




The Role of Depth Perception in XR from a Neuroscience Perspective: A Primer and Survey

V. Hushagen^{1,2,3} , G.C. Tresselt³, N.N. Smit^{2,4}  and K. Specht^{1,2,5} 

¹Department of Biological & Medical Psychology, University of Bergen, Norway

²Mohn Medical Imaging and Visualization Centre, Haukeland University Hospital, Norway

³Tenklabs AS, Bergen, Norway

⁴Department of Informatics, University of Bergen, Norway

⁵Department of Education, University of Tromsø, Norway

Abstract

Augmented and virtual reality (XR) are potentially powerful tools for enhancing the efficiency of interactive visualization of complex data in biology and medicine. The benefits of visualization of digital objects in XR mainly arise from enhanced depth perception due to the stereoscopic nature of XR head mounted devices. With the added depth dimension, XR is in a prime position to convey complex information and support tasks where 3D information is important.

In order to inform the development of novel XR applications in the biology and medicine domain, we present a survey which reviews the neuroscientific basis underlying the immersive features of XR. To make this literature more accessible to the visualization community, we first describe the basics of the visual system, highlighting how visual features are combined to objects and processed in higher cortical areas with a special focus on depth vision. Based on state of the art findings in neuroscience literature related to depth perception, we provide several recommendations for developers and designers. Our aim is to aid development of XR applications and strengthen development of tools aimed at molecular visualization, medical education, and surgery, as well as inspire new application areas.

CCS Concepts

• **General and reference** → *Surveys and overviews*; • **Applied computing** → *Life and medical sciences*; • **Human-centered computing** → *Virtual reality; Mixed / augmented reality*;

1. Introduction

Software tools are more prevalent than ever in medicine, and are in use in education, training, and analysis of images and surgery. Traditionally, such tasks are performed on a PC with a mouse and keyboard setup as input devices. When working with text or 2D image information, this set-up works well. However, with the increasing availability of volume imaging in medicine, depth information is lost in this set-up and needs to be compensated for by additional interaction or visual encoding. Since the 1990s, researchers have investigated the possibility of using Virtual Reality (VR) and Augmented Reality (AR - collectively "XR") in biology and medicine to further enrich interactive visualization possibilities. This article focuses on the head mounted device (HMD) category of XR devices in particular, which offer several advantages. XR technologies can be classified as immersive or semi-immersive, practically referring to to which degree real surroundings are visible to the user or not. "VR" refers to full replacement of external visuals, while "AR" refers to semi-immersive experiences, showing part or whole of the users surroundings.

One clear advantage of such XR devices is stereopsis - the ability of bifocal vision to triangulate visual depth - the fundamental component in depth perception [Par19]. By placing a display in front of each eye, software can digitally manipulate stereopsis to simulate a depth component - simulating how we normally see the world. In this way, tasks that rely on accurate depth information can be better supported.

A second advantage of HMDs is provided by head tracking functionality, combining stereopsis with visuospatial scanning behavior such as head turning and body posture adjustments. In addition, while vision is the main way humans perceive the world, motor abilities and tactile perception are the only way of manipulating the objects in our surroundings. A considerable part of human primary motor cortex is dedicated to hand, and finger control [UMCA06]. To exploit this core aspect of human ability, XR technologies increasingly offer complex and precise interfaces to users' hands, fingers and tactile perceptions [MKC*19]. High perceptual precision combined with the accuracy of the XR motor interface provides an excellent foundation for developing educational and clinical tools.

In addition, the immersive nature of XR applications provides the user freedom to explore data and images in new ways.

The third major advantage of XR technology is the sense of immersion. Due to the low latency hand interaction with virtual objects combined with stereopsis and head tracking, XR users report an illusion of being present in the virtual environment, also known as immersion. The "feeling of immersion" dates back to the first definitions of virtual reality, which described it as the feeling of presence in an artificial world [Ste12]. The term *immersion* in virtual reality may refer to both the subjective user experience as well as the equipment required to generate the virtual environment. In this article, when discussing immersion we refer to the user's experience of being present in a virtual world unless otherwise specified. The immersive aspect of XR is especially suited for surgical training, with the advantage of providing realistic visual and haptic feedback in an environment the user perceives as an operating room.

Despite these promising advantages, early attempts did not succeed in accomplishing widespread adoption in the field. However, recently published reviews in medicine, together with recent advancements in hardware, suggest XR is in a strong position for increased adoption in the biomedical domain [JH20, PNL*17, WD-JGK16]. Developing biomedical applications in XR is a highly active area of research. Numerous recent examples leveraging VR include supporting paramedic skill training [VJR19], liver surgery planning [BHS*21], and protein visualization [CŠR*20]. Since virtual reality grants developers full control over the visual field of a user, a considerable amount of neuronal innervation of the brain can be controlled digitally, and a considerable amount of behavior relevant variables can be manipulated with high fidelity. However, this opportunity also comes with responsibility for the user's experience. Well-designed VR environments and technologies can lead to mesmerizing experiences, enhanced information presentation, situation awareness and a feeling of immersion, while poorly designed environments and technologies may lead to confusion, motion sickness and nausea [MDS17].

Purpose of the survey Several reviews focused on comparing XR with traditional training and visualization methods, reporting mostly positive effects, but with uncertain transfer to practice [Mat15, ALHN16, GALG*18]. In addition, a recent survey provided an overview and taxonomy of spatial interfaces for 3D visualization [BYK*21]. Our survey investigates the basic features of vision and its implications for XR design. Exploration of these elements in visual computing solutions is essential for developing efficient methods and best practices for interacting with different data types. To aid development of the next generation of XR tools in biology and medicine, we explore the basic functioning of vision and the processing of different types of visual information. This paper aims to provide an overview of depth-perception related literature and its relevance when developing XR tools, serious games, and experiences. With this, our goal is to provide guidelines for developing efficient immersive experiences in virtual reality informed by the state of the art in neuroscience. With this, our main contributions are the following:

- **An introduction to Visual Neuroscience:** We provide a basic introduction to the field of visual neuroscience with the goal of

establishing a common understanding of how 2D and 3D objects are processed in the brain.

- **A survey of the state of the art in depth perception for virtual objects:** We provide an overview of neuroscience literature related to object perception, feature detection, interaction with virtual objects, and selected clinical applications of XR.
- **Recommendations for XR Development:** Based on the literature, we provide guidelines and recommendations for XR application design and development.

Organization. In this survey, we provide an overview of volumetric perception of digital objects from a visual neuroscience perspective, with a specific focus on virtual and augmented reality. In Section 2, we present background information on the visual system, and how the brain processes central visual stimuli in visual cortical brain areas. Scientific visualization relies on accurate perception of volumetric information. Therefore, Section 3 presents a literature review of how 3D objects are perceived and processed in the brain. In Section 4 we discuss the implications of central object processing features in light of XR technology development for applications in biology and medicine and beyond. Section 5 consists of a concluding remark and recommendations based on the findings in our survey.

2. Neuroscience Background

Vision is the primary human sense for gathering information about the world and the surrounding environment. Accordingly, VR is a powerful tool for creating realistic immersive experiences [Eks15], and realism and realistic scenarios are essential both medical training and clinical practice. Stereopsis is the foundation of depth perception, and the reason for the immersive feeling of presence in VR. Consequently, when designing applications that replace all visual input, certain biological constraints should be considered, especially when developing medical applications for clinical practice.

2.1. The Visual System

As humans we perceive our surroundings through our senses. Generating virtual experiences relies on mimicking the natural input from our surroundings. To achieve this, we explore which stimuli the brain is used to processing and how this processing occurs. The visual system generates the main sensory representation of the world for humans, see Figure 1. The photoreceptors in the back of the eye (retina) receive about 10 in the 9th bits (125 megabytes) of visual input per second [Kel62, Zha19]. These photoreceptors initiate the process by absorbing electromagnetic radiation between 400 and 780 nm [Pal12]. There are two main classes of photoreceptors: rods and cones. Rods are sensitive to luminance, while three types of cones (R, G, B) are sensitive to different wavelengths of light and form the basis for color vision. Each photoreceptor in the eye is connected to a retinal ganglion cell (RGC, neuron), which transmits information from several photoreceptors based on their pattern of input. These RGCs typically have a centre-surround firing setup. If a RGC is "on centre and off surround" it will fire (initiate action potentials) when stimulated by light in the centre, which is the photoreceptors' signal for stimulation. An illustration of this process is displayed in Figure 1 on the right.

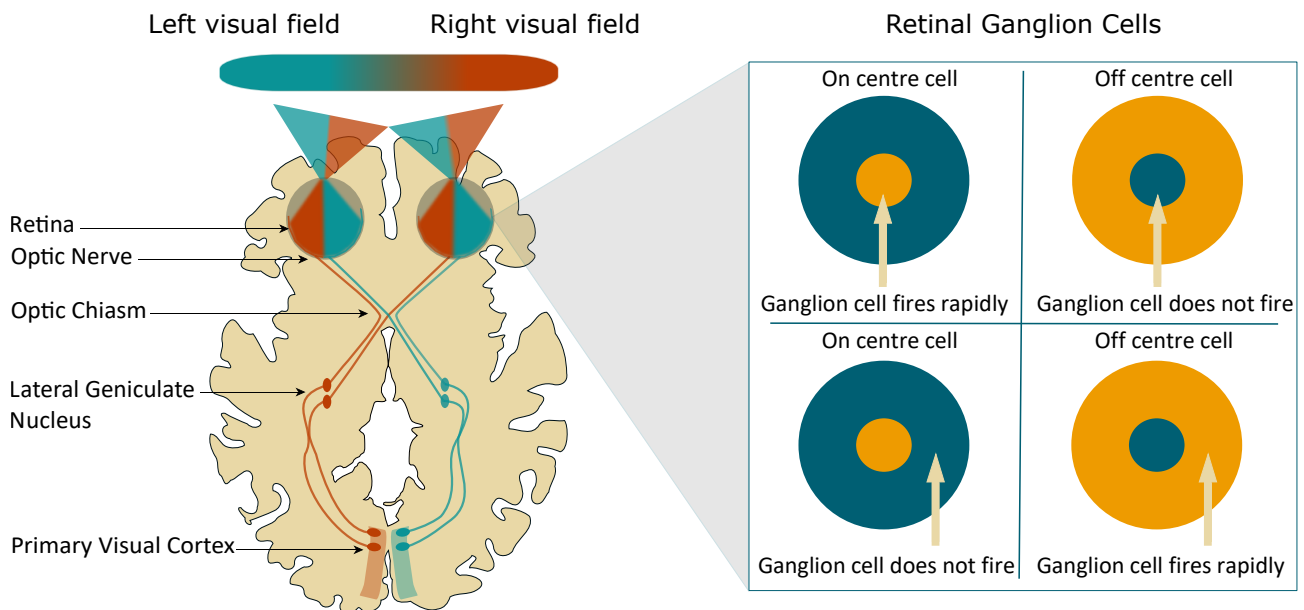


Figure 1: *Left: Overview of the visual system: light stimulates photoreceptor in the retina, signalling to their RGC. Light from the left side of the visual field interacts with photoreceptors in the right side of the retina. Signals from the right side of both retinas travels through the optic nerve to the visual cortex in the right hemisphere. From the retina to the primary visual cortex, the signal passes the optic chiasm and the lateral geniculate nucleus (LGN) of the thalamus, where processing of stimuli properties begins. Right: Retinal Ganglion Cells (RGCs) "on" / "off" receptive fields. RGCs fire rapidly when light stimulates the photoreceptors within the "on" region of the neurons receptive field. They signal to the brain that there is light in this RGCs "on" region.*

The Optic Nerve The axons of the RGCs form a bundle called the optic nerve, one emerges from the back of each eye. The optic nerves cross in the optic chiasm as seen in Figure 1 on the left, with information from the right half of the visual field travelling to the left visual cortex and the left visual field (of each eye) ending up in the right visual cortex. After crossing the optic chiasm, visual stimuli pass through the lateral geniculate nucleus (LGN) in the thalamus [UA15]. The LGN boosts the signal-to-noise ratio from the retinal input, strengthens centre-surround interactions and acts as a gateway for visual stimuli to reach the cortex. The primary visual cortex in each hemisphere receives input from both eyes, but only one visual field, respectively the right and left side of the visual field. For simplicity and the overall purpose of this article, the following sections will focus on the most dominant and most relevant processing streams and hubs of the visual system. The other parts of the visual processing network, which can't be discussed here, are mostly related to reflexive and emotional processing of visual cues.

Primary Visual Cortex In Figure 2, the organization of the primary visual areas in the brain is illustrated. Figure 3 highlights the major connections between these areas. The primary visual cortex, which is located in the back of the brain in the occipital lobe, is also known as V1 or the striate cortex. This name originated from its stripy appearance to the naked eye caused by myelinated axons from the LGN (named: The Line of Gennari). V1 is a hub, it receives all visual input and passes it along to other cortical areas for further processing. Additionally, V1 performs processing of sig-

nals from RGCs. Information from several RGCs is combined in V1 to generate "edge" detectors. These neurons respond to a line of RGC receptive fields creating a macro receptive field shaped like a "bar"(\, /, ...) oriented at a specific angle. Three different types of edge detectors are currently known: "simple cells", "complex cells" and "hypercomplex cells" [HW59, HW68]. Simple cells have a distinct "on" (excitatory) and "off" (inhibitory) area separated by parallel straight lines representing their receptive field. Complex cells are the most common among the receptive cells. They have no on/off region and responds best to a "bar" stimulus presented in a specific orientation. Hypercomplex cells share properties with the complex cell, but in addition if the "bar" is extended beyond the receptive field the activation rate drops noticeably or stops completely. The processing of information in V1 is organized in a retinotopic fashion; adjacent points in the visual field are represented in topographical fashion in the visual cortex [HW68]. The fovea (middle retinal area with highest density of cone receptors) and central degrees of the visual field/retina occupies a disproportionately large area of V1 [DW61]. After receiving and processing visual input V1 transmits information to cortical areas that are higher in the processing hierarchy of the visual system.

V2, which is next to V1, receives the main load of information from V1, performing the next level of processing before transmitting to higher cortical areas. The visual field representation contained in V2 is topographically organized, representing at least 80° of the contralateral visual fields [GGS81]. Further, the visual area V3 represents 30-40° of the central visual field [GSG88].

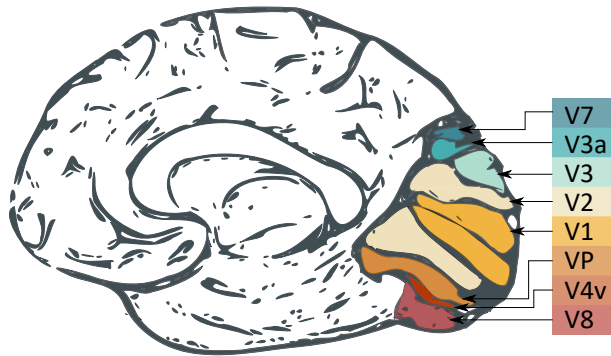


Figure 2: The organization of the primary visual areas projected on the midline slice of the brain. Adapted from Logothetis [Log99].

From V1 until V3 processing of visual information splits progressively into two pathways/streams: the dorsal ("up") and the ventral ("side") stream [GF18, KSB*13]. The streams process different properties of the visual stimuli, which we will come back to after we have discussed the most central properties of visual stimuli.

Visual Stimuli The properties of the visual stimuli that are coded from the retina to V1 are; color (RGB), brightness (centre/surround, RGCs) and their spatial localization in the visual field. The other visual cues mentioned in the next paragraph are second order visual cues derived from primary visual input [Par05].

Color processing mainly occurs in V4/V8, which are located in the lower part of the occipital lobe at the back of the brain (see Figure 2, but processing of color is a continuous process starting in the retina, moving through V1 and V2 before arriving in V4/V8 [HLD*98, WBRW02, WALW08].

Brightness/illumination is necessary for sight, this central cue is processed in V1. There is an important distinction between brightness which is perceived, and illumination which is the actual level of luminosity that interacts with the rods in the retina [RRP96]. Separate groups of neurons in V1 respond to surface/focused luminance, and luminance in the surrounding environment, the latter group corresponds to perceived brightness and the former encodes the level of illumination [KK01a].

Spatial localization is coded from the retina and represented topographically in the primary visual cortex, V1 [Eng01].

Edge Detection occurs in V1, as a sum of the centre/surround RGCs in the retina [HW68].

Second Order Cues The following visual cues are based in one or several of the primary visual cues combined. Here we explore the main second order visual cues, especially the ones relevant to perception of objects and depth.

Motion is primarily processed by the middle temporal area (MT+)/V5. This region is active selectively when viewing moving versus stationary stimuli [TRK*95, WMF*93]. Visual motion detection is a critical cue in way finding, depth segregation, shape from motion, judgement of distance and speed, and judgement of biological activity [CIG11].

Shading is utilized to extract the shape of objects, which follows

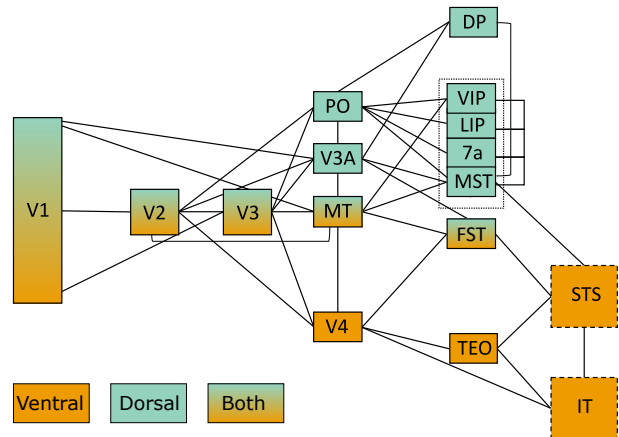


Figure 3: The major connections in the dorsal and ventral visual streams [Ung82, KSB*13, GF18]. The dorsal stream areas are shown in teal and the ventral stream areas in orange. Areas involved in both streams are highlighted in a color gradient. V1, visual area 1; V2, visual area 2; V3, visual area 3; V4, visual area 4, V3A, visual area 3A; PO, parieto-occipital area; MT, Middle temporal; DP, dorsal prelunate area; VIP, ventral intraparietal sulcus; LIP, lateral intraparietal sulcus; 7a, parietal area 7a; MST, medial superior temporal area; FST, fundus of the superior temporal sulcus; STS, anterior complex of the superior temporal sulcus; IT, inferotemporal cortex. Adapted from Distler et al. [DBDU93].

the two assumptions that a single light source illuminates the whole scene and that the light source is placed above the horizon [KR92]. Extraction of shading occurs early in the visual process and may play a role in motion perception. The intra-parietal sulcus (see Figure 3) is commonly activated when viewing 3D surface structures formed by shading [TNIT01].

Texture is the perceived quality of a surface, and is utilized in the perception of depth. Neurobiologically, the areas V4, TEO and V3A are activated by different types of texture [KDWU00, LRRS99].

Familiar size can be applied to judge distance from an object or landmark [CTS19]. Familiar size interacts with angular size, a large object far away activates a larger portion of V1, compared with an object of equal angular size perceived to be small and closer [HB41, MBK06].

Blur is perceived when gaze is fixated in a certain depth plane. Objects further from our focus becoming more blurry compared with the ones close to our fixation plane [Mat97]. This makes blur an effective visual cue for estimating depth in peripheral vision [HCB12].

Occlusion allows identification of an object's location in depth relative to other objects. This process relies on object and figure ground segmentation, these processes are explored further in the following section [BZP97, FS92, HFMK08].

These visual cues are utilized together, and play an important role in differentiating and organizing shapes into objects in the visual scene.

Perceptual Organization Visual perception follows certain cognitive principles, which aid in the extraction of meaning from a scene. Figure ground segmentation involves the perceptual separation of objects from each other and from the background. This is essential when constructing a visual scene from visual cues. The gestalt school attempts to explain how perception of objects and scenes occur, with a focus on "totality". This includes viewing all parts of the visual scene and conscious experience as an interconnected system of dynamic relationships [Wer38]. The main gestalt principles are: Emergence, Reification, Multi-stability, and Invariance [ESS02, Wer38].

- **Emergence** - The whole has properties that the parts do not possess on their own, which emerge only when the parts interact as a larger whole. Typically described along the lines of: "The whole is greater than the sum of its parts".
- **Reification** - A generative aspect of perception which occurs when an individual perceives something that is not present in the visual field as a physical object. An example of could be an intermittent pattern (star) forming a shape that we interpret as said pattern.
- **Multi-stability** - The tendency of ambiguous perceptual experiences to switch back and forth between several alternate interpretations. An example is the Rubin vase, which can be interpreted as either a vase, or two faces looking at each other.
- **Invariance** - A property of perception which allows the perceiver to recognize an object in a different rotation, scale, lighting, and elastic deformation.

These four gestalt principles are built on seven main "laws" of gestalt visual perception. They form the basis for the figure ground organization and perceptual grouping that the gestalt principles attempt to explain [ESS02]. Figure 4 displays five of these seven main "laws". The two final gestalt laws are the law of past experience and the law of common fate. The law of past experience implies that visual stimuli in certain conditions are grouped together based on previous experience with those stimuli. Finally, there is the law of common fate: if two elements of a figure move together in synchrony, they will be perceived as being in the same group.

In modern neuroscience the Gestalt principles have been slightly reworked, but most of them remain relevant and have found support from basic visual research [WEK*11]. Perception of totality and figure ground segmentation is important in depth perception, since depth cues interact with gestalt principles when the brain constructs the spatial organization of a scene.

Depth Perception Perception of depth is essential when perceiving the structure of objects, and creating a representation of the surrounding environment. This process requires contribution from multiple visual cues and gestalt principles, dependent on the organization of the scene. Stimuli contributing to depth perception can be classified as monocular or binocular cues, based on if input from one or both eyes are required to utilize the stimuli [CV95].

Monocular cues enable perception of depth by stimuli from one eye. Monocular perception of depth builds on gestalt cognition and

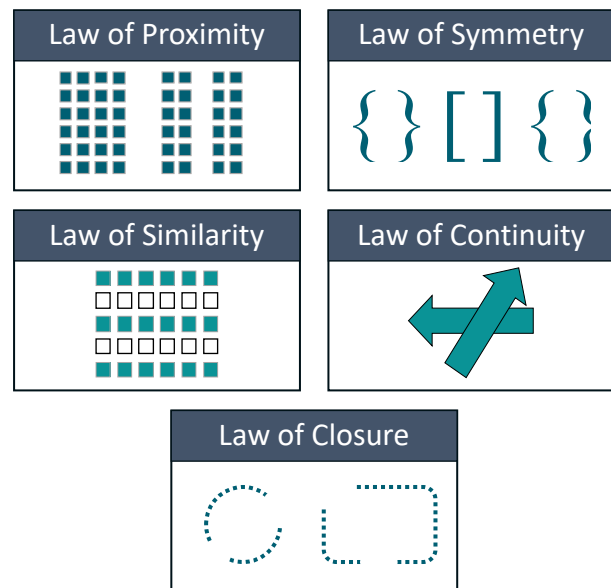


Figure 4: Five of the main visual gestalt laws illustrated.

several of the second order visual cues: motion, shading, texture, size, shadow, blur and occlusion [GTPO08, HPVN06, PCPD*08, PCT*05]. Visualizing volume data on a 2D screen relies on one or several of the monocular cues which were introduced previously.

Binocular stimuli are the foundation of stereopsis, which is the feeling of depth generated by seeing the world summarized from two slightly different images present in each retina. This process is reliant on disparity. The disparity of a point on an object refers to the difference in retinal location of that point between the right and left eye, also called absolute disparity. Disparity can be absolute or relative. Relative disparity refers to the angular difference between where two objects fall on the retina [Par05]. Binocular disparity plays a central role in perception of depth, with distorted binocular disparity leading to dramatically reduced depth perception accuracy [PBK*12]. It is estimated that 7% of the adult population < 60 years is stereo blind. However, recovery of stereopsis is possible in certain cases [CBL19]. Several cortical regions are highly sensitive to relative disparity, both in the dorsal and ventral visual stream which both play a central role in the abstract processing of visual information [BFPH01]. The purpose and organization of the visual streams are introduced in the following subsection.

The Visual Streams Beyond V1 and V2 the processing of visual information separate into two pathways, the dorsal (Latin for "back", meaning up/over neurologically) and ventral (Latin for "belly", meaning down/lower neurologically) [GM92]. The dorsal stream is mainly responsible for "vision for action" and the ventral for "vision for perception". Therefore, they are also often called the "where" and "what" stream.

The dorsal stream is a path of processing responsible for "where" stimuli are in a scene and making that information available to cortical areas involved in motoric movement. The dorsal stream proceeds from V1 up towards parietal areas of the brain.

The main types of information processed in this stream relates to motion and supports motor action. Two areas encountered early in the stream are V3A and V3B/KO. V3A responds systematically to motion and varying disparity [OPF*03, PLKW08], while the area V3B/kinetic occipital (KO) process fusion of disparity and motion depth cues [ABWV19]. The stream branches off to the posterior fusiform sulcus (pFs) which contains neurons that are sensitive to surface gloss [SBDLW15].

A lateral part of the stream proceeds to MT+/V5 which are considered the motion areas. In addition to upstream input, it receives direct input from V1 [GF03]. Another dorsal area that is sensitive to motion, more specifically to the direction of motion is the ventral intraparietal area (VIP) [CDG93]. Further, the neurons of the medial intraparietal area (MIP) are sensitive to reaching movements [CD91]. The intraparietal sulcus (IPS) is a large area at the end of the dorsal stream that is involved in saccades, multisensory integration, visual attention, and visual short term memory [SHM*07].

The ventral stream is mainly responsible for perception of objects, understanding "what" is in the visual scene. The ventral stream originates in V1, moving to V4, and continues to the posterior inferotemporal cortex (TEO), which responds to patterns of luminance, relative motion, and texture differences [Hik97]. The stream continues to the inferior temporal cortex (TE) which codes for particular features that are present in an image or object and together represents the complete representation of the object [OdBV00, Tan96]. Some neurons in TE are also selective to spatial position in the visual field, mostly in the central parts of the visual field.

V4 is connected to V1, V2, V3, MT and TEO, in addition to being involved in both streams, and sensitive to stimuli of different color and shape [UGDG08, WALW08]. Both visual streams utilize binocular disparity in depth processing, but to different extent. The ventral visual areas are specialised in processing of 3D shapes, while dorsal areas are focused on processing of long surfaces, rotation, and differentiation of depth planes [Par05]. It should be noted that the two-stream theory is a useful model for separating general functional areas of the visual system but does not reflect the full complexity of functional and anatomical interconnections in the pathways [GF18, KSB*13]. The following section will introduce some of the methods that have been applied to study these complex relations.

2.2. Central Neuroimaging Methods

The articles we selected from the literature search presented in the next section use a wide variety of experimental and neuroimaging methods. For a better understanding of the selected studies and the implications of their results, this section will briefly introduce some of the most central methods.

Electro Encephalography (EEG) measures the electrical activity in the brain. When neurons are active, they communicate with each other through electro-chemical processes. Through electrodes, which are touching the scalp, changes in brain potentials could be measured. If this method is combined with a structured presentation of stimuli to the research subjects, specific responses

of the neurons could be captured, which are then called **Event Related Potentials (ERP)**. ERP studies have the advantage of a high temporal resolution, since brain responses could be captured in the range of milliseconds. However, the origin and, hence, the localisation of the activated neurons is limited.

Functional magnetic resonance imaging (fMRI) is a non-invasive method to measure brain activation in humans while they, for example, perceive visual stimuli or perform a cognitive task. The method rests on the blood oxygen level dependent (BOLD) response, which was discovered in the 90s, and has since been increasingly utilized to capture functional images of the brain [OLKT90]. The BOLD response is a metabolic response that is triggered by neuronal activity and allows tracing the neuronal activation after the presentation of - in this case - visual stimuli. Compared to EEG and ERP experiments, fMRI has a good spatial resolution through its volumetric imaging method, but the temporal resolution is much lower because of both the underlying imaging method and the slow evolving metabolic signal.

Experimental designs describe how neuroscience studies are conducted. They are typically rooted on psychological or neuroanatomical assumptions and concepts. Several experimental designs for EEG and fMRI studies have been developed to probe the visual network of the brain. We briefly highlight a few examples from our literature search:

- **Priming** occurs when a subject is presented with a stimulus that subconsciously enhances/prepares the brain for processing a specific type of stimuli [GGV*06]. One way to apply priming in the investigation of depth perception is by presenting identical ambiguous stimuli continually and priming the subject to perceive the stimuli as a 2D shape or a 3D object [GGV*06].
- **Random dot stereogram (RDS)** presents one image of random black and white dots to each eye, combining the images leads to the 3D perception of an object through stereopsis [Par05]. RDS can be applied to investigate the effect of binocular disparity in depth perception, without interference from monocular visual cues.
- **Adaptation** is an inhibition of the neuronal response to identical or almost identical stimuli presented repeatedly. This method can be applied to investigate the detection threshold of the human perceptual system [FMKH05].
- **Event-Related paradigms** are those where single stimuli are presented with an inter-stimulus interval of up to several seconds. Those types of experimental designs are the core of ERP studies but are also used in MRI studies, and they may contain 40-100 single stimuli (events).
- **Block-Design** paradigms are the most common experimental designs in fMRI studies. A series of stimuli of the same type (but not identical) are presented over a blocked period of typically 20-40 seconds, alternated with either a period of no stimulation or a control condition. Several of those alternating blocks are forming such a study.

These neuroscience methods form the foundation needed of the survey presented in the following Section. Based on our literature

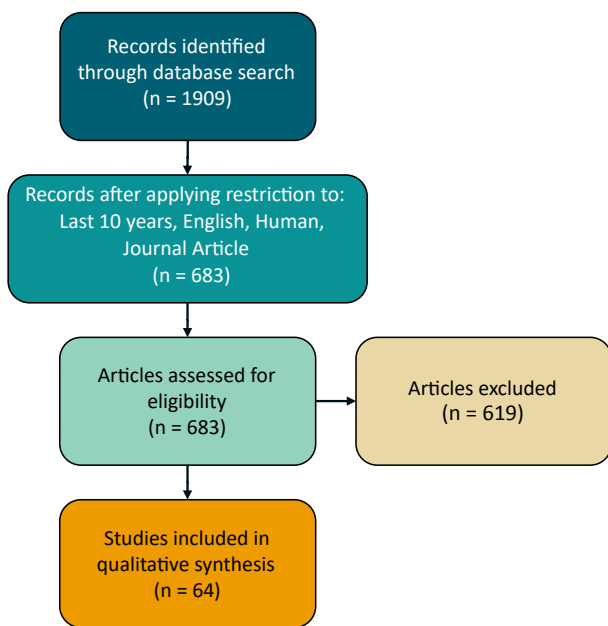


Figure 5: Flow chart illustrating the literature search we performed in our survey.

search we present the results relevant for digital object perception, depth perception and novel interaction with 2D and 3D objects.

3. Survey

In this section, we present our data collection and synthesis of the literature relevant to perception and interaction with 2D and 3D objects in virtual environments. Perception of visual scenes rely on monocular and binocular cues; these form the visual input both when viewing real or digital scenes. To make enjoyable visual experiences, we are dependent on presentation of visual stimuli in a natural way that integrates with the structure of the visual system. When designing digital 2D or 3D scenes, developers and content creators could benefit from an introduction to how the brain extracts features and objects from the neural stream of signalling, which is what we will cover in this section.

First, we describe how we performed our literature search in Section 3.1. Second, we present the state of the art literature on object perception (Section 3.2), third moving on to perception of object features (Section 3.3). Fourth, we look at the integration of these two topics and present interaction with objects in Section 3.4. Fifth, we present the neuroscience literature on virtual reality, as well as selected practical applications of virtual and augmented reality in medicine (Section 3.5).

3.1. Method

In order to provide an overview of the neuroscientific research relevant to depth perception in a virtual reality context, we performed a literature search focused on: depth perception, 3D object interaction, and virtual/augmented reality (see Figure 5). We conducted

our search in PubMed and Google Scholar under the umbrella terms "depth perception", "neuroimaging" and "virtual/augmented reality". The literature search from the two databases/search engines were combined and primarily performed to validate our search strategy. We conducted one search for "depth perception" + "neuroimaging" and one for the XR umbrella term. In PubMed, we used query expansion to capture more relevant results.

Afterwards, we combined results with the "AND" operator in PubMed. This process resulted in 1909 hits in the database. After applying automatic restrictions (publication type, year, type of study, language) and assessing the remaining articles for eligibility.

Search Terms:

- (((stereopsis) OR binocular depth perception) OR pseudo chroma depth) AND (((shape perception) OR edge perception) OR feature combination) AND (((((Neuroimaging) OR Brain imaging) OR fMRI) OR fNIRS) OR EEG) OR ERP)
- (((Virtual reality) OR Augmented reality) OR Mixed Reality) OR XR) AND (((Stereopsis) OR Depth perception) OR Visual processing)

Manual Selection Criteria:

- Title or abstract contains XR terms in combination with vision
- Title or abstract contains information on integration of cortical processing of depth information
- If the article is an experimental study, it should be focused on healthy participants
- The article should be a journal publication
- The article should be a review or an experimental study.

After manual screening of the remaining articles from our search, we included 64 articles in our qualitative synthesis.

Our survey focuses on neuroscience literature primarily and is therefore not intended to be a comprehensive review of all biological and medical applications of XR. An overview summary of all surveyed papers is available in in Table 1.

Table 1: Surveyed paper overview: *Exp:* experiment, *Clin:* clinical study, *Rev:* review, *RCT:* randomized controlled trial, *Dev:* application.

References	Type	Depth Cues		Participants n	Display Medium			Imaging Method		
		Stereo	Mono		XR	2D screen	3D Glasses	fMRI	EEG	MEG
[EYG*17]	Exp	x	x	25(14f)		x	x			
[WDC*06]	Exp	x		5			x	x		
[PKW11]	Exp	x	x	16, 7c		x	x	x		
[ZHF*19]	Exp	x		16 (5f)	VR		x			
[RSKP18]	Exp	x	x	31(26f)	VR					
[SGTO10]	Clin	x	x	46, 25c		x	x			
[WYZ*10]	Exp	x		13(7f)			x	x		
[CMN12]	Exp	x	x	21(7f)			x		x	
[KK00]	Exp		x	11		x		x		
[CKA*16]	Exp	x		10(4f)		x	x		x	x
[BP11]	Exp		x	8(0f)		x			x	
[AMI*19]	Exp	x	x	10(5f)			Real object			
[ACKMS11]	Exp	x	x	11		x				
[SDLB*01]	Exp	x	x	12(9f)		x	x	x		
[LRO*16]	Exp	x	x	20(10f)		x			x	
[DR17]	Exp	x	x	17(6f)		x	x		x	
[JLL*16]	Exp	x	x	30(22f)		x	x			
[DAB*13]	Exp	x	x	12(0f)		x	x	x		
[PDRLO6]	Exp	x	x	14(7f)		x	x		x	
[FHRB11]	Exp	x	x	NaN		x				
[KM01]	Exp	x	x	12(7f)		x	x		x	
[DDD*19]	Exp	x	x	12(10f)		x	x			
[BFPH01]	Exp	x		8		x	x	x		
[NNI*05]	Exp	x		10(6f)		x	x	x		
[MBW13]	Exp	x	x	15(6f)		x	x	x		
[BVD15]	Exp	x	x	15(10f)	VR					
[YCD*01]	Exp	x	x	17		x	x	x		
[GMB*18]	Exp	x	x	50	VR	x				
[SRJ*15]	Exp	x	x	12(6f)	VR		x		x	
[Roe21]	Exp	x		17		x			x	
[ME01]	Exp		x	NaN		x		x		
[ALM06]	Exp	x		10		x	x			
[HBLK08]	Exp	x	x	16(10f)		x	x	x		
[VKH*16]	Exp	x	x	12	VR					
[KEGB03]	Exp	x	x	10		x	x	x		
[CS11]	Exp	x	x	13(10f)		x	x	x		
[FRR15]	Exp	x	x	9		x	x			
[GDUKM02]	Exp	x	x	9(2f)		x	x	x		
[MCSB14]	Exp	x		6		x	x			
[OCR*18]	Exp	x	x	40(24f)		x	x		x	
[LKP17]	Exp	x	x	27(15f)		x	x	x		
[MCSB15]	Exp	x	x	4		x	x			
[GTPO08]	Exp		x	18(10f)		x		x		
[ABH11]	Exp		x	14(10f)		x				
[LJCFY*06]	Exp	x	x	26(14f)		x	x		x	
[GPK*21]	Exp	x	x	22(12f)		x	x	x		
[PBK*12]	Exp	x	x	36(10f)	VR					
[WYZ*10]	Exp	x	x	19(12f)	MR			x		
[PLOE*15]	Exp	x	x	14(3f)			x	fNIRS		
[FR17]	Exp	x	x	72	VR					
[IHN19]	Exp	x	x	26	VR					
[EBDC*19]	Rev			NaN	XR					
[CABL16a]	Exp	x	x	2	AR					
[dCDRCF*16]	Exp	x	x	20	AR					
[SKB17]	RCT	x	x	34	AR					
[XSY*12]	Exp	x	x	2	VR					
[CXW*06]	Dev	x		NaN	AR					
[CMC*12]	Dev	x		NaN	AR					
[KK01b]	Exp		x	8		x		x		
[MT06]	Exp	x	x	6		x	x			
[NSMM09]	Rev			NaN	AR					
[PBLR12]	Dev	x	x	NaN	AR	x				
[SGHMS*15]	Dev	x	x	NaN	AR					
[SA12]	Exp	x	x	70	VR					

3.2. Object Perception

Visual perception is a hierarchical process building from the receptive field of individual photoreceptors to increasingly complex features in the visual cortex, as already described in Section 2. Here we focus on the higher cognitive processes of combining features into sub-shapes and shapes, which enables recognition of multiple shapes as an object. In their fMRI study on depth perception, Welchman et al. demonstrated that the retinotopic areas V1, V2 and V3 detect changes in visual cues, while the higher cortical areas MT/V5 (dorsal stream) and LOC (ventral stream) are sensitive to changes in global 3D shape [WDC*06]. Based on a fMRI adaptation paradigm, studies exploring this 3D structure sensitivity found that the LOC did not adapt to 3D structure changes, but instead displayed adaptation to 2D structure changes [KK00, KK01b, KEGB03, ME01]. This lack of adaptation to 3D objects displays that LOC operates with a volumetric (3D) representation of objects. In addition, a fMRI priming paradigm also revealed LOC activation to 3D compared with 2D stimuli [ME01]. Within LOC, heterogenic adaptation across multiple subregions indicates the presence of different neural populations that could support processing of separate object features. Differentiating between shapes within a larger structure is essential to perceiving the structure accurately. The anterior cortical area with LOC contributes to this process with sensitivity to convex shapes as well as differentiating between convex shapes and non-convex shapes [KEGB03]. The results from several fMRI studies exploring shape processing indicate that the main regions within LOC differentiate between the processing of curved and pointy shapes in objects, primarily favouring curved shapes [ABH11, HBLK08]. In addition, Haushofer et al. found increased activation in LOC for shapes with more curvature and non parallel contours compared with "simpler" objects. This could indicate a visual preference for complex objects. By applying a blocked fMRI experimental design, presenting RDS followed by rest for an extended period, Gilaie-Dotan et al. discovered that the posterior fusiform sulcus (pFs) region within the LOC is specifically sensitive to shapes derived from stereopsis [GDUKM02]. Furthermore, discrimination of fine depth gradients appears to be important when inspecting complex shapes. Neuroanatomically, this is done by the areas V3A, V4 and V6 [CS11, CMN12]. It has also been reported that competing cues for assignment of object edges (such as those visible in the Rubin vase in Figure 6) delay processing time, but they have no following effect on object representation [BP11].

Further, occipital and inferior temporal regions elicit a BOLD response to 3D shape from stereopsis, predominantly in dorsal areas [GPK*21]. The validity of this fMRI finding was strengthened by two control experiments accounting for attention and fixation effects. Object perception occurs in both visual streams, is dependent on assignment of object edges and figure ground segmentation. Perception of whole objects is reliant on the accuracy of perception of their individual features, which are the focus of the next subsection.

3.3. Feature Detection

Swift detection, and grouping of object features is essential in figure ground segmentation and perceptual organization. The temporal processing of monocular and binocular stimuli displays differ-



Figure 6: An example of the "Rubin Vase". Competing cues for assignment of object edges create two possible perceptions derived from the same stimuli.

ent patterns of event related potentials (ERP) (N1 and N2/P3), indicating the involvement of separate neural populations or pathways in their processing [OCR*18, PDRLO6]. Visual task performance indicates that the binocular condition is beneficial for object recognition. Early (P100) ERP components reflect processes that offset binocular input against each other, while later components reflected binocular rivalry (N170) [Roe21]. Binocular rivalry reflects a "competition" between different interpretations of stimuli. This could also be connected with complex structures require more processing. This is seen in the ERP response to 3D shapes, which increases proportionally with object shape complexity [CKA*16, LRO*16]. It is further reported that differences in local object shape are detected around 150ms, and could be interconnected with binocular rivalry (N170). Later, if relative disparity is present in the stimuli, an increase in cognitive processing around 300ms after stimuli onset is observed (N300 ERP component).

Congruent monocular and binocular depth cues reduce ERP signature possibly due to ease of stimuli integration; neuronal specificity is high around 90° and 180° degrees [DDD*19, LKP17]. This could reflect the number of edge detectors available for each orientation degree.

An EEG experiment assessing "cognitive load" from 2D and 3D stimuli found increased cognitive load during the 2D condition [DR17]. In general, challenging tasks in the experiment resulted in higher cognitive load in both conditions, these two observations combined could support the notion that 3D stimuli supports processing of visual information in a more effective way than 2D stimuli.

Further, the absence of binocular disparity reduces the efficiency of monocular perception of 3D structures from shading and texture [SGTO10]. This interaction illustrates the close relationship between processing of monocular and binocular depth cues, and that they to some extent are interdependent.

Distorted binocular disparity leads to reduced depth perception [PBK*12]. In educational or simulation multi-user virtual environments, a leader or instructor is tasked with guiding a group of participants. This could result in improper binocular cues for the participants, since they are not located at the same place as the instructor.

In face recognition, features are typically neglected in favor of holistic processing. However, if a face is upside down (inverted) the processing is more reliant on recognition of individual features. The face inversion effect is established in 2D, and also found similar in 3D, but holistic processing of faces benefit from a 3D presentation [EYG*17]. In a VR paradigm, participants were exposed to brief images of objects in a naturalistic settings [RSKP18]. The binocular stimuli in VR did not provide any large general increase in reaction time, or recall, but provided a significant benefit at the limit of perceptual capacity (5 - 7 units, e.g. digits). This indicates that some feature in the binocular cue provides an increased capacity for short term object retention. Further, task difficulty increased the difference between 2D and 3D performance. Another visual feature that influences reaction time is retinal object size, which is the area an object projects to on the retina. Reaction time is reduced in both conditions with increasing retinal object size, but the 2D condition displays a sharper reduction [JLL*16].

Feature detection and object perception interact in the dorsal cortical areas, providing support for motor function. **Dorsal cortical** areas V3B/KO, V7 and IPS are sensitive to differences in volumetric structure. An adaptation fMRI paradigm found an interaction of monocular (texture and shading) and binocular cues in the adaptation aftereffect [PKW11]. Through a random dot stereogram (RDS) fMRI experiment, Murphy et al. found activation in V3B/KO mainly for representation of depth from texture and disparity [MBW13]. The authors concluded that V3B/KO utilizes this multi-cue input to create a single accurate representation of depth. If the texture and binocular cues are congruent, the depth combination results in a functional activity increase in V3B/KO. A similar adaptation paradigm was employed by Sun et al., finding V3B/KO involved in both monocular and binocular processing of the visual cue surface reflectance (gloss) [SDLB*16].

The stimuli integration in the dorsal stream relays depth information for action, and generally prioritises speed over accuracy when constructing a representation of the environment [KM01]. fMRI investigation of the effect of stereopsis through Random dot stereograms(RDS) displays activity in dorsal areas V3A, V7 and MT/V5, the areas are sensitive to different degrees of disparity [WYZ*10].

Ambiguous shapes can be utilized to study depth perception, while relative depth can be utilized to influence if people see an ambiguous shape in a constant or changing way (shape constancy) [ACKMS11]. Binocular disparity in IPS is sensitive to change in orientation, depth and discriminates between shapes [NNI*05]. Cortical areas that are not specialised in depth processing like the fusiform face area still elicit a BOLD (fMRI) response to changes in size, position and rotation in depth of face-like stimuli [YCD*01].

Functional connectivity between the dorsal and ventral streams

is increased when processing stimuli far away. This could indicate that the dorsal stream has more time and needs information from ventral areas to make a good assessment of the object features [WYZ*10]. As previously noted, object distance to the user influences feature processing, in particular in object interaction.

3.4. Object Interaction

Efficient utilization of digital tools is reliant on human perception of the environment and interaction with the digital content. This section considers the latter, how humans can influence the digital environment and objects in it. The primary way humans interact with their environment is through grasping and hand-movements. Bozzacchi et al. leveraged fMRI to investigate the effect of depth perception on grasping movements [BD15]. They found grasping in the 2D condition limited by missing depth cues. In addition, similar biases were observed between the VR condition and real world condition. This could be due to an almost 10x increase in depth accuracy when stereopsis is included [MT10]. A comparison of path movement and precision between handheld 3D controllers and mouse and keyboard showcased that participant movements are smoother and more precise in the 3D VR condition [GMB*18].

Moving objects on collision course with the observer are often mistakenly perceived as missing, when the virtual object would have collided with the observer [ALM06]. This angular misperception increases with object speed. In addition, direction of motion in depth is systematically misperceived when sensory noise is introduced [FRR15]. These findings represent possible challenges as interaction with objects is dependent on correct perception of the object location. Interaction with objects relies on the distance to the object. This distance is reflected in crossed and uncrossed disparity, and in turn indicates where the visual stimuli fall on the retina. An fMRI experiment by Li et al. revealed that discrimination between crossed and uncrossed disparity activates some of the same cortical areas as stereopsis, V3A and LO in LOC [LZH*17]. Interaction with objects may be more efficient when binocular disparity is neither crossed nor uncrossed (this occurs at a point called the horopter) [BFPH01]. Processing of depth in the posterior parietal cortex does not change when comparing object interaction in 2D and 3D environments, but hand eye coordination is superior in the 3D condition [PLOE*15].

Lateral occipital cortical sites initially elicit separate responses to form and depth features (N150) and afterwards display a combination of the features [KM01]. This process is slightly different for crossed and uncrossed disparity across hemispheres. Crossed and uncrossed disparity could interact with reaching and processing of objects within reaching distance. Both reaching and perception of distance is central when interacting with objects in XR.

3.5. Virtual Reality and Augmented Reality

The introduction of stereopsis is the main visual cue added in HMD XR environments which is unavailable on a regular 2D computer monitors. Head movement and body position are two additional essential cues contributing to immersion and a realistic experience.

However, vergence is still missing in today's hardware. Vergence

is the simultaneous movement of both eyes to focus on an object in a specific depth plane. The absence of realistic vergence in VR is thought to be one of the main contributors to visual fatigue from head mounted displays (HMD) [IHN19]. Introducing defocus blur in XR HMDs is an attempt at reducing visual fatigue, this mimics the distribution of clarity in the human visual field by blurring the off-centre regions of the image [MCSB14]. Defocus blur significantly reduces visual fatigue from wearing HMDs. Furthermore, Maiello et al. tested if the reduced load of stimuli from including defocus blur might improve depth perception but found no indication of enhanced depth acuity [MCSB15].

As mentioned, stereopsis induces an enhanced experience of presence, and in addition, improves performance on spatial navigation tasks (retrieval rate: 85% in 3D vs. 69% in 2D) while increasing cortical activity in motor and visual cortical areas [SRJ*15]. Furthermore, when viewing 3D structures with stereopsis, this influences features of visual brightness perception [AMI*19]. Orientation and spatial frequencies [FHRB11]. In a VR environment our perception of depth requires time to adapt to the new conditions. An extension of this effect is applied in a VR game to recover stereovision in stereo-blind adults [FR17, VKH*16]. This displays depth perception as recoverable in certain cases. An fMRI experiment by Dores et al. found support for the familiar nature of perceiving 3D cues in a paradigm exposing the participants to emotional stimuli [DAB*13]. In this study, the 3D stimuli elicited a greater emotional response than 2D stimuli in the 2D condition.

We have examined how both input and output from the environment and brain may affect each other. Next, we briefly highlight a selection of VR applications in medical practice, education and training.

Virtual reality is currently explored in several application areas that work with virtual 3D objects or require highly accurate interaction, such as anatomy education or endoscopic surgery. Research investigating surgical performance indicates that 2D inspection of medical images hinders critical information from being uncovered, which leads to worse outcomes and more time spent in surgery compared to 3D preparation [SGHMS*15].

In addition to VR applications, AR applications have also been successfully introduced to the biomedical domain. For example, in the diagnosis of congenital heart disease, AR visualization of intracardiac structures improves the precision of doctors' diagnostic performance [XSY*12]. In addition, development of AR methods for surgical navigation and image overlay are thriving [CMC*17, CXW*15, CABL16b]. The inclusion of stereopsis make the surgical navigation more intuitive, and performs within the clinical range of accuracy when tested. Another clinical application of AR is in simulators, which can be used both for training and to prepare for surgeries [dCDRF*16]. A group that performed simulator warm-up before laparoscopic surgery reduced time and improved accuracy measures compared with the non-warm-up group. In a randomized controlled trial, Sorensen et al. tested laparoscopic simulator training, comparing a 2D and 3D condition [SKB17]. The group in the 3D condition spent significantly less time in training to reach a predetermined level, and their skills transferred well to the 2D surgical environment. Another study comparing an AR microscope with 2D and 3D screens found enhanced micro manip-

ulation in the AR condition [PBLR12]. Understanding of the 3D structures seems tightly linked with surgical performance, Selvannder et al. tested this by measuring students surgical performance in training and their stereoacuity [SÅ11]. The study found a positive relationship between stereoacuity and initial surgical training performance. However, stereoacuity's effect on long-term training outcome was not investigated.

One challenge with AR assisted surgery is related to image registration. Organ movement during surgery can make the overlay images inaccurate. However, this could be alleviated through predictive models [NSMM11].

XR may have a promising future in exploration, analysis and presentation of digital objects. Software tailored to work with the visual system could aid researchers, clinicians and industry [EBDC*19, ZHF*19].

4. Discussion

XR holds great potential for interactive visualization of 3D biomedical data. Limitations of displaying such data on a 2D screen with standard input modalities could be remedied by the improved depth perception that XR offers. However, these applications need to be carefully designed in order to maximize benefits while minimizing discomfort. In the following, we briefly summarize the main findings from this survey. We follow up with our interpretation and the implication of these findings. In addition, we highlight several limitations.

4.1. Summary

The aim of this study is to provide an overview of depth perception mechanisms from a neuroscience perspective and their consequences for visualizing digital 3D objects, such as medical images or molecular surface mesh data. Based on our survey of the neuroscience literature, we investigate which benefits 2D and 3D visual environments provide when visualizing digital objects. The visual system processes information from our surrounding environment in a hierarchical manner, from simple features to complex objects placed in a scenic context [Orb11]. Monocular and binocular stimuli influence the accuracy and speed when generating such an environment, with binocular stimuli as key in volumetric and depth judgements [OCR*18]. However, the visual system performs best when both monocular and binocular stimuli are present as the visual cues have a complementary effect during visual processing [GTPO08, GMB*18]. Both visual streams utilize the same stimuli, but for different purposes: spatial localization (dorsal) and physical representation (ventral) [WZM*09]. Research investigating reaction time, selection tasks and visual recognition found advantageous effects in the 3D conditions compared with 2D conditions [JLL*16, PDRL06]. However, the visual recognition task only displayed an advantage for 3D when the task was challenging/pushing the participant to their cognitive/perceptual bandwidth [RSKP18]. Object representation is delayed by competing interpretations of visual cues, but the "winning" object representation takes all input and the final representation is not influenced by the competing processes [BP11].

4.2. Interpretation and Implications

Findings from the neuroscience literature indicate that binocular disparity enhances processing of 3D environments/objects, compared with the efficiency of monocular cues alone. These findings are supported by both fMRI and EEG experiments comparing processing of objects in 2D and 3D environments [KK01b, SRJ*15]. In these studies, the spatial resolution of fMRI aids localization of areas involved in specific tasks, while the temporal resolution in EEG is used to investigate the dynamic propagation of neuronal processes. We found support in the literature for inclusion of both monocular and binocular cues. Task performance is improved and most accurate when monocular and binocular stimuli are congruent [DDD*19, LZH*17]. Viewing 3D objects on a 2D display could lead to increased cue conflict and reduced processing speed. This could be due to binocular and monocular visual cues processing occurring at different levels before integration occurs in higher cortical areas. Based on the literature, we recommend including binocular disparity as a visual stimuli if the goal of an application is optimized visualization of 3D objects. However, depending on application requirements, e.g., designing applications intended for radiology workstations, this may not be possible. In such a case, additional visual cues need to be implemented in order to support depth perception. In addition, immersive environments provide a tangible feeling of presence when the user experiences depth and is enclosed in the environment [ULCV19]. Our interest was mostly in interaction with 3D data, as this is increasingly prevalent due to powerful graphics hardware, 3D scanning, lidar images, and digitalization of medicine [EBDC*19]. While initially computers were mainly used for textual or other 2D information, the current paradigm shift requires new methods for efficient visualization of 3D real-world objects. Increasing perceptual bandwidth is a good reason to choose 3D if working with digital 3D objects [RSKP18].

Our results can support the design of environments for improved visualization of digital volumetric objects, since visualization of digital objects is a central and increasingly important topic in medical education, research, and practice. From our summary of the literature, we hope to support the development of seamlessly integrated volumetric visualizations, which boost the feeling of presence and efficiency in working with digital 3D objects. A natural way for humans to visualize 3D objects is by viewing them in 3D and using 3D movements to grasp and manipulate them.

Our angle for this survey was from a neuroscience point of view. The presented studies highlight that visual perception in general, and object recognition and depth perception in particular, are embedded in a highly complex and hierarchically organised network of different processing hubs that involves several brain areas. Any XR application needs to consider that the human visual system is tuned to navigate in a 3D world but that it also uses specific features for extracting the required information while discarding others. We have briefly discussed which brain areas are processing which piece of information from a visual scene, and we presented studies that have tested some of those hubs specifically. Accordingly, a mismatch between pieces of information from a visual scene that are naturally not co-occurring but could co-occur in an XR world might generate discomfort and confusion and, consequently, a disadvantage of an XR over a 2D application.

4.3. Limitations

A central feature of our survey is our interdisciplinary approach, aiming at building a bridge from neuroscience literature to the visualization community. This set-up necessitated a certain focus regarding the range of topics we investigated. We chose to focus on depth perception, stereopsis, and the effect of depth perception on interaction with digital objects primarily. To accommodate readers interested in a broader approach, we included references to several reviews and focused articles. The broad span of neuroscience; from research on animals and cell cultures to systems neuroscience and theoretical modelling of neural networks required us to limit our search to human studies.

In this survey, we focus primarily on literature from the neuroscience community. While there is a large body of work presenting XR applications developed in a biomedical visualization research context, they are out of scope for this brief literature survey. Summarizing the state of the art in this area would be an interesting opportunity for future work.

The brain is a highly complex organ. As such, there many uncertainties regarding function and processes in the brain we do not fully understand. This includes the visual system, processing, and integration of visual stimuli. In this review we have attempted to summarise the current state of research focused on visual perception and depth processing. However, there is still a lot we do not know about the brain.

5. Recommendations

Based on the literature presented in this survey, we propose the following guidelines for XR designers and developers:

1. **Consider stereoscopic displays for 3D visualization.** 3D object interaction is generally easier in 3D. A 3D visualization and physical simulation of interaction with the object will aid interaction efficiency. 3D visualization on 2D screens can still be efficient with sufficient use of monocular depth cues. However, missing disparity has a detrimental effect on the efficiency of monocular cues. This leads us to recommend providing stereoscopic disparity if volumetric structure is an essential feature in the object of interest [SGTO10]. However, stereo blindness is quite prevalent affecting 4 - 30% of the population, emphasising the need for both good monocular and binocular visual cues [CBL19].
2. **Make sure big objects in the scene are conveying the most relevant information.** Retinal size of objects influences the processing time of the objects, meaning that large objects close to the user's point of view might be distracting [JLL*16].
3. **Consider if 3D would benefit task performance.** Behavioural studies, comparing performance in 2D and 3D conditions found a 5 - 10 times performance increase assessing depth differences between two vertical bars [MT10]. If the application goal is to perform difficult cognitive or visual tasks, stereopsis is recommended as it provides a performance boost at the perceptual limit of the user [RSKP18].

4. **Make sure visual cues are congruent.** The visual system works best with consistent monocular and binocular stimuli. The VR HMD induces a vergence and accommodation conflict. Keeping all other visual cues congruent leads to a smoother visual experience, as incongruent stimuli delays processing and creates competing visual representations [DDD*19, LZH*17]. For example in multi-user XR applications, it can cause confusion if the followers are given input from the leaders point of view.
5. **Place essential objects in the scene a little further away than reaching distance.** Depth perception is most accurate at the "horopter", when the eyes disparity is neither crossed nor uncrossed. This might not be possible in traditional VR environments due to the screen being too close to the eyes (vergence accommodation conflict), but could still indicate an optimal distance for viewing and interacting with objects [BFPH01].

6. Conclusion and Future Work

It is important to consider how humans perceive and interact with the world when designing applications that aim to imitate real world objects and interactions with them. Stereopsis, one of the main benefits with HMD-based XR, provides a large boost to depth perception, which is a key feature if the goal of the application is to interact with 3D objects. In neuroscience, examinations with fMRI paradigms have identified central cortical areas in processing of depth cues, both through priming (3D volume), RDS (pure stereopsis) and adaptation (sensitivity to change in stimuli) paradigms. Findings from such studies can inform application development, but are typically challenging to access for visualization researchers. With this survey, we provide relevant neuroscience background information and a survey of the state of the art in depth perception. Based on the literature, we provide guidelines for XR application design. Most of our recommendations are not controversial, but we hope that an understanding of some of the mechanisms behind them will aid development of XR applications in biology and medicine.

The value of 3D and virtual reality is increasingly appreciated in medical training, education and surgery [CABL16a, EBDC*19]. The depth vision and perceptual benefits of properly designed XR environments are substantial for application in biology and medicine. For example, medical students can practice specific scenarios an unlimited number of times, with no risk for patients involved. Integration of head and hand movement provide a realistic experience, for simulation training and education. Application developers and designers need to carefully consider the implementation of monocular and binocular depth cues to provide an optimized experience. XR also grants the user freedom to interact with digital representations in ways that would not be possible in a real environment, including physically impossible angles, or with visual enhancement. The addition of binocular cues in XR environments has the potential to change interaction with digital objects in a fundamental way. However, vergence is a central visual cue that is still missing. If this could be integrated in future XR hardware implementation, XR could be like stepping in to a new world.

Acknowledgements. This project was funded by NFR and Tenklabs (grant number 310369) and the Trond Mohn Foundation (grant number 811255).

References

- [ABH11] AMIR O., BIEDERMAN I., HAYWORTH K. J.: The neural basis for shape preferences. *Vision Research* 51, 20 (2011), 2198–2206. doi: 10.1016/j.visres.2011.08.015. 8, 9
- [ABWV19] ARMENDARIZ M., BAN H., WELCHMAN A. E., VANDUFFEL W.: Areal differences in depth cue integration between monkey and human. *PLoS biology* 17, 3 (2019), e2006405. doi:10.1371/journal.pbio.2006405. 6
- [ACKMS11] AZNAR-CASANOVA A., KEIL M. S., MORENO M., SUPPER H.: Differential intrinsic bias of the 3D perceptual environment and its role in shape constancy. *Experimental brain research* 215, 1 (2011), 35–43. doi:10.1007/s00221-011-2868-8. 8, 10
- [ALHN16] AÏM F., LONJON G., HANNOUCHE D., NIZARD R.: Effectiveness of virtual reality training in orthopaedic surgery. *Arthroscopy: the journal of arthroscopic & related surgery* 32, 1 (2016), 224–232. doi:10.1016/j.arthro.2015.07.023. 2
- [ALM06] AGUADO B., LOPEZ-MOLINER J.: Perceived speed of motion in depth modulates misjudgements of approaching trajectories consistently with a slow prior. *Vision research* 159 (2019-06), 1–9. doi: 10.1016/j.visres.2019.03.009. 8, 10
- [AMI*19] ARAI T., MASUDA T., IGARASHI Y., OMORI K., AIZAWA Y., MASUDA N.: Depth inversion with a 3D structure influences brightness perception. *PLoS one* 14, 10 (2019), e0224192. doi:10.1371/journal.pone.0224192. 8, 11
- [BD15] BOZZACCHI C., DOMINI F.: Lack of depth constancy for grasping movements in both virtual and real environments. *Journal of neurophysiology* 114, 4 (2015), 2242–2248. doi:10.1152/jn.00350.2015. 10
- [BFPH01] BACKUS B. T., FLEET D. J., PARKER A. J., HEEGER D. J.: Human cortical activity correlates with stereoscopic depth perception. *Journal of neurophysiology* 86, 4 (2001), 2054–2068. doi:10.1152/jn.2001.86.4.2054. 5, 8, 10, 13
- [BHS*21] BOEDECKER C., HUETTL F., SAALFELD P., PASCHOLD M., KNEIST W., BAUMGART J., PREIM B., HANSEN C., LANG H., HUBER T.: Using virtual 3d-models in surgical planning: workflow of an immersive virtual reality application in liver surgery. *Langenbeck's Archives of Surgery* 406, 3 (2021), 911–915. doi:10.1007/s00423-021-02127-7. 2
- [BP11] BROOKS J. L., PALMER S. E.: Cue competition affects temporal dynamics of edge-assignment in human visual cortex. *Journal of cognitive neuroscience* 23, 3 (2011), 631–644. doi:10.1162/jocn.2010.21433. 8, 9, 11
- [BVD15] BOZZACCHI C., VOLCIC R., DOMINI F.: Grasping lacks depth constancy in both virtual and real environments. *Journal of vision* 15, 12 (2015), 1147–1147. doi:10.1152/jn.00350.2015. 8
- [BYK*21] BESANĀĜON L., YNNERMAN A., KEEFE D. F., YU L., ISENBERG T.: The state of the art of spatial interfaces for 3d visualization. *Computer Graphics Forum* 40, 1 (2021), 293–326. doi: 10.1111/cgf.14189. 2
- [BZP97] BAUMANN R., ZWAN R. V. D., PETERHANS E.: Figure-ground segregation at contours: a neural mechanism in the visual cortex of the alert monkey. *European Journal of Neuroscience* 9, 6 (1997), 1290–1303. doi:10.1111/j.1460-9568.1997.tb01484.x. 4
- [CABL16a] CITARDI M. J., AGBETOBA A., BIGCAS J.-L., LUONG A.: Augmented reality for endoscopic sinus surgery with surgical navigation: a cadaver study. *International forum of allergy & rhinology* 6, 5 (2016), 523–528. doi:10.1002/alr.21702. 8, 13

- [CABL16b] CITARDI M. J., AGBETOBA A., BIGCAS J.-L., LUONG A.: Augmented reality for endoscopic sinus surgery with surgical navigation: a cadaver study. In *International forum of allergy & rhinology* (2016), vol. 6, Wiley Online Library, pp. 523–528. doi:10.1002/alr.21702. 11
- [CBL19] CHOPIN A., BAVELIER D., LEVI D. M.: The prevalence and diagnosis of 'stereoblindness' in adults less than 60 years of age: a best evidence synthesis. *Ophthalmic and Physiological Optics* 39, 2 (2019), 66–85. doi:https://doi.org/10.1111/opo.12607. 5, 12
- [CD91] COLBY C. L., DUHAMEL J.-R.: Heterogeneity of extrastriate visual areas and multiple parietal areas in the macaque monkey. *Neuropsychologia* 29, 6 (1991), 517–537. doi:10.1016/0028-3932(91)90008-V. 6
- [CDG93] COLBY C. L., DUHAMEL J.-R., GOLDBERG M. E.: Ventral intraparietal area of the macaque: anatomic location and visual response properties. *Journal of neurophysiology* 69, 3 (1993), 902–914. doi:10.1152/jn.1993.69.3.902. 6
- [CIG11] CHRISTENSEN A., ILG W., GIESE M. A.: Spatiotemporal tuning of the facilitation of biological motion perception by concurrent motor execution. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 31, 9 (2011), 3493–3499. doi:10.1523/jneurosci.4277-10.2011. 4
- [CKA*16] CHO H., KANG M.-K., AHN S., KWON M., YOON K.-J., KIM K., JUN S. C.: Cortical responses and shape complexity of stereoscopic image-a simultaneous EEG/MEG study. *Neurosignals* 24, 1 (2016), 102–112. doi:10.1159/000442617. 8, 9
- [CMC*17] CUTOLO F., MEOLA A., CARBONE M., SINCERI S., CAGNAZZO F., DENARO E., ESPOSITO N., FERRARI M., FERRARI V.: A new head-mounted display-based augmented reality system in neurosurgical oncology: a study on phantom. *Computer assisted surgery* 22, 1 (2017), 39–53. doi:10.1080/24699322.2017.1358400. 11
- [CMC*12] CUTOLO F., MEOLA A., CARBONE M., SINCERI S., CAGNAZZO F., DENARO E., ESPOSITO N., FERRARI M., FERRARI V.: A new head-mounted display-based augmented reality system in neurosurgical oncology: a study on phantom. *Computer assisted surgery (Abingdon, England)* 22, 1 (2017-12), 39–53. doi:10.1080/24699322.2017.1358400. 8
- [CMN12] COTTEREAU B. R., MCKEE S. P., NORCIA A. M.: Bridging the gap: global disparity processing in the human visual cortex. *Journal of neurophysiology* 107, 9 (2012), 2421–2429. doi:10.1152/jn.01051.2011. 8, 9
- [CS11] CARDIN V., SMITH A. T.: Sensitivity of human visual cortical area v6 to stereoscopic depth gradients associated with self-motion. *Journal of neurophysiology* 106, 3 (2011), 1240–1249. doi:10.1152/jn.01120.2010. 8, 9
- [CŠR*20] CASSIDY K. C., ŠEFCÍK J., RAGHAV Y., CHANG A., DURRANT J. D.: Proteinvr: Web-based molecular visualization in virtual reality. *PLoS computational biology* 16, 3 (2020), e1007747. doi:10.1371/journal.pcbi.1007747. 2
- [CTS19] COUTANCHE M. N., THOMPSON-SCHILL S. L.: Neural activity in human visual cortex is transformed by learning real world size. *NeuroImage* 186 (2019), 570–576. doi:10.1016/j.neuroimage.2018.11.039. 4
- [CV95] CUTTING J. E., VISHTON P. M.: Perceiving layout and knowing distances: The integration, relative potency, and contextual use of different information about depth*. In *Perception of Space and Motion*, Epstein W., Rogers S., (Eds.), Handbook of Perception and Cognition. Academic Press, 1995, pp. 69–117. doi:10.1016/b978-012240530-3/50005-5. 5
- [CXW*15] CHEN X., XU L., WANG Y., WANG H., WANG F., ZENG X., WANG Q., EGGER J.: Development of a surgical navigation system based on augmented reality using an optical see-through head-mounted display. *Journal of biomedical informatics* 55 (2015), 124–131. doi:10.1016/j.jbi.2015.04.003. 11
- [CXW*06] CHEN X., XU L., WANG Y., WANG H., WANG F., ZENG X., WANG Q., EGGER J.: Development of a surgical navigation system based on augmented reality using an optical see-through head-mounted display. *Journal of biomedical informatics* 55 (2015-06), 124–131. doi:10.1016/j.jbi.2015.04.003. 8
- [DAB*13] DORES A. R., ALMEIDA I., BARBOSA F., CASTELO-BRANCO M., MONTEIRO L., REIS M., DE SOUSA L., CALDAS A. C.: Effects of emotional valence and three-dimensionality of visual stimuli on brain activation: an fMRI study. *NeuroRehabilitation* 33, 4 (2013), 505–512. doi:10.3233/nre-130987. 8, 11
- [DBDU93] DISTLER C., BOUSSAOU D., DESIMONE R., UNGERLEIDER L. G.: Cortical connections of inferior temporal area teo in macaque monkeys. *Journal of Comparative Neurology* 334, 1 (1993), 125–150. doi:10.1002/cne.903340111. 4
- [dCDRCF*16] DA CRUZ J. A. S., DOS REIS S. T., CUNHA FRATI R. M., DUARTE R. J., NGUYEN H., SROUGI M., PASSEROTTI C. C.: Does warm-up training in a virtual reality simulator improve surgical performance? a prospective randomized analysis. *Journal of surgical education* 73, 6 (2016), 974–978. doi:10.1016/j.jsurg.2016.04.020. 8
- [dCDRF*16] DA CRUZ J. A. S., DOS REIS S. T., FRATI R. M. C., DUARTE R. J., NGUYEN H., SROUGI M., PASSEROTTI C. C.: Does warm-up training in a virtual reality simulator improve surgical performance? a prospective randomized analysis. *Journal of surgical education* 73, 6 (2016), 974–978. doi:10.1016/j.jsurg.2016.04.020. 11
- [DDD*19] DILLON M. R., DUYCK M., DEHAENE S., AMALRIC M., IZARD V.: Geometric categories in cognition. *Journal of experimental psychology. Human perception and performance* 45, 9 (2019), 1236–1247. doi:10.1037/xhp0000663. 8, 9, 12, 13
- [DR17] DAN A., REINER M.: EEG-based cognitive load of processing events in 3d virtual worlds is lower than processing events in 2D displays. *International Journal of Psychophysiology : Official Journal of the International Organization of Psychophysiology* 122 (2017), 75–84. doi:10.1016/j.ijpsycho.2016.08.013. 8, 9
- [DW61] DANIEL P., WHITTERIDGE D.: The representation of the visual field on the cerebral cortex in monkeys. *The Journal of physiology* 159, 2 (1961), 203–221. doi:10.1113/jphysiol.1961.sp006803. 3
- [EBDC*19] EL BEHEIRY M., DOUTRELIGNE S., CAPORAL C., OSTERTAG C., DAHAN M., MASSON J.-B.: Virtual reality: Beyond visualization. *Journal of molecular biology* 431, 7 (2019), 1315–1321. doi:10.1016/j.jmb.2019.01.033. 8, 11, 12, 13
- [Eks15] EKSTROM A. D.: Why vision is important to how we navigate. *Hippocampus* 25, 6 (2015), 731–735. doi:10.1002/hipo.22449. 2
- [Eng01] ENGEL S.: Retinotopic organization in human visual cortex and the spatial precision of functional MRI. *Cerebral Cortex* 7, 2 (1997-03-01), 181–192. doi:10.1093/cercor/7.2.181. 4
- [ESS02] EHRENSTEIN W. H., SPILLMANN L., SARRIS V.: Gestalt issues in modern neuroscience. *Axiomathes* 13, 3 (2002), 433–458. doi:10.1023/B:AXIO.0000007203.44686.aa. 5
- [EYG*17] ENG Z. H. D., YICK Y. Y., GUO Y., XU H., REINER M., CHAM T. J., CHEN S. H. A.: 3D faces are recognized more accurately and faster than 2D faces, but with similar inversion effects. *Vision research* 138 (2017), 78–85. doi:10.1016/j.visres.2017.06.004. 8, 10
- [FHRB11] FLEMING R. W., HOLTMANN-RICE D., BULTHOFF H. H.: Estimation of 3D shape from image orientations. *Proceedings of the National Academy of Sciences of the United States of America* 108, 51 (2011), 20438–20443. doi:10.1073/pnas.1114619109. 8, 11
- [FMKH05] FANG F., MURRAY S. O., KERSTEN D., HE S.: Orientation-tuned fMRI adaptation in human visual cortex. *Journal of Neurophysiology* 94, 6 (2005), 4188–4195. doi:10.1152/jn.00378.2005. 6

- [FR17] FULVIO J. M., ROKERS B.: Use of cues in virtual reality depends on visual feedback. *Scientific reports* 7, 1 (2017), 16009. doi:10.1038/s41598-017-16161-3. 8, 11
- [FRR15] FULVIO J. M., ROSEN M. L., ROKERS B.: Sensory uncertainty leads to systematic misperception of the direction of motion in depth. *Attention, perception & psychophysics* 77, 5 (2015), 1685–1696. doi:10.3758/s13414-015-0881-x. 8, 10
- [FS92] FINKEL L. H., SAJDA P.: Object discrimination based on depth-from-occlusion. *Neural Computation* 4, 6 (1992), 901–921. doi:10.1162/neco.1992.4.6.901. 4
- [GALG*18] GORBANEV I., AGUDELO-LONDOÑO S., GONZÁLEZ R. A., CORTES A., POMARES A., DELGADILLO V., YEPES F. J., MUÑOZ Ó.: A systematic review of serious games in medical education: quality of evidence and pedagogical strategy. *Medical education online* 23, 1 (2018), 1438718. doi:10.1080/10872981.2018.1438718. 2
- [GDUKM02] GILAIIE-DOTAN S., ULLMAN S., KUSHNIR T., MALACH R.: Shape-selective stereo processing in human object-related visual areas. *Human brain mapping* 15, 2 (2002), 67–79. doi:10.1002/hbm.10008. 8, 9
- [GF03] GALLETTI C., FATTORI P.: Neuronal mechanisms for detection of motion in the field of view. *Neuropsychologia* 41, 13 (2003), 1717–1727. doi:10.1016/s0028-3932(03)00174-x. 6
- [GF18] GALLETTI C., FATTORI P.: The dorsal visual stream revisited: stable circuits or dynamic pathways? *Cortex* 98 (2018), 203–217. doi:10.1016/j.cortex.2017.01.009. 4, 6
- [GGS81] GATTASS R., GROSS C., SANDELL J.: Visual topography of v2 in the macaque. *Journal of Comparative Neurology* 201, 4 (1981), 519–539. doi:10.1002/cne.902010405. 3
- [GGV*06] GANEL T., GONZALEZ C. L., VALYEAR K. F., CULHAM J. C., GOODALE M. A., KÖHLER S.: The relationship between fmri adaptation and repetition priming. *Neuroimage* 32, 3 (2006), 1432–1440. doi:10.1016/j.neuroimage.2006.05.039. 6
- [GM92] GOODALE M. A., MILNER A. D.: Separate visual pathways for perception and action. *Trends in Neurosciences* 15, 1 (1992), 20–25. doi:10.1016/0166-2236(92)90344-8. 5
- [GMB*18] GERIG N., MAYO J., BAUR K., WITTMANN F., RIENER R., WOLF P.: Missing depth cues in virtual reality limit performance and quality of three dimensional reaching movements. *PLoS one* 13, 1 (2018), e0189275. doi:10.1371/journal.pone.0189275. 8, 10, 11
- [GPK*21] GEORGIEVA S., PEETERS R., KOLSTER H., TODD J. T., ORBAN G. A.: The processing of three-dimensional shape from disparity in the human brain. *Journal of Neuroscience* 29, 3 (2009-01-21), 727–742. doi:10.1523/jneurosci.4753-08.2009. 8, 9
- [GSG88] GATTASS R., SOUSA A., GROSS C.: Visuotopic organization and extent of v3 and v4 of the macaque. *Journal of neuroscience* 8, 6 (1988), 1831–1845. doi:10.1523/JNEUROSCI.08-06-01831.1988. 3
- [GTPO08] GEORGIEVA, TODD J. T., PEETERS R., ORBAN G. A.: The extraction of 3D shape from texture and shading in the human brain. *Cerebral cortex (New York, N.Y. : 1991)* 18, 10 (2008), 2416–2438. doi:10.1093/cercor/bhn002. 5, 8, 11
- [HB41] HOLWAY A. H., BORING E. G.: Determinants of apparent visual size with distance variant. *The American Journal of Psychology* 54, 1 (1941), 21–37. doi:10.2307/1417790. 4
- [HBLK08] HAUSHOFER J., BAKER C. I., LIVINGSTONE M. S., KANWISHER N.: Privileged coding of convex shapes in human object-selective cortex. *Journal of neurophysiology* 100, 2 (2008-08), 753–762. doi:10.1152/jn.90310.2008. 8, 9
- [HCB12] HELD R. T., COOPER E. A., BANKS M. S.: Blur and disparity are complementary cues to depth. *Current Biology* 22, 5 (2012), 426–431. doi:10.1016/j.cub.2012.01.033. 4
- [HFMK08] HEGDÉ J., FANG F., MURRAY S. O., KERSTEN D.: Preferential responses to occluded objects in the human visual cortex. *Journal of Vision* 8, 4 (2008), 16–16. doi:10.1167/8.4.16. 4
- [Hik97] HIKOSAKA K.: Responsiveness of neurons in the posterior inferotemporal cortex to visual patterns in the macaque monkey. *Behavioural brain research* 89, 1-2 (1997), 275–283. doi:10.1016/s0166-4328(97)00073-9. 6
- [HLD*98] HADJIKHANI N., LIU A. K., DALE A. M., CAVANAGH P., TOOTELL R. B.: Retinotopy and color sensitivity in human visual cortical area v8. *Nature neuroscience* 1, 3 (1998), 235–241. doi:10.1038/681. 4
- [HPVN06] HOU C., PETTET M. W., VILDAVSKI V. Y., NORCIA A. M.: Neural correlates of shape-from-shading. *Vision research* 46, 6 (2006), 1080–1090. doi:10.1016/j.visres.2005.10.017. 5
- [HW59] HUBEL D. H., WIESEL T. N.: Receptive fields of single neurons in the cat's striate cortex. *The Journal of physiology* 148, 3 (1959), 574–591. doi:10.1113/jphysiol.1959.sp006308. 3
- [HW68] HUBEL D. H., WIESEL T. N.: Receptive fields and functional architecture of monkey striate cortex. *The Journal of physiology* 195, 1 (1968), 215–243. doi:10.1113/jphysiol.1968.sp008455. 3, 4
- [IHN19] ISKANDER J., HOSSNY M., NAHAVANDI S.: Using biomechanics to investigate the effect of VR on eye vergence system. *Applied ergonomics* 81 (2019), 102883. doi:10.1016/j.apergo.2019.102883. 8, 11
- [JH20] JAVAID M., HALEEM A.: Virtual reality applications toward medical field. *Clinical Epidemiology and Global Health* 8, 2 (2020), 600–605. doi:https://doi.org/10.1016/j.cegh.2019.12.010. 2
- [JLL*16] JIANG Y., LI S., LI Y., ZENG H., CHEN Q.: Effect of the retinal size of a peripheral cue on attentional orienting in two- and three-dimensional worlds. *Attention, perception & psychophysics* 78, 5 (2016), 1285–1292. doi:10.3758/s13414-016-1089-4. 8, 10, 11, 12
- [KDWU00] KASTNER S., DE WEERD P., UNGERLEIDER L. G.: Texture segregation in the human visual cortex: A functional MRI study. *Journal of Neurophysiology* 83, 4 (2000), 2453–2457. doi:10.1152/jn.2000.83.4.2453. 4
- [KEGB03] KOURTZI Z., ERB M., GRODD W., BULTHOFF H. H.: Representation of the perceived 3-d object shape in the human lateral occipital complex. *Cerebral cortex (New York, N.Y. : 1991)* 13, 9 (2003), 911–920. doi:10.1093/cercor/13.9.911. 8, 9
- [Kel62] KELLY D.: Information capacity of a single retinal channel. *IRE Transactions on Information Theory* 8, 3 (1962), 221–226. doi:10.1109/TIT.1962.1057716. 2
- [KK00] KOURTZI Z., KANWISHER N.: Cortical regions involved in perceiving object shape. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 20, 9 (2000), 3310–3318. doi:https://doi.org/10.1523/JNEUROSCI.20-09-03310.2000. 8, 9
- [KK01a] KINOSHITA M., KOMATSU H.: Neural representation of the luminance and brightness of a uniform surface in the macaque primary visual cortex. *Journal of Neurophysiology* 86, 5 (2001), 2559–2570. doi:10.1152/jn.2001.86.5.2559. 4
- [KK01b] KOURTZI Z., KANWISHER N.: Representation of perceived object shape by the human lateral occipital complex. *Science (New York, N.Y.)* 293, 5534 (2001), 1506–1509. doi:10.1126/science.1061133. 8, 9, 12
- [KM01] KASAI T., MOROTOMI T.: Event-related brain potentials during selective attention to depth and form in global stereopsis. *Vision research* 41, 10 (2001), 1379–1388. doi:10.1016/s0042-6989(01)00067-0. 8, 10
- [KR92] KLEFFNER D. A., RAMACHANDRAN V. S.: On the perception of shape from shading. *Perception & Psychophysics* 52, 1 (1992), 18–36. doi:10.3758/BF03206757. 4

- [KSB*13] KRAVITZ D. J., SALEEM K. S., BAKER C. I., UNGERLEIDER L. G., MISHKIN M.: The ventral visual pathway: an expanded neural framework for the processing of object quality. *Trends in cognitive sciences* 17, 1 (2013), 26–49. doi:10.1016/j.tics.2012.10.011. 4, 6
- [LJCFY*06] LI H., JIA H., CHUNG-FAT-YIM A., JIN L., YU D.: The neural correlates of vertical disparity gradient and cue conflict in panum's limiting case. *Brain and cognition* 114 (2017-06), 11–19. doi:10.1016/j.bandc.2017.02.005. 8
- [LKP17] LI S., KAY S., PORTER S.: A 3D assessment and feedback tool for ankylosing spondylitis from the perspective of healthcare professionals. *Informatics for health & social care* 42, 3 (2017), 274–289. doi:10.1080/17538157.2016.1252375. 8, 9
- [Log99] LOGOTHETIS N. K.: Vision: A window on consciousness. *Scientific American* 281, 5 (1999), 68–75. URL: <http://www.jstor.org/stable/26058485>. 4
- [LRO*16] LEEK E. C., ROBERTS M., OLIVER Z. J., CRISTINO F., PEGNA A. J.: Early differential sensitivity of evoked-potentials to local and global shape during the perception of three-dimensional objects. *Neuropsychologia* 89 (2016), 495–509. doi:10.1016/j.neuropsychologia.2016.07.006. 8, 9
- [LRRS99] LAMME V. A., RODRIGUEZ-RODRIGUEZ V., SPEKREIJSE H.: Separate processing dynamics for texture elements, boundaries and surfaces in primary visual cortex of the macaque monkey. *Cerebral Cortex* 9, 4 (1999), 406–413. doi:10.1093/cercor/9.4.406. 4
- [LZH*17] LI Y., ZHANG C., HOU C., YAO L., ZHANG J., LONG Z.: Stereoscopic processing of crossed and uncrossed disparities in the human visual cortex. *BMC neuroscience* 18, 1 (2017), 80. doi:10.1186/s12868-017-0395-7. 10, 12, 13
- [Mat97] MATHER G.: The use of image blur as a depth cue. *Perception* 26, 9 (1997), 1147–1158. doi:10.1068/p261147. 4
- [Mat15] MATHUR A. S.: Low cost virtual reality for medical training. In *IEEE Virtual Reality (VR)* (2015), pp. 345–346. doi:10.1109/VR.2015.7223437. 2
- [MBK06] MURRAY S. O., BOYACI H., KERSTEN D.: The representation of perceived angular size in human primary visual cortex. *Nature Neuroscience* 9, 3 (2006), 429. doi:10.1038/nn1641. 4
- [MBW13] MURPHY A. P., BAN H., WELCHMAN A. E.: Integration of texture and disparity cues to surface slant in dorsal visual cortex. *Journal of Neurophysiology* 110, 1 (2013), 190–203. 8, 10
- [MCSB14] MAIELLO G., CHESSA M., SOLARI F., BEX P. J.: Simulated disparity and peripheral blur interact during binocular fusion. *Journal of vision* 14, 8 (2014), 13. doi:10.1167/14.8.13. 8, 11
- [MCSB15] MAIELLO G., CHESSA M., SOLARI F., BEX P. J.: The (in)effectiveness of simulated blur for depth perception in naturalistic images. *PloS one* 10, 10 (2015), e0140230. doi:10.1371/journal.pone.0140230. 8, 11
- [MDS17] MUNAFO J., DIEDRICK M., STOFFREGEN T. A.: The virtual reality head-mounted display Oculus Rift induces motion sickness and is sexist in its effects. *Experimental Brain Research* 235, 3 (2017), 889–901. doi:10.1007/s00221-016-4846-7. 2
- [ME01] MOORE C., ENGEL S. A.: Neural response to perception of volume in the lateral occipital complex. *Neuron* 29, 1 (2001), 277–286. doi:10.1016/s0896-6273(01)00197-0. 8, 9
- [MKC*19] MAZUREK J., KIPER P., CIEŚLIK B., RUTKOWSKI S., MEHLICH K., TUROLLA A., SZCZEPAŃSKA-GIERACHA J.: Virtual reality in medicine: a brief overview and future research directions. *Human Movement* 20, 3 (2019), 16–22. doi:10.5114/hm.2019.83529. 1
- [MT10] MCKEE S. P., TAYLOR D. G.: The precision of binocular and monocular depth judgments in natural settings. *Journal of vision* 10, 10 (2010), 5–5. doi:10.1167/10.10.5. 10, 12
- [MT06] MCKEE S. P., TAYLOR D. G.: The precision of binocular and monocular depth judgments in natural settings. *Journal of Vision* 10, 10 (2010-08-06), 5–5. doi:10.1167/10.10.5. 8
- [NNI*05] NAGANUMA T., NOSE I., INOUE K., TAKEMOTO A., KATSUYAMA N., TAIRA M.: Information processing of geometrical features of a surface based on binocular disparity cues: an fMRI study. *Neuroscience research* 51, 2 (2005), 147–155. doi:10.1016/j.neures.2004.10.009. 8, 10
- [NSMM09] NICOLAU S., SOLER L., MUTTER D., MARESCAUX J.: Augmented reality in laparoscopic surgical oncology. *Surgical oncology* 20, 3 (2011-09), 189–201. doi:10.1016/j.suronc.2011.07.002. 8
- [NSMM11] NICOLAU S., SOLER L., MUTTER D., MARESCAUX J.: Augmented reality in laparoscopic surgical oncology. *Surgical oncology* 20, 3 (2011), 189–201. doi:10.1016/j.suronc.2011.07.002. 11
- [OCR*18] OLIVER Z. J., CRISTINO F., ROBERTS M. V., PEGNA A. J., LEEK E. C.: Stereo viewing modulates three-dimensional shape processing during object recognition: A high-density ERP study. *Journal of experimental psychology. Human perception and performance* 44, 4 (2018), 518–534. doi:10.1037/xhp0000444. 8, 9, 11
- [OdBV00] OP DE BEECK H., VOGELS R.: Spatial sensitivity of macaque inferior temporal neurons. *Journal of Comparative Neurology* 426, 4 (2000), 505–518. doi:10.1002/1096-9861(20001030)426:4<505::AID-CNE1>3.0.CO;2-M. 6
- [OPF*03] ORBAN G. A., FIZE D., PEUSKENS H., DENYS K., NELISEN K., SUNAERT S., TODD J., VANDUFFEL W.: Similarities and differences in motion processing between the human and macaque brain: evidence from fMRI. *Neuropsychologia* 41, 13 (2003), 1757–1768. doi:10.1016/s0028-3932(03)00177-5. 6
- [OLKT90] OGAWA S., LEE T.-M., KAY A. R., TANK D. W.: Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *Proceedings of the National Academy of Sciences* 87, 24 (1990), 9868–9872. doi:10.1073/pnas.87.24.9868. 6
- [Orb11] ORBAN G. A.: The extraction of 3d shape in the visual system of human and nonhuman primates. *Annual review of neuroscience* 34 (2011), 361–388. doi:10.1146/annurev-neuro-061010-113819. 11
- [Pal12] PALCZEWSKI K.: Chemistry and biology of vision. *Journal of Biological Chemistry* 287, 3 (2012), 1612–1619. 2
- [Par19] PARKER A. J.: Stereopsis and depth perception, 05 2019. doi:10.1093/acrefore/9780190264086.013.77. 1
- [Par05] PARKER A. J.: Binocular depth perception and the cerebral cortex. *Nature Reviews Neuroscience* 8, 5 (2007-05), 379–391. doi:10.1038/nrn2131. 4, 5, 6
- [PBK*12] POLLOCK B., BURTON M., KELLY J. W., GILBERT S., WINER E.: The right view from the wrong location: depth perception in stereoscopic multi-user virtual environments. *IEEE transactions on visualization and computer graphics* 18, 4 (2012), 581–588. doi:10.1109/tvcg.2012.58. 5, 8, 10
- [PBLR12] PALMA S. R., BECKER B. C., LOBES L. A., RIVIERE C. N.: Comparative evaluation of monocular augmented-reality display for surgical microscopes. In *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (2012), IEEE, pp. 1409–1412. doi:10.1109/EMBC.2012.6346203. 8, 11
- [PCPD*08] PARADIS A. L., CORNILLEAU-PERES V., DROULEZ J., VAN DE MOORTELE P. F., LOBEL E., BERTHOZ A., LE BIHAN D., POLINE J. B.: Visual perception of motion and 3-d structure from motion: an fMRI study. *Cerebral cortex (New York, N.Y. : 1991)* 18, 8 (2000-08), 772–783. doi:10.1093/cercor/10.8.772. 5
- [PCT*05] PEUSKENS H., CLAEYS K. G., TODD J. T., NORMAN J. F., HECKE P. V., ORBAN G. A.: Attention to 3-d shape, 3-d motion, and texture in 3-d structure from motion displays. *Journal of Cognitive Neuroscience* 16, 4 (2004-05), 665–682. doi:10.1162/089892904323057371. 5
- [PDRL06] PEGNA A. J., DARQUE A., ROBERTS M. V., LEEK E. C.:

- Effects of stereoscopic disparity on early ERP components during classification of three-dimensional objects. *Quarterly journal of experimental psychology (2006)* 71, 6 (2018-06), 1419–1430. doi:10.1080/17470218.2017.1333129. 8, 9, 11
- [PKW11] PRESTON T. J., KOURTZI Z., WELCHMAN A. E.: Adaptive estimation of three-dimensional structure in the human brain. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 29, 6 (2009-02-11), 1688–1698. doi:10.1523/jneurosci.5021-08.2009. 8, 10
- [PLKW08] PRESTON T. J., LI S., KOURTZI Z., WELCHMAN A. E.: Multivoxel pattern selectivity for perceptually relevant binocular disparities in the human brain. *Journal of Neuroscience* 28, 44 (2008), 11315–11327. doi:10.1523/JNEUROSCI.2728-08.2008. 6
- [PLOE*15] PAGGETTI G., LEFF D. R., ORIHUELA-ESPINA F., MYLONAS G., DARZI A., YANG G.-Z., MENEGAZ G.: The role of the posterior parietal cortex in stereopsis and hand-eye coordination during motor task behaviours. *Cognitive processing* 16, 2 (2015), 177–190. doi:10.1007/s10339-014-0641-1. 8, 10
- [PNL*17] PELARGOS P. E., NAGASAWA D. T., LAGMAN C., TENN S., DEMOS J. V., LEE S. J., BUI T. T., BARNETTE N. E., BHATT N. S., UNG N., ET AL.: Utilizing virtual and augmented reality for educational and clinical enhancements in neurosurgery. *Journal of Clinical Neuroscience* 35 (2017), 1–4. doi:10.1016/j.jocn.2016.09.002. 2
- [Roe21] ROEBER U.: Neural processing of orientation differences between the eyes' images. *Journal of vision* 12, 13 (2012-12-21), 20. doi:10.1167/12.13.20. 8, 9
- [RRP96] ROSSI A. F., RITTENHOUSE C. D., PARADISO M. A.: The representation of brightness in primary visual cortex. *Science* 273, 5278 (1996), 1104–1107. doi:10.1126/science.273.5278.1104. 4
- [RSKP18] RAILO H., SAASTAMOINEN J., KYLMALA S., PELTOLA A.: Binocular disparity can augment the capacity of vision without affecting subjective experience of depth. *Scientific reports* 8, 1 (2018), 15798. doi:10.1038/s41598-018-34137-9. 8, 10, 11, 12
- [SA12] SELVANDER M., ASMAN P.: Stereoacuity and intraocular surgical skill: effect of stereoacuity level on virtual reality intraocular surgical performance. *Journal of cataract and refractive surgery* 37, 12 (2011-12), 2188–2193. doi:10.1016/j.jcrs.2011.05.048. 8
- [SÅ11] SELVANDER M., ÅSMAN P.: Stereoacuity and intraocular surgical skill: effect of stereoacuity level on virtual reality intraocular surgical performance. *Journal of Cataract & Refractive Surgery* 37, 12 (2011), 2188–2193. doi:10.1016/j.jcrs.2011.05.048. 11
- [SBDLW15] SUN H.-C., BAN H., DI LUCA M., WELCHMAN A. E.: fmri evidence for areas that process surface gloss in the human visual cortex. *Vision research* 109 (2015), 149–157. doi:10.1016/j.visres.2014.11.012. 6
- [SDLB*16] SUN H.-C., DI LUCA M., BAN H., MURYY A., FLEMING R. W., WELCHMAN A. E.: Differential processing of binocular and monocular gloss cues in human visual cortex. *Journal of neurophysiology* 115, 6 (2016), 2779–2790. doi:10.1152/jn.00829.2015. 10
- [SDLB*01] SUN H.-C., DI LUCA M., BAN H., MURYY A., FLEMING R. W., WELCHMAN A. E.: Differential processing of binocular and monocular gloss cues in human visual cortex. *Journal of neurophysiology* 115, 6 (2016-06-01), 2779–2790. doi:10.1152/jn.00829.2015. 8
- [SGHMS*15] SANCHEZ-GOMEZ S., HERRERO-SALADO T. F., MAZA-SOLANO J. M., ROPERO-ROMERO F., GONZALEZ-GARCIA J., AMBROSIANI-FERNANDEZ J.: Improved planning of endoscopic sinonasal surgery from 3-dimensional images with osirix(r) and stereolithography. *Acta otorrinolaringologica espanola* 66, 6 (2015), 317–325. doi:10.1016/j.otorri.2014.10.002. 8, 11
- [SGTO10] SAWAMURA H., GILLEBERT C. R., TODD J. T., ORBAN G. A.: Binocular stereo acuity affects monocular three-dimensional shape perception in patients with strabismus. *The British journal of ophthalmology* 102, 10 (2018-10), 1413–1418. doi:10.1136/bjophthalmol-2017-311393. 8, 9, 12
- [SHM*07] SWISHER J. D., HALKO M. A., MERABET L. B., MCMAINS S. A., SOMERS D. C.: Visual topography of human intraparietal sulcus. *Journal of Neuroscience* 27, 20 (2007), 5326–5337. 6
- [SKB17] SØRENSEN S. M. D., KONGE L., BJERRUM F.: 3d vision accelerates laparoscopic proficiency and skills are transferable to 2d conditions: A randomized trial. *The American Journal of Surgery* 214, 1 (2017), 63–68. doi:10.1016/j.amjsurg.2017.03.001. 8, 11
- [SRJ*15] SLOBOUNOV S. M., RAY W., JOHNSON B., SLOBOUNOV E., NEWELL K. M.: Modulation of cortical activity in 2d versus 3d virtual reality environments: an EEG study. *International journal of psychophysiology : official journal of the International Organization of Psychophysiology* 95, 3 (2015), 254–260. doi:10.1016/j.ijpsycho.2014.11.003. 8, 11, 12
- [Ste12] STEUER J.: Defining virtual reality: Dimensions determining telepresence. *Journal of Communication* 42, 4 (1992-12), 73–93. doi:10.1111/j.1460-2466.1992.tb00812.x. 2
- [Tan96] TANAKA K.: Inferotemporal cortex and object vision. *Annual Review of Neuroscience* 19, 1 (1996), 109–139. doi:10.1146/annurev.ne.19.030196.000545. 6
- [TNIT01] TAIRA M., NOSE I., INOUE K., TSUTSUI K.-I.: Cortical areas related to attention to 3d surface structures based on shading: An fMRI study. *NeuroImage* 14, 5 (2001-11-01), 959–966. doi:10.1006/nimg.2001.0895. 4
- [TRK*95] TOOTELL R., REPPAS J. B., KWONG K. K., MALACH R., BORN R. T., BRADY T. J., ROSEN B. R., BELLIVEAU J. W.: Functional analysis of human mt and related visual cortical areas using magnetic resonance imaging. *Journal of Neuroscience* 15, 4 (1995), 3215–3230. doi:doi.org/10.1523/JNEUROSCI.15-04-03215.1995. 4
- [UA15] USREY W. M., ALITTO H. J.: Visual functions of the thalamus. *Annual review of vision science* 1 (2015), 351–371. doi:doi.org/10.1146/annurev-vision-082114-035920. 3
- [UGDG08] UNGERLEIDER L. G., GALKIN T. W., DESIMONE R., GATTASS R.: Cortical connections of area v4 in the macaque. *Cerebral Cortex* 18, 3 (2008), 477–499. doi:doi.org/10.1093/cercor/bhm061. 6
- [ULCV19] UJI M., LINGNAU A., CAVIN I., VISHWANATH D.: Identifying cortical substrates underlying the phenomenology of stereopsis and realism: A pilot fMRI study. *Frontiers in neuroscience* 13 (2019), 646. doi:doi.org/10.3389/fnins.2019.00646. 12
- [UMCA06] URGESI C., MORO V., CANDIDI M., AGLIOTTI S. M.: Mapping implied body actions in the human motor system. *Journal of Neuroscience* 26, 30 (2006), 7942–7949. doi:doi.org/10.1523/JNEUROSCI.1289-06.2006. 1
- [Ung82] UNGERLEIDER L. G.: Two cortical visual systems. *Analysis of visual behavior* (1982), 549–586. 4
- [VJR19] VAUGHAN N., JOHN N., REES N.: Paravr: Paramedic virtual reality training simulator. In *International Conference on Cyberworlds (CW)* (2019), pp. 21–24. doi:10.12968/jpar.2020.12.12.478. 2
- [VKH*16] VEDAMURTHY I., KNILL D. C., HUANG S. J., YUNG A., DING J., KWON O.-S., BAVELIER D., LEVI D. M.: Recovering stereo vision by squashing virtual bugs in a virtual reality environment. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 371, 1697 (2016). doi:10.1098/rstb.2015.0264. 8, 11
- [WALW08] WADE A., AUGATH M., LOGOTHETIS N., WANDELL B.: fmri measurements of color in macaque and human. *Journal of vision* 8, 10 (2008), 6–6. doi:10.1167/8.10.6. 4, 6
- [WBRW02] WADE A. R., BREWER A. A., RIEGER J. W., WANDELL B. A.: Functional measurements of human ventral occipital cortex:

- retinotopy and colour. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* 357, 1424 (2002), 963–973. doi:10.1098/rstb.2002.1108. 4
- [WDC*06] WELCHMAN A. E., DEUBELIUS A., CONRAD V., BÜLTHOFF H. H., KOURTZI Z.: 3D shape perception from combined depth cues in human visual cortex. *Nature Neuroscience* 8, 6 (2005-06), 820–827. doi:10.1038/nn1461. 8, 9
- [WDJGK16] WANG R., DEMARIA JR S., GOLDBERG A., KATZ D.: A systematic review of serious games in training health care professionals. *Simulation in Healthcare* 11, 1 (2016), 41–51. doi:10.1097/SIH.000000000000118. 2
- [WEK*11] WAGEMANS J., ELDER J. H., KUBOVY M., PALMER S. E., PETERSON M. A., SINGH M., VON DER HEYDT R.: A century of gestalt psychology in visual perception i. perceptual grouping and figure-ground organization. *Psychological bulletin* 138, 6 (2012-11), 1172–1217. doi:10.1037/a0029333. 5
- [Wer38] WERTHEIMER M.: Laws of organization in perceptual forms. In *A source book of Gestalt psychology*. Kegan Paul, Trench, Trubner & Company, 1938, pp. 71–88. doi:10.1037/11496-005. 5
- [WFM*93] WATSON J. D., MYERS R., FRACKOWIAK R. S., HAJNAL J. V., WOODS R. P., MAZZIOTTA J. C., SHIPP S., ZEKI S.: Area v5 of the human brain: evidence from a combined study using positron emission tomography and magnetic resonance imaging. *Cerebral cortex* 3, 2 (1993), 79–94. doi:10.1093/cercor/3.2.79. 4
- [WYZ*10] WANG F., YANG W., ZHANG L., GUNDRAN A., ZHU X., LIU J., LI X., BAO S., GAO S.: Brain activation difference evoked by different binocular disparities of stereograms: An fMRI study. *Physica medica : PM : an international journal devoted to the applications of physics to medicine and biology : official journal of the Italian Association of Biomedical Physics (AIFB)* 32, 10 (2016-10), 1308–1313. doi:10.1016/j.ejmp.2016.07.007. 8, 10
- [WZM*09] WANG R., ZHANG M., MENG X., GENG Z., WANG F.-Y.: 3-d tracking for augmented reality using combined region and dense cues in endoscopic surgery. *IEEE journal of biomedical and health informatics* 22, 5 (2018-09), 1540–1551. doi:10.1109/jbhi.2017.2770214. 11
- [XSY*12] XUE H., SUN K., YU J., CHEN B., CHEN G., HONG W., YAO L., WU L.: Three-dimensional echocardiographic virtual endoscopy for the diagnosis of congenital heart disease in children. *The international journal of cardiovascular imaging* 26, 8 (2010-12), 851–859. doi:10.1007/s10554-010-9649-5. 8, 11
- [YCD*01] YUE X., CASSIDY B. S., DEVANEY K. J., HOLT D. J., TOOTELL R. B. H.: Lower-level stimulus features strongly influence responses in the fusiform face area. *Cerebral cortex (New York, N.Y. : 1991)* 21, 1 (2011-01), 35–47. doi:10.1093/cercor/bhq050. 8, 10
- [Zha19] ZHAOPING L.: A new framework for understanding vision from the perspective of the primary visual cortex. *Current Opinion in Neurobiology* 58 (2019), 1–10. doi:10.1016/j.conb.2019.06.001. 2
- [ZHF*19] ZHOU Q., HAGEMANN G., FAFARD D., STAVNESS I., FELS S.: An evaluation of depth and size perception on a spherical fish tank virtual reality display. *IEEE transactions on visualization and computer graphics* 25, 5 (2019), 2040–2049. doi:10.1109/tvcg.2019.2898742. 8, 11