Imperceptible manipulation of lateral camera motion for improved virtual reality applications

ANA SERRANO*, Universidad de Zaragoza, I3A, Max Planck Institute for Informatics
DANIEL MARTIN*, Universidad de Zaragoza, I3A
DIEGO GUTIERREZ, Universidad de Zaragoza, I3A
KAROL MYSZKOWSKI, Max Planck Institute for Informatics
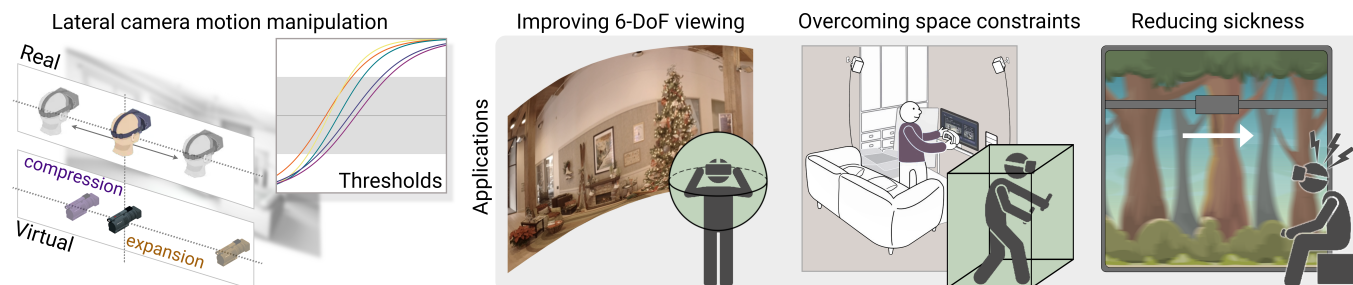BELEN MASIA, Universidad de Zaragoza, I3A

Fig. 1. When viewing VR scenes, lateral motion plays an important role, since it triggers motion parallax and facilitates scene understanding. In this paper, we investigate detection thresholds for the manipulation of lateral camera motion in response to the corresponding head motion in virtual reality, and show how they can be applied to three key open problems in VR. *Left*: We have conducted a psychophysical experiment for measuring translation gain sensitivity for both compression and expansion gains in natural viewing conditions, and we have found that detection thresholds are influenced by scene layout. *Right*: We propose and evaluate three different VR applications, which improve when we apply our resulting imperceptible translation gains: (i) 6-DoF viewing, where our virtual compression reduces visible artifacts, (ii) overcoming space constraints, where our virtual expansion reduces the required physical space to successfully complete a virtual task, and (iii) reducing sickness, where our compression can be used to reduce the amount of vection.

Virtual Reality (VR) systems increase immersion by reproducing users' movements in the real world. However, several works have shown that this real-to-virtual mapping does not need to be precise in order to convey a realistic experience. Being able to alter this mapping has many potential applications, since achieving an accurate real-to-virtual mapping is not always possible due to limitations in the capture or display hardware, or in the physical space available. In this work, we measure detection thresholds for lateral translation gains of virtual camera motion in response to the corresponding head motion under natural viewing, and in the absence of locomotion, so that virtual camera movement can be either compressed or expanded while these manipulations remain undetected. Finally, we propose three applications for our method, addressing three key problems in VR: improving 6-DoF viewing for captured 360° footage, overcoming physical constraints, and reducing simulator sickness. We have further validated our thresholds and evaluated our applications by means of additional user studies confirming that our manipulations remain imperceptible, and showing that (i) compressing virtual camera motion reduces visible artifacts in 6-DoF, hence improving perceived quality, (ii) virtual expansion allows for completion of virtual tasks within a reduced physical space, and (iii) simulator sickness may be alleviated in simple scenarios when our compression method is applied.

CCS Concepts: • **Human-centered computing** → **Virtual reality**.

Additional Key Words and Phrases: Virtual reality, immersive environments, camera manipulation, detection thresholds

1 INTRODUCTION

Users' physical movement in virtual reality (VR) systems increases immersion and can evoke a sense of spatial presence in the virtual world [Slater 2009], while contributing to a more comfortable experience due to visual-vestibular consistency [Koulieris et al. 2019; LaViola 2000; McGill et al. 2017]. Interestingly, large mismatches between such physical movement and its virtual counterpart might remain unnoticed [Nilsson et al. 2018]. This has been recently leveraged for instance in the context of forward locomotion, leading to significant results in redirected walking (RDW) to enable the exploration of virtual worlds larger than the available physical space. However, many VR applications require that the user be seated or standing, with only head and upper body motion available. This is the case of scenarios and games that involve long term use of VR systems, or simply due to limited space. Home setups do usually suffer

---

from severe constraints in physical space; for reference, according to the *Steam Hardware & Software Survey* 2017, 25.3% of SteamVR users use a play area size of just 1m x 1m, and only 1% use an area larger than 3m x 3m. Game designers are aware of this limitation: none of the top-10 VR games in 2019[1] use forward locomotion; these games offer instead teleportation or camera control via controller for larger displacements, and rely on the users' limited motion only for local camera manipulation.

In the absence of forward locomotion, lateral head translation becomes particularly relevant: It is not only a quite natural motion, but it also facilitates scene understanding by triggering motion parallax, which is one of the strongest depth perception cues [Kellnhofer et al. 2016; Nawrot et al. 2014; Steinbach et al. 1991; Yoonessi and Baker 2011]. Moreover, motion parallax in VR systems increases immersion and perceived realism, while reducing discomfort [Anderson et al. 2016; Patney et al. 2018], and contributing to the sense of spatial presence.

In this work, we investigate manipulations of the virtual camera motion in the range of *moderate head translation*. We refer to moderate head translation as the range of natural head movements spawned while exploring content without explicit locomotion (which has been reported to be around ±35 cm [Serrano et al. 2019; Thatte et al. 2016]). Our experiments allow us to obtain *detection thresholds* for *translation gains* (Figure 1, left). In particular, we investigate the range of manipulations that can be carried out before the user notices them, both for compression and expansion of the lateral head motion in VR, in response to the corresponding head motion in the real world. We then show how to apply such thresholds in complex scenes, computing a *translation gain map* which takes into account retinal velocity (the motion of the images projected on the retina) and the viewer's dynamic fixations.

To our knowledge, we are the first to investigate the range of imperceptible translation gains for lateral (left-right) camera motion in response to the corresponding head motion. In particular, we have focused on upper-body motion in the absence of locomotion. We have carefully designed a psychophysical experiment for measuring detection thresholds involving different relative object placements and varied depth coverage, and have discovered a novel dependency of these thresholds with scene depth. Our experiment is different from existing literature investigating real-to-virtual mapping manipulations during forward locomotion, since the different proprioceptive cues (sense of self-movement and body position) might affect the measured thresholds [Patney et al. 2018].

In addition, we leverage our findings in several proof-of-concept applications, addressing three key open problems in VR (Figure 1, right) for which definite solutions do not exist: (i) improving 6-DoF (six degrees of freedom) viewing for captured content, (ii) overcoming physical space constraints, and (iii) reducing simulator sickness. 6-DoF viewing enables motion parallax in 360° content that has been captured from a single viewpoint, providing a more compelling and realistic experience [Hedman and Kopf 2018; Pozo et al. 2019; Serrano et al. 2019]. Our *improving 6-DoF viewing* application allows a scene-adaptive compression of the virtual head translation that minimizes the visible artifacts arising when displaying novel viewpoints

different from the captured ones, thus improving immersion and realism, and contributing to evoking spatial presence. *Overcoming physical space constraints* is crucial when the virtual environment is considerably larger than the available physical space [Bölling et al. 2019; Langbehn et al. 2018; Razzaque et al. 2002; Steinicke et al. 2010], even in the absence of locomotion. We show that lateral motion can be significantly expanded in the virtual space, thus reducing the physical space requirements. Last, our results suggest that applying a compression strategy may help to *reduce simulator sickness* by reducing the illusion of self-motion (vection) resulting from the compound movement of the camera and the user's head. Nevertheless, simulator sickness is a complex phenomenon that can be influenced by many confounding factors, and other forms of motion conflict may arise, requiring further examination. In all cases, we quantify the efficacy of our method through additional subjective experiments, while showing that our manipulations of the virtual camera motion remain unnoticed by the user.

Our contributions can be summarized as follows:

- We measure *detection thresholds* for the manipulation of lateral head motion (both compression and expansion) during natural VR viewing. We further analyze how these thresholds are influenced by scene depth layout.
- We apply our thresholds and insights to arbitrary complex scenes by computing *translation gain maps*, based on the retinal velocity generated by the elements in the scene.
- We show the applicability of our findings in three proof-of-concept applications addressing key open problems in VR: improving 6-DoF viewing of captured content, overcoming physical space constraints, and reducing simulator sickness. Results are further validated by means of additional subjective experiments.

Our data and code are available at *https://ana-serrano.github.io/projects/VR-LateralMotion*.

2 RELATED WORK

The majority of works devoted to camera manipulation thresholds target redirected walking (RDW) [Razzaque et al. 2002] as their main application (see the recent work of Nilsson et al. [2018] for an overview). In this section, we first discuss techniques for continuous camera manipulations, then discuss other works that explicitly exploit the characteristics of the human perceptual system to extend the range of such manipulations.

Continuous manipulations. In these works, subtle camera manipulations are usually applied while the user is *walking* through the virtual environment. When these manipulations are below a certain threshold, the user is not able to detect the redirection. There has been a large body of work devoted to identifying detection thresholds for manipulations of path curvature, rotations, and translations during forward movements. *Path curvature gains* introduce virtual rotations for manipulating (bending) the path trajectory while the user is walking. Steinicke et al. [2010] showed that users can walk on a circular arc while still believing that they are walking straight. Later, Neth et al. [2012] studied velocity-dependent dynamic curvature gains and showed that slower walking speeds lead to lower sensitivities to manipulations, while Langbehn et al. [2017] considered

---

[1]https://www.forbes.com/sites/solrogers/2019/12/20/the-top-ten-vr-games-of-2019/

bent paths for extending curvature detection thresholds. Recently, Bölling et al. [2019] explored adaptation for prolonged exposures to curvature gains, and showed that adaptation allows to apply stronger curvatures. *Rotation* and *translation* gains scale the amount of virtual rotation or translation relative to users' real movement. Steinicke et al. [2010] found that users can be physically rotated 49% more and 20% less than the virtual rotation, and distances can be downscaled by 14% and upscaled by 26%. Jerald et al. [2008] suggested that users are less sensitive to gains applied in the same direction as head rotation. Grechkin et al. [2016] revisited detection thresholds for curvature gains by combining them with translation gains in order to increase redirection effectiveness. Later, Bruder et al. [2011] showed that self-motion illusions can be used to introduce gains by manipulating the users' optic-flow fields. Recently, Matsumoto et al. [2020] studied translation gains in the vertical axis, and found that users are more sensitive to manipulations when crouching than when stretching. Additional efforts have been made to also study the impact of rotation manipulations in different tasks, such as spatial orientation or training transfer [Ragan et al. 2016; Sargunam et al. 2017]. Sun et al. [2016] focused on manipulating the virtual path of a scene to adapt it to the real available space while balancing visual and locomotion fidelity. In the field of telepresence, Nitzsche et al. [2003] studied motion compression for walking in large-scale remote environments, and Zhang et al. [2018] studied detection thresholds for rotation and translation gains in 360° video-based telepresence systems. Interrante et al. [2007] considered overt instead of subtle manipulations, achieving an impressive scaling up to 10x the original movement.

Extended manipulations. Other approaches have explicitly taken advantage of the human visual system (HVS) in order to extend the range of manipulations that can be performed without users noticing. Suma et al. [2011; 2012] proposed to apply redirection by exploiting change blindness. Langbehn et al. [2018] showed that it is possible to introduce additional discrete rotations and translations during eye blinks, while Bolte et al. [2015] exploited saccadic suppression in a similar fashion. Other works have considered the impact of scaling eye height [Williams-Sanders et al. 2019], or the use of distractors: Sun et al. [2018] used image-space modulations for increasing the amount of saccadic redirection they could apply, while Peck et al. [2011] used distractors to guide users away from the boundaries of the physical space. In Table 1 we compare some close works and ours. Different from our work, most of these works target RDW applications, and do not explore the potential dependency with scene layout. As the table shows, we are the first to measure lateral (left-right) thresholds with continuous manipulations. This is a challenging scenario due to the limited proprioceptive cues in the absence of locomotion, and the fact that we allow for natural user interaction (see Section 3.1). The estimated threshold values vary greatly among different studies; this can be potentially explained by different factors, including variations in the task being performed, the axes of the translation, and methodological differences. Finally, we report, for the first time, a dependency of detection thresholds on the scene depth layout.

3 PSYCHOPHYSICAL EXPERIMENTS

We aim to measure detection thresholds for virtual camera translation gains, for the lateral head translations that occur naturally when viewing a VR scene. Similar to previous work [Steinicke et al. 2010], we define a *translation gain* g_T as a mapping between the real head translation T_r and the virtual camera translation T_v, so that $g_T = \frac{T_v}{T_r}$. Values lower than 1 indicate a *compression* of the real motion in the virtual world, while values greater than 1 indicate an *expansion*. Gains for translation movements can be defined on each axis for lateral, vertical, and forward translations. In this experiment, we focus on measuring sensitivity to lateral translation gains.

3.1 Background

To better mimic actual use of VR technology, we carry out our experiments allowing natural head motion under a free-viewing condition. This is a key aspect of our methodology motivated by fundamental research in vision science showing that active and passive observers can perceive three-dimensional structure differently, and that self-motion information is incorporated into visual judgements of 3D structure and distances [Dokka et al. 2011; Panerai et al. 2002; Wexler et al. 2001]. Moreover, recent works have shown that measured thresholds are significantly different when the user is interacting naturally with the environment, compared to rigidly controlled viewing setups (e.g., using chin rests) where the stimuli or the head are artificially displaced. This has been reported for depth perception [Scarfe and Glennerster 2015; Yoonessi and Baker 2011], allocentric position and orientation [Wexler and Van Boxtel 2005], or redirected walking [Schmitz et al. 2018]. To measure our thresholds we follow an experimental procedure that is well-established in the literature [Bölling et al. 2019; Langbehn et al. 2018; Steinicke et al. 2010]. However, allowing for free-viewing implies that the visual input will not be exactly the same for all participants, for example, some accidental rotation is expected to occur. Therefore, we perform a careful analysis of head motion and eye tracking data to ensure that observers are behaving as expected to fulfill the requirements of the experiment. While we include in Section 3.3 the main insights of our experiments, please refer to Section S2 in the supplementary material for the detailed statistical analysis.

In this work, we are interested in the perception of motion with respect to the space, which depends on a combination of different sources of information: retinal velocity, eye movement, head or body movement, and perceived egocentric distance [Swanston and Wade 1988]. Translation gain modifications alter the motion of the virtual camera, therefore, they affect *retinal velocity*, which in turn depends on the relative motion between the observer and the scene [Kim et al. 2016]. Retinal velocity is also conditioned by scene configuration: there is no variation in retinal velocity when fixating at a point on a flat surface, while for curved surfaces retinal velocity changes smoothly with relative motion between scene and observer. However, the most critical case occurs when fixating on a depth discontinuity, or depth edge. In this case, the retinal velocity changes abruptly with relative motion between scene and observer; it is therefore this case where modifications of the translation gain will most likely be detected. Hence, we design our experiment to

Table 1. Comparison of some recent works involving translation gain manipulations. Variations in the estimated thresholds can be due to different factors, including differences in the task being performed, the axes of the translation, and other methodological differences. Some of these works focus on continuously manipulating users' motion, while others introduce discrete translations. For continuous manipulations, negative values correspond to compressions of the movement, and positive ones to expansions; for discrete manipulations, the sign is indicative of the direction of the movement.

	Scene dependency	Translation axes	Measured on free motion	Continuous/discrete manipulations	Unnoticeable manipulations
[Steinicke et al. 2010]	No	Fwd. – Bwd.	Yes	Continuous	Manipulations from -14% to 26%
[Zhang et al. 2018]	No	Fwd.– Bwd.	Yes	Continuous	Manipulations from -5.8% to 9.7%
[Bolte and Lappe 2015]	No	Fwd. – Bwd.	No	Discrete (during saccades)	Translations from -0.63m to 0.53m
[Langbehn et al. 2018]	No	Fwd. – Bwd. Up – Down Left – Right	No	Discrete (during blinks)	(Fwd – Bwd.) Translations from -0.097m to 0.077m (Up – Down) Translations from -0.040m to 0.039m (Left – Right) Translations from -0.039m to 0.052m
[Matsumoto et al. 2020]	No	Up – Down	Yes	Continuous	Manipulations for stretching/crouching from -16%/-17% to 155%/94%
Ours	Yes	Left – Right	Yes	Continuous	Manipulations depending on depth from -58%/-16%/ to 28%/96% (closest/farthest)

analyze this particular case, in order to derive *conservative thresholds* for translation gain.

3.2 Experimental procedure

We conducted a perceptual experiment based on a two-alternative forced-choice (2AFC) task with a method of constant stimuli, i.e., the applied translation gains are not related between trials, but distributed randomly. This experimental procedure is common in RDW research [Bölling et al. 2019; Bolte and Lappe 2015; Langbehn et al. 2018; Steinicke et al. 2010]. In each trial, subjects were presented one stimulus with a certain translation gain applied to the camera, and they had to choose whether the virtual movement was *smaller* or *larger* than their physical movement. Since users are forced to choose one of these two options, unnoticed translation gains will be answered randomly, and therefore they will be correctly answered in 50% of the cases on average. Varying the translation gain results in measured points, to which a psychometric curve will be fitted, modeling detection performance.

Stimuli. Participants viewed a realistic, stereoscopic virtual environment created in Unity, depicting a large room with some furniture to provide depth discontinuities and natural motion parallax cues (see Figure 2). They sat on a rotating stool without wheels, to prevent them from walking. In the virtual world, they appeared in the center of the room, $z_b = 6m$ away from each wall. During each trial, three planes appeared at the same distance, randomly in three out of four predefined locations, between the user and the room walls, parallel to the latter. Behind one of these planes, also chosen randomly, there was a hidden coin. The subjects' task was to find this coin. We introduced this task in order to naturally prompt a small range of head motion while inducing fixation on a depth edge in a natural scenario. We tested seven levels of translation gain $g_T = \{0.4, 0.6, 0.8, 1.0, 1.25, 1.67, 2.5\}$. We sample our translation gain ratios symmetrically so that, for each compression ratio g_T, the
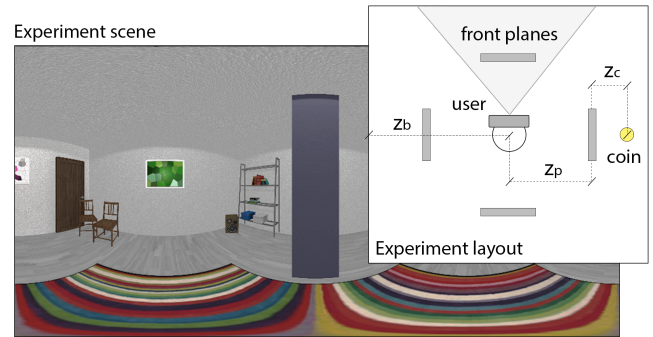

Fig. 2. Experiment scene and schematic representation of the experiment virtual layout. Users are located at the center of a room, three planes are situated at a distance z_p from the center, and the room walls are placed at a distance z_b. One coin is hidden behind one of the three planes at a distance z_c, so that users have to laterally translate their heads to find it.

corresponding expansion ratio is $1/g_T$. A key aspect is that our sampling is such that gains tested involve both sub-threshold and supra-threshold stimuli, thus covering a sufficient range that captures the PSE, as well as the upper and lower detection thresholds. In addition, we varied the distance between the occluding planes and the participants' starting position $z_p = \{1m, 1.75m, 2.5m, 3.25m, 4m\}$, while the coin was always fixed at $z_c = 1m$ behind the plane[2]. When choosing the distances to the occluding plane, we rely on the work of Shibata et al. [2011]. Our stimuli were presented on an HTC Vive Pro, which displays a virtual image at approximately one meter, which determines the eye accommodation state; according to Shibata et al., vergence then remains within the comfort zone for up to

[2]In a previous experiment we found that this distance did not have a significant influence on translation gain detection, therefore we leave it fixed. Refer to Section S1 in the supplementary material for more details.

a virtual object distance of around three meters. While the plane at four meters might be slightly uncomfortable, we explicitly ask about discomfort symptoms at the end of each session (each participant underwent two sessions, as described in the paragraph on *Procedure*): 10% of the users reported some mild discomfort symptoms at the end of the first session, and only 4% of users reported them at the end of the second session. To ensure that the coin could be revealed with moderate head motion, the width of the occluding plane varied as a function of z_p, so it would always require a virtual camera motion of $\pm 35cm$ from the starting point [Serrano et al. 2019; Thatte and Girod 2018] in order to find the coin.

Hardware. The stimuli were presented on an HTC Vive Pro with a nominal field of view of $110°$, with a resolution of 1440×1600 pixels per eye, and a frame rate of 90 fps. To accurately apply our translation gain to the participants' movements, precise head tracking is required. The HTC Vive tracking precision is in the range 0.2mm − 2mm [Groves et al. 2019; Kreylos 2016], well below the just noticeable difference value (JND) for translation detection [Nicolini et al. 2014]. We used three tracking sensors (HTC Vive stations) in a 2m x 2m obstacle-free space to reduce the probability of tracking failure. The experiment was monitored at all times, and no tracking errors were observed by the experimenter nor reported by participants. Additionally, head tracking data was recorded and checked after the experiment to confirm that the real-to-virtual mapping applied for each translation gain was accurate. In particular, we confirmed that participants' side-to-side tracked head amplitude for the different translation gains corresponded to the expected values (see Section S2.3 in the supplementary material for more details). Participants provided their answers with an HTC Vive controller, and the experiment routine was designed and controlled in Unity.

Participants. Thirty participants (11 female and 19 male, average age 21.19 years old, $\sigma = 1.97$) voluntarily took part in the study and provided written consent. They were economically compensated for their time. They were naïve to the purpose of the experiment, and all reported normal or corrected-to-normal vision. Nineteen of them had tried an HMD before, and only three of them used HMDs frequently.

Procedure. We used a full-factorial within-subjects experimental design. The combination of g_T and z_p values yields 35 different conditions. Each condition was repeated five times, so each participant completed $7 \times 5 \times 5 = 175$ trials. These were split in two sessions separated by at least one hour, in order to avoid fatigue; participants were allowed to take additional breaks at any time (only one participant needed to take a break during the first block). All trials were randomized. At the end of each session, participants filled in a questionnaire that included aspects like discomfort or fatigue (Figure 7 in the supplementary). Before the experiment there was a brief introductory session to familiarize participants with VR navigation, and to allow them to adjust the headset (including IPD adjustment dials and straps) until they felt comfortable and their vision was clear. They were shown the same scene as in the trials, and were asked to look around and explore the room, while remaining seated. When they felt ready, the task was explained to them before commencing the actual experiment.

In each trial, three planes would appear randomly at three of the four possible locations, and a coin would appear hidden behind one of the planes. Participants were asked to face the planes by rotating on the stool and then perform a left-right lateral translation with their upper body to search for the occluded coin. Participants were instructed to press the trigger when they found the coin; we set a time limit of 40 seconds per trial for this task, but none of the participants reached this limit. After finding the coin, a screen was presented asking the subjects: *Has your virtual movement been larger or smaller than your real head movement?,* together with the two options to be selected. To choose their answer, users could pre-select their desired option with the controller, which would turn green. They then had to confirm their answer with the trigger. None of the participants reported accidental inputs nor requested corrections. After each trial, subjects were asked to relocate back to a comfortable position in the chair, if needed.

3.3 Results

We collected the following data from each user: head position and orientation at 90 fps, eye tracking data (using a pupil labs HTC Vive add-on eye tracker[3]) also at 90 fps, time taken to complete each task, and the answer to the 2AFC question.

Participants' behavior. Given that our experimental setup features free viewing by design (Section 3.1), we first performed an analysis of head and gaze data to ensure that the participants were behaving as expected during the experiment. In particular, we analyzed the *average head speed* during trials, the *head amplitude* during motion, and the total *time* per trial. We also used the eye tracking data to monitor the items of the scene that participants were looking at throughout each trial. The *maximum head amplitude* reached during each trial indicates that participants behave as expected regarding side-to-side head movement, reaching head amplitudes within the range of natural motion as planned, and allowing them to find the coin in all cases. The *average head speed* during each trial reveals an interesting insight: subjects accommodate the real head speed in order to maintain a constant virtual camera speed. This means that the real head speed is lower for higher translation gains, resulting in an approximately equivalent virtual camera speed (on average 0.2m/s, SEM = 0.001m/s) for the different translation gains. This in turn creates an approximately constant retinal velocity through different translation gains. This range of approximately constant velocities seems to be in accordance with previous work suggesting that the parallactic depth perceptual system achieves its maximum efficiency after reaching a head velocity greater than approximately $0.13m/s$, saturating at higher velocities [Ono and Ujike 2005]. Finally, qualitative analysis of eye tracking data confirms that participants spent a significant amount of time looking at the occluding planes that generate depth discontinuities in search for the coin; these depth discontinuities are the most critical case for detecting our manipulations. Please, refer to Sections S2.3 to S2.6 in the supplementary material for extended statistical analyses and more details.

Fitting psychometric functions. Response data for each participant was fitted to a psychometric function of the form $f(x) =$

---

[3]https://pupil-labs.com/

$\frac{1}{1+e^{ax+b}}$ where a and b are real numbers [Steinicke et al. 2010; Zhang et al. 2018]. The translation gain value at which the participants responded at chance level (50%) for each condition is considered as the *point of subjective equality* (PSE), at which participants perceive their real head motion and the virtual camera motion movements as identical.

Obtaining robust detection thresholds. Detection thresholds (DTs) are defined as the gain values at which participants can just detect a discrepancy between the two conditions. Following standard procedure in psychometric experiments (e.g., [Langbehn et al. 2018; Steinicke et al. 2010]), we set this threshold at 75% (halfway between the PSE and 100%) for correctly answering that the virtual movement was *larger* when $g_T > 1$, and at 25% (halfway between the PSE and 0%) for providing the same answer when $g_T < 1$ (since this means that the opposite answer, *smaller*, was chosen correctly 75% of the time). The interval between the 25% (compression) and 75% (expansion) detection thresholds defines the range of possible translation gains that can be applied.

As expected, there is a certain degree of variability in sensitivity between participants. Since our final goal is to leverage the derived thresholds to control camera manipulations in different VR applications *while keeping such manipulations undetectable*, we take a conservative approach to ensure their general applicability. We discard participants who show very low sensitivity to manipulations in order not to overstate the manipulation ratios that can be achieved in practical scenarios. We use the condition $g_T = 2.5$ (which represents an increased translation of 150% over the original movement) as a sentinel question, and discard participants if they do not consistently identify this virtual movement as larger than the real movement in at least 4/5 of the trials per condition. Following this procedure, eight participants were dropped, while the remaining 22 were used to compute our *robust thresholds*, and to perform the rest of the analysis in this section. From the removed participants, four identified as male and four as female, ages 18 to 23; four had worn an HMD before. We did not find statistically significant differences when analyzing these demographic factors with respect to the non-removed participants. Please refer to Section S2.2 in the supplementary material for an extended analysis including all participants.

Analysis of results. Figure 3 shows the resulting pooled results and fitted psychometric functions, for all different z_p values. The x-axis shows the translation gain, while the y-axis shows the probability that the participants detected a *larger* motion in the virtual environment. As expected, PSEs do not coincide with a gain value of $g_T = 1$. This would indicate that users perceive translations as perfectly identical in real and virtual spaces, which is rarely the case: Biases is PSE are commonly found in the literature [Bölling et al. 2019; Bolte and Lappe 2015; Langbehn et al. 2018; Steinicke et al. 2010; Suma et al. 2012], and are, in turn, indicative of underestimation or overestimation of virtual translations and rotations, an effect that has been extensively studied [Engel et al. 2008; Frenz et al. 2007; Jaekl et al. 2002; Jerald et al. 2008; Lappe et al. 2007]. We show PSE and DT values (both at 25% and 75%) in Figure 4.
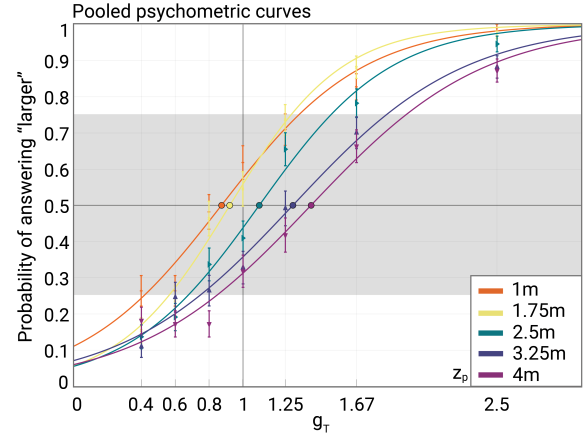
Fig. 3. Measured pooled results (error bars represent the standard error of the mean) and fitted psychometric curves for the detection performance for different values of z_p. The x-axis shows the applied translation gain g_T, and the y-axis shows the probability of estimating that the virtual translation was larger than the physical translation. The point of subjective equality (PSE) for each curve is marked with a colored dot.

	1m	1.75m	2.5m	3.25m	4m
DT 25%	0.42	0.56	0.67	0.74	0.84
PSE	0.87	0.92	1.1	1.29	1.40
DT 75%	1.33	1.28	1.52	1.84	1.96

Fig. 4. Values for PSE and DTs (compression and expansion) obtained from the pooled psychometric functions in Figure 3, and results of pairwise comparisons for the different plane distances (z_p). Values in the same set are statistically indistinguishable.

We have further performed a statistical analysis to assess whether the observed bias for different values of z_p is significantly different. Since our data is not normally distributed ($p < 0.05$ for the Shappiro-Wilk test), we employ a Friedman test (non-parametric equivalent of repeated measures ANOVA) with Dunn-Bonferroni post-hocs, and Bonferroni correction for multiple comparisons. Interestingly, we found significant differences ($p < 0.001$) both for PSE and DT values. This can be explained by the variation of relative velocity as a function of z_p, which affects the perception of relative head movement, and relative depth. This is a key insight that we will apply in our three example applications (improving 6-DoF viewing, overcoming physical space constraints and reducing simulator sickness). Levels included in the same set in Figure 4 are statistically indistinguishable ($p > 0.05$). Detailed pairwise significance values can be found in Section S2.1 in the supplementary material.

Discussion. We have found a statistically significant difference for the PSE for different values of the distance z_p to the plane (Figure 4). This indicates that participants' perception of a natural movement (1:1 mapping of real to virtual movement) varies for different levels of this condition. When this distance increases, participants' PSE is increasingly biased towards larger translation gains.

Our results seem to be in accordance with previous works reporting that distances are underestimated in immersive virtual environments (see Renner et al. [2013] for a complete review on this effect), and that the larger the distance, the stronger this effect is [Foley et al. 2004; Li et al. 2011]. Previous work [Nawrot et al. 2014] has also reported that depth from motion parallax is greatly foreshortened compared to the depth that might be expected, and that the error in determining viewing distance ratios is larger over longer viewing distances. Since underestimation is larger for larger distances, we hypothesize that the PSE may shift as a consequence of this: If users perceive something as closer than it is, they would expect more parallax, which would cause their subjective 1:1 real-virtual mapping (PSE) shift towards larger expansions.

Following this PSE bias, detection thresholds for compression (25% DT) and expansion (75% DT) also vary significantly with z_p (Figure 4). These variations can be explained with the PSE bias, and the slope of the psychometric functions. As the PSE varies as a function of z_p, this affects the associated detection thresholds. Additionally, as one could expect, we observe a decreasing trend in the slopes of the psychometric functions as z_p increases, suggesting that increasing distances would produce even lower slopes, to the point where participants would not be able to detect camera manipulations. The reason for this is that retinal velocity decreases for increasing distances; parallactic depth cannot be perceived until a certain relative image velocity threshold is reached [Ujike and Ono 2001], and, in the limit, for very far distances, not even motion can be perceived [Bonnet 1982].

4 TRANSLATION GAIN FOR COMPLEX SCENES

In the previous section we obtain detection thresholds for translation gain, i.e., we obtain the maximum translation gain that we can apply without the viewer noticing. This section describes how to apply these thresholds to a complex scene, as opposed to isolated points. Note that our thresholds are measured for lateral (left-right) translation gains, therefore, from now on, we apply them to lateral translations with respect to users' tracked head orientation. If users rotate their head, the lateral directions change accordingly. We consider the orientation relative to the head following previous work that shows that during free viewing, the rest of the body will align naturally to the head [Grasso et al. 1996; Hollands et al. 2004].

The depth layout of natural scenes is often complex and it may not be easily parametrizable in terms of simple distances to planes like the scene in our main experiment. Therefore, we resort to a parameterization based on retinal velocity, which allows us to take into account both complex depth layouts, and the dynamic behavior of the viewer. Given a scene, we obtain a translation gain map $\Phi(\mathbf{p})$; this map indicates, for each point $\mathbf{p}$, the maximum translation gain that can be applied without the viewer noticing when they are looking at such point $\mathbf{p}$. Note that, since we have found a dependency of the thresholds with the depth layout of the scene, computing such map will require depth information. In the case of synthetic scenes, ground truth depth is available; in the case of real, captured scenes, an estimated depth map is available in most cases as a result of the acquisition [Anderson et al. 2016], or can be computed even from a single image [Godard et al. 2019].
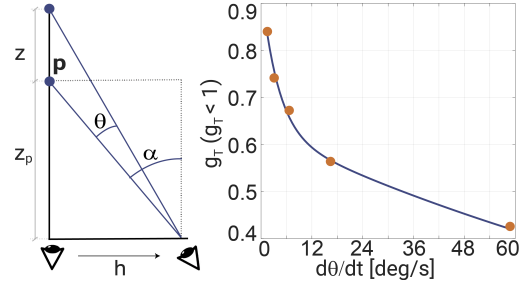

Fig. 5. *Left*: Illustration of retinal velocity for a viewer fixating on a point $\mathbf{p}$ as they translate their head (see text for details, image adapted from Nawrot and Stroyan [2009]). We show a single viewpoint for simplicity; however, note that in all our experiments binocular cues were correctly displayed. *Right*: Measured translation gain thresholds for compression as a function of retinal velocity $d\theta/dt$, and corresponding fitted function ϕ (please refer to the text for more details).

Retinal velocity map Θ. When the viewer observes a scene, the existence of motion is key to the perception of translation gain. We therefore employ a dynamic parameterization, in which we compute, for each point $\mathbf{p}$ in the scene, retinal velocity when the viewer is fixating on such point [Kellnhofer et al. 2016], yielding a retinal velocity map $\Theta(\mathbf{p})$.

Retinal velocity depends both on the relative motion between the viewer and the scene, and on the scene layout at the point the viewer is fixating on, as we described in Section 3. In particular, dependency on the scene layout is given by the *relative depth* between neighboring points. While the static optical geometry of retinal disparity explains the perception of depth from binocularity, our geometrical analysis is based on the work of Nawrot and Stroyan [2009], which describes the impact of dynamic geometry in perception according to the relationship of retinal velocity and pursuit of the eye. When the viewer is fixating on a point $\mathbf{p}$ and translates their head, retinal velocity is defined as $d\theta/dt$ (see Figure 5, left). Note that, in our VR setting, it is the virtual translation that contributes to changes in retinal velocity, since this is the translation that is used for generating the content the viewer sees. It has been shown that retinal velocity approximately follows the relation [Nawrot and Stroyan 2009]:

$$\frac{d\theta}{dt} \simeq \frac{z}{z_p}\frac{d\alpha}{dt}, \tag{1}$$

where $\frac{z}{z_p}$ is the relative depth (in our experiments $z = z_b - z_p$), and α is the rotation of the eye necessary to maintain a stable fixation on such point when translating (see Figure 5, left). For a given virtual eye translation h we have:

$$\alpha = atan(\frac{h}{z_p}). \tag{2}$$

Note that we account both for relative depth at discontinuities (z/z_p in Equation 1), and the distance to the fixated point ($1/z_p$ in Equation 2). From these equations, we can compute a retinal velocity map $\Theta(\mathbf{p})$ that encodes, for each point $\mathbf{p}$, the corresponding retinal velocity as the user translates their head, assuming they are fixating on a point $\mathbf{p}$. Virtual eye translation (equivalent to virtual head

Table 2. Parameter values and goodness of fit for translation gain (compression and expansion) as a function of retinal velocity in deg/s.

	a_1	b_1	a_2	b_2	R^2
Compression ($g_T < 1$)	0.309	-0.251	0.628	-0.007	0.995
Expansion ($g_T > 1$)	1.052	-0.147	1.131	0.003	0.994

translation) h can be obtained in real time from positional tracking, while z and z_p can be obtained from the depth information of the scene. For dynamic scenes, in which point $\mathbf{p}$ may be moving in time, its displacement needs to be taken into account in the computation of h; this can be accurately done in the case of synthetic content, or by means of optical flow in the case of captured content. We leverage the work by Gogel et al. [1982], which reported that the apparent motion of a physically stationary object concomitant with a lateral motion of the head is perceptually indistinguishable from apparent motion resulting from the physical motion of the object. Following this finding, we compute the retinal velocity by taking into account the optical flow [Lucas and Kanade 1981] from the dynamic scenes together with users' head motion.

To account for potential inaccuracies in scene depth, retinal velocity for a point $\mathbf{p}$ is computed with respect to every pixel within a neighborhood $\mathcal{N}$, yielding multiple retinal velocity values $\theta_i^p, i \in \mathcal{N}$. The final value stored in the retinal velocity map for point $\mathbf{p}$ is the maximum within the neighborhood, i.e., $\theta^p = max_i(\theta_i^p), i \in \mathcal{N}$. We use a neighborhood comprising $3°$ of visual angle.

Translation gain map Φ. From the retinal velocity map $\Theta(\mathbf{p})$, we seek to obtain a translation gain map $\Phi(\mathbf{p})$. The maximum unnoticeable translation gain is given by the thresholds we obtained in Section 3, which we plot as a function of retinal velocity in Figure 5, right, for the case of compression (please refer to Section S3 in the supplementary material for the case of expansion). Their behavior is well explained by a sum of exponential functions:

$$\phi\left(\frac{d\theta}{dt}\right) = a_1 e^{b_1 \frac{d\theta}{dt}} + a_2 e^{b_2 \frac{d\theta}{dt}}. \tag{3}$$

The values of a_1, a_2, b_1 and b_2 for compression and expansion, as well as the corresponding measure of goodness of fit, can be found in Table 2. The plot in Figure 5, right, shows how, as retinal velocity tends to zero, there is an asymptotic behavior of g_T. This is consistent with the fact that when retinal velocity is zero, detection of real-to-virtual manipulation is not feasible. In practice, in order to ensure a smooth transition, we assign $g_T(0) = 1$. For values of retinal velocity greater than 60 deg/s we make a conservative approximation and clamp g_T to 0.42. This has been used satisfactorily in our application scenario (description and validation in Section 5.1); nevertheless, more research would be needed to explore the behavior for higher retinal velocities.

To compute the translation gain map of a scene $\Phi(\mathbf{p})$ we use Equation 3, where the values of $d\theta/dt$ are obtained from the retinal velocity map $\Theta(\mathbf{p})$. To avoid drastic variations in translation gain between adjacent points, we filter the translation gain map $\Phi(\mathbf{p})$ by applying a Gaussian kernel (heuristically, we choose $\sigma = 3°$ of visual angle, which matches the spatial support of the previously defined

neighborhood $\mathcal{N}$). Additionally, for dynamic scenes, we smooth the resulting translation gain maps in the temporal dimension using a window of 0.5 seconds (15 frames in our videos) in order to avoid temporal flickering in the applied translation gain.
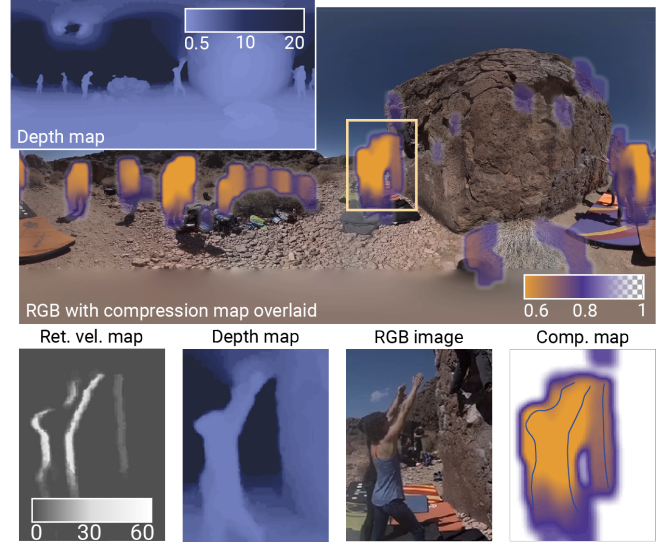

Fig. 6. *Top row*: Translation gain map (compression) for a 360° RGBD panorama, overlaid on the corresponding RGB image for easier interpretation. *Bottom row, from left to right:* Close-ups of the retinal velocity map, the depth map, the RGB image, and the resulting translation gain map for compression. Close-by regions with disocclusion boundaries, where retinal velocity is high, can be significantly compressed (disocclusion boundaries are approximately marked as blue lines in the compression map for reference).

For our applications, we precompute translation gain maps and store them as grayscale image (or video) panoramas (see Section S4.1 in the supplementary material for more details). Then, given the translation gain maps, during real-time viewing we dynamically apply the translation gain depending on the viewer's eye fixations; these can be accurately obtained with eye tracking. If eye tracking is not available, as is common in current hardware, head orientation can be used as a proxy by applying a minor modification to the maps. For computing our translation gain maps, we use head orientation as a proxy: We leverage recent work showing that in VR, given a head orientation, gaze locations will likely fall within a neighborhood around it [Rai et al. 2017a,b; Sitzmann et al. 2018]. In particular, we take this neighborhood as $11.7°$ of visual angle as measured by Sitzmann et al. [2018]. We assume that detection of motion manipulation will be driven by a region around the fixation motivated by the fact that egocentric motion perception relies on the center of the visual field [Brandt et al. 1973], and that the central retina is more specialized than the periphery in *smooth* motion perception [Finlay 1982], which is the nature of the motion studied in our work. We therefore compute, for each point $\mathbf{p}$ in the retinal velocity map $\Theta(\mathbf{p})$, the average perceived retinal velocity in such a neighborhood, and this is the value used for the computation of the translation gain map $\Phi(\mathbf{p})$.

Figure 6 shows a sample translation gain map for a 360° RGBD panorama. Specifically, we show a compression map, which is useful for, e.g., artifact reduction in 6-DoF viewing systems (see Section 5.1). For regions with higher values of retinal velocity (e.g., at close disocclusion boundaries), more compression is permitted. Very distant regions are not highly compressed, however at those distances 6-DoF artifacts are rare.

5 APPLICATIONS

In this section we implement three proof-of-concept applications of our translation gain thresholds that help us both to validate our thresholds in different scenarios by means of additional user studies, and to show that they can be useful for alleviating different current limitations of VR systems. None of the participants recruited for the psychometric experiment were allowed to participate in the validation experiments. The three chosen applications are important problems in VR for which no definitive solution exists. The first one brings an improvement in terms of visual quality to 6-DoF viewing systems, by integrating our translation gain maps (Section 4) for compressing camera motion. A second application shows how we can help overcome physical constraints during VR experiences by limiting lateral translations during local interactions, employing our measured translation gain thresholds for expansion. Finally, a simple, proof-of-concept application tests whether our thresholds can be used to reduce motion sickness during VR viewing by compressing camera motion. All three applications were run in an Oculus Rift headset, to confirm that our thresholds are not constrained to a single head-mounted display (our experiments in Section 3 were completed with an HTC Vive Pro).

We report here the main results of each validation, and refer the reader to Sections S4 through S7 in the supplementary material for further information about the experimental procedure, implementation details, and statistical analysis.

5.1 6-DoF viewing

Common 360° footage lacks motion parallax, which is not only unnatural, but can also break the sense of spatial presence and may induce discomfort as well, when shown on a VR headset [Richardt et al. 2019; Sylwan et al. 2019]. This happens even if the viewer tries to remain mostly static, since accidental motion always occurs to a certain degree. Previous work has already shown that users do notice the missing degrees of freedom when presented with only 3-DoF, and that they prefer 6-DoF for interaction and navigation [Serrano et al. 2019; Thatte and Girod 2018]. Existing methods for enabling 6-DoF viewing (rotation and translation) succeed in providing the necessary motion parallax (e.g., [Hedman and Kopf 2018; Overbeck et al. 2018; Pozo et al. 2019; Serrano et al. 2019]), but often result in artifacts at disoccluded areas of the scene that have not been captured (see Figure 7). These artifacts become more obvious as the head displacement from the captured position increases, strongly degrading the experience and reducing its realism. Therefore, leveraging our thresholds to compress virtual camera motion may help reduce the visibility of such artifacts, improving the quality of the VR experience. In this section, we implement this application by integrating our translation gain maps into a state-of-the-art 6-DoF

viewing system, and evaluate whether the compression improves visual quality while remaining unnoticeable.
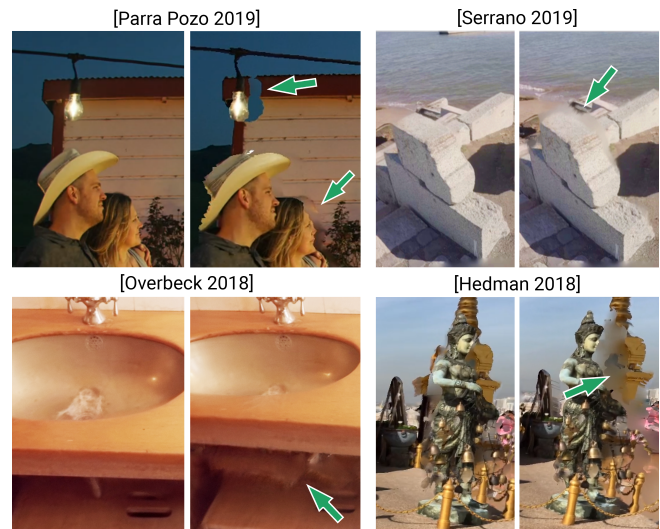

Fig. 7. Representative artifacts common in recent depth image-based rendering approaches enabling 6-DoF VR viewing from captured content. As the rendered viewpoint (*right*) moves away from the captured point (*left*), artifacts of a diverse nature start becoming more noticeable.

Specifically, given a 360° RGBD image or video, we compute the corresponding translation gain maps for compression as described in Section 4, and integrate them into the recent 6-DoF viewing system of Serrano et al. [2019], whose source code is publicly available. Translation gain maps are used by the 6-DoF viewing system to control camera movement, so our method could be integrated into most existing systems. We implement and test the more common scenario of eye tracking not being available during viewing, and during real-time playback we dynamically apply the translation gain, which is encoded in the translation gain map, depending on the viewer's head orientation (which indicates the region of the scene being observed). Further implementation details can be found in the supplementary material.

Validation experiments. We verify whether our compression of camera motion improves the perceived visual quality in 6-DoF viewing, while remaining undetected by users. In addition, we explore how well our measured thresholds (obtained for static scenes) apply to the more general and challenging case of dynamic scenes. Recent work has shown that users do not reliably detect dynamic changes in translation gains [Zhang and Kuhl 2013; Zhang et al. 2014], which suggests that changes in the compression map can be applied during video playback. We therefore perform two experiments, one with static images and one with videos, both with the same procedure: Each participant viewed eight scenes, and each scene was shown twice, resulting in sixteen trials. The stimuli for both experiments were obtained from the same scenes. In each trial, users had to complete a two-interval forced-choice task. They viewed each scene once with compressed camera motion and once without compression, back to back. Both the scene and compression mode were fully
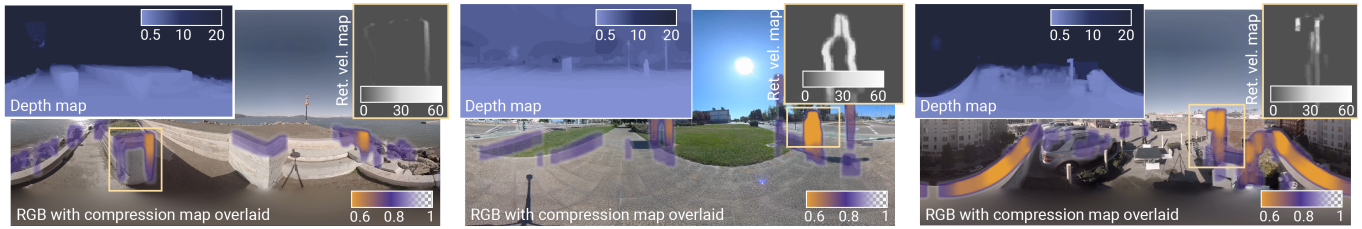
Fig. 8. Three sample scenes from the set used to validate the application of our translation gain maps (compression) to 6-DoF viewing systems. For each scene, we show the translation gain map for compression overlaid on the RGB image, together with the associated depth map and a close-up of the retinal velocity map. For the static experiment, a single frame was taken from each of the videos.
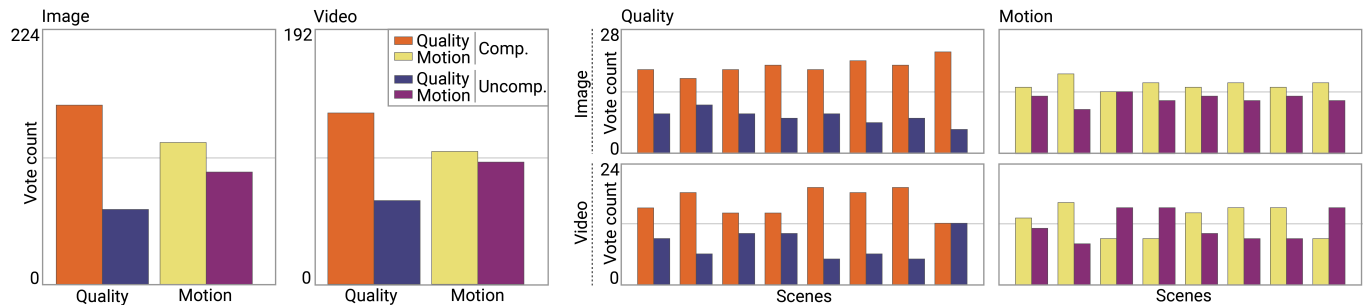

Fig. 9. *Left:* Vote count for the quality preference and the better real-to-virtual motion reproduction for all scenes pooled, for image and video experiments. *Right:* Vote count for both quality and motion reproduction preference for both modes per scene. The x-axis shows the eight different scenes tested. Compressed motion was always preferred in terms of quality, while there were no significant differences in the answers to which method had a better reproduction of the real movement in the virtual world.

randomized, except for not allowing the same scene to appear in two successive trials. After watching each compressed-uncompressed pair, participants were asked to choose which one *had the best image quality*, and which one *better reproduced the real movement in the virtual world.* Figure 8 shows a number of sample scenes used in the experiment, together with their depth maps, retinal velocity maps, and translation gain (compression) maps; the images show how close-by regions with disocclusion boundaries can be significantly compressed, allowing for a reduction of 6-DoF viewing artifacts.

Fourteen participants (4 female and 10 male, average 25.43 years old, σ = 2.96) took part in our image experiment; whereas twelve participants (3 female and 9 male, average 25.33 years old, σ = 2.86) took part in our video experiment. For all the experiments reported in this section, participants voluntarily signed up, and all of them reported normal or corrected-to-normal vision. In all cases the participants were naïve as to the purpose of the experiment and were previously informed that they could stop the experiment at any point if they felt uncomfortable (none requested to stop).

Results. The results of the experiment are presented in Figure 9, showing the number of votes received by the compressed and un-compressed methods. We perform a significance test of the vote count differences following existing literature [David 1963; Rubinstein et al. 2010; Setyawan and Lagendijk 2004], to reveal whether the two methods produced results that were perceived as clearly different (see Section S4 in the supplementary material for details on this analysis). We observe a statistically significant difference

in perceived image quality in favor of our compressed motion. In terms of which method better reproduced the real movement in the virtual world, we found no statistically significant difference in the answers between the compressed and uncompressed methods. Our solution thus helps extending 3-DoF viewing (just rotations) to a more natural, parallax-inducing 6-DoF viewing, for both static and dynamic VR environments.

5.2 Overcoming physical space constraints

Virtual environments can be considerably larger than the available physical space, which is commonly constrained to a standard-sized room. Redirected walking techniques (RDW) try to tackle the problem of locomotion in reduced spaces by unnoticeably manipulating users' physical movements. Existing solutions (e.g., [Bölling et al. 2019; Razzaque et al. 2002; Steinicke et al. 2010]) focus on forward locomotion. However, most popular and consumed VR applications (e.g., skiing or boxing games, and top-played VR games like Beat-Saber, or SuperhotVR), as well as in-flight or training simulators, rely on scenarios that exploit lateral motion, such as searching or dodging actions, and require that the user be mostly standing or seated. Besides, these lateral translations are key to induce motion parallax. While smaller, lateral translations can be hindered by the limited physical space, and an unnoticeable reduction in the physical space needed can allow for reduced risks and a better experience. We therefore present a proof-of-concept application suggesting that our translation gain thresholds can be applied in these scenarios (or

in combination with techniques for forward locomotion) to reduce the actual lateral translation in the physical space.

To test this application, we have devised a simple scenario where a standing user is asked to dodge incoming columns by means of lateral movements (see Figure 10, left). We perform two different experiments, which will allow us to validate (i) whether our real-to-virtual motion manipulation is indeed unnoticeable, and (ii) the actual resulting physical displacements of the users with and without our manipulation.

Validation experiments. Our first experiment aims to validate that users are not able to notice our manipulation of lateral motion. The experiment consists of 17 consecutive trials. Seven had no motion manipulation, seven had the virtual lateral motion expanded within the limits of our thresholds ($g_T = 1.28$ to $g_T = 1.96$, depending on the distance to the incoming column), and the remaining three acted as sentinels, with a clearly noticeable expansion ($g_T = 2.5$). The trials were shown in random order, and we asked the participants to inform each time they detected a manipulation in the camera motion. The second experiment is similar to the first one, except that the speed and positions of the columns are fixed, in order to perform a systematic analysis of the users' lateral motion. Participants again had to dodge the columns with and without our motion manipulation. A total of 11 participants voluntarily took part in our validation experiments: six users (4 male and 2 female, average 23.83 years old, $\sigma = 2.78$) in the first experiment, and five users (4 male and 1 female, average 23 years old, $\sigma = 1.09$) in the second experiment.

Results. In the first experiment, after discarding answers to sentinels to avoid biasing the results, for the trials with no manipulation, manipulation was reported in 43% of the cases; while for the manipulated ones, manipulation was reported in 48% of them. In both cases users reported manipulations at chance level, suggesting that such manipulations remain unnoticed within our measured thresholds. A logistic regression with the users' answers as the dependent variable found no statistically significant effect of the user, the order, nor the presence or absence of manipulation on the answers provided.
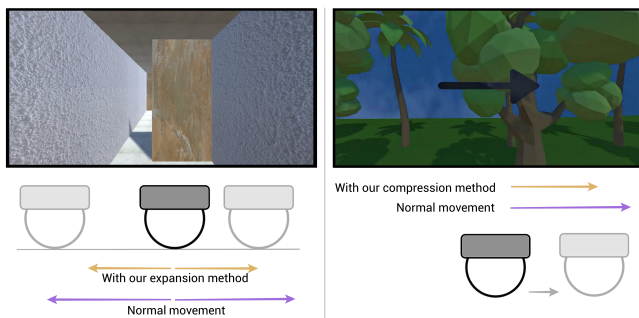

Fig. 10. Example applications of our detection thresholds. *Left:* Overcoming physical space constraints. Users needed on average 12.75% less physical space to move, while not noticing our camera manipulations. *Right:* Reducing motion sickness. Users reported on average 12.71% less discomfort.

For the second experiment, our results show that users translate on average 12.75% less when our manipulation is applied, while being able to complete the same virtual task in a smaller physical space. This improvement is on par with improvements achieved by other works performing manipulation of real-to-virtual mapping in VR [Langbehn et al. 2018; Zhang et al. 2018] (note, however, that we are the first to measure and use translation gain thresholds for *lateral* camera motion). Users remain closer to the starting point when our thresholds are applied, their positions along the x-axis following a Gaussian distribution with $1.96\sigma = 0.70m$ for the manipulated trials and $1.96\sigma = 0.80m$ for the non-manipulated trials. This indicates that users' positions during the task are 10cm closer to the starting point with a 95% confidence when our thresholds are applied.

In summary, the results of our experiments suggest that our motion manipulation is not perceived by the users when applied within the limits of our thresholds, and that it can effectively reduce the physical space required to move laterally in VR.

5.3 Reducing motion sickness

VR systems can induce sickness and discomfort due to the discrepancy between visual and vestibular cues [LaViola 2000; McGill et al. 2017]. A number of works have tried to address this issue by, e.g., optimizing camera trajectories to reduce it [Hu et al. 2019], using so-called rest frames [Cao et al. 2018], or reducing the field of view [Fernandes and Feiner 2016]. Some VR scenarios, such as simulators or stationary VR usage, control the camera motion so that it follows a pre-established trajectory. This generates self-motion illusions, commonly called vection. In those cases, the compound movement resulting from the camera and the user head motion may incur a larger vection, thus increasing the probability and the severity of sickness symptoms.

Validation experiment. To show whether our detection thresholds can help mitigate this, we have devised a perceptual experiment where seated participants see a forest scene made up of three layers: a static background layer at $10m$ representing the sky, and two dynamic layers at $3m$ and $1.5m$ with trees. These layers move at different speeds creating the illusion of depth and movement of the viewer, while they are periodically asked to move their head in the opposite direction, i.e., in the direction that the camera is moving, effectively increasing vection (see Figure 10, right). We compare variations over time in the resulting *discomfort score* [Cao et al. 2018; Fernandes and Feiner 2016] for the baseline case (no manipulation of camera motion) and for the case of compressing camera motion according to our thresholds ($g_T = 0.56$ in our experimental setting). Five users (4 female and 1 male, average 24.2 years old, $\sigma = 4.9$) voluntarily participated in the experiment.

Results. The outcome of this study can be seen in Figure 11. On average, when our motion compression strategy was applied within our thresholds, participants reported 12.71% less discomfort over time. Further, the standard deviation was 7.35% smaller when our method is applied, indicating less variance across participants' ratings. Although sickness may be influenced by many confounding factors, our results suggest that our compression strategy may be
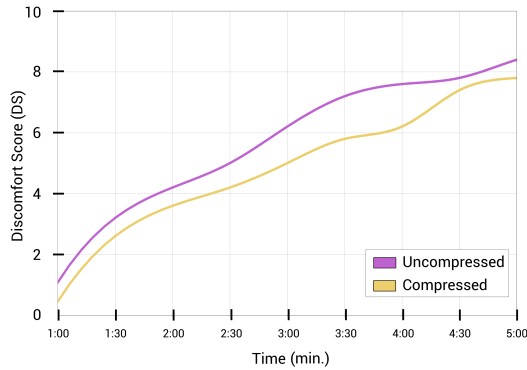
Fig. 11. Reducing motion sickness: discomfort scores over time with and without our motion manipulation. We follow recent works [Cao et al. 2018; Fernandes and Feiner 2016] and ask the participants every 30 seconds to rate a discomfort score between 0 and 10, with 0 being no discomfort and 10 being too much discomfort. When our compression strategy is applied, participants reported 12.71% less discomfort over time.

helpful to mildly alleviate sickness induced by vection, since it reduces sensory conflict, which is one of the known sickness-inducing factors. Nevertheless, we have only tested a proof-of-concept application in a simple scenario. Simulator sickness is a complex phenomenon: further consideration must be given to other forms of motion conflicts that may arise, longer exposure times, or other sources of discomfort [Koulieris et al. 2017]. The applicability of our thresholds in such contexts needs to be further examined.

6 DISCUSSION

Motion perception has been studied by many disciplines, including psychology, neurology, and computer science, and it has proven to be a difficult problem for which no definitive model that explains all mechanisms of neural processing exists.

As we discussed in Section 3, there are two main factors that contribute to retinal velocity, which is key in the perception of motion: scene layout, and observer relative motion with respect to the scene. Several works have shown that the degree of unnoticed manipulations when an observer is actively moving and interacting with the environment is significantly different from measured thresholds with artificial movements of the observer or the stimulus [Scarfe and Glennerster 2015; Schmitz et al. 2018; Wexler and Van Boxtel 2005; Yoonessi and Baker 2011]. One of our main goals is to derive thresholds that can be applied in natural viewing conditions. Due to this freedom, we enable free exploration in our experiment by parameterizing retinal velocity as a function of scene layout (and, specifically, relative depth). Further, we analyze the most critical case of depth edges, since these produce the most abrupt changes in retinal velocity in the presence of relative motion.

During our experiments we measured participants' virtual camera motion, and found it to be quite constant, with an average of $0.2 m/s$ (SEM = $0.001 m/s$). This is consistent with previous work, which already reported that observers typically perform lateral head movements in a similar manner for a given task [Yoonessi and Baker

2011]. We thus use this measured average to compute our thresholds as a function of retinal velocity (Section 4). Previous works seem to suggest that the sensitivity to translation gain expressed in terms of retinal velocity will remain constant across different observer velocities: Results from Nagata et al. [1989] show that parallactic depth thresholds expressed in terms of retinal velocity remain constant across different observer velocities, and findings in the literature suggest that parallactic depth perception and motion perception share the same 'front end', a retinal image motion predictor [Ujike and Ono 2001]. Supported by these findings, we believe our thresholds depend only on the final perceived retinal velocity. Our validation experiments in the different proof-of-concept applications (Section 5), in which users interacted in a natural manner with different scenarios and tasks, are consistent with this observation: Our manipulations remain undetected at different camera motion velocities, provided the adequate translation gain is applied as a function of retinal velocity.

Limitations and future work. Our thresholds are measured under the assumption of head translation in the absence of locomotion, which is one of the most common viewing configurations for VR. However, combining our findings with other scenarios extensively studied in the literature such as those involving full-body locomotion [Bölling et al. 2019; Grechkin et al. 2016; Neth et al. 2012; Steinicke et al. 2010], or introducing additional manipulations during rotations [Bolte and Lappe 2015; Bruder et al. 2011; Jerald et al. 2008; Langbehn et al. 2018] is an interesting topic for future work. As in all studies of this nature, our results are only strictly valid for our measured conditions. For example, our results are strictly valid only for the type of display tested (fixed-focal display); nevertheless, currently, HMDs with a fixed-focal display are the only ones available on the market. Extending our analysis to other types of displays (vari-focal or multi-focal) once they become available would be an interesting avenue for future research.

In our main experiment, only 37% of the recruited users identified as female; therefore, the reader should be aware that our results may be biased in terms of gender. An analysis of our data using a GLMM, which is robust to unbalanced populations, suggests that there are no significant differences between genders while performing our task, although this may be due to our sample size. Previous works [Nguyen et al. 2018; Williams and Peck 2019] suggest that males are more sensitive to rotation and curvature gains; it is thus possible that female users could take advantage of even larger thresholds, which would be beneficial for our example applications. Nevertheless, explicitly quantifying the potential differences for different demographics, such as gender, age range, or VR experience is a crucial aspect in VR research [Peck et al. 2020], and remains an important avenue of future work.

We have validated the application of our computed thresholds in a number of scenarios with varied conditions and tasks. Moreover, our translation gain function behaves in a sensible manner when extrapolated outside the measured range, but future work is required to explore significant departures from this range. In some applications, such as 6-DoF viewing, larger motions lead to more noticeable artifacts; for such cases, more compression could be considered as acceptable, even if noticed by users, in order to prevent

further degradation of image quality. Conversely, very small head motions might not result in perceptible artifacts, and therefore our compression strategy could be bypassed. We leave as future work further investigation on the trade-off between the perception of 6-DoF-induced artifacts, and the application of our compression strategy. For some applications, various aspects of compression are supported by other aspects of dynamic and seamless camera manipulations, such as bringing the region of interest close to the display plane [Kellnhofer et al. 2016]. Some of these manipulations could be relevant for our application; we relegate such considerations for future work.

7 CONCLUSION

To our knowledge, we have presented the first attempt at measuring detection thresholds for lateral translation gains for head motion during natural VR viewing. Such motion is important, yielding precise motion parallax judgments [Steinbach et al. 1991], facilitating depth perception and scene understanding. Our psychophysical experiments reveal that significant compression and expansion of the virtual camera motion is possible, while remaining imperceptible to users. We have further analyzed whether these detection thresholds are influenced by the scene layout, finding an important dependency with retinal velocity, and in particular with the relative depth of the elements in the scene. Our findings are agnostic to the particular VR system used, and can be applied to static and dynamic scenes. In order to apply our thresholds, only head tracking and information about the depth layout is needed (which can be obtained in real time for generated scenes, or precomputed for captured ones). Therefore, they can be safely applied without incurring in performance degradation. We have demonstrated that our thresholds are robust and can be used to address several open problems in VR: improving 6-DoF viewing, overcoming physical constraints, and reducing motion sickness. We believe our work helps further understand the relation between real and virtual motion, and hope future work will leverage our insights in novel applications.

ACKNOWLEDGMENTS

REFERENCES

Robert Anderson, David Gallup, Jonathan T Barron, Janne Kontkanen, Noah Snavely, Carlos Hernández, Sameer Agarwal, and Steven M Seitz. 2016. Jump: virtual reality video. *ACM Trans. Graph.* 35, 6 (2016), 198.

Luke Bölling, Niklas Stein, Frank Steinicke, and Markus Lappe. 2019. Shrinking Circles: Adaptation to Increased Curvature Gain in Redirected Walking. *IEEE Trans. on Vis. and Comp. Graph.* 25, 5 (2019), 2032–2039.

Benjamin Bolte and Markus Lappe. 2015. Subliminal reorientation and repositioning in immersive virtual environments using saccadic suppression. *IEEE Trans. on Vis. and Comp. Graph.* 21, 4 (2015), 545–552.

Claude Bonnet. 1982. Thresholds of motion perception. In *Tut. on Motion Perception.*

Th Brandt, Johannes Dichgans, and Ellen Koenig. 1973. Differential effects of central versus peripheral vision on egocentric and exocentric motion perception. *Exp. brain research* 16, 5 (1973), 476–491.

Gerd Bruder, Frank Steinicke, Phil Wieland, and Markus Lappe. 2011. Tuning self-motion perception in virtual reality with visual illusions. *IEEE Trans. on Vis. and Comp. Graph.* 18, 7 (2011), 1068–1078.

Zekun Cao, Jason Jerald, and Regis Kopper. 2018. Visually-induced motion sickness reduction via static and dynamic rest frames. In *Proc. of IEEE VR.* 105–112.

Herbert Aron David. 1963. *The method of paired comparisons.* Vol. 12. London.

Kalpana Dokka, Paul R MacNeilage, Gregory C DeAngelis, and Dora E Angelaki. 2011. Estimating distance during self-motion: A role for visual–vestibular interactions. *Journal of Vision* 11, 13 (2011), 2–2.

David Engel, Cristóbal Curio, Lili Tcheang, Betty Mohler, and Heinrich H Bülthoff. 2008. A psychophysically calibrated controller for navigating through large environments in a limited free-walking space. In *Proc. of ACM VRST.* 157–164.

Ajoy S Fernandes and Steven K Feiner. 2016. Combating VR sickness through subtle dynamic field-of-view modification. In *Proc. of IEEE 3DUI.* 201–210.

David Finlay. 1982. Motion perception in the peripheral visual field. *Perception* 11, 4 (1982), 457–462.

John M Foley, Nilton P Ribeiro-Filho, and José A Da Silva. 2004. Visual perception of extent and the geometry of visual space. *Vision Research* 44, 2 (2004), 147–156.

Harald Frenz, Markus Lappe, Marina Kolesnik, and Thomas Bührmann. 2007. Estimation of travel distance from visual motion in virtual environments. *ACM Trans. on Applied Perception* 4, 1 (2007).

Clément Godard, Oisin Mac Aodha, Michael Firman, and Gabriel J Brostow. 2019. Digging into self-supervised monocular depth estimation. In *Proc. of IEEE ICCV.*

Walter C Gogel. 1982. Analysis of the perception of motion concomitant with a lateral motion of the head. *Perception & Psychophysics* 32, 3 (1982), 241–250.

Renato Grasso, Stefan Glasauer, Yasuhiko Takei, and Alain Berthoz. 1996. The predictive brain: anticipatory control of head direction for the steering of locomotion. *Neuroreport* 7, 6 (1996), 1170–1174.

Timofey Grechkin, Jerald Thomas, Mahdi Azmandian, Mark Bolas, and Evan Suma. 2016. Revisiting detection thresholds for redirected walking: Combining translation and curvature gains. In *Proc. of the ACM Symp. on Applied Perception.* 113–120.

Leah A Groves, Patrick Carnahan, Daniel R Allen, Rankin Adam, Terry M Peters, and Elvis CS Chen. 2019. Accuracy assessment for the co-registration between optical and VIVE head-mounted display tracking. *Intern. journal of computer assisted radiology and surgery* 14, 7 (2019), 1207–1215.

Peter Hedman and Johannes Kopf. 2018. Instant 3D photography. *ACM Trans. Graph.* 37, 4 (2018), 101.

Mark A Hollands, Nausica V Ziavra, and Adolfo M Bronstein. 2004. A new paradigm to investigate the roles of head and eye movements in the coordination of whole-body movements. *Experimental brain research* 154, 2 (2004), 261–266.

Ping Hu, Qi Sun, Piotr Didyk, Li-Yi Wei, and Arie E Kaufman. 2019. Reducing simulator sickness with perceptual camera control. *ACM Trans. Graph.* 38, 6 (2019), 1–12.

Victoria Interrante, Brian Ries, and Lee Anderson. 2007. Seven league boots: A new metaphor for augmented locomotion through moderately large scale immersive virtual environments. In *Proc. of IEEE 3DUI.*

Philip M Jaekl, Robert S Allison, Laurence R Harris, Urszula T Jasiobedzka, Heather L Jenkin, Michael R Jenkin, James E Zacher, and Daniel C Zikovitz. 2002. Perceptual stability during head movement in virtual reality. In *Proc. of IEEE VR.* 149–155.

Jason Jerald, Tabitha Peck, Frank Steinicke, and Mary Whitton. 2008. Sensitivity to scene motion for phases of head yaws. In *Proc. of the Symp. on Applied Percep. in Graph. and Vis.* 155–162.

Petr Kellnhofer, Piotr Didyk, Tobias Ritschel, Belen Masia, Karol Myszkowski, and Hans-Peter Seidel. 2016. Motion parallax in stereo 3D: Model and applications. *ACM Trans. Graph.* 35, 6 (2016), 176.

HyunGoo R Kim, Dora E Angelaki, and Gregory C DeAngelis. 2016. The neural basis of depth perception from motion parallax. *Philosophical Trans. of the Royal Society B: Biological Sciences* 371, 1697 (2016), 20150256.

George Koulieris, Kaan Akşit, Michael Stengel, Rafał K Mantiuk, Katerina Mania, and Christian Richardt. 2019. Near-Eye Display and Tracking Technologies for Virtual and Augmented Reality. In *Computer Graphics Forum*, Vol. 38. 493–519.

George Koulieris, Bee Bui, Martin S Banks, and George Drettakis. 2017. Accommodation and comfort in head-mounted displays. *ACM Trans. Graph.* 36, 4 (2017).

Oliver Kreylos. 2016. *Analysis of Valve's 'Lighthouse' Tracking System Reveals Accuracy.* https://www.roadtovr.com/analysis-of-valves-lighthouse-tracking-system-reveals-accuracy/. Last accessed on 2020-04-25.

Eike Langbehn, Paul Lubos, Gerd Bruder, and Frank Steinicke. 2017. Bending the curve: Sensitivity to bending of curved paths and application in room-scale VR. *IEEE Trans. on Vis. and Comp. Graph.* 23, 4 (2017), 1389–1398.

Eike Langbehn, Frank Steinicke, Markus Lappe, Gregory F Welch, and Gerd Bruder. 2018. In the blink of an eye: leveraging blink-induced suppression for imperceptible position and orientation redirection in virtual reality. *ACM Trans. Graph.* 37, 4 (2018), 66.

Markus Lappe, Michael Jenkin, and Laurence R Harris. 2007. Travel distance estimation from visual motion by leaky path integration. *Exp. Brain Research* 180, 1 (2007).

Joseph J. LaViola. 2000. A Discussion of Cybersickness in Virtual Environments. *ACM SIGCHI Bulletin* 32, 1 (2000), 47–56.

Zhi Li, John Phillips, and Frank H Durgin. 2011. The underestimation of egocentric distance: Evidence from frontal matching tasks. *Attention, Perception, & Psychophysics* 73, 7 (2011), 2205.

Bruce D. Lucas and Takeo Kanade. 1981. An Iterative Image Registration Technique with an Application to Stereo Vision. In *Proc. of the Intern. Joint Conf. on Artificial Intelligence* (Vancouver, BC, Canada) *(IJCAI'81).* San Francisco, CA, USA, 674–679.

Keigo Matsumoto, Eike Langbehn, Takuji Narumi, and Frank Steinicke. 2020. Detection Thresholds for Vertical Gains in VR and Drone-based Telepresence Systems. In *Proc. of IEEE VR.* 101–107.

Mark McGill, Alexander Ng, and Stephen Brewster. 2017. I am the passenger: how visual motion cues can influence sickness for in-car VR. In *Proc. of the ACM CHI Conf. on Human Factors in Computing Sys.* 5655–5668.

Shojiro Nagata. 1989. How to reinforce perception of depth in single two-dimensional pictures. (1989).

Mark Nawrot, Michael Ratzlaff, Zachary Leonard, and Keith Stroyan. 2014. Modeling depth from motion parallax with the motion/pursuit ratio. *Frontiers in Psychology* 5 (2014), 1103.

Mark Nawrot and Keith Stroyan. 2009. The motion/pursuit law for visual depth perception from motion parallax. *Vision Research* 49, 15 (2009), 1969–1978.

Christian T Neth, Jan L Souman, David Engel, Uwe Kloos, Heinrich H Bulthoff, and Betty J Mohler. 2012. Velocity-dependent dynamic curvature gain for redirected walking. *IEEE Trans. on Vis. and Comp. Graph.* 18, 7 (2012), 1041–1052.

Anh Nguyen, Yannick Rothacher, Bigna Lenggenhager, Peter Brugger, and Andreas Kunz. 2018. Individual differences and impact of gender on curvature redirection thresholds. In *Proc. of the ACM Symp. on Applied Perception.* 1–4.

Carlo Nicolini, Carlo Fantoni, Giovanni Mancuso, Robert Volcic, and Fulvio Domini. 2014. A framework for the study of vision in active observers. In *Human Vision and Electronic Imaging XIX*, Vol. 9014. Intern. Society for Optics and Photonics, 901414.

Niels Christian Nilsson, Tabitha Peck, Gerd Bruder, Eri Hodgson, Stefania Serafin, Mary Whitton, Frank Steinicke, and Evan Suma Rosenberg. 2018. 15 years of research on redirected walking in immersive virtual environments. *IEEE Computer Graph. and App.* 38, 2 (2018), 44–56.

Norbert Nitzsche, Uwe D Hanebeck, and Günther Schmidt. 2003. Motion compression for telepresent walking in large-scale remote environments. In *Helmet-and Head-Mounted Displays VIII: Technologies and Applications*, Vol. 5079. Intern. Society for Optics and Photonics, 265–276.

Hiroshi Ono and Hiroyasu Ujike. 2005. Motion parallax driven by head movements: Conditions for visual stability, perceived depth, and perceived concomitant motion. *Perception* 34, 4 (2005), 477–490.

Ryan S. Overbeck, Daniel Erickson, Daniel Evangelakos, Matt Pharr, and Paul Debevec. 2018. A System for Acquiring, Processing, and Rendering Panoramic Light Field Stills for Virtual Reality. *ACM Trans. Graph.* 37, 6, Article 197 (2018), 15 pages.

Francesco Panerai, Valerie Cornilleau-Pérès, and Jacques Droulez. 2002. Contribution of extraretinal signals to the scaling of object distance during self-motion. *Perception & Psychophysics* 64, 5 (2002), 717–731.

Anjul Patney, Marina Zannoli, Joohwan Kim, Robert Konrad, Frank Steinicke, and Martin S. Banks. 2018. Applications of Vision Science to Virtual and Augmented Reality. In *ACM SIGGRAPH Courses* (Vancouver, British Columbia, Canada) *(SIGGRAPH '18).* Article 8, 28 pages.

Tabitha C Peck, Henry Fuchs, and Mary C Whitton. 2011. An evaluation of navigational ability comparing Redirected Free Exploration with Distractors to Walking-in-Place and joystick locomotio interfaces. In *Proc. of IEEE VR.* 55–62.

T. C. Peck, L. E. Sockol, and S. M. Hancock. 2020. Mind the Gap: The Underrepresentation of Female Participants and Authors in Virtual Reality Research. *IEEE Trans. on Vis. and Comp. Graph.* 26, 5 (2020), 1945–1954.

Albert Parra Pozo, Michael Toksvig, Terry Filiba Schrager, Joyce Hsu, Uday Mathur, Alexander Sorkine-Hornung, Rick Szeliski, and Brian Cabral. 2019. An integrated 6DoF video camera and system design. *ACM Trans. Graph.* 38, 6 (2019), 1–16.

Eric D Ragan, Siroberto Scerbo, Felipe Bacim, and Doug A Bowman. 2016. Amplified head rotation in virtual reality and the effects on 3d search, training transfer, and spatial orientation. *IEEE Trans. on Vis. and Comp. Graph.* 23, 8 (2016), 1880–1895.

Yashas Rai, Jesús Gutiérrez, and Patrick Le Callet. 2017a. A dataset of head and eye movements for 360 degree images. In *Proc. of the ACM Multim. Sys. Conf.* 205–210.

Yashas Rai, Patrick Le Callet, and Philippe Guillotel. 2017b. Which saliency weighting for omni directional image quality assessment?. In *Proc. of IEEE QoMEX.* 1–6.

Sharif Razzaque, David Swapp, Mel Slater, Mary C. Whitton, and Anthony Steed. 2002. Redirected Walking in Place. In *Proc. of the Workshop on Virtual Envs. 2002* (Barcelona, Spain) *(EGVE '02).* Eurographics Association, 123–130.

Rebekka Renner, Boris Velichkovsky, and Jens R Helmert. 2013. The perception of egocentric distances in virtual environments-a review. *Comput. Surveys* 46, 2 (2013).

Christian Richardt, Peter Hedman, Ryan S Overbeck, Brian Cabral, Robert Konrad, and Steve Sullivan. 2019. Capture4VR: From VR Photography to VR Video. In *ACM SIGGRAPH Courses.* https://richardt.name/Capture4VR/

Michael Rubinstein, Diego Gutierrez, Olga Sorkine, and Ariel Shamir. 2010. A comparative study of image retargeting. In *ACM Trans. Graph.*, Vol. 29. 160.

Shyam Prathish Sargunam, Kasra Rahimi Moghadam, Mohamed Suhail, and Eric D Ragan. 2017. Guided head rotation and amplified head rotation: Evaluating semi-natural travel and viewing techniques in virtual reality. In *Proc. of IEEE VR.* 19–28.

Peter Scarfe and Andrew Glennerster. 2015. Using high-fidelity virtual reality to study perception in freely moving observers. *Journal of Vision* 15, 9 (2015), 3–3.

Patric Schmitz, Julian Hildebrandt, André Calero Valdez, Leif Kobbelt, and Martina Ziefle. 2018. You spin my head right round: Threshold of limited immersion for rotation gains in redirected walking. *IEEE Trans. on Vis. and Comp. Graph.* 24, 4 (2018), 1623–1632.

Ana Serrano, Incheon Kim, Zhili Chen, Stephen DiVerdi, Diego Gutierrez, Aaron Hertzmann, and Belen Masia. 2019. Motion parallax for 360° RGBD video. *IEEE Trans. on Vis. and Comp. Graph.* 25, 5 (2019).

Iwan Setyawan and Reginald L Lagendijk. 2004. Human perception of geometric distortions in images. In *Security, Steganography, and Watermarking of Multim. Contents VI*, Vol. 5306. Intern. Society for Optics and Photonics, 256–267.

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

Vincent Sitzmann, Ana Serrano, Amy Pavel, Maneesh Agrawala, Diego Gutierrez, Belen Masia, and Gordon Wetzstein. 2018. Saliency in VR: How do people explore virtual environments? *IEEE Trans. on Vis. and Comp. Graph.* 36, 4 (2018).

Mel Slater. 2009. Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philosophical Trans. of the Royal Society B: Biological Sciences* 364, 1535 (2009), 3549–3557.

Martin J Steinbach, Hiroshi Ono, and Marla E Wolf. 1991. Motion parallax judgements of depth as a function of the direction and type of head movement. *Canadian Journal of Psychology / Revue Canadienne de Psychologie* 45, 1 (1991), 92.

Frank Steinicke, Gerd Bruder, Jason Jerald, Harald Frenz, and Markus Lappe. 2010. Estimation of detection thresholds for redirected walking techniques. *IEEE Trans. on Vis. and Comp. Graph.* 16, 1 (2010), 17–27.

Evan A Suma, Seth Clark, David Krum, Samantha Finkelstein, Mark Bolas, and Zachary Warte. 2011. Leveraging change blindness for redirection in virtual environments. In *Proc. of IEEE VR.* 159–166.

Evan A Suma, Zachary Lipps, Samantha Finkelstein, David M Krum, and Mark Bolas. 2012. Impossible spaces: Maximizing natural walking in virtual environments with self-overlapping architecture. *IEEE Trans. on Vis. and Comp. Graph.* 18, 4 (2012).

Qi Sun, Anjul Patney, Li-Yi Wei, Omer Shapira, Jingwan Lu, Paul Asente, Suwen Zhu, Morgan Mcguire, David Luebke, and Arie Kaufman. 2018. Towards virtual reality infinite walking: dynamic saccadic redirection. *ACM Trans. Graph.* 37, 4 (2018), 67.

Qi Sun, Li-Yi Wei, and Arie Kaufman. 2016. Mapping virtual and physical reality. *ACM Trans. Graph.* 35, 4 (2016), 64.

Michael T Swanston and Nicholas J Wade. 1988. The perception of visual motion during movements of the eyes and the head. *Perception & Psychophysics* 43, 6 (1988).

Sebastian Sylwan, Michael Interbartolo, Liz Warren, and Felix Lajeunesse. 2019. "Space Explorers: Life in Orbit": Filming VR in Microgravity. In *ACM SIGGRAPH Production Sessions* (Los Angeles, California) *(SIGGRAPH '19).* Article 6, 1 pages.

Jayant Thatte, Jean-Baptiste Boin, Haricharan Lakshman, and Bernd Girod. 2016. Depth augmented stereo panorama for cinematic virtual reality with head-motion parallax. In *Proc. of IEEE ICME.* 1–6.

Jayant Thatte and Bernd Girod. 2018. Towards perceptual evaluation of six degrees of freedom virtual reality rendering from stacked omnistereo representation. *Electronic Imaging* 2018, 5 (2018), 352–1.

Hiroyasu Ujike and Hiroshi Ono. 2001. Depth thresholds of motion parallax as a function of head movement velocity. *Vision Research* 41, 22 (2001), 2835–2843.

Mark Wexler, Francesco Panerai, Ivan Lamouret, and Jacques Droulez. 2001. Self-motion and the perception of stationary objects. *Nature* 409, 6816 (2001), 85–88.

Mark Wexler and Jeroen JA Van Boxtel. 2005. Depth perception by the active observer. *Trends in Cognitive Sciences* 9, 9 (2005), 431–438.

Niall L Williams and Tabitha C Peck. 2019. Estimation of rotation gain thresholds considering fov, gender, and distractors. *IEEE Trans. on Vis. and Comp. Graph.* 25, 11 (2019), 3158–3168.

Betsy Williams-Sanders, Tom Carr, Gayathri Narasimham, Tim McNamara, John Rieser, and Bobby Bodenheimer. 2019. Scaling Gain and Eyeheight While Locomoting in a Large VE. In *Intern. Conf. on Human-Computer Interaction.* 277–298.

Ahmad Yoonessi and Curtis L Baker. 2011. Contribution of motion parallax to segmentation and depth perception. *Journal of Vision* 11, 9 (2011), 13–13.

Jingxin Zhang, Eike Langbehn, Dennis Krupke, Nicholas Katzakis, and Frank Steinicke. 2018. Detection Thresholds for Rotation and Translation Gains in 360° Video-Based Telepresence Systems. *IEEE Trans. on Vis. and Comp. Graph.* 24, 4 (2018), 1671–1680.

Ruimin Zhang and Scott A Kuhl. 2013. Human sensitivity to dynamic rotation gains in head-mounted displays. In *Proc. of the ACM Symp. on Applied Perception.* 71–74.

Ruimin Zhang, Bochao Li, and Scott A Kuhl. 2014. Human sensitivity to dynamic translational gains in head-mounted displays. In *Proc. of the ACM SUI.* 62–65.