A Design Space for Gaze Interaction on Head-Mounted Displays

Teresa Hirzle
Institute of Media Informatics,
Ulm University, Germany

Jan Gugenheimer
Institute of Media Informatics,
Ulm University, Germany

Florian Geiselhart
Institute of Media Informatics,
Ulm University, Germany

Andreas Bulling
Institute for Visualisation and
Interactive Systems,
University of Stuttgart, Germany

Enrico Rukzio
Institute of Media Informatics,
Ulm University, Germany

ABSTRACT

Augmented and virtual reality (AR/VR) has entered the mass market and, with it, will soon eye tracking as a core technology for next generation head-mounted displays (HMDs). In contrast to existing gaze interfaces, the 3D nature of AR and VR requires estimating a user's gaze in 3D. While first applications, such as foveated rendering, hint at the compelling potential of combining HMDs and gaze, a systematic analysis is missing. To fill this gap, we present the first design space for gaze interaction on HMDs. Our design space covers human depth perception and technical requirements in two dimensions aiming to identify challenges and opportunities for interaction design. As such, our design space provides a comprehensive overview and serves as an important guideline for researchers and practitioners working on gaze interaction on HMDs. We further demonstrate how our design space is used in practice by presenting two interactive applications: EyeHealth and XRay-Vision.

CCS CONCEPTS

• **Human-centered computing → Interface design prototyping**; *Mixed / augmented reality*; *Virtual reality*;

KEYWORDS

design space, 3D gaze, gaze interaction, head-mounted displays, interaction design, augmented reality, virtual reality

ACM Reference Format:
Teresa Hirzle, Jan Gugenheimer, Florian Geiselhart, Andreas Bulling, and Enrico Rukzio. 2019. A Design Space for Gaze Interaction on Head-Mounted Displays. In *CHI Conference on Human Factors in Computing Systems Proceedings (CHI 2019), May 4–9, 2019, Glasgow, Scotland UK*. ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3290605.3300855

1 INTRODUCTION

Augmented and virtual reality head-mounted displays are currently experiencing rising consumer adoption. Eye tracking is expected to be an integral part of these devices given that it both provides possibilities to expand and enhance current interaction techniques [27, 48] and has the potential to address unique challenges of HMD interaction (e.g. foveated rendering [45]).

Prior work in gaze interaction has typically focused on 2D gaze, most likely because current interactive systems mainly support two-dimensional display-based interaction, e.g. on smartphones [22, 43], smartwatches [15] or ambient displays [55, 61, 62]. That is, gaze is usually considered to be a 2D point on a screen. In contrast, AR and VR are inherently three-dimensional: they either extend the physical world with digital information by overlaying it with virtual content (AR) or create a new, also three dimensional, virtual world (VR). They therefore require 3D gaze information rather than a 2D gaze point on a device screen. While this may seem like a small shift in perspective, the implications, as shown in this paper, are compelling and far-reaching.

To fully understand the interaction space arising from the combination of HMDs and 3D gaze, a systematic analysis of properties of both technologies is necessary, but currently missing. To fill this gap, we propose the first design space for gaze interaction on HMDs. The aim is to identify key challenges and characterize the potential for future interaction design for the combination of these technologies.

Our design space is presented as a two-dimensional matrix (also known as Zwicky box [64]), which is spanned by

---

CHI 2019, May 4–9, 2019, Glasgow, Scotland UK

ACM ISBN 978-1-4503-5970-2/19/05…$15.00
https://doi.org/10.1145/3290605.3300855

the two dimensions: **D1** technical properties of HMDs and **D2** properties of human depth perception. In a first step we describe the resulting matrix based on a pure technological comparison. In a second step we show three different perspectives on our design space (technological, application-based and interaction-based). Additionally, we discuss usage implications of the design space for interaction design from a user-centered perspective, and identify basic interaction possibilities that serve as basic modules for the development of gaze-based applications for HMDs.

With the design space we offer an approach and perspective for designers, researchers and practitioners to explore potential interaction techniques and applications arising from the combination of HMD and 3D Gaze. We strive to inspire readers of our work to build upon the presented design space aiming to create new possibilities for interaction design.

The contributions of our work are two-fold: First, we develop a design space for gaze interaction on head-mounted displays, offering a new approach and perspective to derive possible interaction techniques and applications for the combination of HMDs and 3D gaze. Second, we show how to apply our design space by demonstrating how to design and implement applications and presenting a prototypical implementation of two interactive applications: EyeHealth and XRay-Vision.

2 BACKGROUND

Our work mainly builds upon four general fields of research: *Design Spaces in HCI*, *Classifications of HMDs*, *Human Depth Perception* and *Measuring Gaze Depth*.

Design Spaces in HCI

In the field of Human-Computer Interaction (HCI) taxonomies [16] and design spaces [9] have been used to understand and explore the potential of existing (e.g. smartphone [2]) and upcoming technologies (e.g. shape changing interfaces [26]). Foley et al. [16] showed that taxonomies are a useful way to systematize knowledge about input devices and interaction techniques for graphical user interfaces. Covering most of Foley et al.'s presented input devices, Card et al. [8] proposed a taxonomy to structure the huge amount of arising input devices for desktop computing in the late 80s. They showed that interaction can be modeled as an artificial language among three parts: *human*, *device* (user dialogue machine) and *application*. In a second work Card et al. [9] demonstrated that morphological analysis, as presented by Zwicky [64], can be applied for the creation of design spaces and revelation of new interaction designs in the field of HCI. Morphological analysis aims to create a multidimensional matrix that contains all possible combinations of parameters that are relevant to a specific problem. In the case of user interfaces these parameters are often certain properties of devices (e.g. physical properties such as *rotate* or *position* of input devices [2, 9]). As pointed out by Ballagas et al. [3] the resulting matrix indicates promising families of solutions by heavily populated cells, as well as a possible lack of solutions by unpopulated areas of the matrix.

As such, building a design space to gain understanding of the fusion of HMDs and 3D gaze is a first step to uncover its potential. Inspired by Card et al. [8], building our design space we also focused on the three parts *human* (**D2** properties of depth perception), *device* (**D1** technical properties of HMDs) and *application* (implications for interaction design).

Classifications of Head-mounted displays

Ivan Sutherland's "Sword of Damocles" is often considered to be the first HMD created [53]. Sutherland wanted to emphasize the 3D nature of this new technology and called it initially a "head-mounted three dimensional display". In the following years researchers started to gain deeper understanding of this type of display and started to define properties [1, 5, 63] and propose taxonomies [37] and classifications [38, 60]. Milgram et al. proposed one of the fundamental classifications by categorizing HMDs based on a set of criteria along a reality-virtuality-continuum [37, 38]. One of the big insights of Milgram et al. was to divide the perspective on HMDs into a real and a virtual component. As such HMD devices can be positioned on the continuum according to the relation between their real and virtual component. This fundamental perspective allowed later to define properties of AR devices [1, 63] and VR devices [5, 60] respectively. AR was characterized by Azuma [1] to have the following characteristics: *combines real and virtual*, *interactive in real time* and *registered in 3D*. Zhou et al. [63] presented an overview of 20 years AR research showing a particular focus on: *interaction*, *tracking* and *display technology*. VR can here be seen as an extreme case on the reality-virtuality continuum consisting of solely virtual content. Therefore, Billinghurst [4] named a set of key characteristics for VR as: *3D stereoscopic display*, *wide field of view* and *low latency head tracking*. Additionally to the reality-virtuality continuum Milgram et al. [38] offer three more classification criteria of HMDs: *extent of world knowledge*, *reproduction fidelity* and *extent of presence metaphor*.

Based on these classifications and taxonomies we defined parameters for our first dimension **D1** to classify HMD technology in our design space. Based on Milgram et al.'s [38] continuum and Azuma's [1] characteristics we defined the parameter **device type** with the values *AR* and *VR*. Based on Zhou et al.'s [63] overview of display technology we defined the parameter **display type** with the values *monoscopic* and

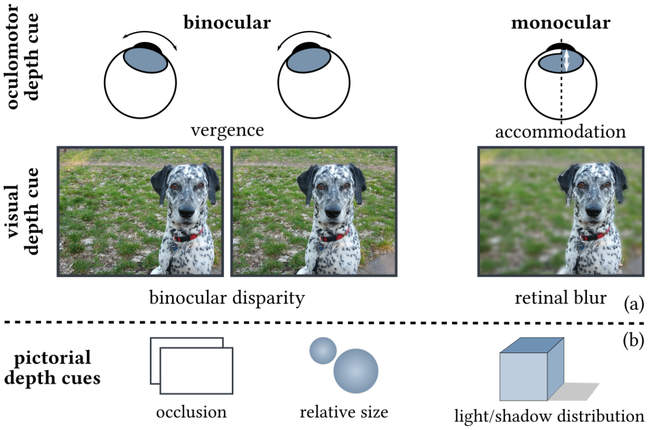


Figure 1: (a) overview of oculomotor and visual depth cues. Vergence is mainly driven by binocular disparity and accommodation is mainly driven by retinal blur (adapted from [24]). (b) three examples for pictorial depth cues.

stereoscopic. Finally, based on Milgram et al.'s extent of world knowledge, Azuma's characteristics and Zhou et al.'s tracking we defined the parameter **world knowledge** with the values *full* and *none*.

Human Depth Perception

The human visual system perceives depth through several *visual depth cues* that send independent signals to the brain to create a three dimensional image [29]. These visual cues can be classified into *monocular* and *binocular* cues. *Monocular visual depth* cues are either static or dynamic. Static refers to pictorial cues, such as the relative size of objects, occlusion or light and shadow distribution [50], or retinal blur, which refers to the sensed magnitude of focal blur on the retina [33], and can imply depth even in flat images (see Fig. 1 (b)). There is only one binocular depth cue, which is *binocular disparity* and refers to the two slightly different images perceived by the eyes due to their horizontal separation [24].

The perception of these visual depth cues evokes an *oculomotor response* of the eyes to ensure that a visual scene is perceived sharply, which is referred to as *oculomotor depth cues*. These are vergence, accommodation and pupillary constriction [50]. Here, *vergence* is the simultaneous inward rotation of the eyes to locate the projected images per eye on the fovea, which is the part on the retina with the highest spatial acuity, allowing the brain to fuse them into a single percept [13]. *Accommodation* refers to the alteration of the lens to maintain the area of interest on the fovea and the *pupil constricts* with near vergence/accommodation to compensate for a narrow depth of field [29] (see Fig. 1 (a)).

Whereas vergence is mostly driven by binocular disparity (binocular depth cue), accommodation is evoked by retinal blur (monocular depth cue) [24]. Oculomotor cues are cross correlated and mostly in agreement with each other under normal viewing conditions. Stereoscopic displays often only allow for a few depth cues to be realized [11], which leads to conflicting depth information (e.g. vergence-accommodation conflict [20, 51]). In the following we discuss how these oculomotor depth cues can be applied for measuring gaze depth.

Measuring Gaze Depth

Since we are interested in the 3D position of a user's gaze, we focus on oculomotor depth cues using either *vergence* or *accommodation*. Whereas previous work on 3D gaze estimation mainly focuses on vergence-estimates, we also shortly discuss the measurement of accommodation.

Vergence. Most related work that measures gaze depth applies vergence-estimates, mostly referred to as 3D gaze tracking. On average the closest point one can converge to is at about 20cm distance from the eyes and it is generally agreed on that vergence can reliably (with a difference in visual angle of greater than 1°) be measured for a maximum distance of about 1.5m from the eyes (e.g. [39]). In the following we will briefly discuss how gaze depth can be measured vergence-based.

Pfeiffer [46] classified 3D gaze tracking algorithms into *geometry-based* and *holistic* estimation. The former relies on a geometric model of the target objects, which is intersected with the visual axis of at least one eye resulting in a 3D point of regard. Holistic estimation is independent of a geometric model, here a 3D point is calculated based on the information of the observer only. The main difference is that geometry-based estimation is object centered, i.e. the estimated 3D gaze point is positioned on a target object, whereas holistic estimation is world-centered positioning a 3D gaze point absolutely in the 3D world. For Pfeiffer's geometry-based systems it is sufficient to calculate a 2D gaze point on a screen and intersect it with a model of the environment. For this, monocular [18, 35, 41] or binocular [44, 47] gaze estimation can be applied. There are also several approaches for the holistic estimation of 3D gaze. Whereas monocular holistic approaches require a 3D calibration procedure [30, 34, 52], binocular holistic approaches do not necessarily rely on it [12, 39, 58]. General challenges for 3D gaze estimation are the estimation of a gaze point on target objects that are smaller than the accuracy of the eye tracking system and can thus not be measured reliably. A problem of monocular tracking is that only what one eye sees is considered, if the sight of the not-measured eye is occluded by an object this cannot be represented in the 3D gaze tracking model. A further general problem of 3D gaze estimation is that due to the physiology of the eyes, attention should rather be seen as a 3D volume in space than a single point in space [46].

There is a large number of works that applies eye tracking for explicit or implicit interaction on HMDs (e.g. solely [13, 23] or as additional modality [27, 48]). Whereas these works mainly rely on 2D gaze tracking, for 3D gaze and HMDs only few works have been presented. Elmadjian et al. [14] point out that 3D interaction on AR/VR HMDs can be refined using 3D gaze, e.g. by resolving where a user is looking at when objects partially occlude. Weier et al. [59] presented an estimator for gaze depth using an eye tracker inside a VR HMD, applying multiple features, including vergence measures. Lee et al. [31] point out that for remote collaboration AR applications it is important to know whether the user is looking at the virtual image or real world objects.

Accommodation. Accommodation is usually given in diopters (D = 1/meters) and describes the power of the lens, i.e. the range for which the human eye can put objects into focus. On average the closest point one can accommodate to is at about 25cm distance from the eyes (the farthest point is infinity [54]). The human eye is further considered to have a depth-of-field (DoF) of about 0.3D [7], which influences both vergence [12] and accommodation measurement.

Accommodation is measured with an autorefractor. These are standard devices in ophthalmology, but have not been used often in HCI research. They are used in the field of gaze-contingent rendering: Padmanaban et al. [42] used an autorefractor to measure how display prototypes influence accommodation and Koulieris et al. [25] evaluated whether gaze-contingent DoF and several multi-focal display approaches drive the actual accommodation of the eye using measures from an autorefractor attached to an HMD.

We integrated the different possibilities to measure gaze depth in the design space for the definition of our second dimension **D2**. For this we selected the following two parameters to classify human depth perception and its application for obtaining a 3D gaze presentation: **oculomotor depth cue** with the values *vergence* and *accommodation* and **ocularity** with the values *monocular* and *binocular*.

3 DESIGN SPACE

The unique properties of human depth perception and the specific technical requirements of current head-mounted displays call for a structured analysis to identify key challenges and characterize the potential for future interaction design. With our design space we give an approach for such a structured analysis, which is currently missing. The design space aims to make researchers and designers aware of and help them to address challenges of future gaze interfaces on HMDs.

			vergence		accommodation	
			monocular	binocular	monocular	binocular
VR	stereoscopic	full				▨
		none	▨	▨	▨	▨
	monoscopic	full	▨	▨	▨	▨
		none	▨	▨	▨	▨
AR	stereoscopic	full				▨
		none				▨
	monoscopic	full				▨
		none				▨

Figure 2: The matrix as obtained by the morphological approach containing all combinations of parameters. The combinations that make technically no sense, or are irrelevant for the design space are highlighted in gray.

The two dimensions that span the design space are (**D1**) technical properties of HMDs and (**D2**) properties of human depth perception. **D1** is hereby defined by three parameters that were selected according to generally accepted technical classifications of HMD technology (see section 2 classifications of HMDs): **device type**, **display type** and **world knowledge**. **D2** is defined by two parameters that were selected based on an analysis of human depth cues (see section 2 human depth perception) and their application for measuring gaze depth (see section 2 measuring gaze depth): **oculomotor depth cue** and **ocularity**. Each of the parameters has two binary values. These are for **D1**: device type (**AR**/**VR**), display type (**monoscopic**/**stereoscopic**) and world knowledge (**none**/**full**) and for **D2**: oculomotor depth cue (**vergence**/**accommodation**) and ocularity (**monocular**/**binocular**). To reduce complexity we restricted the parameters to have binary values. This does not map the whole spectrum of possibilities, however the set of values can be expanded in future.

Following the approach of morphological analysis we combined all of these values in a multidimensional matrix, also known as Zwicky box [64], which is a well-established tool for ideation and the creation of design spaces (e.g. [3]). According to the approach of morphological analysis the such created matrix contains all combinations of parameters that are relevant for a given problem and helps to identify promising families of solutions, as well as a possible lack of solutions by the quantity of solutions in specific cells. The resulting matrix with all combinations is shown in Figure 2, where **D1** is positioned on the y-axis and **D2** on the x-axis.

In the following we describe the parameters and their set of values for each dimension in more detail. We will then discuss the content of the resulting cells in the matrix from different perspectives and give approaches on how to use and apply the design space.

Dimensions, Parameters and Values

The parameter with their individual set of values were obtained from a literature review and are described in the following, starting with the values for **D1** and then those for **D2** accordingly.

Device Type (D1). is divided into the values **AR** and **VR** HMDs.

Display Type (D1). is divided into **monoscopic** and **stereoscopic** displays. On monoscopic displays only monocular visual depth cues (e.g. motion parallax or pictorial cues, such as occlusion or light and shadow distribution) can be displayed. On stereoscopic displays we have additionally binocular disparity, i.e. both eyes perceive a slightly different image, as such "real depth" can be induced.

World Knowledge (D1). is divided into **full** and **none**. The parameter is inspired by Milgram et al. [38], who described the extent of world knowledge as a continuum reaching from the "unmodeled world" extreme, where nothing is known about the world, to the "world completely modeled" extreme, which is described as the computer having complete knowledge about each object in the world (including semantics). We do not depict the whole continuum, but limit it to the two values *none* and *full*. Having *no* world knowledge implies not having any information about the surrounding world. We define *full* knowledge as having a 3D representation (e.g. mesh) of the surrounding environment (corresponding to the current state of the art of devices such as HoloLens [36] and MagicLeap [32]). This does not include semantic knowledge, but is limited to the physical representation, since we wanted to focus on current devices for now. For AR the surrounding environment refers to the physical world, which is enriched with virtual content. For VR the surrounding environment is by definition completely virtual. We also include positional and rotation tracking into our definition of world knowledge, i.e. having positional and/or rotational tracking of the user belongs to *full* world knowledge, while with *no* world knowledge we have no additional tracking.

Oculomotor Depth Cue (D2). is divided into **vergence** and **accommodation**. Both cues can be applied to obtain an estimate for a 3D gaze presentation. Whereas the measurement of vergence (i.e. simultaneous inward rotation of the eyes) results in a **3D gaze point**, the measurement of accommodation (i.e. bending of the lens) results in a **3D gaze depth level**, which can be imagined as a plane instead of a point in space. As pointed out in section 2, vergence can be measured for about 0.2m to 1.5m distance from the user. The measurement of accommodation is influenced by the depth-of-field of the human eye, i.e. the range for which objects appear sharply when accommodating to a certain depth. Since this increases with increasing distance (for a far away placed object the distance range for which objects appear in focus around the fixated object increases exponentially), the accuracy for accommodation measurements decreases accordingly.

Ocularity (D2). is divided into **monocular** and **binocular**. This refers to whether the measurement of one eye suffices to obtain a 3D gaze representation or if the estimation of both eyes is required. Since vergence is a binocular depth cue, in general both eyes have to be measured. However, there are systems that obtain vergence values based on one eye only. These approaches require a depth calibration procedure, i.e. mapping gaze values to several depths, whereas binocular vergence estimation does not necessarily rely on a depth calibration. Accommodation is a monocular depth cue and as such it is sufficient to measure one eye.

We excluded the combinations of values that do not make sense technically (*monoscopic VR*) or that are irrelevant for the resulting design space, e.g. *binocular accommodation* estimation. Accommodation is a monocular depth cue and as such requires only one eye to be measured [54]. We also excluded the column for *no world knowledge* in VR, since according to the definition of world knowledge for VR we have always *full world knowledge*. The resulting matrix is indicated by the white cells in Figure 2, **D1** represented on the y-axis and **D2** on the x-axis.

Views on the Design Space

The resulting design space contains cells that are each composed by an HMD part (**D1**) and a representation of 3D gaze (**D2**). In the following we show the broad applicability of our design space by giving three exemplary views on how the cells of the design space can be filled. These are *technology*-, *application*- and *interaction-based*. A cell is hereby given as (n,m), where n is derived by **D1** (y-axis) and m by **D2** (x-axis).

Technology-based View. This view refers to filling the design space with technical devices in combination with eye tracking devices and gaze depth algorithms.

The Microsoft HoloLens with an attached Pupil Labs eye tracking add-on (capable of obtaining a 3D gaze point) can for example be placed in cell (2,2) (see Fig. 3). At this **D1** defines to which row the Hololens belongs: it is an *AR* device with a *stereoscopic* display, having *full* world knowledge. **D2** defines to which row the Pupil Labs/eye tracking algorithm belongs: the headset is able to obtain a 3D gaze point based on *binocular vergence* estimates.

Another example is the HTC Vive [57], also with an attached Pupil Labs add-on. Here **D1** defines row 1 for the HTC Vive, since it is a *stereoscopic VR* device by definition

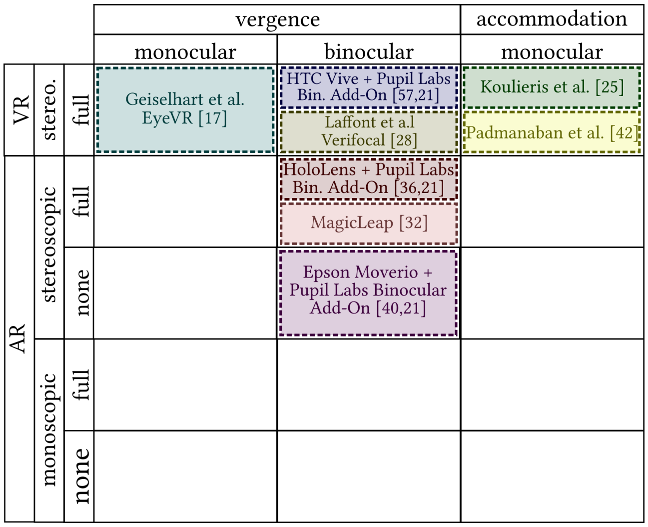


Figure 3: Technology-based view on the design space: it can be filled with device types and 3D gaze tracking approaches/eye tracking devices. Here each device type can be positioned in one row of the design space and each eye tracking device/3D gaze estimation technique can be positioned in one column.

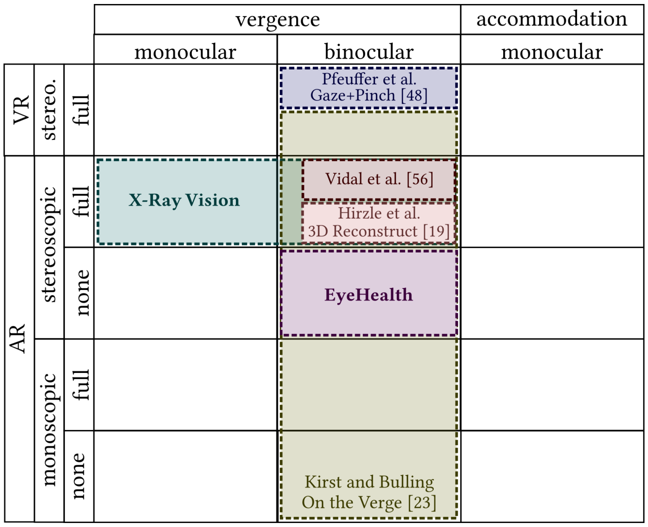


Figure 4: Application-based view on the design space: it can be filled with concrete applications. When an application is positioned in only one cells it means that it can exclusively implemented with the according devices from the technology-based view. Applications that are positioned in more than one cell indicate alternative implementations/device types.

having *full* world knowledge. **D2** defines again column 2 of the 3D gaze part, resulting in cell (1,2). In this way all cells can be filled accordingly. There are some combinations that are technically not possible yet (e.g. measuring accommodation with an AR device). Because of this our design space is intentionally built to be expandable in future. We show some more examples of technical allocations in Figure 3.

Application-based View. This view refers to filling the design space with applications that combine a 3D gaze-based interaction approach with HMDs.

One example for this view was presented by Hirzle et al. [19]. In their work they implemented an application that creates a 3D scan of a gazed-at object. For **D1** they used a *stereoscopic AR* device with *full* world knowledge (row 2). For **D2** the application relies on *binocular* vergence estimate to calculate a 3D gaze point in space, as such the application results in cell (2,2) of the design space

Another example is presented by Kirst and Bulling [23]. They used voluntary vergence eye movements to perform a timely precise and accurate input gesture. Although their study was conducted with a display, the same eye movement could be applied for a selection task on AR or VR displays. As such only **D2** defines the classification of this application in our design space. This results in a a whole column rather than a single cell (see Fig. 4). Content that can be put into more than one cell or even into a whole row/column are defined as being independent of the content of the according dimensions, as presented in this example. More precisely for this current example it means the technical implementation is independent of the type of HMD used. In other words: the application can be implemented with all kinds of HMDs occurring in our design space, but only depends on the 3D gaze tracking algorithm. Cells can also be filled with various applications, as shown exemplary in cell (2,2) or (3,2).

Interaction-based View. This view refers to filling the design space with general interaction possibilities. These possibilities can then be used for/implemented in concrete applications, as presented in the section before.

An example for this category is to use a user's 3D gaze point in space for the correct positioning of augmented content in the real world. This interaction possibility requires for **D1** *full* world knowledge and a *stereoscopic* display to be able to correctly display depth information. For **D2** it requires the estimation of a 3D gaze point. As such it can be positioned in cells (2,1) and (2,2) as shown in Figure 5 (red).

Another interaction possibility is to use gaze information to differ whether the user is currently looking at the real world or at augmented content. This possibility can be implemented with various devices/tracking algorithms. For example one could use accommodation-estimates to recognize whether the user is accommodating to the virtual plane, which would correspond to all cells in the rightmost column (see Fig. 5 (yellow)). The same possibility could be implemented using an AR device with *full* world knowledge and thus can be positioned in all cells of the second row. This

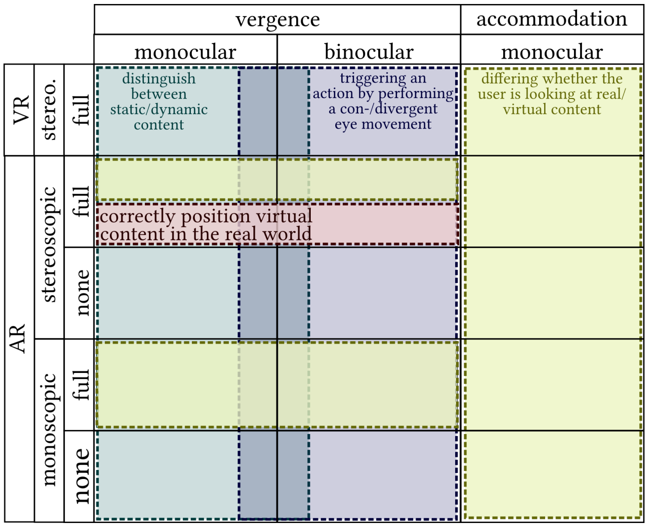


Figure 5: Interaction-based view on the design space: it can be filled with basic interaction possibilities, which can then be applied/implemented in concrete applications.

specific example shows that a possibility is not exclusively limited to one cell or one concrete implementation derived by **D1** or **D2**. It can rather be implemented by a set of values from both dimensions.

With our design space we aim to give a starting point to analyze the interaction space of HMD and 3D gaze technology. The design space was built to give some examples on how to tackle problems in this area and to start filling it with exemplary content. This shows that the contribution of our design space is not the actual listing of content or its allocations, but rather the conceptual framework, we offer to think about the combination of HMDs and 3D gaze. We further want to highlight that the views we presented on the design space are not unique, i.e. the design space was designed openly and allows for more definitions of views and content. It therefore provides the opportunity to derive a much broader set of device types, applications and interaction possibilities than explicitly presented here. One of the results of applying views on the design space is the identification of clusters of promising solutions, as visible in cell (2,2), because this one is already heavily populated (which holds for all three here presented views). A lack of solutions can also already be seen for example in cell (5,3), which is only sparsely populated. This is open for interpretations of causality. Meaning, it could on the one hand mean that this specific combination of technologies is meaningless in the context of this design space and is not applicable for interaction design. On the other hand this lack could mean that solutions have just not been found or developed yet.

4 USAGE OF DESIGN SPACE

We propose two ways how our design space can be used in practice to derive new *technological combinations*, *applications* and *interaction possibilities*. The **classification-based approach** aims to position applications inside the space to identify technological requirements. The usage of the design space as **ideation tool** aims to derive new interaction possibilities, applications and even new device types based on fundamental components of the design space.

Classification-based Technique

The classification-based technique of using the design space aims to identify technological requirements for given content and can be seen as a top-down approach. It is mainly directed towards designers and practitioners that have specific application scenarios or interaction possibilities for gaze-interaction on HMDs in mind. The aim is to identify devices and concrete implementations they could use. This approach also provides the possibility of identifying *interaction possibilities* that have to be fulfilled, similar to a requirements analysis. By filling several cells, this approach helps to identify alternative implementations or device types that could be used. In the following we illustrate this form of usage with an application example:

Example. We demonstrate the classification-based technique with the example application *EyeHealth*, which is composed of several exercises that aim to train the eye muscles and help with eye redness, fatigue and tension. In this section we concentrate on the classification of the application, a more detailed description of the single eye exercises follows in section 5. We started by decomposing existing *EyeCare* applications (mainly based on the smartphone application *Eye Care Plus* [6]) into the basic interaction possibility they rely on. Since the aim of those exercises is mainly to train the eye muscles and trigger refocusing of the eyes, we found the main *interaction possibility* "triggering an action by performing a con-/divergent eye movement" as shown in Figure 5. In the next step it had to be identified on which specific technological requirements the *interaction possibility* relies on in this case. Since the application is based on stimuli that are presented for both eyes (to be able to induce depth) a stereoscopic display is required. The presentation of the stimuli is further fixed to the display and therefore does not require *full world knowledge*. The combination of these requirements point out a cell, where the *EyeHealth* application is positioned inside the design space (cell (3,2)) indicating what specific technologies (e.g. type of HMD) and implementations (e.g. mono-/binocular eye tracking) have to be used to implement the application (see Fig. 4).

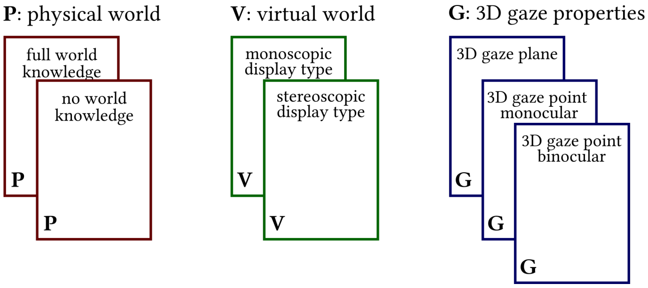


Figure 6: Here the different components are shown that are available for the ideation techniques.

Ideation Tool

Using the design space as an ideation tool is mainly directed towards researchers, designers and practitioners with the aim to derive new device types, interaction possibilities and application ideas expanding the content of the design space.

For this, we define three types of components (**P**, **V** and **G**) that cover a user's view of the world in the context of the design space. **P** refers here to knowledge that is available about the physical world, **V** refers to virtual world knowledge and **G** refers to knowledge that is available about the 3D gaze representation. This definition of component types is inspired by Milgram et al.'s [38] view on HMDs, who defined them to have a physical and virtual part. We combine this perspective with a representation of the user's gaze in 3D. An important realization is that **G** can exist inside both, the virtual **V** and physical **P** world but not at the same time. This realization is important to keep in mind when designing applications. For VR systems **P** is meaningless and only **V** and **G** apply, since VR by definition relies on virtual content only. The component types are strongly influenced by the parameters occurring in the design space (see Fig. 6). We present two different components for **P** and **V** respectively and three for **G**. However, these sets are not exhaustive and users are encouraged to expand them in the future.

Inspired by Card et al.'s [8] operators we then define exemplary **aggregations** that are applied to transfer the properties of each component into new expressions of device types or basic interaction possibilities. We choose two aggregations, which we will describe in more detail: **addition** and **substitution**.

Addition: adding two or three components results for example in a basic interaction possibility. An example would be adding *full* world knowledge for type **P** and a 3D gaze point for type **G** as shown in Figure 7 (left). Here the gaze point exists in the physical world **P** and can therefore be applied for refining the knowledge that we have about the physical world. One could for example recognize when the user follows a moving object with the eyes and as such detects something that is not available in the 3D representation of **P**. This information can then be applied to refine the physical 3D representation. This interaction possibility could be put in cells (2,2) and (4,2) of the design space.

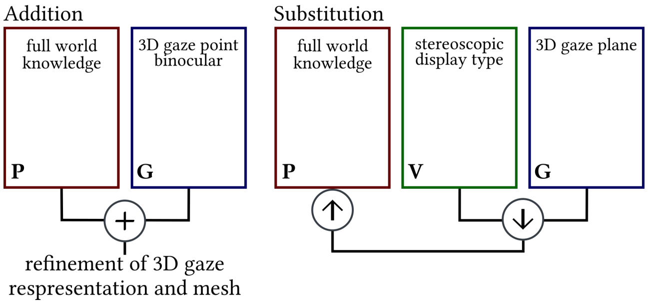


Figure 7: Here we graphically demonstrate the two aggregation techniques addition and substitution, which we propose as ideation techniques.

Substitution: here we start with a concrete set of three components (one of each type). Then we remove one of them and try to substitute for it's functionality using the remaining two components. An example is shown in Figure 7 (right): here the set of components consists of *full* world knowledge for type **P**, a *stereoscopic* display type for type **V** and *accommodation*-based 3D gaze estimation for type **G**. The component that should be substituted is *full* world knowledge. This could be done by recognizing whether the user is looking at the virtual display indicated by accommodation measurement. If the accommodation values do not correspond to the virtual plane, the user is probably looking at real world content. Thus the lack of world knowledge can be compensated for by the other two components. In this case, the functionality is not available yet because an autorefractor integrated into an HMD would be required. However, we want to use it for showing the power and feasibility of our design space that also applies to future device types. (Also ideation aims to generate ideas, not to discard them due to technical restrictions.)

Example. We demonstrate the *ideation tool* technique with the example *X-Ray Vision*, which was derived by the concept *addition*: We started with the components *P: full world knowledge* (as 3D mesh) and *G: 3D gaze binocular*, the combination of which results in three options: 3D gaze lies *behind*, *on* or *in front of* the mesh. Therefore, 3D gaze can be used to focus on objects that lie *on*, *behind* or *in front of* an object. A first implementation showed that it is almost impossible to focus on an imaginary point. Therefore, in the next iteration we added the third component *V: stereoscopic display type* to be able to display a virtual construct to guide the gaze, resulting in a *scaffolding pattern* that helps the user to shift focus (see section 5 for a more detailed description).

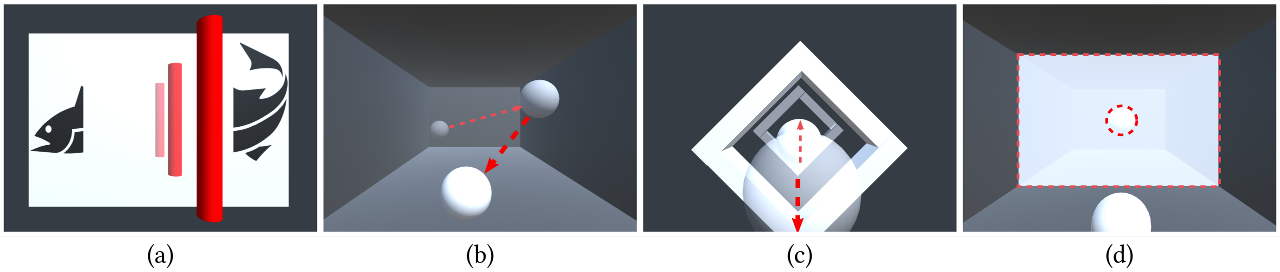

(a) (b) (c) (d)

Figure 8: Here we present four eye health exercises that help to reduce eye strain and tension by training the eye muscles. These are: (a) "Split Images", (b) "Follow the Bouncing Ball", (c) "Focus Shift" and (d) "Look Through the Wall".

The set of aggregations we proposed is expandable. For example Ramesh Raskar's [49] idea hexagon could also be applied for the components. We showed two examples of how our design space can be used to generate new ideas. In the next section we present two example applications that were derived by using these techniques.

5 PROTOTYPE AND APPLICATIONS

In the following we present two example applications that we derived using the *classification-based technique* of the design space. With the specific implementation we want to show the feasibility and applicability of our design space on the one hand. On the other hand we want to give insights that we derived during the implementation of the applications, which is what we present first before describing the concrete applications.

Insights

We started with using voluntary con-/divergent eye movements as an interaction method. Here we realized that while performing voluntary convergent eye movements is achievable with some amount of training, it is extremely difficult to fixate on a concrete but invisible point in space (e.g. fixating on a point behind a wall). To support the user in doing so, we explored different approaches to provide visual guidance. One of these is using a grid of semi transparent cubes as a *scaffold* (see Fig. 9 (c)). Another application of scaffolding is to help users resolve depth conflicts: since current HMDs like the HoloLens have a fixed accommodation plane it is sometimes difficult to distinguish whether a virtual object is positioned in front of or behind a physical object. For this we colored the cubes of the scaffold in two different colors indicating whether it lies in front of or behind a physical object. We explored some variations of the representation with the goal of the scaffold to be as unintrusive as possible for the user (e.g. size, transparency, frequency). One possible adoption is that the scaffold only appears around UI elements that need depth as an interaction instead of being currently scattered all over the physical space. However, future research can explore the individual benefits of different scaffolding techniques for 3D gaze interaction.

We implemented two example applications, where we practically implemented the insights. For this, we used the Microsoft HoloLens with an eye tracking add-on from Pupil Labs [21] to implement the applications that we present in the following. To calculate the 3D gaze point we used the 3D model tracking algorithm from Pupil Labs, which relies on a model-fitting algorithm approximating the two eyeballs as spheres and measuring their rotation to measure a 3D gaze point. We further applied the 1€ low-pass filter for data smoothing [10]. In section 4 we explained how the two example applications were derived using the presented usage techniques of the design space. In this section we concentrate on the specific content and implementation of the applications.

EyeHealth

As indicated in section 4, for this application we implemented four exercises that aim to train the eye muscles and help with eye redness, fatigue and tension. They are also designed to help with spasm of accommodation, which refers to a condition where the eye remains in a constant state of contraction. The exercises are based on the *Eye Care Plus* application [6], which also includes the medical intentions.

Split Images: Here a so called "split image" is presented on a virtual plane (see Fig. 8 (a)). A split image refers hereby to an image which is split in half, and each of its components is shown on one side of the virtual plane. In front of the image a red bar is positioned, which can be moved forward/backward. When focusing on the red bar it turns green, as such the system indicates that it keeps track of the user's eye movements. When focusing on the forward moving bar the two split images in the background merge to one percept. This exercise aims to improve focusing and stimulate the vision center of the brain among others.

Follow the Bouncing Ball: A sphere is moving through space and the user has to follow it with her eyes. The sphere turns green when a fixation is detected (see Fig. 8 (b)). This application aims to reduce eye stress and tension and tries to make the eyes more focused.

Focus shift: this application aims to train the eye muscles by "forcing" the eyes to refocus between forwards and backwards moving objects. Hereby a sphere and a cube are presented on the display, one of which moves forward or backward while the other does not move. The user has to follow the moving object with her eyes, which is again indicated by turning green (see Fig. 8 (c)).

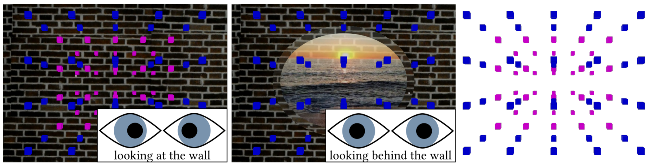
Look Through the Wall: Here performing a voluntary, divergent eye movement is trained. A 3D object is positioned behind a slightly transparent wall. The user has to focus the object behind the wall. Once this is detected the wall in front of the object disappears. Once the user refocuses on an object in front of the wall the wall appears again. This can be tested with different degrees of transparency (see Fig. 8 (d)) and also aims to train the eye muscles.

X-Ray Vision

This application was derived by the design space following the *ideation tool* technique as described in section 4. For this application, some hidden virtual content is triggered by the user when they focus at a point behind the wall (see Fig. 9). This is difficult, because our eyes are not used to focus on an invisible point in space. Therefore we provide a "scaffolding pattern", as shown in Figure 9 (c). This pattern is meant to support the user in refocusing, i.e. performing voluntary convergent and divergent eye movements, by applying pictorial depth cues (e.g. points in the front are bigger than points in the background). When the user successfully triggered a divergent movement with the eyes, i.e. established a predefined threshold from the 3D gaze point to the wall, the hidden virtual content is displayed (see Fig. 9 (b)). At this the horizon in the image helps to keep the eyes fixated at a certain distance.

6 CONCLUSION

In this work we presented the first design space for 3D gaze interaction on head-mounted displays. We identified two dimensions that combine technical properties of current HMD technology (**D1**) with properties of human depth perception (**D2**). The design space is directed towards researchers, designers and practitioners to identify opportunities and challenges for this new combination of two arising technologies. We demonstrated with three *views* on the design space how it can be filled with different types of content (technological, applications and interaction possibilities). We further showed the feasibility of our design space by proposing two forms of usage, one to classify applications and a second to derive new interaction ideas by using the design space as an ideation tool. In a last step, we used our design space to design and implement two exemplary interactive applications.



(a) (b) (c)

Figure 9: X-Ray Vision enables a user to reveal hidden virtual content behind a static physical object ((a) and (b)), e.g. a wall, by focusing on a point that lies behind the object. Since this is difficult to achieve without training we provide a "scaffolding pattern" (c), which is composed of small cubes to support the user in focusing at points on different depth levels. The colors and size of the scaffold in the figure are depicted exaggeratedly to make them visible to the reader.

Future Work and Limitations

Our design space is not limited to current device types or gaze tracking algorithms. The aim was to present a design space that gives an approach on how to think about problems and especially opportunities that arise from this combination of two technologies. The parameters we presented here were chosen to cover most of the current device types, however we want to emphasize that the framework allows for (and even gears towards) extension.

An example to expand the design space is adding light field or multi-focal displays to **D1**, which would result in more interaction possibilities for accommodation-based estimates.Another extension point is the parameter *world knowledge*: right now we treat this parameter as a binary one, assuming to have either no knowledge or to have a 3D representation of the environment with user tracking. However, we are aware that especially this parameter is expandable. On the one hand one could classify it more detailed considering more than two values and even extend it towards adding semantic object knowledge, where a representation of the environment would not only include a 3D map, but also knowledge about the objects that exist in it (similar to the extent of world knowledge as defined by Milgram et al. [38]).

ACKNOWLEDGMENTS

This work was supported by the Emmy Noether research group "Mobile Interaction with Pervasive User Interfaces" funded by German Research Foundation (DFG) and by the Landesgraduiertenförderung (LGFG) Scholarship for PhD students.

REFERENCES

[1] Ronald T Azuma. 1997. A survey of augmented reality. *Presence: Teleoperators & Virtual Environments* 6, 4 (1997), 355–385.

[2] Rafael Ballagas, Jan Borchers, Michael Rohs, and Jennifer G Sheridan. 2006. The smart phone: a ubiquitous input device. *IEEE Pervasive Computing* 1 (2006), 70–77.

[3] Rafael Ballagas, Sarthak Ghosh, and James Landay. 2018. The Design Space of 3D Printable Interactivity. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 2 (2018), 61.

[4] Mark Billinghurst. 2016. Virtual Reality: Sensing the Possibilities. Retrieved September 9, 2018 from https://de.slideshare.net/marknb00/virtual-reality-sensing-the-possibilities

[5] Mark Billinghurst. 2018. Introduction to Virtual Reality [Lecture]. Retrieved September 12, 2018 from https://de.slideshare.net/marknb00/lecture1-introduction-to-virtual-reality

[6] Brainbooth. 2013-2016. Eye Care Plus Exercises. Retrieved September 19, 2018 from http://www.eyeexamtest.com/en/

[7] Fergus W Campbell. 1957. The depth of field of the human eye. *Optica Acta: International Journal of Optics* 4, 4 (1957), 157–164.

[8] Stuart K Card, Jock D Mackinlay, and George G Robertson. 1990. The design space of input devices. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 117–124.

[9] Stuart K Card, Jock D Mackinlay, and George G Robertson. 1991. A morphological analysis of the design space of input devices. *ACM Transactions on Information Systems (TOIS)* 9, 2 (1991), 99–122.

[10] Géry Casiez, Nicolas Roussel, and Daniel Vogel. 2012. 1€ filter: a simple speed-based low-pass filter for noisy input in interactive systems. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2527–2530.

[11] David Drascic and Paul Milgram. 1996. Perceptual issues in augmented reality. In *Stereoscopic displays and virtual reality systems III*, Vol. 2653. International Society for Optics and Photonics, 123–135.

[12] Andrew T Duchowski, Donald H House, Jordan Gestring, Robert Congdon, Lech Świrski, Neil A Dodgson, Krzysztof Krejtz, and Izabela Krejtz. 2014. Comparing estimated gaze depth in virtual and physical environments. In *Proceedings of the Symposium on Eye Tracking Research and Applications*. ACM, 103–110.

[13] Andrew T Duchowski, Donald H House, Jordan Gestring, Rui I Wang, Krzysztof Krejtz, Izabela Krejtz, Radosław Mantiuk, and Bartosz Bazyluk. 2014. Reducing visual discomfort of 3D stereoscopic displays with gaze-contingent depth-of-field. In *Proceedings of the acm symposium on applied perception*. ACM, 39–46.

[14] Carlos Elmadjian, Pushkar Shukla, Antonio Diaz Tula, and Carlos H Morimoto. 2018. 3D gaze estimation in the scene volume with a head-mounted eye tracker. In *Proceedings of the Workshop on Communication by Gaze Interaction*. ACM, 3.

[15] Augusto Esteves, Eduardo Velloso, Andreas Bulling, and Hans Gellersen. 2015. Orbits: Gaze interaction for smart watches using smooth pursuit eye movements. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology*. ACM, 457–466.

[16] James D Foley, Victor L Wallace, and Peggy Chan. 1984. The human factors of computer graphics interaction techniques. *IEEE computer Graphics and Applications* 4, 11 (1984), 13–48.

[17] Florian Geiselhart, Michael Rietzler, and Enrico Rukzio. 2016. EyeVR: low-cost VR eye-based interaction. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct*. ACM, 277–280.

[18] Jan Hendrik Hammer, Michael Maurus, and Jürgen Beyerer. 2013. Real-time 3D gaze analysis in mobile applications. In *Proceedings of the 2013 Conference on Eye Tracking South Africa*. ACM, 75–78.

[19] Teresa Hirzle, Jan Gugenheimer, Florian Geiselhart, Andreas Bulling, and Enrico Rukzio. 2018. Towards a Symbiotic Human-Machine Depth Sensor: Exploring 3D Gaze for Object Reconstruction. In *The 31st Annual ACM Symposium on User Interface Software and Technology Adjunct Proceedings*. ACM, 114–116.

[20] David M Hoffman, Ahna R Girshick, Kurt Akeley, and Martin S Banks. 2008. Vergence–accommodation conflicts hinder visual performance and cause visual fatigue. *Journal of vision* 8, 3 (2008), 33–33.

[21] Moritz Kassner, William Patera, and Andreas Bulling. 2014. Pupil: an open source platform for pervasive eye tracking and mobile gaze-based interaction. In *Proceedings of the 2014 ACM international joint conference on pervasive and ubiquitous computing: Adjunct publication*. ACM, 1151–1160.

[22] Mohamed Khamis, Florian Alt, Mariam Hassib, Emanuel von Zezschwitz, Regina Hasholzner, and Andreas Bulling. 2016. Gazetouchpass: Multimodal authentication using gaze and touch on mobile devices. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*. ACM, 2156–2164.

[23] Dominik Kirst and Andreas Bulling. 2016. On the Verge: Voluntary Convergences for Accurate and Precise Timing of Gaze Input. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*. ACM, 1519–1525.

[24] Robert Konrad, Nitish Padmanaban, Keenan Molner, Emily A Cooper, and Gordon Wetzstein. 2017. Accommodation-invariant computational near-eye displays. *ACM Transactions on Graphics (TOG)* 36, 4 (2017), 88.

[25] George-Alex Koulieris, Bee Bui, Martin S Banks, and George Drettakis. 2017. Accommodation and comfort in head-mounted displays. *ACM Transactions on Graphics (TOG)* 36, 4 (2017), 87.

[26] Matthijs Kwak, Kasper Hornbæk, Panos Markopoulos, and Miguel Bruns Alonso. 2014. The design space of shape-changing interfaces: a repertory grid study. In *Proceedings of the 2014 conference on Designing interactive systems*. ACM, 181–190.

[27] Mikko Kytö, Barrett Ens, Thammathip Piumsomboon, Gun A Lee, and Mark Billinghurst. 2018. Pinpointing: Precise Head-and Eye-Based Target Selection for Augmented Reality. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, 81.

[28] Pierre-Yves Laffont, Ali Hasnain, Pierre-Yves Guillemet, Samuel Wirajaya, Joe Khoo, Deng Teng, and Jean-Charles Bazin. 2018. Verifocal: a platform for vision correction and accommodation in head-mounted displays. In *ACM SIGGRAPH 2018 Emerging Technologies*. ACM, 21.

[29] Marc Lambooij, Marten Fortuin, Ingrid Heynderickx, and Wijnand IJsselsteijn. 2009. Visual discomfort and visual fatigue of stereoscopic displays: A review. *Journal of Imaging Science and Technology* 53, 3 (2009), 30201–1.

[30] Ji Woo Lee, Chul Woo Cho, Kwang Yong Shin, Eui Chul Lee, and Kang Ryoung Park. 2012. 3D gaze tracking method using Purkinje images on eye optical model and pupil. *Optics and Lasers in Engineering* 50, 5 (2012), 736–751.

[31] Youngho Lee, Choonsung Shin, Alexander Plopski, Yuta Itoh, Thammathip Piumsomboon, Arindam Dey, Gun Lee, Seungwon Kim, and Mark Billinghurst. 2017. Estimating Gaze Depth Using Multi-Layer Perceptron. In *Ubiquitous Virtual Reality (ISUVR), 2017 International Symposium on*. IEEE, 26–29.

[32] Inc Magic Leap. 2018. Magic Leap. Retrieved Septmeber 21, 2018 from https://www.magicleap.com/

[33] Andrew Maimone, Gordon Wetzstein, Matthew Hirsch, Douglas Lanman, Ramesh Raskar, and Henry Fuchs. 2013. Focus 3D: Compressive accommodation display. *ACM Trans. Graph.* 32, 5 (2013), 153–1.

[34] Mohsen Mansouryar, Julian Steil, Yusuke Sugano, and Andreas Bulling. 2016. 3d gaze estimation from 2d pupil positions on monocular head-mounted eye trackers. In *Proceedings of the Ninth Biennial ACM Symposium on Eye Tracking Research & Applications*. ACM, 197–200.

[35] Christopher McMurrough, Jonathan Rich, Vangelis Metsis, An Nguyen, and Fillia Makedon. 2012. Low-cost head position tracking for gaze point estimation. In *Proceedings of the 5th International Conference on PErvasive Technologies Related to Assistive Environments*. ACM, 22.

[36] Microsoft. 2018. Microsoft HoloLens. Retrieved Septmeber 21, 2018 from https://www.microsoft.com/de-de/hololens

[37] Paul Milgram and Fumio Kishino. 1994. A taxonomy of mixed reality visual displays. *IEICE TRANSACTIONS on Information and Systems* 77, 12 (1994), 1321–1329.

[38] Paul Milgram, Haruo Takemura, Akira Utsumi, and Fumio Kishino. 1995. Augmented reality: A class of displays on the reality-virtuality continuum. In *Telemanipulator and telepresence technologies*, Vol. 2351. International Society for Optics and Photonics, 282–293.

[39] Esteban Gutierrez Mlot, Hamed Bahmani, Siegfried Wahl, and Enkelejda Kasneci. 2016. 3D Gaze Estimation using Eye Vergence.. In *HEALTHINF*. 125–131.

[40] Moverio. 2018. MOVERIO Smart Glasses. Retrieved Septmeber 21, 2018 from https://epson.com/moverio-smart-glasses-augmented-reality-devices-headsets

[41] Susan M Munn and Jeff B Pelz. 2008. 3D point-of-regard, position and head orientation from a portable monocular video-based eye tracker. In *Proceedings of the 2008 symposium on Eye tracking research & applications*. ACM, 181–188.

[42] Nitish Padmanaban, Robert Konrad, Tal Stramer, Emily A Cooper, and Gordon Wetzstein. 2017. Optimizing virtual reality for all users through gaze-contingent and adaptive focus displays. *Proceedings of the National Academy of Sciences* (2017), 201617251.

[43] Lucas Paletta, Helmut Neuschmied, Michael Schwarz, Gerald Lodron, Martin Pszeida, Stefan Ladstätter, and Patrick Luley. 2014. Smartphone eye tracking toolbox: accurate gaze recovery on mobile displays. In *Proceedings of the Symposium on Eye Tracking Research and Applications*. ACM, 367–68.

[44] Lucas Paletta, Katrin Santner, Gerald Fritz, and Heinz Mayer. 2013. 3D recovery of human gaze in natural environments. In *Intelligent Robots and Computer Vision XXX: Algorithms and Techniques*, Vol. 8662. International Society for Optics and Photonics, 86620K.

[45] Anjul Patney, Marco Salvi, Joohwan Kim, Anton Kaplanyan, Chris Wyman, Nir Benty, David Luebke, and Aaron Lefohn. 2016. Towards foveated rendering for gaze-tracked virtual reality. *ACM Transactions on Graphics (TOG)* 35, 6 (2016), 179.

[46] Thies Pfeiffer. 2012. Measuring and visualizing attention in space with 3D attention volumes. In *Proceedings of the Symposium on Eye Tracking Research and Applications*. ACM, 29–36.

[47] Thies Pfeiffer and Patrick Renner. 2014. EyeSee3D: a low-cost approach for analyzing mobile 3D eye tracking data using computer vision and augmented reality technology. In *Proceedings of the Symposium on Eye Tracking Research and Applications*. ACM, 369–376.

[48] Ken Pfeuffer, Benedikt Mayer, Diako Mardanbegi, and Hans Gellersen. 2017. Gaze+ pinch interaction in virtual reality. In *Proceedings of the 5th Symposium on Spatial User Interaction*. ACM, 99–108.

[49] Ramesh Raskar. 2012. Idea Hexagon. Retrieved Septmeber 20, 2018 from https://de.slideshare.net/cameraculture/raskar-ink-2012-in-pune-india-future-of-imaging-and-idea-hexagon

[50] Stephan Reichelt, Ralf Häussler, Gerald Fütterer, and Norbert Leister. 2010. Depth cues in human visual perception and their realization in 3D displays. In *Three-Dimensional Imaging, Visualization, and Display 2010 and Display Technologies and Applications for Defense, Security, and Avionics IV*, Vol. 7690. International Society for Optics and Photonics, 76900B.

[51] Takashi Shibata, Joohwan Kim, David M Hoffman, and Martin S Banks. 2011. The zone of comfort: Predicting visual discomfort with stereo displays. *Journal of vision* 11, 8 (2011), 11–11.

[52] Sheng-Wen Shih and Jin Liu. 2004. A novel approach to 3-D gaze tracking using stereo cameras. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics* 34, 1 (2004), 234–245.

[53] Ivan E. Sutherland. 1968. A Head-mounted Three Dimensional Display. In *Proceedings of the December 9-11, 1968, Fall Joint Computer Conference, Part I (AFIPS '68 (Fall, part I))*. ACM, New York, NY, USA, 757–764. https://doi.org/10.1145/1476589.1476686

[54] FM Toates. 1972. Accommodation function of the human eye. *Physiological Reviews* 52, 4 (1972), 828–863.

[55] Mélodie Vidal, Andreas Bulling, and Hans Gellersen. 2013. Pursuits: spontaneous interaction with displays based on smooth pursuit eye movement and moving targets. In *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing*. ACM, 439–448.

[56] Mélodie Vidal, David H Nguyen, and Kent Lyons. 2014. Looking at or through?: using eye tracking to infer attention location for wearable transparent displays. In *Proceedings of the 2014 ACM International Symposium on Wearable Computers*. ACM, 87–90.

[57] HTC Vive. 2018. VIVE VR system. Retrieved Septmeber 21, 2018 from https://www.vive.com/us/product/vive-virtual-reality-system/

[58] Rui I Wang, Brandon Pelfrey, Andrew T Duchowski, and Donald H House. 2014. Online 3D gaze localization on stereoscopic displays. *ACM Transactions on Applied Perception (TAP)* 11, 1 (2014), 3.

[59] Martin Weier, Thorsten Roth, André Hinkenjann, and Philipp Slusallek. 2018. Predicting the gaze depth in head-mounted displays using multiple feature regression. In *Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications*. ACM, 19.

[60] David Zeltzer. 1992. Autonomy, Interaction, and Presence. *Presence: Teleoper. Virtual Environ.* 1, 1 (Jan. 1992), 127–132. https://doi.org/10.1162/pres.1992.1.1.127

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

[62] Yanxia Zhang, Jörg Müller, Ming Ki Chong, Andreas Bulling, and Hans Gellersen. 2014. 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*. ACM, 559–563.

[63] Feng Zhou, Henry Been-Lirn Duh, and Mark Billinghurst. 2008. Trends in Augmented Reality Tracking, Interaction and Display: A Review of Ten Years of ISMAR. In *Proceedings of the 7th IEEE/ACM International Symposium on Mixed and Augmented Reality (ISMAR '08)*. IEEE Computer Society, Washington, DC, USA, 193–202. https://doi.org/10.1109/ISMAR.2008.4637362

[64] Fritz Zwicky. 1967. The morphological approach to discovery, invention, research and construction. In *New methods of thought and procedure*. Springer, 273–297.