L, C, R\}$ in a sliding window of W frames within time span $[t, t+W-1]$. We process all images frame by frame from the start. For each incoming frame, we collect new observations denoted as $O_{current}$. The set of measurements $O_{1:W}$ corresponding to time span $[t+1, t+W]$ in the buffer is updated. We only consider valid observations when both eye corners and centres are detected. As shown in Figure 6.3, an activation is triggered when the same gaze direction is detected consecutively in the sliding window.

We distinguish two different activation approaches: *discrete* and *continuous*. Discrete activation clears all measurements $O_{1:W}$ in the sliding window after an event is triggered. This causes delay as a new stream of measurements in the sliding window need to be collected. Continuous activation only updates the last measurement in the sliding window with every incoming frame, hence, allowing for fast response.

6.2 User Study: Selecting, Scrolling and Sliding

The interaction model proposed for the SideWays system is to treat gaze straight at the display's centre as a default state in which the eyes do not trigger any action, while "sideways" glances to the left or right are foreseen for user input. We designed a user study to evaluate our system and the proposed interaction model on three generic tasks: *Selecting*, *Scrolling* and *Sliding*. For each task, we run a separate experiment to evaluate different aspects of our system.

Selecting was always conducted first as it was designed to fundamentally characterise the interface in terms of correct classification of input depending on size of the sliding window used in the process. A smoothing window of five frames was used but data was collected for post-hoc analysis of detection accuracy versus speed (shorter smoothing windows). The other two experiments were conducted in counter-balance. Scrolling tested the users' ability to use our system for discrete scrolling through a list of items. A window size of four was used with discrete activation, so that a scroll step was executed only if the user's gaze dwelled for four valid frames on the left/right control. Sliding tested control of a continuous slider and the users' ability to move a slider to accurately hit a target position. A window size of three was chosen with continuous activation, which meant that that a sliding step was executed in each frame, for as long as the detected gaze direction matched the previous two frames. Three different speeds of the slider

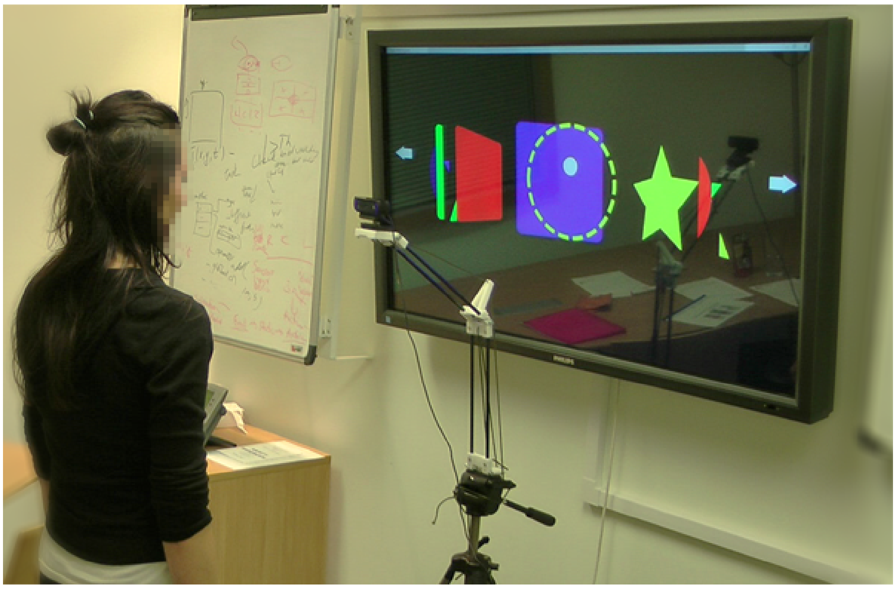


Figure 6.4: A snapshot of a study session.

were used, and data captured to analyse how often users needed to change sliding direction to reach the target.

6.2.1 Setup

Fourteen paid participants (six female, eight male), with body heights ranging from 1.65m to 1.96m (*M*=1.77, *SD*=0.10), aged between 21 and 47 years (*M*=28.79, *SD*=7.27), and various eye colours took part in the study. Three participants wore contact lenses during the study.

The hardware setup for our study consisted of a 55 inch (121cm×68.5cm) LCD display from Philips with a resolution of 1920x1080 pixels, mounted on the wall at 120cm height (lower bezel). A Logitech HD Pro Webcam C920 with a resolution of 1280x720 pixels and a video frame rate of 30Hz was mounted on a stand and positioned 60cm in front of the screen (see Figure 6.4). The real-time image processing and gaze estimation software was implemented in OpenCV, and ran on a laptop with a 2.67GHz processor and 4GB of RAM.

The study was conducted in office space under normal lighting conditions. We asked participants to stand at a distance of 1.3m in front of the display (visual angle of the display 49.9° horizontal, 29.5° vertical). A marker on the floor indicated where the participant should stand. However, during the user study, participants were free to fine tune the distance for their own comfort. The distance between the camera and the user was 70cm±5cm. The captured image resolution was 300x300 for faces, and 80x70 for eye images, slightly varying across users. In a real world deployment, cameras would typically be mounted on the display but we positioned it closer to the user as we aimed to evaluate interaction with our system, not the limits of eye pupil and corner detection.

6.2.2 Procedure

Each session lasted for approximately 45 minutes. Participants were first introduced to the system and allowed to complete one trial of each task. All participants then first completed the Selecting experiment, while the remaining two experiments were counter-balanced. After each experiment, user feedback was collected with a questionnaire. The questionnaire asked for the participants' subjective experience, problems they encountered, and strategies to overcome issues of the system.

Experiment 1: Selecting

In the selecting task, participants had to look at either the left or the right region of the display. Participants were asked to initially focus on the centre region of the display. Once the system had detected the participants' gaze, the system indicated the desired gaze region using a green arrow pointing either left or right. In addition,

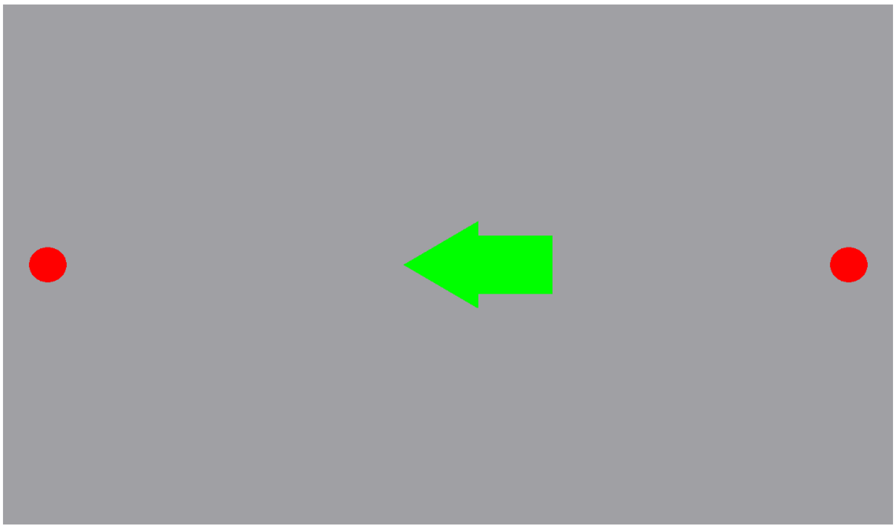


Figure 6.5: Selection task

a red circle was shown in both the left and right region to assist participants in fixating on that region (see Figure 6.5). The system continuously estimated the gaze direction with a smoothing window of five frames. Upon detection of gaze on the correct target, the target colour would change from red to green. Participant's were instructed to return their gaze to the centre after each completion of a trial. With a short delay, the next trial would be triggered by display of an arrow. In total, this was repeated twelve times (six for each direction, randomised).

Experiment 2: Scrolling

In the scrolling task, participants were asked to scroll through a list of objects using their gaze and to find the object that matched a predefined target. We used a combination of four shapes (circle, square, triangle, and star) and four colours (red, green, blue, and yellow) to represent a set of sixteen scrolling objects (see Figure 6.6). At the beginning, the sixteen objects were randomly placed horizontally at equal distances, arranged as a flow with the display as viewport showing the current selection in the middle, and one adjacent object on either side.

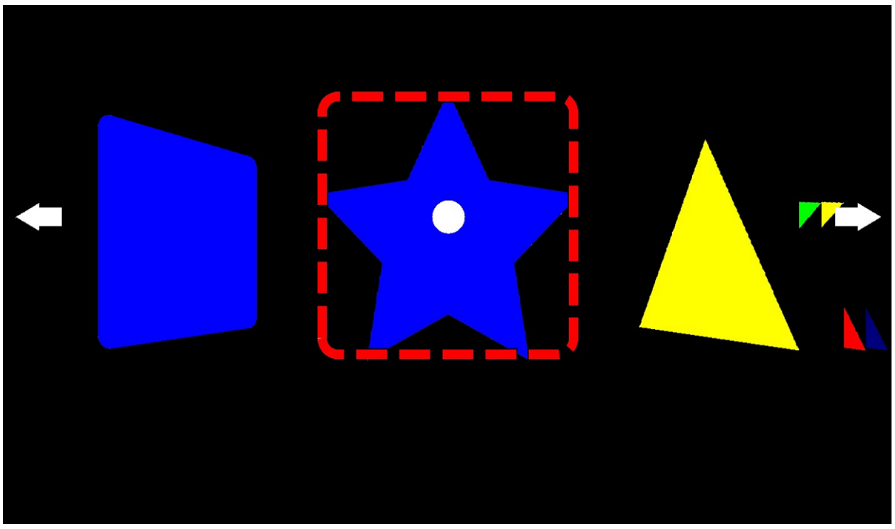


Figure 6.6: Scrolling task

Participants then had to scroll through the objects to find a preselected object that was indicated by a coloured dash-bordered shape and shown at the centre of the display. Participants had to look left or right of the display to scroll items from that direction toward the centre. Arrows were displayed on both sides of the display to help participants fixate. The task was repeated six times, and each time with a different target shape. For each iteration, the browser starting position alternated between the left-most and the right-most positions of the object collection.

Experiment 3: Sliding

In the sliding task, the participant's objective was to control a horizontal slider by moving the slider indicator either left or right onto a target position with their eyes. For this task, the display showed a horizontal slider widget in the centre region (see Figure 6.7). The slider target was represented by a black line, and it contained a red circle as slider indicator. Green arrows on the left and right display regions represented the slider's controls. At the start of the task, the indicator was placed at either the left-most or the right-most slider position, and the distance to

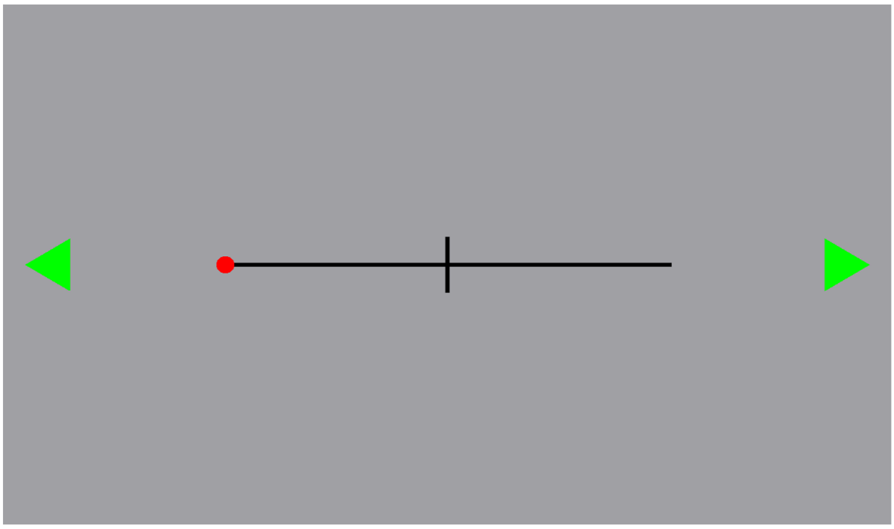


Figure 6.7: Sliding task

the target was always 480 pixels. When participants looked at the left controller, the slider would progress one step width to the left in each frame, and vice versa for the right control. Sliding speed was increased over the trials, with three step widths representing 0.01, 0.025 and 0.04 of total screen width (i.e. 19.2, 48, and 76.8 pixels), requiring 25, 10 or 6 steps respectively to reach the target. When participants were satisfied that they had reached the target, they returned their gaze to the centre of the display to complete the trial. The task was repeated twice for each step size, for a total of six trials per participant.

6.3 Results

Four of the 14 participants required eyeglasses for correct vision and three wore contact lenses while the fourth removed his glasses. One of them reported that the contact lenses had affected her speed in fast and frequent eye movements, and that they caused discomfort after using the system for a while.

Two participants experienced asymmetric system performance in left and right directions. One of them explained that she had had an eye operation, which

	$\mathbf{w = 2}$	$\mathbf{w = 3}$	$\mathbf{w = 4}$	$\mathbf{w = 5}$
Correct	135	151	160	162
Wrong	12	5	2	0
Missed	21	11	6	6

Table 6.1: Number of correct, wrong (detected as opposite direction) and missed (detected as centre) detections of the total 168 trials of selecting tasks using different window sizes.

	$\mathbf{t < 2(s)}$	$\mathbf{t < 3(s)}$	$\mathbf{t < 4(s)}$	$\mathbf{t < 5(s)}$	**overall**
$\mathbf{w = 2}$	76.8%	78.6%	79.2%	80.4%	80.4%
$\mathbf{w = 3}$	80.4%	86.9%	89.3%	89.9%	90.4%
$\mathbf{w = 4}$	81.5%	87.5%	90.5%	91.1%	95.2%
$\mathbf{w = 5}$	72.0%	84.5%	88.1%	91.1%	96.4%

Table 6.2: Accuracy for selection over different window sizes and timeout thresholds.

effected gaze to the right. The other participant reported better performance for gaze to the right but her reason was unknown. However, she described that she compensated by turning her head slightly towards the opposite direction. The participant who removed his glasses was far-sighted. He often squinted his eyes while looking at the display, which drastically slowed down the system's detection speed (more frames were discarded for lack of pupil detection, and the smoothing window would fill up more slowly). Several participants reported that blinking also reduced the system's detection speed. This was not surprising given that blinking caused the eyelids to occlude the eye pupils, so the system could not determine their gaze direction.

6.3.1 Selecting Task

Participants performed a total of 168 trials (12 each) of looking left and right for selection. Table 6.1 summarises the results with post-hoc analysis for different window sizes. Eye gaze matching the target was counted as correct, eye gaze on the opposite target counted as wrong (faulty selection), and eye gaze in the centre as missed (no selection). Window size 5, as experienced by the users in the study, results in the least number of errors. This was expected, as increasing the window size also increases certainty. Analysis of recorded data revealed that three of the errors (missed selections) were from one participant due to squinting. On average, users needed 1.78s (*SD*=0.3s) per selection, but the user who squinted required an average of 2.91s. Although increasing the window size provides better detection accuracy, it inherently increases the time required to detect a correct selection. Table 6.2 provides an analysis of correct detection rates depending on window size and time thresholds. The results indicate that a window size of four frames is optimal, but windows of three frames perform almost as well for detection in limited time.

Participants reported that using the system for left and right selections was “intuitive”, “easy to perform”, and “suitable for touch-free interactions”. However, two mentioned problems, such as inconsistency in system response time, one of them noting that to improve the system, he had tried not to blink and kept his eyes wide open. Several participants mentioned that although it wasn’t tiring to use the system they would prefer faster response times.

6.3.2 Scrolling Task

We observed 84 trials for the scrolling task, six per participant. In each trial we counted how many scrolling steps users required to complete the trial and compared this with the minimally required steps. On average, participants required 1.2 (*SD*=3.4) extra steps to complete a scrolling trial. In 61 out of the 84 (72.6%)

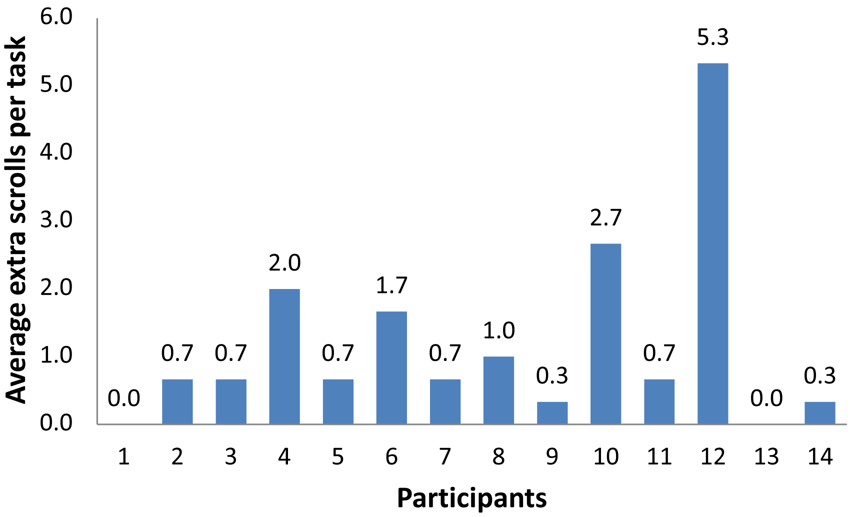


Figure 6.8: Extra scroll counts for 14 participants over six scrolling trials.

trials the scrolling task was completed efficiently with the minimum number of steps. In general, participants were able to correct mistakes with very few extra steps. In 20 out of the 23 error cases, the participants only changed the scrolling direction once for correction (12 for overshooting one step away from the target, eight for a distance of two steps from the target).

Figure 6.8 illustrates the scrolling accuracy of each participant. Two participants finished every task without any extra scroll steps, and evidently learnt to use the system very quickly. Seven participants made one error (an extra scroll). However, participant 12 took 28 extra scrolls for his first trial, explaining that he lost focus during the first trial as the target shape was at the very end and the system was not responsive. Our post-study analysis showed that his head orientation drifted from the centre towards the scrolling direction, where new shapes were coming from. This caused the system to classify his gaze as central, instead of triggering a scroll step. The participant learnt to re-centre his head when the system was not responsive, and completed the remaining trials with only minor errors. Overall, we observed that six participants tended to turn their head towards the scrolling direction. Since the focus region is on the far end of the display, people intuitively turned their head towards the region to examine upcoming information. However, when they noticed the scrolling stopped, they returned their head orientation back to the centre.

System errors mainly resulted from delays in stopping the scrolling. Four participants mentioned that the system was not sensitive enough for stopping and that they had to look at the centre region already before the target object reached in the centre. In particular, participants 10 and 6 found it difficult to stop scrolling, while participant 6 found it hard to judge the colour of objects in the centre with peripheral vision, while gazing to the left or right for scrolling.

We observed the participants' behaviours and strategies in scrolling. Six participants fixated on the arrow indicator to scroll continuously. The participants were able to finish 40 out of 84 (47.6%) trials without stopping to scroll before reaching the target. Some participants mentioned that frequent stopping and checking the information in the centre helped them to perform better, but those who scrolled without stopping did not cause more errors (six errors in 40 trials). These results show that the participants were able to handle frequently changing information when the shapes are moving. The participants devised strategies to avoid mistakes. Some participants noticed the delay in triggering single scroll/stop actions (the system was set to discrete activation after four frames), and exploited this for brief glances to the centre without causing to stop the scrolling action.

Overall, participants were satisfied with their experience of using SideWays for scrolling. Most felt that the system reacted to their left and right gazes correctly, and that the system provided sufficient precision for real applications. They also

	Mean (s)	*Std* (s)	*Minimum* (s)
Large	14.2	12.0	2.4
Medium	16.2	11.5	6.0
Small	26.7	18.8	7.1

Table 6.3: Average and minimum completion time in seconds for three different step widths in the sliding task.

felt that it was convenient to search objects using only their eyes, and suggested that the system is suitable for controlling "objects beyond reach". The participants enjoyed the experience of searching in a smooth flow, "without clicking". Most remarked it was easy and natural to use their peripheral vision for searching in this task. Given the big object size and simple content, they were able to see what was in the centre while looking at the scroll arrows. However, several participants mentioned that they needed to keep their head still, which they found difficult for longer scrolling. Exaggerated eye movements (e.g. changing from left to right) to correct mistakes caused fatigue. In addition, the participants preferred faster triggering time.

6.3.3 Sliding Task

This task tested the participants' ability to accurately move a slider to a target position. Participants performed six trials, two for each of three different slider speeds, and we collected data on a total of 84 trials, 28 per slider step widths. Table 6.3 shows the participants' average completion time. The last column (minimum) indicates the fastest time that the participants achieved. However, the average completion time was much higher. Most of the time was spent on position fine tuning. Many participants missed the slider target, thus requiring longer time to correct the slider's position.

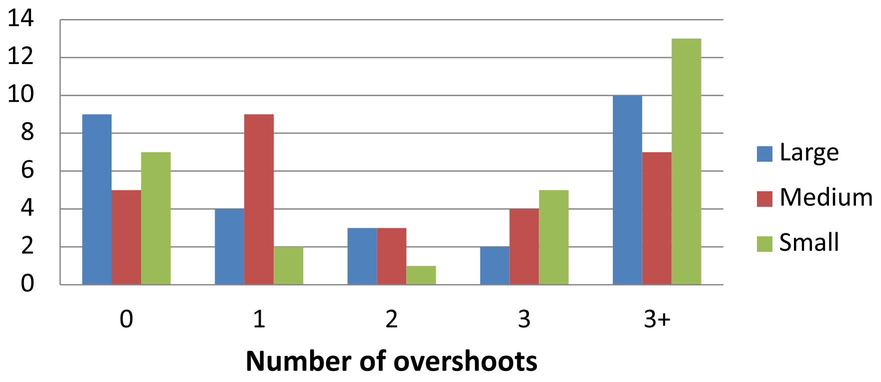


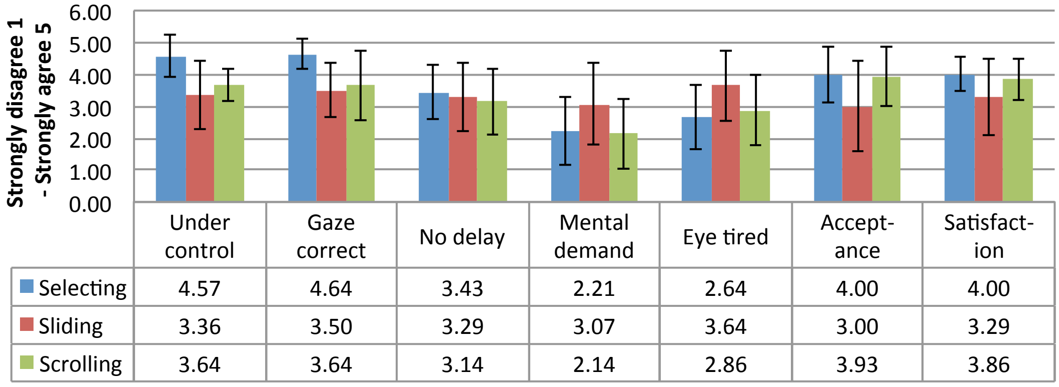
Figure 6.9: Histogram shows the count of the number of overshoots (0, 1, 2, 3 and > 3) for three different step widths in the sliding task.

We define overshooting as the number of instances when the indicator had jumped past the target location. Figure 6.9 provides a histogram that summarises observed overshooting. In 25% of all trials, participants managed to slide directly to the target without overshooting. Repeated overshoots indicate problems with accurate control, and occurred more frequently with small and large step width. The mean of overshoots was 2.7 for medium step width, but 4.3 for large and 5.0 for small step width. Note that trials always started with small step width and that results will be influenced by a learning effect.

The errors caused by the fast speed (large step width) were mainly caused by system delay. Six participants criticised that the system was not fast enough to react when gaze direction changes rapidly. In addition, several participants also mentioned that the fast speed was too fast for their eyes. One participant was having difficulty with small step widths (12.5 overshoots/trial) while performing well with the medium and fast speed (1.5 overshoots for medium steps, 2.5 for large steps). The participant was short-sighted and removed her glasses during the study. She was not familiar with stopping with the initial trials, and followed the indicator moving left and right repeatedly crossing the target.

In general, the participants found it difficult to control the sliding indicator precisely with their eyes. Their strategy was to first bring the indicator near the target location as close as possible. This was done by staring at the control arrow for continuous sliding, while using their peripheral vision to approximate the indicator's location. Once the indicator was near the target, the participants looked at the centre region to stop the sliding. They fine tuned the movement by looking at the arrow control and the centre region back and forth. For the slow and medium speeds, in many trials, the participants were able to stop before the indicator reached the target (out of 28 trials, 18 times for slow and 16 times for medium); however, with faster speed they struggled more to do that (10 out of 28 trials). On the other hand, the slow speed caused issues in fine tuning. The participants reported that it was difficult to control with precision using peripheral vision. Since the distance of each jump was small, it was difficult to judge when exactly it reached the target. Some participants struggled to use their peripheral vision, and could not see both the arrows and the target together simultaneously.

Overall, most participants found the sliding task challenging when using the fast and the slow speeds. A few participants suggested the system could be useful for moving objects in out-of-reach distance. Several participants liked the fast and accurate response of the system. They found it easy to control the direction of the sliding object by using left and right eye movement. The majority felt that they were not able to control the system for fine tuning positions, especially using small step width. Half of the participants mentioned that it was unnatural to use their peripheral vision to see detail in the centre, while looking left or right. Some participants felt that they needed to concentrate and be patient to use SideWays for



	Under control	Gaze correct	No delay	Mental demand	Eye tired	Accept-ance	Satisfact-ion
🟦 Selecting	4.57	4.64	3.43	2.21	2.64	4.00	4.00
🟥 Sliding	3.36	3.50	3.29	3.07	3.64	3.00	3.29
🟩 Scrolling	3.64	3.64	3.14	2.14	2.86	3.93	3.86

Figure 6.10: Participants' subjective feedback on use of Sideways for the selecting, scrolling and sliding tasks. Users were asked: did they feel in control; did the system respond correctly to their gaze; did the system respond without delay; did they find the task mentally demanding; did they find it tiring their eyes; would they accept the system for the task; and were they overall satisfied with use of the system for the task.

sliding. Also, the participants experienced fatigue due to frequent eye movements changing between left and right for position fine tuning. A few participants disliked the long time required to precisely move an object to the target.

6.3.4 Subjective Feedback

For each task, we asked seven questions with regards to the participants' subjective feedback. The results are presented in Figure 6.10. We run the Friedman Test on the subjective feedback data. Post-hoc analysis with Wilcoxon Signed-Rank Tests was conducted with a Bonferroni correction applied.

We asked the participants whether they felt they were in direct control of the SideWays system. One participant particularly commented "I can feel a real power, starting from my brain and ending on the screen." We found a significant difference between the three tasks, $\chi^2(2)$=17.07, p<.001. A post-hoc analysis showed that the selecting vs sliding (p=0.002) and selecting vs scrolling (p=0.002) pairs were significantly different. The participants felt that they were most in control when using our system for selecting objects.

We found a significant difference in whether the participants perceived the system as responding correctly to their gaze, $\chi^2(2)=14.70$, $p<.002$. The selecting vs sliding ($p=0.001$) and selecting vs scrolling ($p=0.012$) pairs were significantly different. The participants felt that the system was most responsive when used for selecting objects.

We also found a significant difference in tiredness of eyes in using *SideWays*, $\chi^2(2)=7.54$, $p<.023$. The participants felt significantly more tired when using our system for sliding than for selecting ($p=0.006$).

Participants provided comments in comparison of scrolling and sliding. Sliding was found demanding as it required target status check in the centre of the screen while concentrating gaze on either left or right. For scrolling, participants noted that they can see what is coming while they are controlling, and they found it less demanding to use peripheral vision to check what is in the centre, as it was displayed more largely than in the sliding task.

Participants also commented on possible applications of the system. Since SideWays is touch-free, the interaction is sanitary and therefore suitable for public environments, such as airports, libraries and shopping malls. Some participants suggested that it could benefit disabled people with paralysis. Several participants criticized the lack of visual feedback for the detection of gaze directions. This is important as it provides indication of whether the system interpreted the user's input correctly.

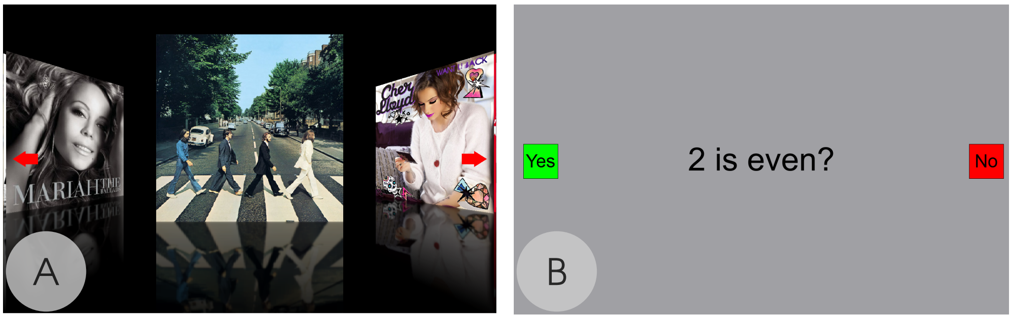


Figure 6.11: Two SideWays applications used in the user study: *album cover browser* allows users to navigate music albums; *gaze quiz* presents a question in the centre and uses the side inputs for "Yes/No" answers.

6.4 Applications

After completing the previous three tasks, participants interacted with two SideWays applications (see Figure 6.11): an *album cover browser* and a *gaze quiz*. The album cover browser acts as an interface of a music jukebox. A user browses for music by scrolling left and right, and the centre region represents the music album to play. The gaze quiz application is an interactive quiz game. A user first reads a question displayed in the centre, and then answers the question by selecting yes or no, which was placed on the left and right positions of the screen, respectively. We used a window size 3 for the album browser and a window size 4 for the gaze quiz. Both applications used discrete activation. Only qualitative feedback of the system (e.g. preference and suggestions) was collected.

Participants were allowed to use the interfaces freely. We gave no instruction of how to interact with the interfaces, but the participants were still able to use the applications. For the media browser application, the users were able to navigate through all the music album covers and check what music was available. Sometimes, they scrolled back/forward to stop at the one they were interested in.

During the gaze quiz application, the participants read twelve questions displayed in the centre and provided answers for all the questions. All participants understood how to search the media album covers and make selections. One participant encountered the Midas Touch Problem in the gaze quiz [87]. While we displayed one question with a long sentence, the participant accidentally chose the answer "No" (on the right end) as he was reading the sentence. Thus, information in the centre should not be extended to the far left or right regions of the display.

Participants further suggested that SideWays could be applied in situations where a display is obstructed by a glass wall or window, such as shop displays for pedestrians. Another suggestion was for controlling television, e.g. adjusting volume or switching channels.

6.5 Discussion

Our study validates that SideWays enables eye gaze as input for interactive displays, without the need of prior calibration or specialist hardware. This is significant in a number of ways. First, achieving robust gaze control, albeit coarse-grained, without need for calibration means that our system is person-independent. Any user can walk up to a display fitted with our system, and interact with it using their eyes only. Secondly, as we overcome calibration, users will be able to interact immediately (in principle) which is important for serendipitous and short-lived interactions that don't warrant preparation phases. Thirdly, we achieve gaze control with an off-the-shelf camera. This means that displays can be made gaze-aware at low cost, potentially on pervasive scale.

6.5.1 Effects of System Parameters

The threshold T_r (used in equation 6.2) defines the size of the central region, and in our prototype was set to corresponds to a horizontal visual angle of 40°. Most participants were aware how far they needed to look left/right as SideWays provided guidance by displaying the visual control stimulus. Increasing the threshold essentially increases the visual angle. If the threshold is small, the central region becomes narrower, and the system also becomes more sensitive to small eye movements of looking left/right. However, if the threshold is large, the user will need to look left/right further to trigger input which might cause discomfort and fatigue. The optimal threshold will depend on application, and designers need to consider the distance between the user and the display and the size of the display.

The window size determines how much evidence is collected before an action is triggered and trades off between accuracy and speed. A large window size improves the accuracy of gaze detection, but causes longer delays and slower response. This is suitable for discrete actions (e.g. selecting an object), where it is more important that the system detects the correct object. A small windows size speeds up response, but increases likelihood of errors caused by noise. This will be reasonable for continuous actions (e.g. sliding) where faster response is important and where the effect of occasional misclassifications will quickly corrected by continuous updates.

Designers can map input state to discrete versus continuous actions to fit the nature of the task. For example, if the content is visual (e.g. a photo album), continuous action may be chosen for fast scrolling as our study participants found that they can scroll to larger distinctive objects with peripheral vision. If the content requires attention (e.g., flicking through book pages), a discrete action mapping is better suited.

6.5.2 Limitations and Design Considerations

Responsiveness. When the system is not able to detect eyes images of sufficiently quality for computing eye centre and eye corners, the interface responsiveness decreases. This happens when users blink and squint or when the eyes get occluded in any other way, and can also be caused by larger head movement. When the system does not respond correctly or fast enough, no manual intervention is needed to reset SideWays. Participants reported that they reinitialized SideWays by closing their eyes, or by adjusting head positions or distance to the screen, indicating a good understanding of what causes misfunction and how to recover.

Head orientation. Our system requires users to keep their head oriented toward the centre of the display and only move their eyes. Some participants commented that this was unnatural, because they often subconsciously turn their head towards the direction of their visual focus. As a result, the detection of gaze direction becomes unreliable. This poses a limitation of user interaction. Although restricted head movement was commented as unnatural, in general, the participants were able to recover from loss of gaze detection by quickly correcting their head orientation, in order to achieve their tasks. To minimize head turning, designer should pay attention to the display region where information changes. Dynamic movement on the control regions can attract user attention; hence, causing head turning.

Provide detection feedback. Users need feedback when an event is triggered, to understand whether the system has detected their gaze. Visual feedback can be explicit confirmation of users input, for example by highlighting a displayed control that was triggered, or implicit in the behaviour of application, for example by updating the content displayed in the centre of the screen. However, when users glance sideways to trigger a control, it can be difficult for user to acquire feedback that effects only the centre of the display.

Avoid fine-tuning. Results from our sliding task showed that users often overshoot a target with SideWays. Users felt least in control in that task. When precision is required, the user needs to look left and right rapidly for fine tuning. Rapidly changing gaze directions can cause fatigue and discomfort. Designers should avoid using SideWays for tasks that require adjustment for precision. Instead of having a precise target location, it is better to set a target region.

6.6 Conclusion

In this chapter, we presented SideWays, a novel eye gaze interface for spontaneous interaction with a display. It leverages left and right regions of the display for gaze controls, while keeping the central region for display content. We conducted a user study to evaluate SideWays on three interactive tasks. The results show that people are able to use SideWays to interact with a display, and we have gained insights of people's gaze control strategies.

This chapter addresses the question of how to design gaze interfaces to support spontaneous interaction that do not require accurate eye tracking systems. We demonstrated that coarse gaze tracking is sufficient to allow people to control contents on a display using their eyes. The proposed techniques enable intuitive, fast and hands-free interactions that are suitable for many pervasive display applications, such as public exhibitions.

Person-independence and interaction without preparation are critical steps toward genuinely spontaneous interaction with displays we encounter in public environment. While our evaluation shows that our system achieves both, it does so under the constraints of a controlled study designed to systematically test and characterise the interaction techniques. Deployment in real-world contexts naturally raises a range of further challenges. For example, although a calibration phase is avoided, there will still be a gulf in how users can readily obtain and use gaze control over a display they encounter. However, with a lab study of our system, we now have a foundation for addressing deployment challenges, as well as insights on user performance and strategies that can inform application design. For ecological validity, it will is important to inform further development by understanding of how users interact with displays “in the wild".

7

Gaze Interaction for Public Display

In this chapter, we address the question of how to deploy a gaze interface in public environments. In pervasive contexts, users encounter displays spontaneously and usually have no prior knowledge of displays' interactivity. Their interactions are usually unplanned and of short phase. In addition, human assistance is not as available as that in lab settings. The objective of this chapter is to overcome this design challenge and investigate how to enable gaze interaction without any human assistance.

Gaze input has long been confined to controlled settings but there is an increasing interest in using gaze for public display contexts. Gaze has the advantage that it naturally indicates user interest, and that users can express gaze input fast [196], over a distance, and without the need of external devices. Recently, a variety of techniques have been developed that leverage eye movement in novel ways for ad hoc interaction with displays, easing and overcoming the need for calibration to individual users [176, 201, 137]. In spite of these advances, deployment of gaze tracking with public displays has remained limited to passive observation of user attention [187].

A question that arises for the design of gaze interfaces for public display is how passers-by can discover and use gaze control “on the spot". Gaze interface design conventionally assumes that users know how the interface works and how to position themselves to use it, as a result of prior training or because they are guided by an expert (e.g., in a usability lab). However, passers-by attend spontaneously to public displays and usually without prior awareness of what interactions a display supports. Gaze interfaces offer no physical affordance and their availability is not directly visible to users. While we can often quickly glean how an interface works by observing the actions of others using it, this is not possible with gaze interfaces that are controlled by subtle eye movements. The challenge of making passers-by aware of the interactive affordances of a display has been highlighted previously, in studies of public multi-touch and gesture interfaces [98, 124, 181]. Other related work considered social factors in enticing users and groups to approach displays and begin interaction [30, 133, 116]. Yet, no previous research has explored how gaze interaction can be bootstrapped in a public display context.

In this chapter, we present *GazeHorizon*, a vision system designed to enable passers-by to interact with a public display by gaze only (illustrated in Figure 8.1). The system supports multiple interaction phases and adapts its behaviour when a user approaches the display (leveraging the proxemic interaction concept [66, 179]). It detects when users walk up to a display and provides interactive feedback to guide users into appropriate proximity for use of the gaze interface, and to provide

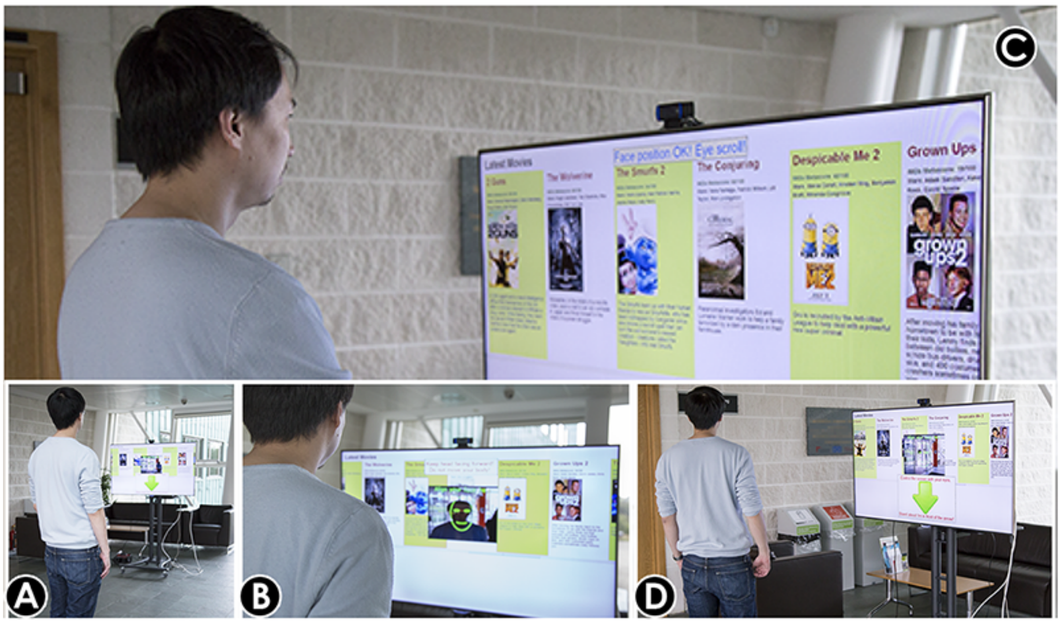


Figure 7.1: GazeHorizon enables passers-by to interact with public displays by gaze. The system detects users walking up to a display (A), provides interactive guidance to bootstrap use of gaze (B) and lets users navigate displayed information using horizontal eye movement (C). When a user walks away, the system is immediately ready for the next user (D).

cues to help users operate the interface. The gaze interface itself is purposely simple but robust, using horizontal eye movement of the user as relative input and an intuitive mapping to navigate display content (contrasting SideWays where eye movement to left or right was mapped to screen regions [201]). When a user turns away from the display, it is immediately ready for the next user. GazeHorizon has been developed through iterative field studies, eliciting insight into the guidance needed for users to be able to use the gaze interface and testing the efficacy of visual cues and interactive feedback. The system was then deployed in the wild for four days to evaluate how passers-by would interact with it.

7.1 Design of GazeHorizon

The goal of GazeHorizon is to create a system that allows any passes-by to walk up to a display, and to navigate the displayed information using only their eyes.

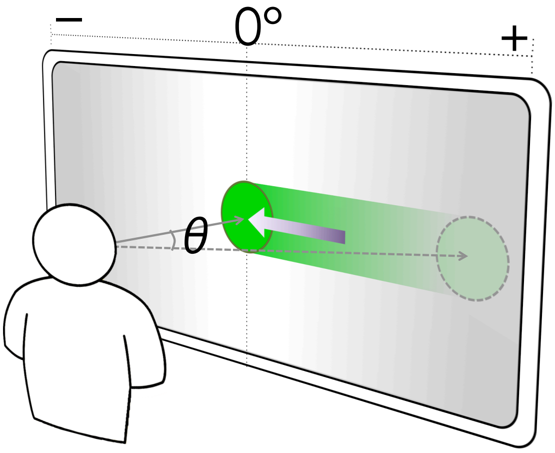


Figure 7.2: GazeHorizon maps gaze direction for rate-controlled scrolling with a self-centering effect. If the user looks at an object on the right, the display will scroll to the left; as the user's gaze naturally follows the object-of-interest, the scrolling speed will decrease and bring the object to a halt in the centre.

7.1.1 Interaction Model

The underline eye tracking mechanism is based on the PCR technique described in Chapter 4. The PCR technique is a generic device that provides relative input; it does not capture a gaze position, but a gaze direction. Conventional eye gaze mappings are absolute, where a gaze direction is directly associated with a point on the screen. In GazeHorizon, we use a relative mapping instead, for rate-controlled navigation of horizontally organised information on the screen (as illustrated in Figure 7.2).

The mapping is designed to leverage implicit navigation of the displayed content, where the scrolling action is triggered by attention to the content, and not perceived as an explicit command. The design follows the principle such that an object moves to the centre of the display when a user looks at it. This results in the scrolling effect of moving content towards the centre. If the user follows the content as it becomes centred, the scrolling speed decreases and the content comes to a halt in the centre of the display. The control users gain over the display is limited to navigation along one dimension, but the mapping provides a robust user experience where the gaze control task is naturally aligned with the visual perception of the displayed content.

In order to provide a stable display experience, we do not map PCR uniformly to scrolling speed, but define lower and upper thresholds for visual angle relative to the display centre. This allows us to set scrolling speed to zero for a wider region in the centre of the display, and to a maximum constant for larger visual angles. Note that the system is based on gaze direction and does not actually predict what the user is looking at. However, the interaction design provides a robust illusion of response to the user's point of regard. We suggest this mapping of gaze is useful for exploration of larger information spaces that can be presented in one dimension along the horizontal axis. A good example is a timeline, of which the display reveals a part at any time. The further the user looks toward the edge of the display, the faster it will scroll to the effect of revealing new information "in the direction of interest".

7.1.2 System Implementation

The system is designed for deployability in the wild, and requires only a single off-the-shelf web camera – placed either in front or on top of a display – to capture the presence of users (at a frame rate of 30Hz and a resolution of 1280×720px), and supports interaction of one user at a time. From a system's perspective, the interaction involves the following phases (see Figure 7.3):

Face detection and tracking : For every thirty frames, the system scans for the presence of users looking towards the display, by detecting frontal faces. If a group of users approach the system, only the person standing in the centre of the screen (i.e. the face positioned in the central region) is tracked continuously.

Eye tracking: When the image resolution of the tracked face is larger than 200×200px, the system extracts eye feature points from the face image. The size restraint corresponds to a user standing at a distance of approximately 1.3 meters away from the camera.

Gaze interaction: The system uses the extracted eye feature points for computing horizontal gaze direction. The gaze direction is mapped to rate controlled scrolling of the display. When the user looks to the left, the display is scrolled to the right (and vice versa), and the scroll speed varies based on how far away from the centre that the user looks.

Reset: If the system detects the user's face has disappeared from its field of view, the system resets back to the initial *Face detection and tracking* phase.

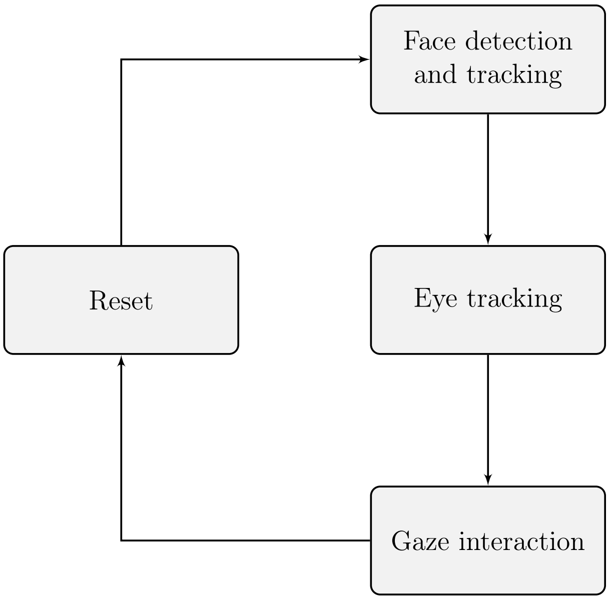


Figure 7.3: System diagram of GazeHorizon

7.2 Methodology

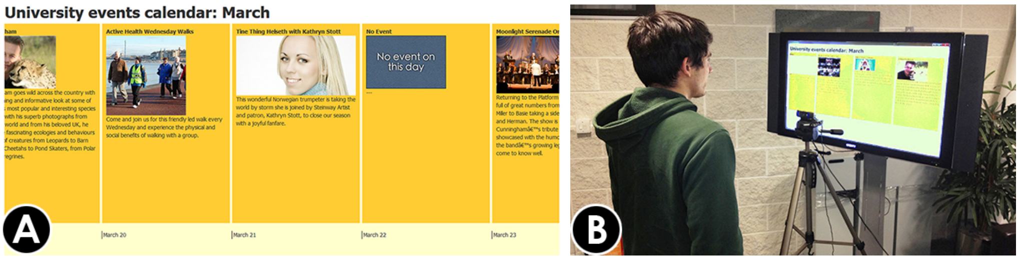
Gaze is currently an uncommon modality. General public are unfamiliar with gaze interaction. Our aim is to create a standalone interface that communicates gaze interactivity to novice users, so they can comprehend the interaction model of GazeHorizon.

We conduct a series of studies to evolve the interface design of GazeHorizon. In the first study, we prompt users to find out what levels of guidance they require to use the system. Then we integrate different levels of guidance to provide visual cues in the interface. We test the effectiveness of the visual cues through a second field study. Finally, we deploy the system in the wild to observe how passers-by interact with GazeHorizon without any human assistance, and we also interview them to get subjective feedback.

7.3 Field Study 1: Requirements for Interactive Guidance

The aim of this study is to understand what information novice users need in order to use GazeHorizon; hence, the level of guidance they require to figure out the interaction model. We thus conducted the study in the field by deploying our system in a public area.

We conducted the study over a two-day period, in the reception area of a university building(see Figure 7.4). The building hosts a computer science department, a café, and numerous technology companies. People with various technology



University events calendar: March

ham
am goes wild across the country with
ing and informative look at some of
most popular and interesting species
with his superb photographs from
world and from his beloved UK, he
fascinating ecologies and behaviours
f creatures from Leopards to Barn
Cheetahs to Pond Skaters, from Polar
egrines.

Active Health Wednesday Walks
Come and join us for this friendly led walk every Wednesday and experience the physical and social benefits of walking with a group.
March 20

Tine Thing Helseth with Kathryn Stott
This wonderful Norwegian trumpeter is taking the world by storm she is joined by Steinway Artist and patron, Kathryn Stott, to close our season with a joyful fanfare.
March 21

No Event
No event on this day

March 22

Moonlight Serenade Or
Returning to the Platform
full of great numbers from
Miller to Basie taking a side
and Herman. The show is
Cunninghamâ€™s tribute
showcased with the humc
the bandâ€™s growing le
come to know well.
March 23

Figure 7.4: Field Study 1: (A) GazeHorizon event browser interface. (B) An illustration of the study setup.

background passed by everyday. The area was illuminated by ambient day light, and ceiling light was turned on during late afternoon.

7.3.1 Procedure

During the study, a researcher invited passers-by to take part. The researcher introduced the system as an interactive display that showed local events, but never revealed the system was eye-based during the introduction. Thereafter, the participant was encouraged to experience the system.

To test for intuitiveness, we evaluated the amount of instructions (or guidance) that users needed to comprehend the operation correctly. We predefined the instructions into five stages. Table 7.1 lists the instruction protocol and the number of participants who needed the levels. The researcher started off by inviting participants to stand in front of the display (hence, giving *L1* instruction) and then prompted the participant for their perceived interaction model. If the participant answered incorrectly or required further guidance, the researcher gradually revealed the next level and prompted the participants again. This continued until the participant realised the correct operation or the whole set of instructions was revealed.

After experiencing our system, we interviewed the participants for qualitative

Instruction Levels and Hints		Count
L1	Stand in front of the display.	10
L2	The system reacts to eyes.	11
L3	Keep head still, face towards the display, and move eyes only.	5
L4	Look specifically at each event object.	0
L5	Look at an event object and follow it. The object stops in the centre.	0
*	Failed to use the system after all five levels were revealed	4

Table 7.1: (Field Study 1) The five levels of guidance we provided to our participants, in ascending order. The count column indicates the number of participants who needed up to that level of instruction to determine the interaction of GazeHorizon.

feedback. We first asked a set of predefined questions, and then further prompted them with probing questions for detailed explanation.

7.3.2 Results

In total, we invited 30 passers-by (8 female). Seven participants wore glasses, and five of them removed their glasses during the study. The participant stood at an average distance of 120cm (*SD*=10) away from the display.

Levels of Guidance Required

In total, twenty-six participants successfully comprehended the operation, and twenty-four of them acknowledged that it was easy to figure out. Ten participants required only level 1 instruction. They commented that the system was "*easy*", "*self-explanatory*", "*just look at the events... and focus on what I want to read*". They explained that they realised the system was eye-based when they noticed the movement of the content corresponded to the movement of their eyes. For instance, a participant reported that "*[the picture] moved when my eyes moved*".

Eleven participants needed up to the level 2 instruction. They explained that gaze input was uncommon. For example, a participant said "*I felt something is moving; however, I am not sure what [the system] reacts to.*" Six participants first attempted to touch the screen, wave their hands or use body gestures. They imagined the system was "*motion-based*" because they were aware of existing motion capture technology (e.g., Kinect). However, after we revealed that the system was eye-based, the interaction became "*obvious*". The participants mentioned that the level 2 instruction was important, as it eliminated them from attempting to use their body for interactions.

Five participants needed up to level 3 instruction. When they looked at the displayed information, they turned their head to face towards it. After they were told to keep their head still and face towards the centre of the display, the participants realised the system reacted to eye movements.

Four participants failed to use our system. Two of them could not understand the interaction model, even full instructions were given. Post study analysis revealed that one participant failed because the system failed to detect his pupil centre. Also, one participant declined to retry after a failed attempt.

Three Patterns of User Operation

We observed three different ways of how the participants used our application. The majority of the participants' perceived concept were in line with the interaction model we designed. They preferred to read detailed information from the centre of the display. One participant related this to his desktop; he arranged items (e.g., windows and icons) in the centre for a better viewing.

The second pattern was acknowledged by six participants. They sometimes read short sentences immediately as they entered the display from the side. However, a few participants found that "*it was disturbing*" trying to read moving information, which easily caused them to lose focus. Even though the information was moving, as the participants followed it towards the centre, the information slowed down and eventually became stationary. They can then read the information. Also, if users turn their head to look towards the side, the scrolling will halt, because the system could no longer detect the presence of a frontal face looking towards the screen. This inherently allows the participants to read information on the side by turning their head.

The third pattern was identified by two participants. They found that the system was difficult to control, even though they were fully informed of the instructions. They got distracted easily by moving objects, so they often moved their eyes to look for new information on the display. As a result, the interaction model failed because the participants did not fixate on the centre to stop the scrolling. This is essentially the Midas Touch Problem. The scrolling action is triggered unintentionally and causes user frustration.

Alternative Interactions Discovered by Users

Surprisingly, five participants discovered that they can fixate on the centre of the display and slightly turn their head; "*the movement of the picture is synchronised with head movement*". This is similar to Mardanbegi *et al.*'s head gestures technique [115]. After a few trials, the participants acknowledged that "*[the system] was less responsive to head turning*" and they could not focus on what they wanted to see. Also, a participant suggested that she could stop scrolling by looking downward, such as staring at the event's date at the bottom. Looking downward caused her eyes to be occluded by her eyelids and eye lashes, so the the system stopped detecting her eyes.

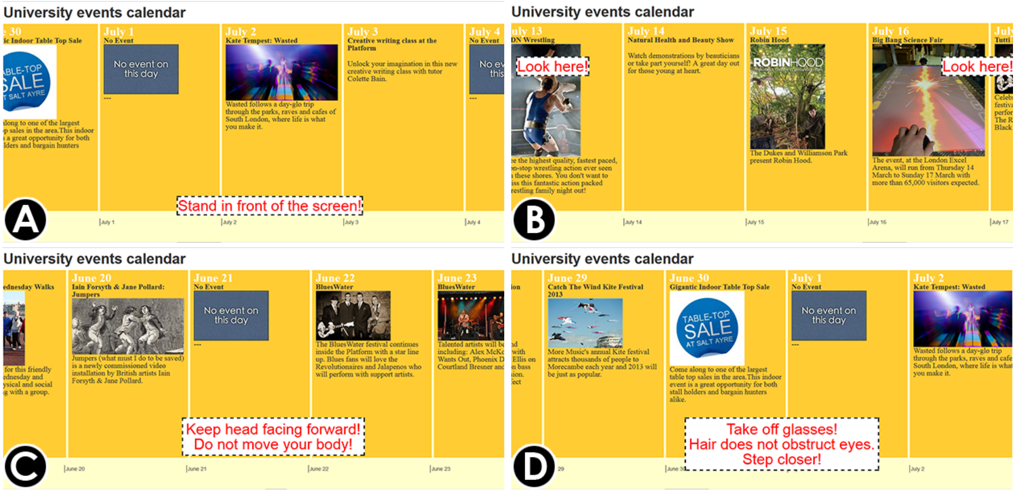
Summary

From this study, we found that many people are unaware of gaze as an input modality for large display navigation. Our results revealed that *L2* & *L3* instructions are vital for communicating the gaze interactivity. Thus, we translated these levels into visual cues, and embedded the guidance in the interface.

7.4 Field Study 2: Testing Interactive Guidance

Our aim is to design a standalone application, where users interpret the interaction solely based on information given on the interface. From field study 1, we learned that users needed three levels of guidance: (1) *position* (stand in front of the display), (2) *eye input* (the system reacts to users' eyes), and (3) *head orientation* (keep head facing forward and move eyes only). We embedded these three levels as visual cues on the interface for instructing users.

Previous research showed that textual information was very effective in enticing interaction on public displays [98], so we added the instructions as text labels. To further attract users' attention, we added pulsing effect (where labels enlarged and reduced continuously) [123]. The system presents visual cues in multiple stages for interactive assistance (see Figure 7.5): In the first stage, the interface displays a message to invite users to stand in front of the display; Once the system detects the presence of a person, it displays "look here" labels, which indicates at where the user should look; If the system detects the user is not facing forward (e.g. the head is turned), the system displays a "keep head facing forward" message; Also,


University events calendar
A
University events calendar
B
University events calendar
C
University events calendar
D

Stand in front of the screen!

Look here!

Look here!

Keep head facing forward!
Do not move your body!

Take off glasses!
Hair does not obstruct eyes.
Step closer!

Figure 7.5: Field Study 2: Stages of visual cues. (A) Position (B) Eye input (C) & (D) Corrective guidance

if the system detects the user's face, but not the eyes, the system assumes that something is occluding the eyes or the user is too far. The interface suggests the user to ensure their eyes are not occluded and to step closer when the detected face is too small.

7.4.1 Procedure

We conducted a second field study to test whether users can translate the visual cues into user operation. We deployed our system in the reception area of a university research building. We only invited novice users who did not participate in our earlier study. The conversations between the researchers and the participants were strictly limited to invitation, and we provided no assistance.

7.4.2 Results

We conducted the field study over a two-day period. In total, 35 passers-by (aged between 18 to 41, plus a child) tested our interface; 6 failed to use the system due to: strabismus (crossed eyes), myopia (blurred vision without corrective lenses), wearing tinted glasses, standing too close to the screen, not noticing the visual cues, and a child whose height was too short. Interviews revealed that most of the participants found the visual cues informative – especially the "*Look Here*" label, but suggested that it was only needed for a short duration – and helped them to realise the interaction very quickly. They commented that the instructions were "*clear*" and "*self-explanatory*". Two users mentioned that the "Hair should not obstruct eyes" label was helpful for people with long hair fringe. Also, the pulsing effect naturally drew their attention, and it was very efficient for communicating interactivity.

We found the majority of the participants followed the displayed instruction correctly. In general, the first two levels of information (see Figure 7.5(A) & 7.5(B)) were crucial for apprehending the correct interaction. If no errors occurred, the visual cues for correcting guidance (see Figure 7.5(C) & 7.5(D)) did not appear. A few users commented that some of the textual messages were too long, and suggested further improvement could include graphical guidance (e.g. pictures) or simplified textual information (e.g. shortening phrases or highlighting keywords).

Floor Marker vs. On-screen Distance Information

As described in an earlier section, the distance of how far a user stands from the camera affects the tracking of the user's eyes. During our pilot test, the initial interface gave no precise information of how far and where the users should stand (see Figure 7.5 (A)). We noticed that our pilot users often stood too far (over two meters) away from the screen, so the system failed to detect their presence and remained non-interactive. This never occurred in the previous study as the setup was different and a researcher gave assistance. To help users positioning

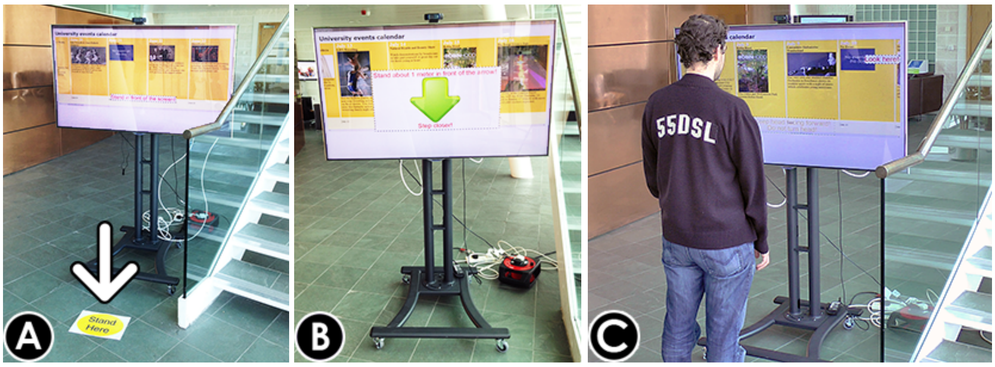


Figure 7.6: Field Study 2: User positioning. (A) Floor Marker (B) On-screen Distance Information (C) A user testing our interface.

themselves, this study also tested two approaches (see Figure 7.6): (a) using a *floor marker*, by placing a "*Stand Here*" sign on the floor; (b) providing an *explicit distance information* on-screen, where the display showed a label informing users to stand at a distance of one meter away.

During the first day of the study, we used a floor marker for helping users to position themselves. Some participants initially ignored the floor marker and only realised it later when they looked down. This indicates that people easily noticed the visual cues on the display, but not the cues on the floor level. On the second day, we removed the floor marker and the interface explicitly displayed the distance. This was more effective. All the participants noticed the cue, but required longer time for adjusting themselves to the correct distance and position.

Summary

This study confirmed our translation of the minimum required levels from Field Study 1 to interactive visual cues on the user interface. Our study showed that many people were able to follow a sequence of guidance labels on the display to figure out the interaction of GazeHorizon. Furthermore, we also learned that all visual guidance should be shown on the display (on the same level as the user's field of vision), otherwise labels placed outside the display could be unnoticed.

7.5 Field Study 3: GazeHorizon in the Wild

To understand how people use our system without human assistance, the objective of this study is to determine the general effects of GazeHorizon on passers-by in an ecologically valid setting. To maintain validity, we neither invited, interfered nor advised the passers-by; instead, participants were enticed purely by the interface.

7.5.1 Procedure

We implemented GazeHorizon as a browser of latest movies, and we deployed the system in the lobby of a university building, in Germany. Many people passed through this area everyday. They were mainly university students, staff and visitors. We used a 45-inch display, positioned at a height of 170cm above ground, and we mounted a web-camera on top of the display.

During deployment, the system logged anonymous data of users' eye images and timestamped system events. We placed a video recorder opposite to the screen for capturing user behaviours. After the users finished their interaction, a member of the research team approached the users for feedback. In the post study analysis, two researchers independently analysed the log data, the recorded videos and the interview recordings.

We adopted Schmidt et al.'s two-phase deployment approach [148]: *Optimisation* and *Testing*. During day 1 and 2 (optimisation), we performed several iterations of improving the interface according to our observations and users' feedback.

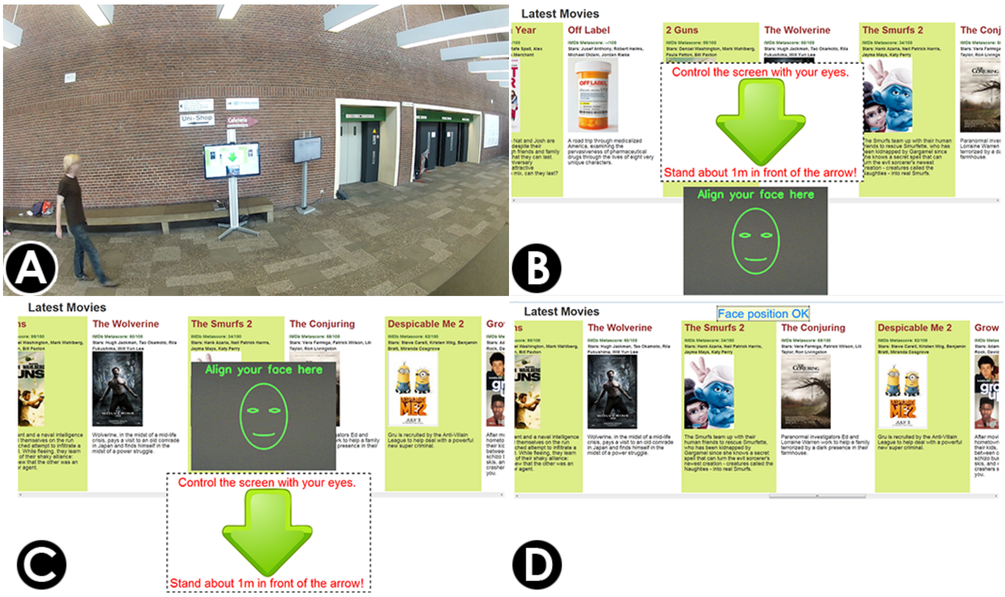


Figure 7.7: Field Study 3: in the wild deployment. (A) We deployed GazeHorizon in a lobby of a university building. (B) Initial interface before optimisation. (C) After initial optimisation, we moved the mirrored video feed to the central region. (D) We also amended the interface for constant face alignment feedback.

During day 3 and 4 (testing), the prototype was not modified, and we conducted detailed interviews with users.

Interface Optimisation

During the first two days, we interviewed 46 users for qualitative feedback. We asked the users of the issues that they encountered and for suggestions for improvement. Based on the issues reported by the users and those we observed, we made amendments to the interface. This was done iteratively until a sufficient number of users were able to use the system.

To assist users positioning themselves, we added a mirrored video feed, overlaid with a face outline, on the interface (Figure 7.7). Using video feed helped to communicate interactivity of the display [124]. In our initial interface, the mirrored video feed was positioned at the bottom of the screen (see Figure 7.7(B)). When the users looked down, their eye lashes/lids often occluded their eyes, which prevented the system from detecting their pupil centres and eye corners. This inherently slowed down the system detection. We resolved this by moving the video feed to the centre of the display (see Figure 7.7(C)).

Also, the original video feed was constantly shown at the bottom of the screen. Some users criticised that it was distracting. We changed the video feed to disappear after a user's face was aligned correctly. However, without the video, the users reported that they were unsure of whether their face was still aligned correctly. To provide constant feedback, we added a "Face position OK" label on the top region of the screen (see Figure 7.7(D)). This label only disappeared if the user's face was out of alignment. Other minor changes include changing the colour of labels to give higher contrast.

7.5.2 Results

The log data revealed a total of 129 interaction instances, where each instance contains a full episode of uninterrupted use, either by one or more users [133]. Of the instances, 107 triggered continuous scrolling, with a mean interaction time of 67.1s (*SD*=54.2s). Figure 7.8 shows a histogram of the interaction time. Most users interacted with our system for between 20 to 80 seconds. We were surprised that one user spent over five minutes. She explained that she really enjoyed movies, and she spent most of the interaction time reading the synopsis. From the moment when the system detected users' presence, on average, the users required 4.8s (*SD*=8.5s) to align their face into the correct position and 7.2s (*SD*=11.0s) to perform a scroll (also measured from the same beginning moment). Over the entire interaction duration, the users spent 27.0% (*SD*=15.1%) of the time for

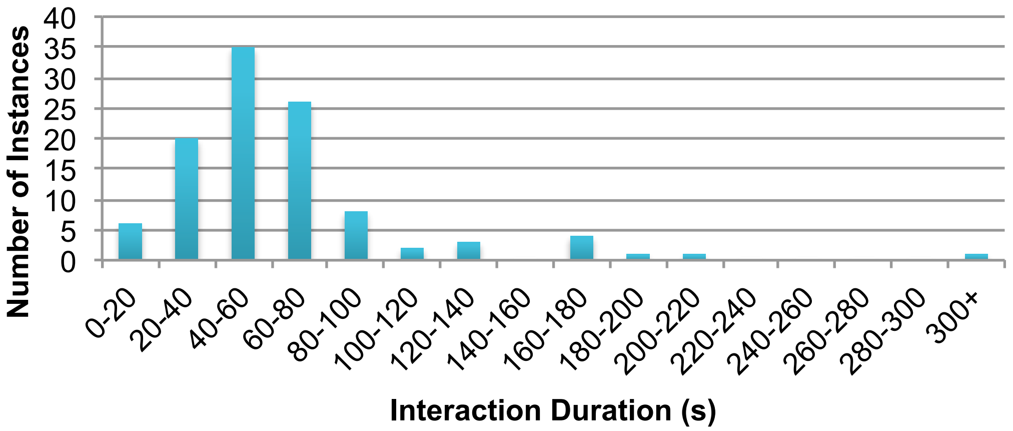


Figure 7.8: Field Study 3: A histogram of overall users' interaction time over four days of deployment.

scrolling the content.

During the optimisation phase, we interviewed 46 users, and 35 of them (76.0%) reported that they were able to use the system for scrolling information. This rate increased after optimisation. We interviewed 41 users during the testing phase, and 35 of them (85.4%) reported that they were able to scroll the content. Over the four day period, we observed 20 users who wore glasses, and 9 of them were still able to use our system without removing their glasses.

Group Behaviour and Sharing Experience

Passers-by sometimes approached our system in groups, but usually one person interacted with our system at a time. People were more willing to try if they saw another person successfully used the system. If a user was able to comprehend the interaction, the user would encourage other members to experience the system, so the group were more likely to try. Also, people in a group helped others by pointing to or reading out the displayed instructions.

We observed the *honeypot effect* [30]. Passers-by became curious after noticing someone using GazeHorizon (see Figure 7.9). Spectators first positioned them-

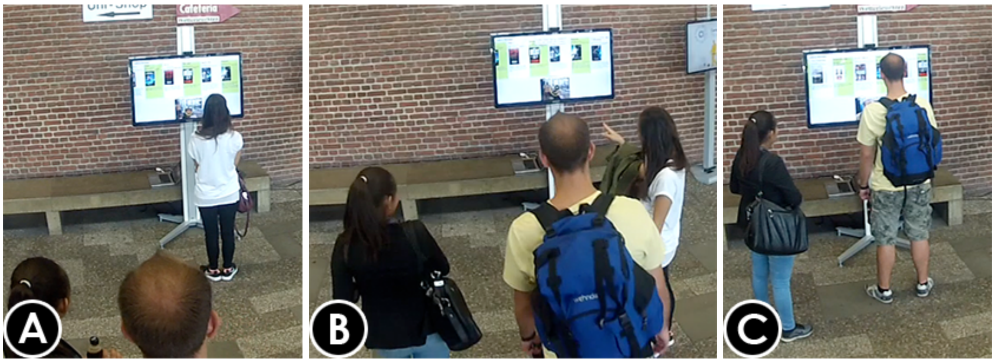


Figure 7.9: Field Study 3: Honeypot effect. (A) Two passers-by observed a user. (B) The user explained the interaction. (C) The passers-by tried the system.

selves behind the user and observed from a distance, without disturbing the user. When the user noticed people were observing, the user often explained the interaction and invited the observers to try. We noticed instances where strangers engaged in short conversations to discuss about the operation of our system.

Interacting with GazeHorizon Display

The textual label "Control the screen with your eyes" gave an obvious hint to users that the system is gaze interactive. The majority of users realised the interaction by noticing the movement of content when they looked at the display. Several people explained that the movie images first attracted their attention, and then they realised the interaction model when the pictures moved. For example, "*followed it [a movie's picture] until the article [the synopsis] got slower ... in the centre it stopped*". Some users commented that they got comfortable with the system very quickly, and after 2 to 3 scrolling attempts they realised that they did not need to stare at the "look here" label for scrolling content.

A few users explained that they were attracted by the novelty effect of "*using eyes to control*" scrolling. They were mainly interested in experiencing the new

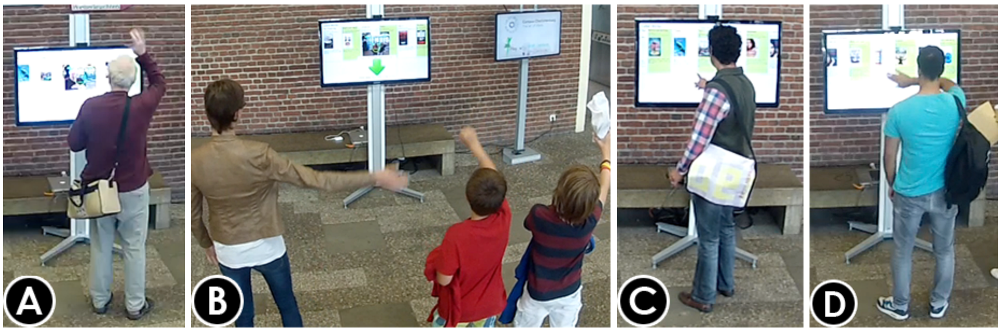


Figure 7.10: (Field Study 3) Users' expected interactions. (A) & (B) Users attempted to interact with the system by waving their hands. (C) A user attempted to make a selection by touching the screen. (D) A user attempted to scroll content by performing a touch and swipe gesture.

form of interaction and attempted the system to seek for different effects, for example, "*it scrolls faster when look more to the sides*". Other users who were interested in movies usually browsed through the entire content and interacted with the system for much longer.

Currently, gaze is still an uncommon modality, and many people are unfamiliar with using their eyes for interaction. When our users first approached the system, they sometimes did not read all the displayed cues, so they did not immediately know that the system was controlled by eye movement. Without knowing it was eye-based, some people waved their hands to attempt to interact with the system (see Figure 7.10(A) & 7.10(B)). After the users noticed no responses, they would then read the displayed cues and follow the given instructions. However, some users were impatient and abandoned further attempts after seeing no responses.

Although our system was designed only for scrolling content, some people expected the content to be selectable. We noticed that, after some users had successfully scrolled the content, they touched the screen in an attempt to trigger a selection (see Figure 7.10(C)). Interviewees suggested using double touch, head nodding, facial expression (e.g. open mouth), blinks or winks, as well as stare at an object for a few seconds (i.e. dwell time), to activate a selection. Also, for some users, even though they knew the interaction was eye-based, they attempted to use touch and swipe gesture to control the content (see Figure 7.10(D)). Two users suggested "*vertical scrolling*" for more text when the content stops in the centre.

Self Positioning

We noticed that several failure instances were caused by people standing at the location out of our system's tracking range (see Figure 7.11(A) & 7.11(B)). For instance, some users stood too far away from the camera; some noticed the video feed but did not want to be recorded, so they stood out of the camera focus. Either way, the camera failed to detect the users' presence. Another case was that, instead moving eyes, users turned their head towards the sides (see Figure 7.11(C)).

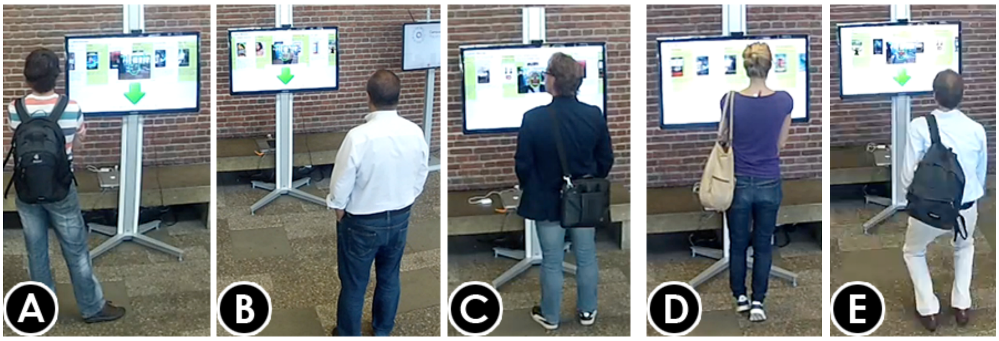


Figure 7.11: (Field Study 3) Examples of users' patterns of interaction. *Common causes of failure:* (A) Standing on the side of the display. (B) Standing too far from the display. (C) Turned head to look at the side. *Height adjustment:* (D) A short user lifted her feet. (E) A tall user bent his knees.

For the majority of novice users, they intuitively realised that they needed to align their face to the outline of the video feed. Although a label explicitly informed the user to stand one metre away from the display, interviews revealed that users had no reference for estimating the one-meter length. Instead, they judged their position by aligning their face. The users explained that they used the video feed for reference, as it provided realtime feedback.

When users positioned themselves, they were often able to stand correctly in the middle but too far away from the display. People stepped back and forth to adjust their distance, and then fine-tuned by leaning/tilting their body, without moving their feet. Tall users tended to bend their upper body or their knees to lower their height, while shorter users lifted their heels (see Figure 7.11(D) & 7.11(E)). We observed an instance where children jumped up and down to align their faces. To accommodate different height, the system can use a camera with a wider vertical angle for detecting users.

Users' Feedback

We generally received positive feedback such as "*very promising*", "*useful*","*good for when hands are dirty and busy*" and "*great for the disabled*". The majority of users felt the system was "*really easy*", "*very fast to get used to how it works*", and the instructions were "*clear and helpful*". Some users commented that the scrolling interaction was "*logical*". They felt the system managed to "*captured [their] attention*" and the content "*changed with [their] view*".

Some users pointed out that the system was "*a bit slow with glasses*", "*works better when glasses were removed, but not effective as not able to read*". This was expected, as the shape of glasses frame can affect the face and eye detection. Other users also mentioned "*need to be patient*", "*takes too much time to make it [the content] move to the centre*". Delays varied between persons and also depends on lighting condition. The system works best in a bright environment.

Privacy Concerns

In the testing phase, we prompted the users about privacy concerns while using GazeHorizon. The majority (34/41) reported that they were comfortable with using their eyes for scrolling information in public and did not perceive any privacy risks. Amongst those who were concerned (4 reported "Yes" and 3 "uncertain"), they noticed the web camera and they were worried about how the captured images were stored and used. They explained that it is acceptable if the data and their identity were not revealed publicly, but they preferred to have an option for opting out from being recorded. Also, the displayed content can impact their sense of privacy. One person particularly mentioned that information about movies was acceptable; however, other types of content (added with gaze information) may reveal personal interests unintentionally.

Summary

From the deployment, we confirmed that by providing intuitive guidance novice users were able to control a GazeHorizon display without prior training, and we learned that:

- Letting users know that the system reacts to eye movement at first glance is crucial. Knowing this up front helps users to eliminate attempts of other modalities, such as touch or gestures, and makes it easier for users to interpret the control of GazeHorizon during their first attempt.
- We observed that users lose patience very quickly. Some people abandoned further attempts if they see no immediate system response from their first action, and this is similar to the findings reported by Marshall *et al.* [116].

- The mirror image was more effective than text labels to assist users positioning themselves, as it provides real-time reference for users to perceive their position.

7.6 Discussion

7.6.1 Relative Gaze Mapping for Rate-controlled Navigation

While conventional eye tracking methods map gaze to absolute screen locations, we employed a relative mapping approach that provides different interaction experiences. Although relative mapping does not detect where on the screen users look at, our users' feedback revealed that they felt the system captured their view. This confirms our design that relative mapping can provide a robust illusion of display response to what the user looks at.

In absolute mapping, the reference is device-centric to the screen space. Any error in estimated gaze direction will affect the user experience, as the display response will be relative to a screen position that differs from what the user actually looks at. In contrast, a relative mapping as adopted in GazeHorizon provides a user experience that is robust to inaccuracy in gaze estimation. An error in the estimate effects the scrolling speed but the user's illusion of content-of-interest moving to the centre of the display is robustly maintained.

7.6.2 Bootstrapping Gaze Interaction with Interactive Guidance

Eye movements are subtle. From our observations during the studies, users cannot learn gaze-based interaction by purely observing other users; instead, the learning process requires guidance. The guidance could be provided by either an experienced user explaining the interaction, or interface guidance on the display. An experienced user could provide direct feedback and explanations; however, this relies on the experienced user understanding the interaction correctly. An alternative is via interactive guidance. We avoided to add explicit instructions; instead we provided guided assistance when the system detects an anomaly. We believe that this is more effective and potentially reduces the cognitive load of users, as they discover the interaction model by exploring the interface at their own pace, and the guided assistance can help to prevent misconceptions and to correct user errors.

We learned that the "look here" label naturally captured users' attention. Although the intention of the users was primarily to look at the label, the action activated scrolling as an after-effect with no extra cost. From a novice user's perspective the scrolling can be seen as an effect of his eye movement, which helps the user to conceptualise the activation of scrolling. We believe that the initial user experience was rather implicit; however, the interaction may become more explicit once the user understands the interaction. A few interviewees explained that once they learned the interaction, they explicitly moved their eyes to the sides for scrolling. Even though our interface did not provide any guidance for stopping the scrolling, somehow all of our participants self-discovered this operation.

From our field studies, we realised that there are many unpredictable factors that could hinder the tracking of users' eyes, such as unpredictable user behaviours. Causes of failure were often due to users standing too far away, in an incorrect position, wearing glasses or their eyes were occluded by their hair. They could be corrected by giving appropriate interactive guidance based on specific aspects. We realised that if users are aware of a particular reason that causes the system to stop tracking their eyes, the users are generally cooperative and willing to adjust themselves, like removing glasses, stepping closer, etc. However, we observed an interesting behaviour that sometimes after users noticed the display, they would step away or stand on one side of the display to observe for a period of time. We consider this behaviour as *mirror image avoidance*: although users were curious to experience the system, they might deliberately position themselves to avoid being shown on the "mirror image". In some cases users even moved around while kept looking at the display. This could be due to the users not knowing that the mirror image will disappear and they did not want to be recorded.

7.7 Conclusion

This chapter answers the question of how to deploy a gaze interface in pervasive contexts. The work shows that we can design an interface where users can walk up to a display and comprehend the gaze interaction without any human assistance. By having an appropriate design, novice users can be guided interactively to overcome the limitation of the gaze sensing systems. This is important for pervasive display applications because users' interactions are often unprepared and spontaneous.

The contribution comprises insights from the design and deployment of GazeHorizon. By providing appropriate interactive guidance, novice users can be made aware of using only their gaze to control a public display without any expert assistance, and are able to adjust themselves to match GazeHorizon's vision tracking requirements. Whereas conventional eye trackers mainly detect eyes, our vision-based system also detects other useful information based on the users' actions. For example, GazeHorizon tracks whether a user is approaching our system, and whether the user's head is turned. Our system interprets this context information to present dynamic guidance to assist the user in realtime.

Our work is the first to demonstrate that gaze input can be used in public settings. Our studies show that novice users can easily apprehend the interaction of GazeHorizon. However, we have only explored the use of gaze; future work can compare or combine gaze with different types of input for public displays, such as combining gaze with head orientation.

We implemented GazeHorizon as a single user application. In our in the wild deployment, we observed that users in groups took turns to interact with the display individually. Nonetheless, applications in public spaces pose an issue of sharing an interaction space among multiple users. GazeHorizon could overcome this by distinguishing individual users from their faces and pairs of eyes, and the screen could be divided into multiple regions to support simultaneous interaction of multiple users in parallel. However, this inherently prohibits collaboration where users share the same screen space.

A challenge is thus to design interaction for group collaboration and to minimise confusions and conflicts between users. For example, a user might mistake that an action on the display was triggered by his gaze input, while the action was in fact triggered by another user. This leads to several open questions: How can we design interactions that give users a better perception of gaze control among multiple users? Also, when multiple users are involved, they are not aware of each other's eye movement; how can we design an interface that promotes eye-based group collaboration?

8

Gaze for Shared Display

The goal of this chapter is to answer how gaze can be useful for shared displays. In the previous chapters, we investigated gaze interfaces designed for single user interaction. On the other hand, pervasive displays can also be used by more than one person for collaborative tasks. Prior research only considered using gaze to support individual's interaction on large displays. The question of how group interaction can be enhanced if pervasive displays can sense multiple users' gaze remains unanswered, which we aim to answer in this chapter.

There are increasing numbers of high-density information large displays installed in public and work places. These displays afford group activities, because a large display itself can act as a shared source of information used by multiple persons [145]. In a meeting, for example, a team of geologists can gather around a large map on a shared display to plan an upcoming trip.

Mutual gaze awareness is important in communication and collaboration in

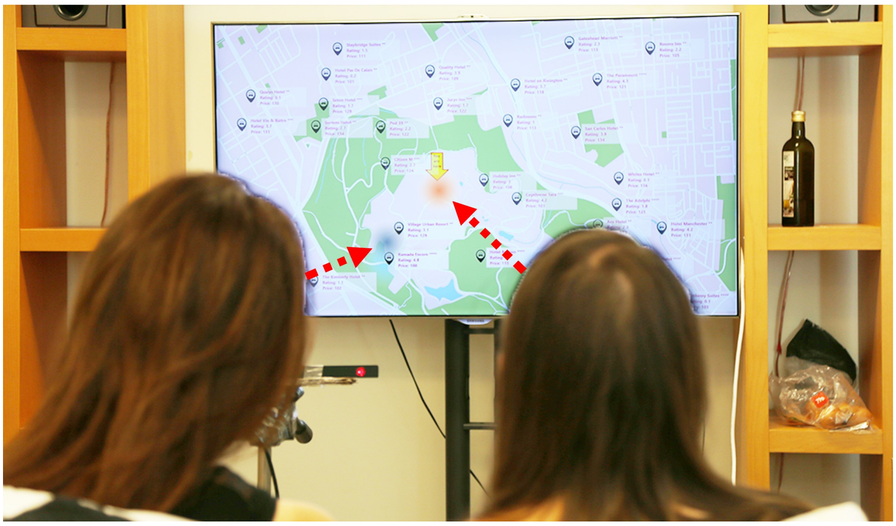


Figure 8.1: Gaze-assisted co-located collaborative search (arrows indicate gaze directions)

group activities. For example, in "backing away" scenarios when two users sit or stand at a distance from large displays to view the entire display and look for information together (see Figure 8.1), gaze cues (e.g., eye contact and joint attention) provide rich context information that other body cues cannot reveal. To understand how gaze can enhance collaborative activities on a large shared display, we propose to provide visual representations of mutual gaze awareness into the design of a shared display interface.

Prior works proposed different ways to convey gaze cues visually. These include the use of video images of the partner's face and head [150, 173], gaze cursors [29], shared visual space (e.g., focused objects) [40] and scan paths overlaid on a screen [141]. These designs provide different gaze cues and are mostly targeted at remote settings. However, it is not clear what gaze cues are useful for in co-located collaboration. In addition, integrating gaze as visual representations on a shared user interface could potentially clutter the interface and interfere with group activities. This essentially raises another open question of how to present gaze cues effectively to benefit collaboration.

To address the above research questions, this chapter presents an exploratory study to understand how gaze cues can enhance collaboration between two users in front of a large shared display. Gaze cues are visualisations of where users are looking, and they require absolute 2D gaze positions on the display. This makes the PCR method (presented in Chapter 4) less applicable, as PCR only provides horizontal relative gaze information. Instead of adopting our custom-built gaze estimation system, we use a commercial eye tracker to capture precise gaze positions.

This chapter first presents an implementation of our system that supports gaze visualisation of two users and the design of four gaze representations. We then present two empirical studies. In the first study, we examine how different gaze representations affect user performance and people's preferences in an abstract collaborative visual search task, where participants search for a specific object on a display with high-density information. The results show that people prefer a subtle and less explicit gaze representation to reduce distractions, but there is a trade-off between visibility and distractions. We further improve our gaze representation design based on findings from the first study and integrate it into a tourist map application (see Figure 8.1). In the second study, we aim to understand the usage of gaze representation and subjective experience of the gaze-enhanced map application. We learn that gaze indicators can ease communication. However, some people are reluctant to share their gaze due to privacy concerns.

8.1 Conveying Gaze Cues in Collaboration

Based on prior findings in observation studies, we learn that multiple gaze cues can benefit collaboration on a large shared display.

Gaze has been considered as a valuable communication resource [97]. It naturally provides moment-by-moment information about a collaborator's focus, which can facilitate the interpretation of the partner's utterance because they can see the object that their partners are attending to. Seeing where the speaker is looking at has been found to make disambiguation of their referring expressions early [157]. In particular, collaborations on a large display often involve members frequently referring to a specific piece of information on the shared display that is related to their discussion. The action of identifying on-screen objects is often carried out verbally, but when information is in high-density, unstructured, and cannot be described using simple phrases, people may resort to body languages, such as pointing. Gaze can be a natural source of input information that benefits collaboration.

Another aspect in our face-to-face communication that gaze enables is to establish joint attention. [1] Achieving joint attention is critical for successful collaborative activities where groups reach a common ground in decision-making [42, 185]. As users gather around a large shared display, the eye contact and gaze cues can easily get lost due to different body orientation and focus changes between individual and group tasks [183, 145]; for example, when people stand or sit side by side in front of the display. This can make the process of establishing a joint attention

[1] *Joint attention* is when participants are mutually oriented to a common part of their shared visible environment and are aware that their conversational partners are also looking at it [185].

challenging (see Figure 8.1). Similar issue has been reported in previous study on collaborative data analysis on shared displays [83]. Their results revealed that participants commonly overlaid their mouse cursors to the joint focus area to show joint attention on a specific information item under discussion. Group members in their study further requested additional visual aids for drawing attention to mouse cursors.

Additionally, gaze can provide information that other body cues cannot reveal, such as ongoing cognitive activities (e.g., scanning, interests towards an object, and comparisons of different objects) [191, 157]. These can potentially improve collaboration, as observing another person's gaze patterns might reveal the task status of the partner and gain information about other's intention.

8.1.1 Mechanisms for Shared Gaze

Task types	Setup	Role	Mechanisms
Video conference	Remote [150, 173]	Regulate conversation	Faces and head
Problem solving	Remote [62, 157]	Understanding comprehension	Gaze cursor; Scan Path
Referential instruction	Remote [40]; Co-located [117]	Joint attention	Gaze cursor; Visual space
Visual search	Remote [29]	Spatial reference	Gaze cursor
Our work	Co-located	Communication; Coordination	Four gaze representations

Table 8.1: Shared gaze in collaboration

Prior research have proposed various ways of conveying gaze cues (see Table 8.1 for a classification of existing work). For example, video mediated communication systems show video images of the user's face to compensate for eye contact [150, 173]. Another common approach is to present users' gaze (i.e., shared gaze) as a cursor or focused object in the shared visual space, which helps them to be aware of their partner's focus [157, 40, 127]. Maurer *et al.* proposed the use of co-driver's gaze cursor as a possible way of sharing information and fostering collaboration between driver and co-driver [117]. Dynamic eye movements (e.g., scan paths) have also been found to enhance sharing of mental states [141, 62]. Enhancing gaze awareness in collaborative activities has been mostly investigated in remote settings (see Section 5.2.6 in Chapter 5).

The benefits of shared gaze in remote collaboration motivate our research. While previous works focused on remote settings, we further extend this notion in co-located collaboration on a large screen (see Table 8.1). Based on existing designs for shared gaze, we investigate how to provide gaze cues (e.g., direct visual attention and real-time eye movements) effectively and what effects they have on the collaboration.

8.2 System Design and Implementation

We implement our system using C# in Windows 8. Figure 8.2 illustrates the architecture of our system. We connect two Tobii EyeX/Rex eye trackers to a laptop (2.7 GHz, 16GB RAM) that runs the system application, and the laptop is connected to an external large display (120cm × 70cm, 1080p resolution) for output. The eye trackers detect users' gaze at a minimum frequency of 30 Hz (i.e. every 33 milliseconds).

When the eye trackers receive gaze data (Figure 8.2), the system processes it

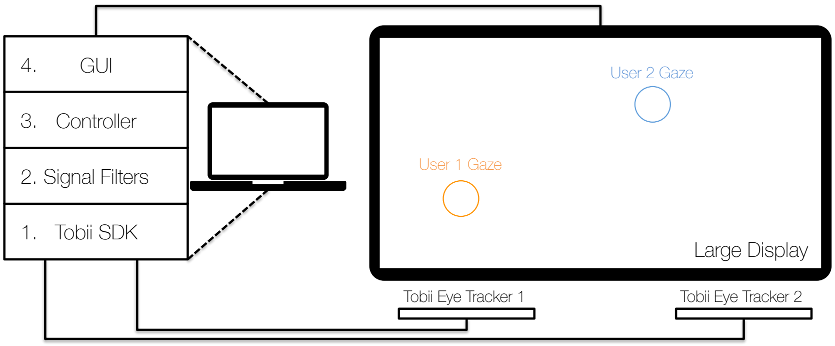


Figure 8.2: The application receives gaze data from two eye tracking devices. Upon receiving the gaze data, the application first preprocesses the data and then informs the controller to update the positions of users' gaze visualisation on the user interface.

in the following four stages:

Stage 1 Tobii SDK: We use the Tobii Gaze SDK to extract raw gaze data from the eye trackers. The SDK provides gaze points (x, y coordinates with reference to the display), eye positions, head positions and presence data. The data is then sent to the next stage to determine the users' fixation points. For each eye tracker, the system runs a dedicated process to receive gaze data. The gaze data values are sent via the signalR packages to the main Windows 8 Store App "controller" which is used to calculate the smoothed gaze data.

Stage 2 Signal Filters: Human eyes jitter during fixations because our eyes naturally make small involuntary movements (e.g., micro saccades). Hence, raw gaze data is inherently noisy [152]. To smoothen raw gaze data, we filter out saccade movements by calculating the real-time distance between gaze points. First we compute the x- and y-axis displacements between current and previous detected gaze positions. Any gaze displacement (i.e. eye movement) that is above the distance threshold of 120 pixels is classified as a saccade, and otherwise is classified as a continuous fixation.

To further stabilise the fixation data, we use a weighted average to smooth the gaze data. Similar to [99], we calculate a fixation point in a time window of i frames (i.e. equivalent to approximately 500ms of gaze data) by using the following equations:

$$x_t = \frac{i * x_{t-1} + (i-1) * x_{t-2} + ... + 2 * x_{t-(i-1)} + x_{t-i}}{i + (i-1) + (i-2) + ... + 3 + 2 + 1} \tag{8.1}$$

$$y_t = \frac{i * y_{t-1} + (i-1) * y_{t-2} + ... + 2 * y_{t-(i-1)} + y_{t-i}}{i + (i-1) + (i-2) + ... + 3 + 2 + 1} \tag{8.2}$$

, where i represents the window size ($i = 15$ in our case).

The current fixation point is sent to the controller as an event to update the previous fixation.

Stage 3 Controller: When the controller component receives a fixation point, it updates the position of the corresponding gaze object (e.g. a cursor). In other words, if new gaze data is received from eye tracker 1, then the gaze object for tracker 1 is updated. This changes the x and y coordinates of the gaze object on the cartesian plane of the display.

Stage 4 GUI: Lastly the application informs the system to render any updated gaze-controlled objects on the display at 10Hz. We do this to maintain a smooth refresh rate due to irregularity from the fixation data.

During our pilot trials, we test several configurations of thresholds and window frames. Although the current implementation has a delay of one frame (i.e. 33 milliseconds), it enables a more stable focus point representation and also allows fast shifts between fixations.

8.2.1 Gaze Representation Design

In this work, we present four types of gaze representations that aim to support users in co-located collaborative tasks based on existing designs summarised in

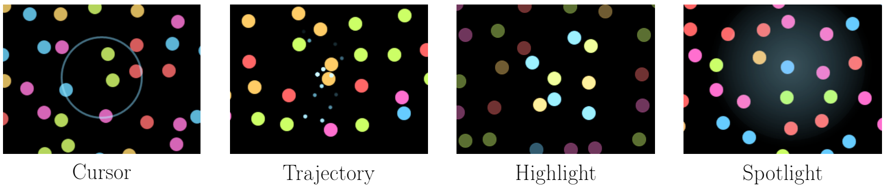
Table 8.1 (Figure 8.3):

- *Cursor*: Gaze is displayed as a coloured circular ring with a radius of 60 pixels. This type of gaze representation is similar to having an onscreen cursor following a user's gaze. This is consistent with the gaze cursor in Table 8.1.

- *Trajectory*: Gaze data within the last 3 seconds is plotted as a trajectory. Each sample is displayed as a small circle, and its opacity decreases with time. Hence, the most recent gaze data has the highest opacity. Trajectory is a representation of the scan path in Table 8.1.

- *Highlight*: Displayed objects within a 60-pixel radius from the gaze point are highlighted by increased brightness. Any objects that are nearby the user's gaze will be automatically made more visible or selected. This is similar to the visual space on focused objects in Table 8.1

- *Spotlight*: This simulates a torch shining effect (shown as a bright Gaussian-blurred disk) that follows the user's gaze location. The resolution is full in the central fovea within 2 degrees of visual angle and falls gradually towards 3 degrees beyond the periphery. This simulates human visual perception. Its resolution is much higher at the fovea focus than the periphery [135], hence Spotlight's opacity gradually fades from fovea to periphery. This is similar to the visual space on focused objects in Table 8.1.



Cursor Trajectory Highlight Spotlight

Figure 8.3: Four types of gaze representations

8.3 Study 1: Effects of Gaze Representation

In this study, we aim to evaluate how people perceive the usefulness of the four gaze representations as communication and coordination cues on a shared display. The goal is to investigate how different representations of gaze help collaboration. We selected a visual search task adapted from Brennan *et al.* [29]. Participants collaboratively search for an oval object amongst a large set of non-overlapping circular objects. They are required to make a joint decision to confirm or reject whether the oval object exists. The task has similar elements as real-world collaborative visual search tasks, where people would need to look for information together in front of a high-density display, such as locating a specific building on a campus map, or finding a particular product in a shopping catalogue.

In our study we aim to understand the following research questions:

- Can gaze representations improve users' performance in collaborative search tasks?
- Can gaze representations influence people's perception of communication and coordination in collaborative tasks?
- Do people feel distracted or attentive when seeing different gaze representation designs? How do they influence collaboration?

We hypothesize that providing gaze information of collaborators can help them to become more aware of each other's attention, and thus better facilitate their communication to reach a common ground. We further hypothesize that gaze history in temporal space (like gaze trajectory) would provide collaborators with revealing additional information of their partner's attention and search strategy; thus better coordinate their search actions.

8.3.1 Participants and Setup

We recruited 16 participants (13 male and 3 female, with a mean age of 27.9 years *SD* 4.7 years), as 8 pairs to take part in the study. We used a 55-inch display (120cm × 70cm, 1080p resolution), with the bottom bezel positioned at a height of 115cm above ground. Each pair of participants stood side by side and at a distance of 2m in front of the display, with a view angle of 46.4° horizontally and 28.1° vertically. Two eye trackers were placed at a distance of 140cm in front of the display, each tracking one user's eyes. One eye tracker was placed at 30cm to the left of the screen's centre; the other one was placed at 30cm to the right. The eye trackers were aligned at a height of 5cm above the bottom of the screen. We conducted a pilot study to fine-tune setup parameters, such as the sizes of gaze representation. We found that a 60-pixel radius (3 degrees of visual angle) is the optimal size.

8.3.2 Task and Procedure

The participants' task is to make a joint decision of whether they find a coloured oval target (0.8° in height and 0.95° in width) among 364 non-overlapping coloured circles (0.8° visual angle). Each task consists of one of two conditions: *target-*

present or *target-absent*. The target-present condition consists of one oval target, placed in a random non-overlapping location amongst other circular dots. In the target-absent condition, all dots are circles.

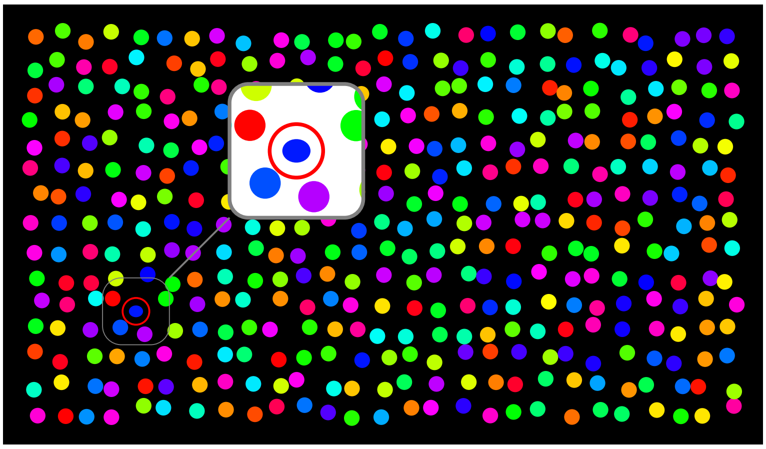


Figure 8.4: Study 1: visual search stimulus

We adopt a within-subjects design for five conditions: *without gaze, gaze cursor, gaze trajectory, objects highlighting* and *spotlight* (see Figure 8.3). In the without gaze condition, the display provides no gaze visualisation. In the other four conditions, both participants see where they are looking at in real-time on the screen, and the gaze visualisation is colour-coded for the respective users (orange, blue). The order of the five conditions was counterbalanced. Each study session consisted of 60 trials (hence 12 trials per gaze visualisation condition), and half of the trials were target-present.

Prior to the study, the eye trackers were calibrated individually to each participant. Participants were allowed sufficient time to practise. A 3-minute break was given after completion of each condition (i.e. 12 trials).

The participants were asked to complete the task as fast and accurately as possible. They were allowed to converse freely with their partner, without restrictions on strategy or communication. After the first participant responded, they received feedback about the correctness. Each session lasted approximately 60 minutes.

8.3.3 Data Collection

We collected quantitative and qualitative data. During the study sessions, the system logged the participants' completion time of each trial and the number of errors made for each condition. After completing each condition, the participants answered questionnaires which made up of 7-point Likert scale questions and open-ended questions for their subjective experience. We balanced the Likert scale questions with both positive and negative questions.

The questionnaire consists of multiple parts. The first part focuses on how people perceived the quality of collaboration and the mental and physical effort required to use gaze indicators for collaboration; for example, how gaze representation helps them to make joint decisions, as well as assists communication and coordination between partners. The second part focuses on the effectiveness of gaze feedback, and we ask questions that are related to distractions, usefulness, and whether and how gaze indicators hinder collaboration.

The questionnaire also asks participants about the strategies that they adopt for collaborating with their partner to complete the task, such as the types of difficulties that they encountered, what types of information that the participants gain from seeing the partner's gaze indicators, and how they feel about the value of seeing the gaze indicators.

Lastly, the experimenter conducted a short interview with the participants (as a pair together) for feedback and suggestions for improvement about the effects of different gaze representations.

8.3.4 Results

Group Performance

We measured the overall search time and accuracy for each visualisation condition. Figure 8.5 illustrates the average search times for the target-present and target-absent trials. The results of average search accuracy across the different gaze representations are presented in Table 8.2. The average search accuracies for different conditions are similar.

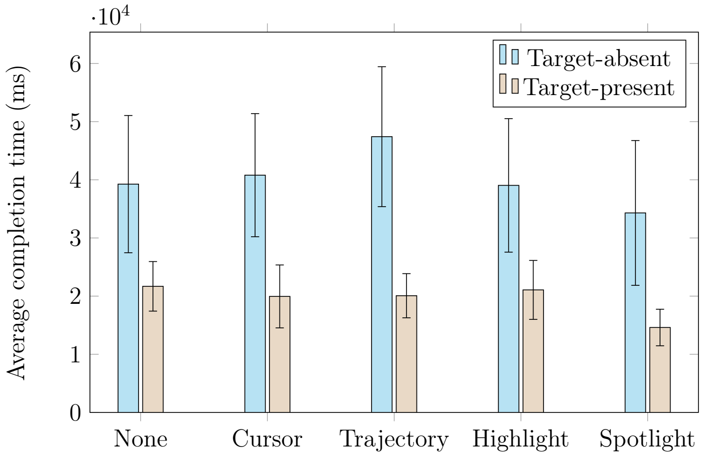


Figure 8.5: Average of the overall search time. Error bars represent the 95% confidence interval of the mean.

	None	Cursor	Trajectory	Highlight	Spotlight
Mean	81.7%	83.3%	80.8%	81.7%	80.8%
Std	17.7%	17.7%	24.2%	14.2%	19.2%

Table 8.2: Average search accuracy

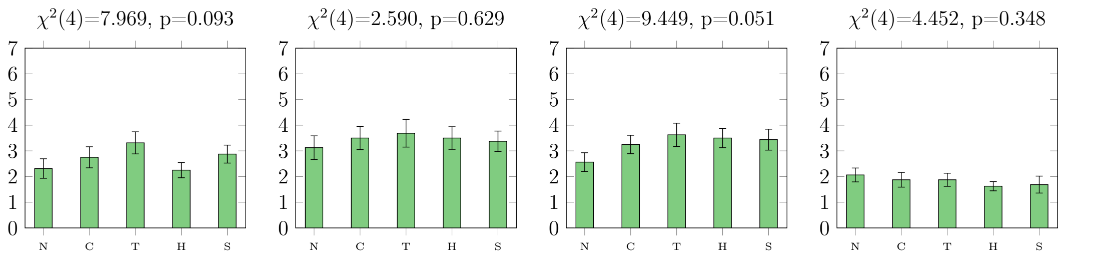
A repeated measure ANOVA analysis showed a significance for completion time across the five conditions in the target-absent (F(4,28)=2.728, $p<0.05$) trials and in the target-present trials (F(4,28)=2.762, $p<0.05$). However, pairwise comparisons showed no pairs with a significant difference in the target-absent trials. Spotlight achieved the shortest completion time in target-present conditions. A significant result (p<0.05) was obtained in target-present trials, with the Spotlight ($M = 14.6s$) being faster than the None ($M = 21.7s$) condition. Our data showed that gaze information can improve the speed of the collaboration task; however, the way of presenting gaze feedback can influence people's performance in speed.

Gaze Role: Feedback and Observations

Gaze for communicating spatial information: Half of the participants (8/16) mentioned that seeing the gaze indicator was helpful and it became *"easier to explain to each other where the target was"*. Gaze was more convenient than speech to describe a target position (such as pointing out a particular display region and colour). After getting used to having gaze visualisation, some participants commented that *"it was strange not to have any indicator of my partner's gaze"* in the *None* condition. Subjective feedback also revealed that users found the gaze indicator useful to indicate the location of a target. Without gaze information, people needed to speak more to explain the location of a target, and they found it *easier* to communicate with gaze indicators. For some participants, gaze information was particularly useful when they needed to confirm or come to an agreement with their partner.

Gaze for coordination: The participants had diverse ways for coordinating the search strategies. When users searched together, they first started with establishing rules by verbal communication. For example, the majority of our participants started with splitting the screen in two regions, like *"I start right, you start left"* or *"I [go] left to right and my partner [goes] top to bottom"*.

An interesting observation we noticed is that, when gaze information was shown, people tended to avoid looking at the same region together at the same

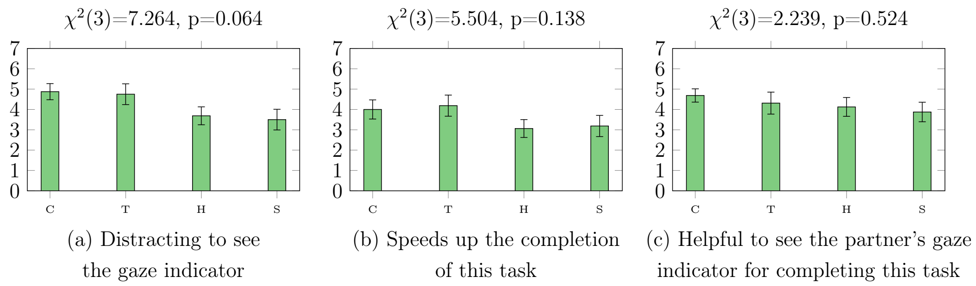


(a) Difficult to complete the task (b) Mentally demanding (c) Physically demanding (d) Difficult to communicate and coordinate

Figure 8.6: Subjective feedback on collaboration experience to complete the search task (1-Strongly disagree to 7-Strongly agree). The error bars in all figures stand for the standard error of the mean. N(None), C(Cursor), T(Trajectory), H(Highlight), S(Spotlight)

time, and this was usually done without explicit verbal communication. For example, if a user saw that his partner was searching the top-right region, the user would choose another region to search. One of our participants explained, *"the gaze indicator showed where my partner was looking, so I could look at other parts of the display."*. This minimised the chance of both users doing the same thing simultaneously, as gaze indicators made them aware of their partner's progress. Other times, users synchronised their actions with the partner, for example, *" First we focused on different sides (left and right) next we scanned the middle part together"*. Thus, they first split the workload and then combined.

The questionnaire data also reflected that the users were monitoring their partner's focus and attended areas through the partner's gaze indicators (e.g., by their peripheral vision). The intention of keeping themselves aware of the partner's gaze was mainly due to the participant adapting their search strategies to cooperate with the partner. Some participants mentioned that they defined a strategy beforehand, hence to gain progress by checking where their partner was looking. For instance, in between if they found their partner's gaze indicators appearing in their half and they would wonder if the partner was properly searching his half and if he *"should check his[the partner] half too"*.



(a) Distracting to see the gaze indicator (b) Speeds up the completion of this task (c) Helpful to see the partner's gaze indicator for completing this task

Figure 8.7: Subjective feedback on effects of the gaze feedback (1-Strongly disagree to 7-Strongly agree). The error bars in all figures stand for the standard error of the mean. N(None), C(Cursor), T(Trajectory), H(Highlight), S(Spotlight)

Gaze for attention guide: Users occasionally lost track of their searching location due to distraction or tiredness. In the gaze trajectory condition, several participants expressed how they used their gaze indicators as a guide for finding where they were scanning. Our participants commented, *"sometimes I got confused about where I was, but because of this indicator, I can quickly continue from where I [got/was] lost"*. The tail of gaze trajectory provided implicit information of the user's scanning process, so when the user was distracted they could quickly refer back to the trajectory tail to continue.

Effects of The Gaze Feedback

The majority of participants did not consider that the task was difficult to complete collaboratively with their partner in the *None*, *Highlight* and *Spotlight* conditions (see Figure 8.6). A third of the participants agreed that the *Trajectory* condition made the task more difficult than the other conditions. Similarly, the *Trajectory* condition was consistently rated higher for physical demand than the *None* condition. Our questionnaire data suggests that the physical demand was mainly induced by eye fatigue. However, a Friedman Test on users' responses (with regard to difficulty to complete the task, mental demand and difficulty in communicating and coordination on all conditions) did not reveal a significant difference (see Figure 8.6).

When we asked the participants about problems and difficulties that they encountered, we learned that the major difficulty was from the presence of the gaze indicator during the normal viewing process, which often distracted them from visually searching. When looking at the user feedback about the effects of different gaze feedback, there is no significant result found in any particular representation winning over the other technique (compared using the Friedman Test; See Figure 8.7). Participants agreed that seeing the gaze indicators was distracting in *Cursor*, *Trajectory* conditions, while the object *Highlight* and *Spotlight* conditions were less distracting.

In the *Cursor* condition, eight participants mentioned that they felt the gaze cursor was distracting although they found it easy to make an agreement in this condition. One problem encountered by many participants was the occlusion by the gaze cursor which made it hard to judge the oval target shape. Other problems include that the cursor was *"inaccurate"* and *"moving too much"* which was caused by instability of human fixation, and the cursor *"size [was] too big"*.

In the *Trajectory* condition, five participants found this representation very distracting which made the search task difficult. They commented that *"the movement [of the trajectory] is very distracting"*, in particular, when two tails (from two users) crossed each other. The side effect was that the participants could not accurately and precisely infer where the other was looking at, rather being unintentionally chasing the other's gaze from time to time. In some cases, the participants even tried to scan faster than the cursor to evade the problem. It seems that the advantage of using gaze for spatial referencing decreased in the *Trajectory* condition, as this type did not provide precise representation of current focus location. Hence, participants felt that it only indicated a rough region and they still needed to perform a further search to locate the target. On the other hand, three participants found this type of gaze indicator helpful as it revealed the partner's search speed, so that they could adjust to cooperate.

In the *Highlight* and *Spotlight* conditions, the majority of the participants felt the indicator was less distractive, e.g., *very subtle and not distracting.* They felt that they could focus on searching and still know what their partner was looking at. The only problem encountered for the *Highlight* feedback was the glimmer effect (mentioned by two participants). In the *Spotlight* condition, two participants mentioned that they felt the indicator was like *"a proper element that was on top"* which sometimes caused them to focus on the gaze feedback rather than the stimulus. As these two types of gaze feedback were more subtle with less visibility, the effects of assisting target referencing were less prominent (see Figure 8.7(b)&(c)).

8.3.5 Lessons Learned

When is gaze useful?: From this study, we learned that that gaze information can be useful in the collaborative search task in a co-located setup on a shared screen, e.g., for referring a remote target, being aware of a partner's focus and guiding their own attention. The gaze information would benefit in particular when people need to corporate and coordinate with their partner. Although participants mentioned that it was useful and interesting to keep an eye on where their partner was looking, gaze was found to be less useful during the normal searching and viewing process.

It is still unclear whether users would need the gaze information all the time during their collaboration or whether it would distract them more from their individual goal.

Avoid gaze trajectory: Our results suggest that the *Trajectory* feedback should be avoided in scenarios where frequent target referencing is required. The main difficulty came from the irregularity of the generated gaze trajectory patterns. The characteristics of eye movements (e.g., saccades) were different from continuous pointer movement such as mouse. Thus, the created trajectories varied in shapes and lengths depending on the amplitude and speed of the eye movements. This non-uniform representation confused users and was less useful in both cases for assisting spatial reference and communicating attention.

Subtle gaze feedback (visibility v.s. distraction): One of the biggest challenges we realised is the conflict between visibility and distraction of the gaze indicators. High visibility gaze indicators (e.g, cursor and trajectory) provided fast and accurate target reference, however caused more distraction. Users preferred subtle representation of gaze feedback in the object highlighting and spotlighting representation. Representing gaze as an object (e.g. a cursor) can distract users. However, when the visibility decreases, the gaze indicator loses its power for spatial referencing and maintaining focus and attention awareness during the collaboration.

8.4 Study 2: Tourist Map Application

Our second study investigates people's qualitative experience in a more realistic setup. We built a tourist map application like those in information centres, train stations, or museums (Figure 8.8). Two users communicate and find a hotel on the map that they both agree and approve to.

Our application integrates two gaze visualisations. From the previous study, we learned that people prefer gaze visualisations that are subtle and less conspicuous, e.g. the highlight and the spotlight gaze representations. We combine the two types of visualisations into a single gaze indicator as illustrated in Figure 8.8(B).

We also added a foot control that enables users to switch the gaze visualisation on or off. Our first study showed that gaze indicators can be distracting from time to time, and we thought to provide the user with more control of their gaze visuals. We chose a foot control so that the user's hands are kept free, enabling natural use of hands for body language during discussion, and to potentially hold on to private items during the activities (in contrast to hand-based control such as mouse/keyboard).

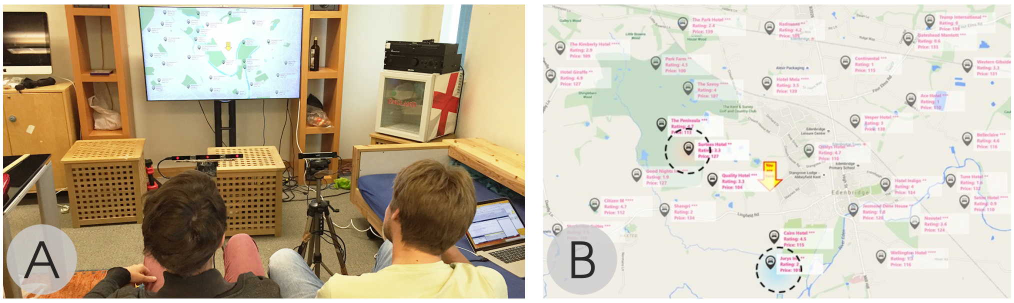


Figure 8.8: Setup: (A) A pair of participants sat in front of a large screen, with an eye tracker facing each person to capture their eye movement. (B) The application interface showing the gaze indicators of two users (the dashed circles are not part of the interface; only added for visibility).

8.4.1 Study Design

We recruited 20 participants (10 pairs, 16 male, 4 female, age from 21 to 43, M=29.7 SD=5.8) from our research department. The setup was similar to the first study, except this time the participants were seated instead of standing (Figure 8.8(A)).

Prior to the study, we demonstrated our tourist map application to the participants and allowed them sufficient time to calibrate the eye trackers, to experience the interaction, and to get comfortable with the system. The system presents a map with 30 hotels (chosen randomly from a pool of 75 hotels) scattered across the screen (Figure 8.8(B)). Each hotel is attached with its name, hotel quality rating (i.e. number of stars), price, location, and average customers rating (on a scale out of five).

During the study, we explained to the participants that they should assume that they are tourists who are travelling together and looking for a hotel. The participants were free to discuss with each other. Their task was open-ended, and the only requirement was that they must come to an agreement of selecting a hotel. To stimulate discussion, each participant was advised to look for hotels that satisfied specific conditions. For example, one participant would look for nearby hotels that are close to where they are (indicated by a "You Are Here" maker), while the other participant would seek for hotels with a good reputation (e.g., user rating).

On average, a study session lasted for approximately 30 minutes, and every session consisted of eight trials. For each trial, a random map was loaded with new hotel information. After four trials, the default settings inverted. After completing the eight trials, the participants filled in an exit questionnaire with their subjective feedback. Half of the participants started with gaze indicators being switched on by default, and the other half with gaze indicators switched off initially. This helps us to learn when users would invoke the gaze indicator and in what situations they would want to make the indicators hidden or visible.

8.4.2 Data Collection

We collected system logs and qualitative feedback through a two-part questionnaire. The first part focused on the participants' collaboration experience. We elicited their feedback by asking questions about how the gaze indicators assisted them to collaborate with their partner. In conjunction, we used an adaptation of the desirability toolkit [24]; we provided the participants with a list of adjectives and asked them to select five or more that most closely matched their personal reactions to the system. The method of selecting adjectives is ideal to elicit a participant's reactions and attitudes, as it provides a quick high-level indication of their reactions. The selection of words then acts as a basis for further explanation and elaboration about why they chose those words.

The second part of the questionnaire focused on how the participants controlled the visibility of their gaze indicators. We asked questions on what caused the user to turn their gaze indicators on and off, as well as what caused them to avoid toggling the gaze indicator. This can help us to find out when the participants perceive gaze indicators as useful or counter-productive. Lastly, we asked the participants to identify any problems that they encountered during the study, the types of applications that they thought gaze indicators would be useful, as well as suggestions for future improvement.

8.4.3 Results

Our participants were positive on the use of gaze for collaboration. Most of them state that it was "*convenient*" to see their partner's gaze location, because it made them aware of which location that their partner was referring to during discussion, and gaze also makes pointing at a map location simpler. Gaze enabled the participants to spend more effort on discussion instead of thinking of words to describe a specific location, as the users can simply point by staring. One participant mentioned that he preferred to describe a map location by referencing nearby landmarks, but acknowledged that gaze indicators are useful in "*quiet*" locations that had no nearby reference landmarks. The participants also mentioned that having the gaze indicators in different colours made them easily distinguishable and reduced confusion. However, a participant stated that any patches of background that had similar colour to the gaze indicator colour could make spotting the indicator difficult.

Several issues were reported, e.g. inaccuracy, which was caused by eye tracking detection errors. Some participants experienced a small distance offset between their focus and their gaze indicators for which they compensated by slightly looking off target. A few users also found their partner's rapid gaze indicators to be distracting, and needed to be conscious not to follow them. They suggested that gaze indicators should be less conspicuous and only be revealed on demand. While some people preferred less apparent gaze indicators, some actually preferred them to be larger and more visible. They explained that increasing visibility would help to explicitly catch other's attention.

Reactions to Gaze Indicators

The participants agreed that having gaze indicators for collaboration was interesting (20/20) and the majority considered it pleasant (10/20) because the interface was "*easy to learn*" and provided a "*straightforward experience*". The participants also stated that the gaze indicators made the task more efficient (15/20) as it provided an "*extra layer of information between [the partners] ... by just looking at [the target]*", and smooth (8/20) because the "*[gaze indicator] followed the eyes ...and saved complicated location description*". Several people mentioned that the experience could be stressful (5/20) because of the distraction of the gaze indicator, so the users needed to "*[focus] on the pointer all the time*". The experience could also be frustrating (3/20) due to inaccuracy which caused the interaction to be "*chunky*", "*jerky*" and "*slow to get the pointer to the exact location*".

Gaze Indicators for Collaboration

The participants frequently described their experience of having gaze indicators for collaboration as helpful (14/20). The primary benefits pointed out by the participants are that using the system was time-saving (12/20) and it speeded up the interactions. At the same time, participants also felt the collaboration experience to be fun (9/20) and entertaining (8/10). One participant even summarised his experience as "*a tedious and potentially worrisome task made easy, pleasant and efficient*".

The participants considered that the interface was simple (13/20) and intuitive (7/20). They acknowledged that gaze indicators can enhance communication, as the users are made aware of their partner's interests. They also recognised that the gaze indicator reduces effort and shortens verbal description, since the gaze already acts as an immediate pointer.

At the same time, gaze indicators also helped the users to gain an idea on whether their partner was paying attention to what they were talking about. We observed an instance where one participant stepped on his partner's foot control to turn on the partner's gaze indicator, so he could know where the partner was looking. Several participants also felt that using the system was frustrating (3/20) and overwhelming (2/20). Sometimes it was because the participants needed a while to realise which gaze indicator belonged to whom. Other times it was caused by requiring attention to divert other's focus to their gaze indicator while not following the partner's gaze indicator.

On/off Toggle Behaviour

We observed two phases of collaboration. In the first phase, *scanning*, the participants individually looked for hotel options in parallel. Some participants considered that having the gaze indicators switched on during this phase could cause distractions. The second phase consisted of *discussion*. The participants often needed to refer to different hotel options on the screen and also to direct their partner's attention to where they were looking. In the second phase, gaze indicators were frequently used, and the participants often switched the gaze indicator on to ensure that it was available. We also observed cases of frequent toggles of the gaze indicators when the participants wanted to refer to different on-screen targets during their discussion.

Three quarters (15/20) of the participants left their gaze indicators on and never switched them off. They explained that the gaze visualisation helped them to focus on picking a hotel option and also made it easier for their partner to see their preferences. Infrequently, five participants switched their gaze indicator off and explained that this was due to fatigue and distraction or they simply no longer wanted to search anymore. Inherently, switching the indicator off can be a social sign to inform the partner that they want to finish the task.

We also observed that some people switched their gaze indicator off for a brief moment and immediately turned it back on. This happened during their discussion of hotel options, where the participants realised although distracting, the indicator was needed for a more efficient communication (like pointing at a hotel). There are also occasions that, when the gaze was off, people toggled their gaze indicator on for a brief moment and immediately toggled back. These instances happened when the participants wanted to use their indicators to quickly direct their partner's attention to what they were looking at when they want to pick another hotel option together. Several participants intentionally switched their indicator off because they were not comfortable and reluctant to let their partner see where they were looking.

8.5 Discussion

As collaborative activities often happen around a large shared display (e.g., surface hub, digital board), we believe that many collaborative applications can benefit from our studies. Our results show that having gaze with visual representations as implicit indicators of visual attention displayed on a shared display enables co-located partners to be aware of each other's focus and indeed helps them to communicate during collaborative tasks. We also show that different types of visualisation of gaze indicators can impact collaboration.

In this work we learned that:

- The subtlety of visual representation of gaze indicators influences the quality of collaboration. Highly visible visualisation can lead to distractions and hampers collaboration. Subtle and less explicit gaze representation is preferred.
- Displaying gaze indicators improves the efficiency of collaborative tasks, as users can refer to a specific on-screen location by looking. This eliminates the verbose process of describing the location verbally.
- Revealing gaze information enhances group synchrony and avoids duplication, as users are aware of collaborators' focus. A gaze indicator also helps to establish joint attention, which benefits collaborators' communication and understanding between partners.

8.5.1 Comparison with Existing Works

In conventional desktop settings, to convey users' focus of attention in shared workspaces, previous research proposed the use of visual representation of mouse movements (e.g., telepointers [65]) and integrating a variety of awareness widgets into the user interface. However, mouse cursors do not represent the users' focus, as cursors can be stationary while the users are paying attention to another location. In other words, cursors do not provide an accurate representation of user attention. Visual awareness widgets (e.g., radar view), which are also determined from mouse cursor positions, require additional space of the shared workspace [71]. What we proposed in this work is to harness gaze as a natural information source of user attention to assist collaboration, which requires no extra user actions. In addition to presenting users' attention, gaze is also a natural pointer, so people can use it to provide spatial references and establish joint attention.

Similar to previous findings in workspace awareness research [69, 138], making actions more perceivable aids maintaining awareness. However, presenting more additional information can increase distractions. We encountered similar problems in our study. Although we found people in general prefer subtle gaze feedback (e.g. highlighting objects), in some cases people actually preferred obvious representations (e.g. spotlight). This happened because making gaze indicators obvious can be useful for spatial referencing and invoking the other's attention.

Our choice of task that is similar to Brennan *et al.* [29]. Brennan *et al.* focused on coordination aspects of gaze sharing, with respect to speech communication in a remote visual search task [29, 127]. Gaze was found to be superior to speech in terms of communicating spatial references. Interestingly, they found that using speech with shared gaze was substantially less efficient than using shared gaze alone due to the coordination cost of speech communication. On the contrary, with a different setup, in co-located settings, gaze enhances communication and coordination with body languages or voice cues. Also, we found that collaborators' gaze provides awareness information so that users would divide their tasks. What we often observed is that use of speech and the gaze indicator worked simultaneously to assist collaboration. Sometimes, speech was used to provide explicit instructions to coordinate action, while gaze was used as an implicit cue to decide the working area or to monitor the other's progress. Other times, gaze was used to initiate attention from the partner, whereas speech was used to confirm he is in the right place. However, the simultaneous use of the gaze indicator with hand gestures has been seen infrequently. This is probably because the gaze and hand gestures can similarly act as a pointer.

We further contribute the user experience aspects of sharing gaze in collaborative activities that have not been covered in previous research. Our results indicate that users had a positive experience with our shared gaze interface. The results are encouraging and our work opens further research opportunities for studying how gaze cues can be integrated into large displays to support more complex collaborative tasks. In future, we intend to study how gaze enhances other activities. For example, in a multi-device ecology, we often find many co-located collaboration opportunities (e.g. cross-device interaction). We predict that gaze can show further benefits in scenarios when hands are occupied with manual input devices (e.g., mobile devices) and there require frequent changes of focus between group and individual devices/tasks.

8.5.2 Lessons Learned and Design Considerations

Our proposed design is simple to implement and can be applied in many shared display applications. We encourage interface designers to consider our approach to use gaze for multi-user collaborative applications. In the following section, we provide lessons learned from this work and limitations of applying our approach.

Trust and Privacy of Shared Gaze

In collaborative tasks, people often first agree upon a divide-and-conquer strategy, so that each person works on an individual region (e.g. one person focuses on the left, while the other focuses on the right). We observed that some people cross over and deviate to their partner's region for double-checking. Having gaze indicators switched on can negatively impact partnership. Seeing a partner's gaze on a non-allocated region can be implied as a lack of trust or that the person is not following agreed instructions.

People naturally look at objects that they are interested in. By observing users' gaze indicators, it is possible to infer their interests. This poses a privacy concern, and users may not be willing to reveal their gaze focus, especially to strangers or to people whom they are not familiar with. In our second study, we provided a control feature for people to hide their gaze indicators. We observed that people would turn off the gaze indicator if they were uncomfortable about letting their partner know what they are looking at. Keeping gaze indicators on throughout the interaction may be acceptable when working with a trusted partner; however, the situation could differ if it is in a public environment. This inherently opens the question of *under what context and constraints are inappropriate to reveal gaze indicators*?

Augmented Gaze Representation

Integrate Semantic Information Similar to [138], the identity problems can cause distraction and confusion, especially with conspicuous gaze indicators that people often need to check which indicators belong to whom. This issue could be alleviated by adding identifiable denotation using strategies similar to telepointers [65], like attaching names, assigning different shapes, photos or arbitrary information to each user's gaze indicator.

Additional Visualisation Control In our design of our application, we only provide a function to toggle the visibility of the gaze indicator. Our observations helped us to realise that gaze indicators provide multiple benefits in assisting collaboration. Sometimes users prefer explicit and use the gaze indicator actively. But from time to time, people use it rather passively for monitoring the other's attention. It may be necessary to empower the users with some level of control over adjusting their gaze presentations. One solution could be, similar to the control of virtual embodiments in tabletop groupware systems [138], allowing users to actively adjust the opacity of visual representations.

Issues of Eye Tracking

Going Beyond a Pair In the setup of our study, we used an eye tracker for each user, because current commercial eye trackers can only support gaze detection of an individual user. This inherently constrains the number of simultaneous users. We envision that in the near future eye trackers can support simultaneous gaze tracking of multiple users. This essentially raises a new research question of *what happens if the interface presents many gaze indicators*? From the studies we learned that users get distracted easily from simply two gaze indicators. Increasing the number of indicators can intensify distractions. Although our users suggested that they prefer to have customised and distinguishable indicators to reduce confusion, finding the right balance between the number of simultaneous gaze indicators and the design of subtlety is an important aspect for future gaze-assisted co-located collaboration.

Stability of Gaze Representation Our experience informed us that eye movement patterns (using trajectories) are difficult to interpret in real-time. In addition, one of the biggest distractions, compared to visual representation used in other groupaware work [65, 71, 138, 70], is actually from the jitteriness of the visual representation. In our work, we showed a simple threshold-filtering technique to remove saccades and to smoothen gaze raw data. We anticipate that more sophisticated fixation and saccade detection algorithms can improve the gaze stability.

8.6 Conclusion

In this chapter, we investigated the use of gaze for collaborative search applications. We presented two users' gaze locations (using four different representations) on the same display, to help collaboration between partners. Our results show that gaze can enhance co-located collaboration and help users' to coordinate their search strategies to minimise chances of doing the same work. However, there is a trade off between visibility of gaze indicators and user distraction. Users preferred subtle feedback such as using object highlighting and blurred gradient visual representations. Although gaze cursor and moving trajectory provided gaze information with high visibility, they seemed to be more distractive and less preferred by the users.

With a gaze representation design that combined both object highlighting and blurred gradient visual representations, users acknowledged that seeing gaze indicators eases communication, because it makes them aware of their partner's interests and attention. Users found gaze is helpful and timesaving when collaborating with partners. Users also perceive the use of gaze for communication easy and intuitive. We believe that the advantage of supporting gaze in co-located collaborative tasks can be further improved by appropriate design and considering how best to present gaze information to balance visibility and distraction.

Application designers should also take into account the issues of trust and privacy for gaze sharing. Besides interface aspects, users can be reluctant to share their gaze information due to privacy, as gaze behaviour is hard to fake and potentially divulges their interests.

9

Concluding Remarks

In this thesis, we investigated eye tracking for pervasive displays. From the literature review, we highlighted multiple advantages of using gaze as an input. However, we realised that existing gaze applications are still mainly restricted in controlled environments. This thesis contributed new insights for enabling gaze input for pervasive displays. With the computing paradigm shifting from desktops to ubiquitous computing, the knowledge gained from this thesis informs new technical developments and interface designs for advancing eye tracking in daily life applications.

In this concluding chapter, we first summarise the contributions of this thesis. Then we reflect on the lessons learned and the issues encountered during the course of this work. Based on our experience, we provide suggestions for technical improvements, gaze interaction design, and identify the limitations of applying eye tracking for pervasive display applications. Lastly, we identify future opportunities of gaze for ubiquitous computing.

9.1 Summary of Contributions

This thesis contributes both technical and design aspects for advancing eye tracking for pervasive displays. The first part of this thesis contributed two novel gaze sensing solutions. It shows that robust eye tracking can be achieved without specialised hardware and individual calibrations. Chapter 3 introduced an appearance-based method that tracks coarse gaze directions, with eye images captured from a normal web camera. The method uses a supervised data-driven approach and is built on prior training using large image datasets, which overcame the challenge of calibration. The method is ideal for scenarios where a web camera is available (e.g., homes and offices), and can be applied when fine-grained gaze tracking is not required, such as channel switching on a smart TV.

Many pervasive displays are installed in pubic environments. The proposed appearance-based method is less applicable, as it requires large amounts of labelled data for prior training to achieve person-independency. Chapter 4 introduced PCR, a feature-based method which does not require prior datasets. PCR uses eye features extracted from facial video images and exploits the symmetry of eye movements to achieve robust calibration-free gaze estimation. The method is suitable for spontaneous interaction with large displays; however, it does not estimate fine-grain gaze positions that traditional gaze interfaces require. This thesis proposed design solutions (Chapter 6 and Chapter 7) that use coarse-grain gaze detection for walk-up-and-use applications in pervasive contexts.

The second part of this thesis contributes novel design of gaze interfaces, evaluations of the interfaces, and new understanding of their usage in uncontrolled environments. In Chapter 6, we proposed SideWays for selecting, scrolling and sliding screen content based on estimations of rough gaze regions by the PCR method. We demonstrated that low-fidelity eye tracking enables people to control content on a remote large display, e.g., browsing music albums.

We further deepen our understanding of gaze interaction in public environments through multiple field deployments of GazeHorizon (Chapter 7). Our studies revealed that people cannot learn gaze interactions by observing other users, as eye movements are subtle. By providing appropriate interactive guidance, we can bootstrap novice users to comprehend gaze interaction and to overcome the limitations of the vision tracking system without any human assistance.

Pervasive displays can also be shared by more than one user. In Chapter 8, we explored gaze in shared displays. We showed that gaze enhanced co-located collaborations by easing communications and assisting in establishing and maintaining shared attention. Our results also identified that distraction and privacy remain open challenges for the design of multi-user gaze interfaces.

The work of this thesis shows that we can devise new gaze estimation techniques and design complementary user interfaces that are suitable for pervasive displays. Nevertheless, open challenges remain for the use of eye tracking in ubiquitous computing. In the following sections, we discuss our lessons learned from the course of this research and propose potential solutions to overcome the challenges.

9.2 Challenges and Lessons Learned

From conducting the research of this thesis, we learned several lessons:

Customise eye tracking for different applications: The application design of this thesis is system driven, where we optimised our interface and interaction to overcome the limitations of the system. System requirements can be defined by interactions and applications. For instance, using relative eye movements as an input, the interactions (e.g., gaze gestures, scroll interaction in GazeHorizon) do not need estimations of onscreen gaze position. Gaze sensing can be simplified by tracking users' eye positions, which eliminates the calibration process. Additionally, for applications that use rough gaze regions, such as SideWays and existing attentive user interfaces, a low-fidelity solution is adequate, reduces hardware requirements and can be achieved using off-the-shelf components (e.g., web cameras). Hence, high-fidelity eye tracking devices are not always needed and designers can devise novel interactions that are also useful with low-fidelity eye tracking solutions.

Large display interaction: Large displays are becoming more interactive. Previous research largely focused on using touch or using external devices (e.g. mobile phones). With the tremendous development of sensing technology for natural user behaviours, there has been increasing work in the past few years that enable interaction through the user's body, e.g., body position, body posture and hand gestures [17]. Compared to common touch input, the use of gaze would be less appropriate in applications that require fine grained manipulation input, because gaze input is not as accurate and expressive as gesture input on a planar surface, such as multi-touch. Users dislike using their eyes for extensive motor control, e.g., long fixations, frequent voluntary saccades with large amplitudes. In addition, gaze is found to be intuitive, effortless and effective in our user studies.

The intuitiveness of eye movements provide advantages over using hand gestures, as gesture recognitions depend on cultural background and the specific context. It is worth noting that as eye movements are subtle, people felt less conscious with using gaze for interaction than other body movements in public. Lastly, gaze is a promising modality that can make interactive public displays more accessible to disabled people, in particular for people who are in wheelchairs.

Gaze interaction for pervasive contexts: There are several things that researchers need to consider to apply gaze in pervasive contexts. *Eye movements are subtle* and make the transfer of the interaction knowledge challenging in pervasive contexts. In home or office environments, people can learn gaze interactions or their usage via communicating with experienced users; however, assistance from an experienced user might not be available in public contexts. In Chapter 7, we proposed one way to teach user interaction through interactive guidance. Researchers can explore other strategies to overcome this challenge. For example, we could design standards for gaze interactions, similar to existing gesture sets for touch interfaces (e.g., pinch to zoom). In this way, people can transfer the knowledge from other interfaces or past experiences.

Gaze is involved in the majority of tasks that we do in daily life. We suggest designers to employ gaze *at its natural occurrence*. Our eyes have special characteristics and it is difficult to control their motion intentionally, unlike other body parts (e.g., hand, head and arm). We encourage researchers to consider the natural affordance of gaze input, for example, how gaze is used in daily life. Designers should also consider using gaze in the way that reflects its natural meaning (e.g., interests, attention or social signals), such that other modalities do not replace gaze input quickly.

Privacy is another concern because eye movements might reveal personal information. Eye movements and gaze are closely related to human internal cognitive and mental states. Extensive research works have shown the link of gaze data to other sensitive personal information, such as age, gender, interests and preferences, health, and focus level. Our eye movements might reveal private information unconsciously that we are not aware of. We believe that extensive attention should be paid to how gaze data is collected, stored, and analysed in the pervasive world [108].

System improvements: Although eye tracking systems have made a tremendous progress over the last few years, there are several ways that the technology can be improved. Existing commercial eye tracking devices provide high accuracy, but their tracking range is still limited to approximately 50cm to 90cm due to the use of glints [1]. Thus it cannot accommodate people with different heights or if they are standing far away from the tracking device. A few research papers reported capabilities of remote eye tracking (e.g., usually used zoom lens, pan and tilt mechanics and high levels IR illuminations), but we still do not have a reliable remote eye tracker available on the market. We would hope more work on remote gaze tracking will be carried out in the near future. Another common eye tracking problem is caused by the reflection of users' glasses or contact lenses. Tracking accuracy deteriorates largely with thick lenses, and sometimes even becomes unusable. We suggest future research to take into account of lens reflectivity and also include subjects with glasses or contact lenses when building gaze estimation models.

9.3 Future Directions

In recent years, we have seen increasing interest in eye tracking application in both the academic context and industry. We predict that in the near future eye tracking will be a necessity to assist our daily life, for example, to support human navigation, identify information, and perceive the world around us. We highlight the following directions which we think eye tracking would be a good complement and may enable new applications.

Implicit Gaze Interaction with Pervasive Displays

One question remains open is that if pervasive displays can sense users' gaze, how can implicit gaze information (e.g., users' interests, attention) be harnessed to improve services provided by these displays?

Currently, many large displays are becoming digital with dynamic content (e.g., food menu, product catalogue). We hope, in the future, gaze capability can be built into these displays. Users can interact with the displays using their natural behaviour such as gaze combined with other modalities (e.g., speech, skin input, body movements).

Pervasive displays can implicitly analyse users' gaze behaviour to infer their interests and preferences. The implicit feedback can be used to adapt content displayed on the screen and improve usability of the pervasive displays, e.g., effectiveness of advertisements. Also, eye movement patterns can be used to differentiate user behaviour, for example, whether users are just browsing or looking for a particular product. This implicit feedback can be used to customise advertising and to refine people's information seeking on their personal devices. We imagine that, in clothing stores, a catalogue display gathers the information of products that people look at. It can analyse the colour, the texture and style that the customers look at in real-time. The results could be used to prompt further suggestions on items people might be interested in. We also imagine that, in supermarkets, users' gaze is analysed to infer products of their interests. Gaze information can be used to prompt personalised information to notify available date, discounts and offers. The similar concepts can be also applied to other scenarios or objects in our daily life.

Gaze Estimation in the Real World

Another open challenge is how gaze can be exploited in real world environments beyond displays, for example, navigation in cities and communications with people around us. Although remote low-cost commercial eye trackers have emerged, users' movements are still restricted and they cannot move around freely. Gaze estimation has still largely focused on a 2D planar surface, but we do not have a good estimation where people are focusing on in the real world. 3D gaze information is highly correlated with user intentions in the context of navigation and manipulation of our surroundings. To estimate people's gaze in non-restricted space is a challenging task, however, we envision gaze estimation in the real world can have numerous applications. We imagine that in future shops customers' gaze will be tracked and their reactions to different products can be automatically recorded. Gaze can be combined together with other behaviour data, such as facial expression, body gestures or heart rate gathered from wearable sensors, to interpret user behaviour. This allows understanding of natural behaviour in realistic scenes in marketing and user research.

Similar applications can be useful in social signal processing and robotics. We envision future robotic systems with cognitive and social capabilities will be able to instruct, play and communicate with humans. Robotic systems can imitate how human visual systems work by observing human's gaze behaviour. Gaze estimation can be used for automatic annotation in who and what people are paying attention to in group interaction, group involvement, rapport and engagement, e.g., meeting, education environments. Another context in which people can benefit from this is in the outdoor environments. For example, when people require remote assistance, in a repair task (e.g., road breakdown), gaze would be beneficial to communicate with remote instructors to provide efficient guidance and collaboration.

Bibliography

[1] http://www.eyegaze.com/eye-tracking-research-studies/. [Cited on pages 4 and 183]

[2] http://www.mrpalsmy.com/?cat=3. [Cited on page 14]

[3] http://webvision.med.utah.edu/book/part-i-foundations/simple-anatomy-of-the-retina/. [Cited on page 15]

[4] http://www.audiologyonline.com/articles/ics-impulse-revolutionizing-vestibular-assessment-12003. [Cited on page 16]

[5] http://www.chronos-vision.de/medical-engineering-produkte.html. [Cited on page 16]

[6] http://www.eyegaze.com/. [Cited on page 18]

[7] François Meunier, Ph. D. and ing. (2009). On the Automatic Implementation of the Eye Involuntary Reflexes Measurements Involved in the Detection of Human Liveness and Impaired Faculties, Image Processing, Yung-Sheng Chen (Ed.), ISBN: 978-953-307-026-1, InTech, DOI: 10.5772/7054. Available from: http://www.intechopen.com/books/image-processing/on-the-automatic-implementation-of-the-eye-involuntary-reflexes-measurements-involved-in-the-detecti. [Cited on page 21]

[8] http://www.2020mag.com/ce/TTViewTest.aspx?LessonId=109646. [Cited on page 22]

[9] http://www.xuuk.com/eyebox2/. [Cited on page 27]

[10] http://www.logitech.com/lang/pdf/ib-rightlight_EN.pdf. [Cited on page 64]

[11] http://opencv.org/. [Cited on page 64]

[12] https://kin450-neurophysiology.wikispaces.com/Saccades. [Cited on page 74]

[13] Gregory D. Abowd and Elizabeth D. Mynatt. Charting past, present, and future research in ubiquitous computing. *ACM Trans. Comput.-Hum. Interact.*, 7(1):29–58, March 2000. [Cited on pages 1, 3, and 5]

[14] Nicholas Adams, Mark Witkowski, and Robert Spence. The inspection of very large images by eye-gaze control. In Stefano Levialdi, editor, *Proc. AVI 2008*, pages 111–118. ACM Press, 2008. [Cited on pages 78 and 79]

[15] Florian Alt, Alireza Sahami Shirazi, Albrecht Schmidt, and Julian Mennenöh. Increasing the user's attention on the web: Using implicit interaction based on gaze behavior to tailor content. In *Proc. NordiCHI 2012*, pages 544–553. ACM, 2012. [Cited on pages 73 and 79]

[16] Richard Andersson, Marcus Nyström, and Kenneth Holmqvist. Sampling frequency and eye-tracking measures: how speed affects durations, latencies, and more. *Journal of Eye Movement Research*, 3(3:6):1–12, 2010. [Cited on page 15]

[17] Carmelo Ardito, Paolo Buono, Maria Francesca Costabile, and Giuseppe Desolda. Interaction with large displays: A survey. *ACM Comput. Surv.*, 47(3):46:1–46:38, February 2015. [Cited on page 181]

[18] Michael Ashmore, Andrew T. Duchowski, and Garth Shoemaker. Efficient eye pointing with a fisheye lens. In *Proc. GI 2005*, pages 203–210, 2005. [Cited on page 77]

[19] Reynold Bailey, Ann McNamara, Nisha Sudarsanam, and Cindy Grimm. Subtle gaze direction. *ACM Trans. Graph.*, 28(4):100:1–100:14, September 2009. [Cited on page 80]

[20] S. Baluja and D. Pomerleau. Non-intrusive gaze tracking using artificial neural networks. Technical report cmu-cs-94-102, Carnegie Mellon University, 1994. [Cited on pages 23, 25, 31, 35, and 52]

[21] Rafael Barea, Luciano Boquete, Jose Manuel Rodriguez-Ascariz, Sergio Ortega, and Elena López. Sensory system for implementing a human-computer interface based on electrooculography. *Sensors*, 11(1):310, 2010. [Cited on page 17]

[22] Patrick Baudisch, Doug DeCarlo, Andrew T. Duchowski, and Wilson S. Geisler. Focusing on the essential: Considering attention in display design. *Commun. ACM*, 46(3):60–66, March 2003. [Cited on pages 2, 73, and 80]

[23] Victoria Bellotti, Maribeth Back, W. Keith Edwards, Rebecca E. Grinter, Austin Henderson, and Cristina Lopes. Making sense of sensing systems: Five questions for designers and researchers. In *Proc. CHI 2002*, pages 415–422, New York, NY, USA, 2002. ACM. [Cited on pages 1 and 3]

[24] J. Benedek and T. Miner. Measuring desirability: New methods for evaluating desirability in a usability lab setting. In *Proc. UPA 2002*, 2002. [Cited on page 166]

[25] D. Beymer and M. Flickner. Eye gaze tracking using an active stereo head. In *Proc. CVPR 2003*, volume 2, pages II–451–8 vol.2, June 2003. [Cited on pages 4 and 21]

[26] Stanley T. Birchfield and Sriram Rangarajan. Spatiograms versus histograms for region-based tracking. In *Proc. CVPR 2005*, pages 1158–1163, 2005. [Cited on page 34]

[27] Richard A. Bolt. Gaze-orchestrated dynamic windows. In *Proc. SIGGRAPH 1981*, pages 109–119, New York, NY, USA, 1981. ACM. [Cited on pages 71, 73, and 80]

[28] Richard A. Bolt. Eyes at the interface. In *Proc. CHI 1982*, pages 360–362. ACM, 1982. [Cited on pages 2, 73, and 80]

[29] Susan E. Brennan, Xin Chen, Christopher A. Dickinson, Mark B. Neider, and Gregory J. Zelinsky. Coordinating cognition: The costs and benefits of shared gaze during collaborative search. *Cognition*, 106(3):1465 – 1477, 2008. [Cited on pages 85, 145, 148, 153, and 172]

[30] Harry Brignull and Yvonne Rogers. Enticing people to interact with large public displays in public spaces. In *Proc. INTERACT 2003*, pages 17–24. IOS Press, 2003. [Cited on pages 116 and 133]

[31] F. Broz, H. Lehmann, C.L. Nehaniv, and K. Dautenhahn. Mutual gaze, personality, and familiarity: Dual eye-tracking during conversation. In *RO-MAN, 2012 IEEE*, pages 858–864, Sept 2012. [Cited on page 83]

[32] Andreas Bulling, Jamie A. Ward, Hans Gellersen, and Gerhard Tröster. Eye Movement Analysis for Activity Recognition Using Electrooculography. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 33(4):741–753, April 2011. [Cited on pages 15 and 17]

[33] Andreas Bulling, Christian Weichel, and Hans Gellersen. Eyecontext: Recognition of high-level contextual cues from human visual behaviour. In *Proc. CHI 2013*, CHI '13, pages 305–308, New York, NY, USA, 2013. ACM. [Cited on pages 2, 4, and 27]

[34] Georg Buscher, Susan T. Dumais, and Edward Cutrell. The good, the bad, and the random: An eye-tracking study of ad quality in web search. In *Proc. SIGIR 2010*, pages 42–49, New York, NY, USA, 2010. ACM. [Cited on page 70]

[35] W. Buxton. Integrating the periphery and context: A new taxonomy of telematics. In *Proc. GI 1995*, pages 239–246. Morgan Kaufman, 1995. [Cited on page 1]

[36] R.H.S. Carpenter. *Movements of the eyes.* Pion, 1988. [Cited on page 13]

[37] Juan J. Cerrolaza, Arantxa Villanueva, and Rafael Cabeza. Study of polynomial mapping functions in video-oculography eye trackers. *ACM Trans. Comput.-Hum. Interact.*, 19(2):10:1–10:25, July 2012. [Cited on page 25]

[38] Jixu Chen and Qiang Ji. Probabilistic gaze estimation without active personal calibration. In *Proc. CVPR 2011*, pages 609–616. IEEE Computer Society, 2011. [Cited on pages 4, 13, 27, and 45]

[39] R. Chen and O. Kalinli. Interface using eye tracking contact lenses, November 8 2012. US Patent App. 13/101,868. [Cited on pages 15 and 16]

[40] Mauro Cherubini, Marc-Antoine Nüssli, and Pierre Dillenbourg. Deixis and gaze in collaborative work at a distance (over a shared map): A computational model to detect misunderstandings. In *Proc. ETRA 2008*, pages 173–180. ACM, 2008. [Cited on pages 85, 145, 148, and 149]

[41] Dong-Chan Cho and Whoi-Yul Kim. Long-range gaze tracking system for large movements. *Biomedical Engineering, IEEE Transactions on*, 60(12):3432–3440, Dec 2013. [Cited on page 18]

[42] Herbert H. Clark and Deanna Wilkes-Gibbs. Referring as a collaborative process. *Cognition*, 22(1):1 – 39, 1986. [Cited on page 147]

[43] Sarah Clinch, Mateusz Mikusz, Miriam Greis, Nigel Davies, and Adrian Friday. *Mercury: an application store for open display networks*, pages 511–522. ACM, 2014. [Cited on page 3]

[44] Mark Cook. Gaze and mutual gaze in social encounters: How long and when we look others "in the eye" is one of the main signals in nonverbal communication. *American Scientist*, 65(3):pp. 328–333, 1977. [Cited on page 83]

[45] Laura Cowen, Linden J. Ball, and Judy Delin. *An Eye Movement Analysis of Webpage Usability.* Springer-Verlag Ltd., 2002. [Cited on page 70]

[46] Tarik Crnovrsanin, Yang Wang, and Kwan-Liu Ma. Stimulating a blink: Reduction of eye fatigue with visual stimulus. In *Proc. CHI 2014*, CHI '14, pages 2055–2064, New York, NY, USA, 2014. ACM. [Cited on pages 27 and 79]

[47] J. Daugman. New methods in iris recognition. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 37(5):1167–1175, 2007. [Cited on page 32]

[48] Connor Dickie, Jamie Hart, Roel Vertegaal, and Alex Eiser. Lookpoint: An evaluation of eye input for hands-free switching of input devices between multiple computers. In *Proc. OZCHI 2006*, pages 119–126, New York, NY, USA, 2006. ACM. [Cited on pages 2 and 78]

[49] Jakub Dostal, Per Ola Kristensson, and Aaron Quigley. Subtle gaze-dependent techniques for visualising display changes in multi-display environments. In *Proc. IUI 2013*, IUI '13, pages 137–148, New York, NY, USA, 2013. ACM. [Cited on pages 27 and 78]

[50] Heiko Drewes, Alexander De Luca, and Albrecht Schmidt. Eye-gaze interaction for mobile phones. In *Proceedings of the 4th International Conference on Mobile Technology, Applications, and Systems and the 1st International Symposium on Computer Human Interaction in Mobile Technology*, Mobility '07, pages 364–371, New York, NY, USA, 2007. ACM. [Cited on page 81]

[51] Heiko Drewes and Albrecht Schmidt. Interacting with the computer using gaze gestures. In *Proc. INTERACT 2007*, INTERACT'07, pages 475–488, Berlin, Heidelberg, 2007. Springer-Verlag. [Cited on page 76]

[52] Heiko Drewes and Albrecht Schmidt. The magic touch: Combining magic-pointing with a touch-sensitive mouse. In *Proc. INTERACT 2009*, INTERACT '09, pages 415–428, Berlin, Heidelberg, 2009. Springer-Verlag. [Cited on pages 76 and 77]

[53] Morten Lund Dybdal, Javier San Agustin, and John Paulin Hansen. Gaze input for mobile devices by dwell and gestures. In *Proc. ETRA 2012*, ETRA '12, pages 225–228, New York, NY, USA, 2012. ACM. [Cited on page 76]

[54] Marc Eaddy, Gabor Blasko, Jason Babcock, and Steven Feiner. My own private kiosk: Privacy-preserving public displays. In *Proc. ISWC 2004*, pages 132–135. IEEE Computer Society, 2004. [Cited on pages 71 and 91]

[55] Alireza Fathi, Yin Li, and James M. Rehg. Learning to recognize daily actions using gaze. In *Proc. ECCV 2012*, ECCV'12, pages 314–327, Berlin, Heidelberg, 2012. Springer-Verlag. [Cited on page 2]

[56] P. M. Fitts, R. E. Jones, and J. L. Milton. Eye movements of aircraft pilots during instrument-landing approaches. *Aeronautical Engineering Review*, 9(2):24–29, 1950. [Cited on pages 13, 69, and 70]

[57] David Fono and Roel Vertegaal. Eyewindows: Evaluation of eye-controlled zooming windows for focus selection. In *Proc. CHI 2005*, pages 151–160, New York, NY, USA, 2005. ACM. [Cited on pages 5 and 78]

[58] L.A. Frey, Jr. White, K.P., and T.E. Hutchison. Eye-gaze word processing. *IEEE Transactions on Systems, Man and Cybernetics*, 20(4):944–950, Jul 1990. [Cited on pages 2 and 71]

[59] K.A. Funes Mora and J. Odobez. Gaze estimation from multimodal kinect data. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2012 IEEE Computer Society Conference on*, pages 25–30, June 2012. [Cited on pages 13, 23, and 24]

[60] K.A. Funes Mora and J.-M. Odobez. Geometric generative gaze estimation (g3e) for remote rgb-d cameras. In *Proc. CVPR 2014*, pages 1773–1780, June 2014. [Cited on pages 4, 13, 22, and 23]

[61] Kenneth Alberto Funes Mora, Florent Monay, and Jean-Marc Odobez. Eyediap: A database for the development and evaluation of gaze estimation algorithms from

rgb and rgb-d cameras. In *Proc. of ETRA 2014*. ACM, March 2014. [Cited on page 7]

[62] Darren Gergle, Robert E. Kraut, and Susan R. Fussell. Language efficiency and visual technology: Minimizing collaborative effort with visual information. *Journal of Language and Social Psychology*, pages 491–517, 2004. [Cited on pages 84, 148, and 149]

[63] Jan-Mark Geusebroek, Rein van den Boomgaard, Arnold W.M. Smeulders, and Hugo Geerts. Color invariance. *IEEE Trans. Pattern Anal. Mach. Intell.*, 23(12):1338–1350, 2001. [Cited on page 49]

[64] T. Gevers. Color in image search engines. *Principles of Visual Information Retrieval*, page 35, 2001. [Cited on page 32]

[65] Saul Greenberg, Carl Gutwin, and Mark Roseman. Semantic telepointers for groupware. In *Proc. OZCHI 1996*, OZCHI '96, pages 54–, Washington, DC, USA, 1996. IEEE Computer Society. [Cited on pages 171, 174, and 175]

[66] Saul Greenberg, Nicolai Marquardt, Till Ballendat, Rob Diaz-Marino, and Miaosen Wang. Proxemic interactions: The new ubicomp? *interactions*, 18(1):42–50, January 2011. [Cited on page 116]

[67] Brian Guenter, Mark Finch, Steven Drucker, Desney Tan, and John Snyder. Foveated 3d graphics. In *ACM Transactions on Graphics*. ACM SIGGRAPH Asia, November 2012. [Cited on pages 73 and 80]

[68] E.D. Guestrin and E. Eizenman. General theory of remote gaze estimation using the pupil center and corneal reflections. *IEEE Transactions on Biomedical Engineering*, 53(6):1124–1133, June 2006. [Cited on page 21]

[69] Carl Gutwin and Saul Greenberg. Design for individuals, design for groups: Tradeoffs between power and workspace awareness. In *Proc. CSCW 1998*, CSCW '98, pages 207–216, New York, NY, USA, 1998. ACM. [Cited on page 172]

[70] Carl Gutwin and Saul Greenberg. A descriptive framework of workspace awareness for real-time groupware. *Comput. Supported Coop. Work*, 11(3):411–446, November 2002. [Cited on page 175]

[71] Carl Gutwin, Mark Roseman, and Saul Greenberg. A usability study of awareness widgets in a shared workspace groupware system. In *Proc. CSCW 1996*, pages 258–267. ACM, 1996. [Cited on pages 171 and 175]

[72] Dan Witzner Hansen and Arthur E. C. Pece. Eye tracking in the wild. *Comput. Vis. Image Underst.*, 98(1):155–181, April 2005. [Cited on pages 18, 22, 26, and 61]

[73] Dan Witzner Hansen, Henrik H. T. Skovsgaard, John Paulin Hansen, and Emilie Møllenbach. Noise tolerant selection by gaze-controlled pan and zoom in 3d. In *Proc. ETRA 2008*, ETRA '08, pages 205–212, New York, NY, USA, 2008. ACM. [Cited on pages 76, 78, 79, and 80]

[74] D.W. Hansen and Q. Ji. In the eye of the beholder: A survey of models for eyes and gaze. *IEEE Trans. Pattern Anal. Mach. Intell.*, 32(3):478–500, 2010. [Cited on pages 2, 3, 4, 15, 18, 20, 23, 26, 32, 40, 61, and 74]

[75] John Paulin Hansen, Anders Sewerin Johansen, Dan Witzner Hansen, and Kenji Itoh. Command without a click: Dwell time typing by mouse and gaze selections. In M. Rauterberg, editor, *Proc. INTERACT 2003*, pages 121 – 128. IOS Press, 2003. [Cited on page 71]

[76] John M. Henderson. Human gaze control during real-world scene perception. *Trends in Cognitive Sciences*, 7(11):498–504, 2003. [Cited on pages 2 and 71]

[77] Craig Hennessey and Jacob Fiset. Long range eye tracking: Bringing eye tracking into the living room. In *Proc. ETRA 2012*, ETRA '12, pages 249–252, New York, NY, USA, 2012. ACM. [Cited on page 28]

[78] David Holman. Gazetop: Interaction techniques for gaze-aware tabletops. In *Proc. EA CHI 2007*, CHI EA '07, pages 1657–1660, New York, NY, USA, 2007. ACM. [Cited on page 82]

[79] Anthony Hornof, Anna Cavender, and Rob Hoselton. Eyedraw: A system for drawing pictures with eye movements. In *Proc. Assets 2004*, Assets '04, pages 86–93, New York, NY, USA, 2004. ACM. [Cited on pages 2 and 71]

[80] Edmund Burke Huey. The psychology and pedagogy of reading. 1908. [Cited on pages 13 and 69]

[81] T.E. Hutchinson, Jr. White, K.P., W.N. Martin, K.C. Reichert, and L.A. Frey. Human-computer interaction using eye-gaze input. *Trans. Sys. Man Cyber Part C*, 19(6):1527 –1534, 1989. [Cited on pages 2, 71, and 75]

[82] Aulikki Hyrskykari, Howell Istance, and Stephen Vickers. Gaze gestures or dwell-based interaction? In *Proc. ETRA 2012*, pages 229–232. ACM Press, 2012. [Cited on page 76]

[83] P. Isenberg, S. Carpendale, A. Bezerianos, N. Henry, and J. Fekete. Coconuttrix: Collaborative retrofitting for information visualization. *Computer Graphics and Applications, IEEE*, 29(5):44–57, Sept 2009. [Cited on page 148]

[84] Yoshio Ishiguro, Adiyan Mujibiya, Takashi Miyaki, and Jun Rekimoto. Aided eyes: Eye activity sensing for daily life. In *Proc. AH 2010*, AH '10, pages 25:1–25:7, New York, NY, USA, 2010. ACM. [Cited on pages 2, 4, 18, 27, and 81]

[85] Hiroshi Ishii and Minoru Kobayashi. Clearboard: A seamless medium for shared drawing and conversation with eye contact. In *Proc. CHI 1992*, pages 525–532. ACM, 1992. [Cited on page 84]

[86] R. J. K. Jacob and K. S. Karn. Eye tracking in human-computer interaction and usability research: Ready to deliver the promises. *The Mind's Eye: Cognitive and Applied Aspects of Eye Movement Research*, pages 573–603, 2003. [Cited on pages 2, 5, 44, and 71]

[87] Robert J. K. Jacob. The use of eye movements in human-computer interaction techniques: what you look at is what you get. *ACM Trans. Inf. Syst.*, 9(2):152–169, 1991. [Cited on pages 2, 71, 73, 75, 76, 86, and 110]

[88] Robert J. K. Jacob. Eye movement-based human-computer interaction techniques: Toward non-command interfaces. In *Advances in Human-Computer Interaction*, pages 151–190. Ablex Publishing Co, 1993. [Cited on pages 2, 72, 73, and 76]

[89] Robert J. K. Jacob. Eye tracking in advanced interface design. In Woodrow Barfield and Thomas A. Furness, III, editors, *Virtual Environments and Advanced Interface Design*, pages 258–288. Oxford University Press, Inc., New York, NY, USA, 1995. [Cited on page 76]

[90] Marcel Adam Just and Patricia A. Carpenter. A theory of reading: from eye fixations to comprehension. *Psychological review*, 87(4):329, 1980. [Cited on pages 2, 13, 71, and 73]

[91] Jari Kangas, Deepak Akkil, Jussi Rantala, Poika Isokoski, Päivi Majaranta, and Roope Raisamo. Gaze gestures and haptic feedback in mobile devices. In *Proc. CHI 2014*, CHI '14, pages 435–438, New York, NY, USA, 2014. ACM. [Cited on page 81]

[92] Robert V. Kenyon. A soft contact lens search coil for measuring eye movements. *Vision Research*, 25(11):1629 – 1633, 1985. [Cited on pages 15 and 16]

[93] Dagmar Kern, Paul Marshall, and Albrecht Schmidt. Gazemarks: gaze-based visual placeholders to ease attention switching. In Elizabeth D. Mynatt, Don Schoner, Geraldine Fitzpatrick, Scott E. Hudson, W. Keith Edwards, and Tom Rodden, editors, *Proc. CHI 2010*, pages 2093–2102. ACM, 2010. [Cited on page 80]

[94] Jeongseok Ki and Yong-Moo Kwon. 3d gaze estimation and interaction. In *3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video, 2008*, pages 373–376, May 2008. [Cited on page 80]

[95] Stefan Kohlbecher, Stanislavs Bardinst, Klaus Bartl, Erich Schneider, Tony Poitschke, and Markus Ablassmeier. Calibration-free eye tracking by reconstruction of the pupil ellipse in 3d space. In *Proc. ETRA 2008*, ETRA '08, pages 135–138, New York, NY, USA, 2008. ACM. [Cited on page 21]

[96] László Kozma, Arto Klami, and Samuel Kaski. Gazir: Gaze-based zooming interface for image retrieval. In *Proc. ICMI-MLMI 2009*, ICMI-MLMI '09, pages 305–312, New York, NY, USA, 2009. ACM. [Cited on page 73]

[97] Robert E. Kraut, Susan R. Fussell, and Jane Siegel. Visual information as a conversational resource in collaborative physical tasks. *Hum.-Comput. Interact.*, 18(1):13–49, June 2003. [Cited on pages 85 and 147]

[98] Hannu Kukka, Heidi Oja, Vassilis Kostakos, Jorge Gonçalves, and Timo Ojala. What makes you click: exploring visual signals to entice interaction on public displays. In *Proc. CHI 2013*, pages 1699–1708. ACM Press, 2013. [Cited on pages 116 and 126]

[99] Manu Kumar, Jeff Klingner, Rohan Puranik, Terry Winograd, and ndreas Paepcke. Improving the accuracy of gaze input for interaction. In *Proc. ETRA 2008*, ETRA '08, pages 65–68, New York, NY, USA, 2008. ACM. [Cited on pages 77 and 150]

[100] Manu Kumar, Andreas Paepcke, and Terry Winograd. Eyepoint: Practical pointing and selection using gaze and keyboard. In *Proc. CHI 2007*, CHI '07, pages 421–430, New York, NY, USA, 2007. ACM. while the speed of a gaze-based pointing was comparable to the mouse, error rates were significantly higher. [Cited on pages 5, 74, and 77]

[101] Manu Kumar and Terry Winograd. Gaze-enhanced scrolling techniques. In *Proc. UIST 2007*, pages 213–216. ACM Press, 2007. [Cited on pages 5 and 78]

[102] Chih-Chuan Lai, Sheng-Wen Shih, Hsin-Ruey Tsai, and Yi-Ping Hung. 3-d gaze tracking using pupil contour features. In *Proc. ICPR 2014*, pages 1162–1166, Aug 2014. [Cited on page 21]

[103] Michael F. Land. Eye movements and the control of actions in everyday life. *Progress in Retinal and Eye Research*, 25(3):296 – 324, 2006. [Cited on pages 2, 71, and 73]

[104] Chris Lankford. Effective eye-gaze input into windows. In *Proc. ETRA 2000*, ETRA '00, pages 23–27, New York, NY, USA, 2000. ACM. [Cited on pages 5, 75, and 76]

[105] Yves Le Grand. *Physiological optics.* Berlin ; New York : Springer-Verlag, 1980. Translation of La dioptrique de l'oeil et sa correction, published as v. 1 of the author's Optique physiologique. [Cited on page 19]

[106] Hyeon Chang Lee, Won Oh Lee, Chul Woo Cho, Su Yeong Gwon, Kang Ryoung Park, Heekyung Lee, and Jihun Cha. Remote gaze tracking system on a large display. *Sensors*, 13(10):13439, 2013. [Cited on page 28]

[107] Ji Woo Lee, Hwan Heo, and Kang Ryoung Park. A novel gaze tracking method based on the generation of virtual calibration points. *Sensors*, 13(8):10802, 2013. [Cited on page 21]

[108] Daniel J. Liebling and Sören Preibusch. Privacy considerations for a pervasive eye tracking world. In *Proc. UbiComp Adjunct 2014*, UbiComp '14 Adjunct, pages 1169–1177, New York, NY, USA, 2014. ACM. [Cited on page 183]

[109] Feng Lu, Takahiro Okabe, Yusuke Sugano, and Yoichi Sato. A head pose-free approach for appearance-based gaze estimation. In *Proc. BMVC 2011*, pages 126.1–126.11. BMVA Press, 2011. [Cited on page 23]

[110] Feng Lu, Y. Sugano, T. Okabe, and Y. Sato. Adaptive linear regression for appearance-based gaze estimation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 36(10):2033–2046, Oct 2014. [Cited on pages 23 and 25]

[111] I. Scott MacKenzie and Xuang Zhang. Eye typing using word and letter prediction and a fixation algorithm. In *Proc. ETRA 2008*, pages 55–58. ACM Press, 2008. [Cited on page 71]

[112] J. J. Magee, M. Betke, J. Gips, M. R. Scott, and B. N. Waber. A human-computer interface using symmetry between eyes to detect gaze direction. *Trans. Sys. Man Cyber. Part A*, 38(6):1248–1261, 2008. [Cited on pages 71 and 91]

[113] Päivi Majaranta, Scott MacKenzie, Anne Aula, and Kari-Jouko Räihä. Effects of feedback and dwell time on eye typing speed and accuracy. *Univers. Access Inf. Soc.*, 5(2):199–208, July 2006. [Cited on pages 71 and 72]

[114] J. Malmivuo and R. Plonsey. *Bioelectromagnetism - Principles and Applications of Bioelectric and Biomagnetic Fields.* Oxford University Press, 1995. [Cited on page 17]

[115] Diako Mardanbegi, Dan Witzner Hansen, and Thomas Pederson. Eye-based head gestures. In *Proc. ETRA 2012*, pages 139–146. ACM Press, 2012. [Cited on page 125]

[116] Paul Marshall, Richard Morris, Yvonne Rogers, Stefan Kreitmayer, and Matt Davies. Rethinking'multi-user': an in-the-wild study of how groups approach a walk-up-and-use tabletop interface. In *Proc. CHI 2011*, pages 3033–3042. ACM, 2011. [Cited on pages 116 and 138]

[117] Bernhard Maurer, Sandra Trösterer, Magdalena Gärtner, Martin Wuchse, Axel Baumgartner, Alexander Meschtscherjakov, David Wilfinger, and Manfred Tscheligi. Shared gaze in the car: Towards a better driver-passenger collaboration. In *Adjunct Proc. AutomotiveUI*, pages 1–6. ACM, 2014. [Cited on pages 148 and 149]

[118] Emiliano Miluzzo, Tianyu Wang, and Andrew T. Campbell. Eyephone: Activating mobile phones with your eyes. In *Proceedings of the Second ACM SIGCOMM*

Workshop on Networking, Systems, and Applications on Mobile Handhelds, MobiHeld '10, pages 15–20, New York, NY, USA, 2010. ACM. [Cited on pages 18, 27, 71, 81, and 82]

[119] Darius Miniotas, Oleg Špakov, and I. Scott MacKenzie. Eye gaze interaction with expanding targets. In *Proc. EA CHI 2004*, CHI EA '04, pages 1255–1258, New York, NY, USA, 2004. ACM. [Cited on page 77]

[120] Carlos H. Morimoto and Marcio R. M. Mimica. Eye gaze tracking techniques for interactive applications. *Comput. Vis. Image Underst.*, 98(1):4–24, 2005. [Cited on pages 4, 13, 21, 25, 27, 45, 52, and 86]

[121] C.H Morimoto, D Koons, A Amir, and M Flickner. Pupil detection and tracking using multiple light sources. *Image and Vision Computing*, 18(4):331 – 335, 2000. [Cited on page 21]

[122] Omar Mubin, Tatiana Lashina, and Evert Loenen. How not to become a buffoon in front of a shop window: A solution allowing natural head movement for interaction with a public display. In *Proc. INTERACT 2009*, pages 250–263. Springer-Verlag, 2009. [Cited on pages 90 and 91]

[123] Jörg Müller, Florian Alt, Daniel Michelis, and Albrecht Schmidt. Requirements and design space for interactive public displays. In *Proc. MM 2010*, pages 1285–1294. ACM Press, 2010. [Cited on pages 3 and 126]

[124] Jörg Müller, Robert Walter, Gilles Bailly, Michael Nischt, and Florian Alt. Looking glass: a field study on noticing interactivity of a shop window. In *Proc. CHI 2012*, pages 297–306. ACM Press, 2012. [Cited on pages 116 and 131]

[125] Takashi Nagamatsu, Michiya Yamamoto, and Hiroshi Sato. Mobigaze: Development of a gaze interface for handheld mobile devices. In *Proc. EA CHI 2010*, CHI EA '10, pages 3349–3354, New York, NY, USA, 2010. ACM. [Cited on pages 71, 81, and 82]

[126] Yasuto Nakanishi, Takashi Fujii, Kotaro Kiatjima, Yoichi Sato, and Hideki Koike. Vision-based face tracking system for large displays. In *Proc. UbiComp 2002*, pages 152–159. Springer-Verlag, 2002. [Cited on pages 28 and 90]

[127] MarkB. Neider, Xin Chen, ChristopherA. Dickinson, SusanE. Brennan, and GregoryJ. Zelinsky. Coordinating spatial referencing using shared gaze. *Psychonomic Bulletin & Review*, 17(5):718–724, 2010. [Cited on pages 85, 149, and 172]

[128] Karlene Nguyen, Cindy Wagner, David Koons, and Myron Flickner. Differences in the infrared bright pupil response of human eyes. In *Proc. ETRA 2002*, ETRA '02, pages 133–138, New York, NY, USA, 2002. ACM. [Cited on page 22]

[129] Jakob Nielsen and Kara Pernice. *Eyetracking web usability.* New Riders, Berkeley, CA, 2010. [Cited on page 70]

[130] B. Noris, K. Benmachiche, and A.G. Billard. Calibration-free eye gaze direction detection with gaussian processes. In *Proc. VISAPP 2008*, pages 611–616. INSTICC, 2008. [Cited on pages 23, 24, 25, 27, 45, and 52]

[131] Maja Pantic, Alex Pentland, Anton Nijholt, and Thomas Huang. Human computing and machine understanding of human behavior: A survey. In *Proc. ICMI 2006*, ICMI '06, pages 239–248, New York, NY, USA, 2006. ACM. [Cited on page 1]

[132] Kevin A. Pelphrey, James P. Morris, and Gregory McCarthy. Neural basis of eye gaze processing deficits in autism. *Brain : a journal of neurology*, 128(5):1038–1048, 2005. [Cited on page 2]

[133] Peter Peltonen, Esko Kurvinen, Antti Salovaara, Giulio Jacucci, Tommi Ilmonen, John Evans, Antti Oulasvirta, and Petri Saarikko. It's mine, don't touch!: interactions at a large multi-touch display in a city centre. In *Proc. CHI 2008*, pages 1285–1294. ACM Press, 2008. [Cited on pages 116 and 132]

[134] H. Peng, F. Long, and C. Ding. Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, pages 1226–1238, 2005. [Cited on page 35]

[135] Jeffrey S. Perry and Wilson S. Geisler. Gaze-contingent real-time simulation of arbitrary visual fields. In *In Human Vision and Electronic Imaging, SPIE Proceedings*, pages 57–69, 2002. [Cited on page 152]

[136] Ken Pfeuffer, Jason Alexander, Ming Ki Chong, and Hans Gellersen. Gaze-touch: Combining gaze with multi-touch for interaction on the same surface. In *Proc.*

UIST 2014, UIST '14, pages 509–518, New York, NY, USA, 2014. ACM. [Cited on page 82]

[137] Ken Pfeuffer, Melodie Vidal, Jayson Turner, Andreas Bulling, and Hans Gellersen. Pursuit calibration: Making gaze calibration less tedious and more flexible. In *Proc. UIST 2013*, pages 261–270, New York, NY, USA, 2013. ACM. [Cited on page 116]

[138] David Pinelle, Miguel Nacenta, Carl Gutwin, and Tadeusz Stach. The effects of co-present embodiments on awareness and collaboration in tabletop groupware. In *Proc. GI 2008*, GI '08, pages 1–8, Toronto, Ont., Canada, Canada, 2008. Canadian Information Processing Society. [Cited on pages 172, 174, and 175]

[139] Alex Poole and Linden J. Ball. Eye tracking in human-computer interaction and usability research: Current status and future. In *Prospects?, Chapter in C. Ghaoui (Ed.): Encyclopedia of Human-Computer Interaction. Pennsylvania: Idea Group, Inc*, 2005. [Cited on page 70]

[140] Matti Pouke, Antti Karhu, Seamus Hickey, and Leena Arhippainen. Gaze tracking and non-touch gesture based interaction method for mobile 3d virtual spaces. In *Proc. OzCHI 2012*, pages 505–512, New York, NY, USA, 2012. ACM. [Cited on page 80]

[141] Pernilla Qvarfordt, David Beymer, and Shumin Zhai. Realtourist: A study of augmenting human-human and human-computer dialogue with eye-gaze overlay. In *Proc. INTERACT 2005*, pages 767–780, 2005. [Cited on pages 79, 145, and 149]

[142] Pernilla Qvarfordt and Shumin Zhai. Conversing with the user based on eye-gaze patterns. In *Proc. CHI 2005*, pages 221–230. ACM, 2005. [Cited on page 85]

[143] Carl Edward Rasmussen and Christopher K. I. Williams. *Gaussian Processes for Machine Learning*. The MIT Press, 2005. [Cited on page 53]

[144] D.A. Robinson. A method of measuring eye movemnent using a scieral search coil in a magnetic field. *Bio-medical Electronics, IEEE Transactions on*, 10(4):137–145, Oct 1963. [Cited on pages 15 and 16]

[145] Yvonne Rogers and Siân Lindley. Collaborating around vertical and horizontal large interactive displays: which way is best? *Interacting with Computers*, 16(6):1133–1152, 2004. [Cited on pages 144 and 147]

[146] Javier San Agustin, John Paulin Hansen, and Martin Tall. Gaze-based interaction with public displays using off-the-shelf components. In *Proc. Ubicomp Adjunct 2010*, pages 377–378. ACM Press, 2010. [Cited on pages 28, 71, and 91]

[147] Anthony Santella, Maneesh Agrawala, Doug DeCarlo, David Salesin, and Michael Cohen. Gaze-based interaction for semi-automatic photo cropping. In *Proc. CHI 2006*, pages 771–780, New York, NY, USA, 2006. ACM. [Cited on page 80]

[148] Constantin Schmidt, Jörg Müller, and Gilles Bailly. Screenfinity: extending the perception area of content on very large public displays. In *Proc. CHI 2013*, pages 1719–1728. ACM Press, 2013. [Cited on page 130]

[149] Alexander C. Schütz, Doris I. Braun, and Karl R. Gegenfurtner. Eye movements and perception: A selective review. *Journal of Vision*, 11(5):9, 2011. [Cited on page 13]

[150] Abigail J. Sellen. Remote conversations: The effects of mediating talk with technology. *Hum.-Comput. Interact.*, 10(4):401–444, December 1995. [Cited on pages 83, 145, 148, and 149]

[151] Laura Sesma, Arantxa Villanueva, and Rafael Cabeza. Evaluation of pupil center-eye corner vector for gaze estimation using a web cam. In *Proc. ETRA 2012*, pages 217–220. ACM Press, 2012. [Cited on page 22]

[152] Linda E. Sibert and Robert J. K. Jacob. Evaluation of eye gaze interaction. In *Proc. CHI 2000*, pages 281–288. ACM Press, 2000. [Cited on pages 5, 73, 75, and 150]

[153] Andreas Sippl, Clemens Holzmann, Doris Zachhuber, and Alois Ferscha. Real-time gaze tracking for public displays. In *Proc. AmI 2010*, pages 167–176. Springer-Verlag, 2010. [Cited on pages 28 and 90]

[154] Brian A. Smith, Qi Yin, Steven K. Feiner, and Shree K. Nayar. Gaze locking: Passive eye contact detection for human-object interaction. In *Proc. UIST 2013*, pages 271–280, New York, NY, USA, 2013. ACM. [Cited on pages 7, 27, 28, 45, 82, 83, and 90]

[155] John D. Smith, Roel Vertegaal, and Changuk Sohn. Viewpointer: lightweight calibration-free eye tracking for ubiquitous handsfree deixis. In *Proc. UIST 2005*, pages 53–61. ACM Press, 2005. [Cited on pages 2 and 82]

[156] India Starker and Richard A. Bolt. A gaze-responsive self-disclosing display. In *Proc. CHI 1990*, CHI '90, pages 3–10, New York, NY, USA, 1990. ACM. [Cited on pages 2, 71, 73, and 79]

[157] Randy Stein and Susan E. Brennan. Another person's eye gaze as a cue in solving programming problems. In *Proc. ICMI 2004*, pages 9–15. ACM, 2004. [Cited on pages 147, 148, and 149]

[158] Sophie Stellmach and Raimund Dachselt. Designing gaze-based user interfaces for steering in virtual environments. In *Proc. ETRA 2012*, ETRA '12, pages 131–138, New York, NY, USA, 2012. ACM. [Cited on page 80]

[159] Sophie Stellmach and Raimund Dachselt. Investigating gaze-supported multimodal pan and zoom. In *Proc. ETRA 2012*, ETRA '12, pages 357–360, New York, NY, USA, 2012. ACM. [Cited on page 78]

[160] Sophie Stellmach and Raimund Dachselt. Look & touch: gaze-supported target acquisition. In *Proc. CHI 2012*, pages 2981–2990. ACM Press, 2012. [Cited on page 81]

[161] Sophie Stellmach and Raimund Dachselt. Still looking: Investigating seamless gaze-supported selection, positioning, and manipulation of distant targets. In *Proc. CHI 2013*, CHI '13, pages 285–294, New York, NY, USA, 2013. ACM. [Cited on page 81]

[162] William Steptoe, Robin Wolff, Alessio Murgia, Estefania Guimaraes, John Rae, Paul Sharkey, David Roberts, and Anthony Steed. Eye-tracking for avatar eye-gaze and interactional analysis in immersive collaborative virtual environments. In *Proc. CSCW 2008*, pages 197–200. ACM, 2008. [Cited on page 83]

[163] R. Stiefelhagen, J. Yang, and A. Waibel. Tracking eyes and monitoring eye gaze. In *Proceedings of the Workshop on Perceptual User Interfaces*, pages 98–100, 1997. [Cited on pages 23 and 31]

[164] Y. Sugano, Y. Matsushita, and Y. Sato. Calibration-free gaze sensing using saliency maps. In *Proc. CVPR 2010*, pages 2667–2674. IEEE Computer Society, 2010. [Cited on pages 13, 23, 25, 27, 45, and 52]

[165] Y. Sugano, Y. Matsushita, and Y. Sato. Learning-by-synthesis for appearance-based 3d gaze estimation. In *Proc. CVPR 2014*, pages 1821–1828, June 2014. [Cited on pages 4, 7, 13, 23, 24, and 26]

[166] Y. Sugano, Y. Matsushita, Y. Sato, and H. Koike. An incremental learning method for unconstrained gaze estimation. *European Conference on Computer Vision (ECCV)*, pages 656–667, 2008. [Cited on pages 18 and 31]

[167] K.H. Tan, D.J. Kriegman, and N. Ahuja. Appearance-based eye gaze estimation. In *Proceedings of WACV*, pages 191–195. IEEE, 2002. [Cited on pages 4 and 23]

[168] A. Tsukada, M. Shino, M. Devyver, and T. Kanade. Illumination-free gaze estimation method for first-person vision wearable device. In *Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on*, pages 2084–2091, Nov 2011. [Cited on page 27]

[169] Akihiro Tsukada and Takeo Kanade. Automatic acquisition of a 3d eye model for a wearable first-person vision device. In *Proc. ETRA 2012*, ETRA '12, pages 213–216, New York, NY, USA, 2012. ACM. [Cited on pages 18 and 22]

[170] Jayson Turner, Jason Alexander, Andreas Bulling, Dominik Schmidt, and Hans Gellersen. Eye pull, eye push: Moving objects between large screens and personal devices with gaze & touch. In *Proc. INTERACT 2013*, pages 170–186, 2013. [Cited on page 81]

[171] Jayson Turner, Andreas Bulling, Jason Alexander, and Hans Gellersen. Cross-device gaze-supported point-to-point content transfer. In *Proc. ETRA 2014*, pages 19–26. ACM Press, 2014. [Cited on page 81]

[172] Roberto Valenti and Theo Gevers. Accurate eye center location and tracking using isophote curvature. In *Proc. CVPR 2008*, pages 1–8. IEEE Computer Society, 2008. [Cited on page 49]

[173] Roel Vertegaal. The gaze groupware system: Mediating joint attention in multiparty communication and collaboration. In *Proc. CHI 1999*, pages 294–301. ACM, 1999. [Cited on pages 83, 84, 145, 148, and 149]

[174] Roel Vertegaal, Aadil Mamuji, Changuk Sohn, and Daniel Cheng. Media eyepliances: using eye tracking for remote control focus selection of appliances. In *Proc. EA CHI 2005*, pages 1861–1864. ACM Press, 2005. [Cited on pages 27 and 82]

[175] Roel Vertegaal, Jeffrey S. Shell, Daniel Chen, and Aadil Mamuji. Designing for augmented attention: Towards a framework for attentive user interfaces. *Computers in Human Behavior*, 22(4):771–789, 2006. [Cited on pages 2, 73, 82, and 90]

[176] Mélodie Vidal, Andreas Bulling, and Hans Gellersen. Pursuits: spontaneous interaction with displays based on smooth pursuit eye movement and moving targets. In *Proc. UbiComp 2013*, pages 439–448. ACM Press, 2013. [Cited on pages 81, 90, and 116]

[177] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. In *Proc. CVPR 2001*, pages 511–518. IEEE Computer Society, 2001. [Cited on pages 33 and 47]

[178] Sarah A. Vitak, John E. Ingram, Andrew T. Duchowski, Steven Ellis, and Anand K. Gramopadhye. Gaze-augmented think-aloud as an aid to learning. In *Proc. CHI 2012*, CHI '12, pages 2991–3000, New York, NY, USA, 2012. ACM. [Cited on page 70]

[179] Daniel Vogel and Ravin Balakrishnan. Interactive public ambient displays: Transitioning from implicit to explicit, public to personal, interaction with multiple users. In *Proc. UIST 2004*, pages 137–146. ACM, 2004. [Cited on page 116]

[180] Oleg Špakov, Poika Isokoski, and Päivi Majaranta. Look and lean: Accurate head-assisted eye pointing. In *Proc. ETRA 2014*, ETRA '14, pages 35–42, New York, NY, USA, 2014. ACM. [Cited on page 77]

[181] Robert Walter, Gilles Bailly, and Jörg Müller. Strikeapose: revealing mid-air gestures on public displays. In *Proc. CHI 2013*, pages 841–850. ACM Press, 2013. [Cited on page 116]

[182] D. Walther and C. Koch. Modeling attention to salient proto-objects. *Neural Networks*, 19(9):1395–1407, 2006. [Cited on page 33]

[183] H. Wang and E. Blevis. Concepts that support collocated collaborative work inspired by the specific context of industrial designers. In *Proc. CSCW 2004*, CSCW '04, pages 546–549, New York, NY, USA, 2004. ACM. [Cited on page 147]

[184] D. J. Ward and D. J. C. MacKay. Fast hands-free writing by gaze direction. *Nature*, 418(6900):838, 2002. [Cited on pages 71 and 72]

[185] Steve Whittaker and Brid O'Conaill. *The Role of Vision in Face-to-Face and Mediated Communication*, chapter The role of vision in face-to-face and mediated communication, pages 23–49. Lawrence Erlbaum Associates, video-mediated communication edition, 1997. [Cited on page 147]

[186] O. Williams, A. Blake, and R. Cipolla. Sparse and semi-supervised visual mapping with the sˆ 3gp. In *Proc. CVPR 2006*, pages 230–237. IEEE, 2006. [Cited on pages 23, 24, 25, 28, 31, 32, 52, and 61]

[187] DavidS. Wooding, MarkD. Mugglestone, KevinJ. Purdy, and AlastairG. Gale. Eye movements of large populations: I. implementation and performance of an autonomous public eye tracker. *Behavior Research Methods, Instruments, & Computers*, 34(4):509–517, 2002. [Cited on page 116]

[188] L.Q. Xu, D. Machin, and P. Sheppard. A novel approach to real-time non-intrusive gaze finding. In *British Machine Vision Conference*, pages 428–437, 1998. [Cited on pages 23, 31, and 35]

[189] Michiya Yamamoto, Munehiro Komeda, Takashi Nagamatsu, and Tomio Watanabe. Development of eye-tracking tabletop interface for media art works. In *Proc. ITS 2010*, ITS '10, pages 295–296, New York, NY, USA, 2010. ACM. [Cited on page 82]

[190] Hirotake Yamazoe, Akira Utsumi, Tomoko Yonezawa, and Shinji Abe. Remote gaze estimation with a single camera based on facial-feature tracking without special calibration actions. In *Proc. ETRA 2008*, ETRA '08, pages 245–250. ACM, 2008. [Cited on pages 4, 18, 22, and 83]

[191] A. L. Yarbus. *Eye Movements and Vision.* Plenum. New York., 1967. [Cited on pages 2, 13, 15, 16, 71, 74, and 148]

[192] ByungIn Yoo, Jae-Joon Han, Changkyu Choi, Kwonju Yi, Sungjoo Suh, Dusik Park, and Changyeong Kim. 3d user interface combining gaze and hand gestures for large-scale display. In *Proc. CHI EA 2010*, CHI EA '10, pages 3709–3714, New York, NY, USA, 2010. ACM. [Cited on pages 80 and 82]

[193] Dong Hyun Yoo, Bang Rae Lee, and Myoung Jin Chung. Non-contact eye gaze tracking system by mapping of corneal reflections. In *Proc. FGR 2002*, page 101. IEEE Computer Society, 2002. [Cited on page 21]

[194] LaurenceR. Young and David Sheena. Survey of eye movement recording methods. *Behavior Research Methods & Instrumentation*, 7(5):397–429, 1975. [Cited on pages 13 and 20]

[195] Shumin Zhai. What's in the eyes for attentive input. *Commun. ACM*, 46(3):34–39, March 2003. [Cited on page 73]

[196] Shumin Zhai, Carlos Morimoto, and Steven Ihde. Manual and gaze input cascaded (magic) pointing. In *Proc. CHI 1999*, pages 246–253. ACM Press, 1999. [Cited on pages 2, 5, 71, 72, 74, 77, 86, and 115]

[197] Xucong Zhang, Yusuke Sugano, Mario Fritz, and Andreas Bulling. Appearance-based gaze estimation in the wild. In *Proc. CVPR 2015*, June 2015. [Cited on pages 7, 13, 23, 24, and 26]

[198] Yanxia Zhang, Andreas Bulling, and Hans Gellersen. *Context-Aware Systems: Methods and Applications*, chapter Calibration-free Remote Eye Tracking based on Eye Symmetry. Springer. [Cited on page 11]

[199] Yanxia Zhang, Andreas Bulling, and Hans Gellersen. Discrimination of gaze directions using low-level eye image features. In *Proc. PETMEI 2011*, pages 9–14. ACM, 2011. [Cited on page 11]

[200] Yanxia Zhang, Andreas Bulling, and Hans Gellersen. Towards pervasive eye tracking using low-level image features. In *Proc. ETRA 2012*, pages 261–264. ACM Press, 2012. [Cited on pages 11, 23, 25, 52, and 54]

[201] Yanxia Zhang, Andreas Bulling, and Hans Gellersen. Sideways: A gaze interface for spontaneous interaction with situated displays. In *Proc. CHI 2013*, pages 851–860. ACM Press, 2013. [Cited on pages 11, 51, 116, and 117]

[202] Yanxia Zhang, Andreas Bulling, and Hans Gellersen. Pupil-canthi-ratio: A calibration-free method for tracking horizontal gaze direction. In *Proc. AVI 2014*, pages 129–132. ACM Press, 2014. [Cited on page 11]

[203] Yanxia Zhang, Ming Ki Chong, Jörg Müller, Andreas Bulling, and Hans Gellersen. Eye tracking for public displays in the wild. *Personal and Ubiquitous Computing*, 19(5-6):967–981, 2015. [Cited on page 11]

[204] Yanxia Zhang, Hans Jörg Müller, Ming Ki Chong, Andreas Bulling, and Hans Gellersen. Gazehorizon: Enabling passers-by to interact with public displays by gaze. In *Proc. UbiComp 2014*, pages 559–563, 2014. [Cited on page 11]

[205] Dingyun Zhu, Tom Gedeon, and Ken Taylor. Exploring camera viewpoint control models for a multi-tasking setting in teleoperation. In *Proc. CHI 2011*, CHI '11, pages 53–62, New York, NY, USA, 2011. ACM. [Cited on page 78]

[206] Jie Zhu and Jie Yang. Subpixel eye gaze tracking. In *Proceedings of the Fifth IEEE International Conference on Automatic Face and Gesture Recognition*, pages 124–129, May 2002. [Cited on pages 18, 22, 46, and 63]

[207] Z. Zhu and Q. Ji. Eye gaze tracking under natural head movements. In *Proc. CVPR 2005*, pages 918–923. IEEE Computer Society, 2005. [Cited on pages 21 and 28]

[208] Zhiwei Zhu and Qiang Ji. Eye and gaze tracking for interactive graphic display. *Mach. Vision Appl.*, 15(3):139–148, 2004. [Cited on pages 25 and 52]

[209] Zhiwei Zhu and Qiang Ji. Novel eye gaze tracking techniques under natural head movement. *IEEE Transactions on Biomedical Engineering*, 54(12):2246–2260, 2007. [Cited on pages 18, 21, and 26]