Article

CIRO: The Effects of Visually Diminished Real Objects on Human Perception in Handheld Augmented Reality

Hanseob Kim ^{1,2} , Taehyung Kim ² , Myungho Lee ³ , Gerard Jounghyun Kim ² and Jae-In Hwang ^{1,*} ¹ Center for Artificial Intelligence, Korea Institute of Science and Technology (KIST), Seoul 02792, Korea; khseob0715@kist.re.kr² Department of Computer Science and Engineering, Korea University, Seoul 02841, Korea; taehyung_kim@korea.ac.kr (T.K.); gkim@korea.ac.kr (G.J.K.)³ School of Computer Science and Engineering, Pusan National University, Busan 46241, Korea; myungho.lee@pnu.edu

* Correspondence: hji@kist.re.kr

Abstract: Augmented reality (AR) scenes often inadvertently contain real world objects that are not relevant to the main AR content, such as arbitrary passersby on the street. We refer to these real-world objects as content-irrelevant real objects (CIROs). CIROs may distract users from focusing on the AR content and bring about perceptual issues (e.g., depth distortion or physicality conflict). In a prior work, we carried out a comparative experiment investigating the effects on user perception of the AR content by the degree of the visual diminishment of such a CIRO. Our findings revealed that the diminished representation had positive impacts on human perception, such as reducing the distraction and increasing the presence of the AR objects in the real environment. However, in that work, the ground truth test was staged with perfect and artifact-free diminishment. In this work, we applied an actual real-time object diminishment algorithm on the handheld AR platform, which cannot be completely artifact-free in practice, and evaluated its performance both objectively and subjectively. We found that the imperfect diminishment and visual artifacts can negatively affect the subjective user experience.

Keywords: handheld augmented reality; diminished reality; perceptual issue; visual perception; distraction; inpainting; dynamic object removal; user experience

Citation: Kim, H.; Kim, T.; Lee, M.; Kim, G.J.; Hwang, J.-I. CIRO: The Effects of Visually Diminished Real Objects on Human Perception in Handheld Augmented Reality. *Electronics* **2021**, *10*, 900. <https://doi.org/10.3390/electronics10080900>

Academic Editor: Juan M. Corchado

Received: 21 March 2021

Accepted: 4 April 2021

Published: 9 April 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Augmented reality (AR) allows users to see a mixed environment in which virtual objects are inserted into and superimposed on the real environment [1]. Such augmentations are often targeted for predesignated objects that are recognized and tracked by the AR system [2]. However, real environments are dynamic such that new objects may appear and inadvertently interfere with the main content [3–6]. For instance, the Pokémon Go application overlays the Pokémon character onto the video captured environment, making it appear as if it is standing on the ground, but a passerby can break such an illusion (see Figure 1b). We refer to these real-world objects as content-irrelevant real objects (CIROs). CIROs may distract users from focusing on the AR content and bring about perceptual problems [7,8], such as depth distortion [9,10] and physicality conflict [11]. One way to address this issue is to recognize arbitrary CIROs and even allow natural and plausible physical interactions [12,13] with the main AR content [4,11].

Instead, we applied diminished reality (DR), that is, to visually diminish the appearance of the CIROs in hopes of reducing the aforementioned perceptual issues in the handheld AR scenes. Diminished reality (DR) is a methodology to reduce the visual gaps between real and virtual scenes by erasing or eliminating out some objects (in this case, the CIROs who are the cause of the perceptual problems) from the scenario to varying degrees [14].

Figure 1. The varying degrees of diminished representation of the content-irrelevant real object: (a) semi-transparent, (b) default (as is), (c) transparent (ideal/ground truth), and (d) removed and inpainted by a real-time algorithm. The work of [7] compared conditions (a–c); and this work compares (b–d), respectively.

However, in handheld devices, DR may introduce another issue related to the inherent dual-view [15]. For example, even if the CIROs are visually eliminated from the scene, as shown through the small screen of the handheld AR device, they are still visible outside the screen in the real world, causing perceptual inconsistency. Thus, complete elimination might not be the best solution. In a prior work [7], we carried out a comparative experiment investigating the effects of DR on user perception by the degree of the visual diminishment of the CIRO (see Figure 1a–c). However, this work assumed the ideal ground truth in which the diminishment was perfect without any noticeable artifacts.

In this work, we applied an actual implementation of removing a dynamic object (such as a pedestrian) and inpainting it with the hidden background in real-time on a handheld AR platform (see Figure 1d). We report on its objective performance and also the results of an experiment, similar to the one reported in [7], in which subjective perceptual performance was assessed. For the experiment, we prepared a scenario in which a pedestrian, that is, the CIRO, walked through the AR interaction space while a user was interacting with a virtual pet. The user would observe the pedestrian invading AR scenes and experience distorted depth sensations due to the absence of occlusion effects.

This paper is organized as follows: In Section 2, we review previous research. We also provide a brief summary of the aforementioned prior experiment [7] to better set the stage for this work. Sections 3 and 4 outline the basic experimental set-up, and the real time inpainting algorithm implemented for the handheld platform and its objective performance, respectively. Section 5 details the subjective perceptual experiment, followed by an in-depth discussion and implications of the results in Section 6. Section 7 concludes the paper and discusses future research directions.

2. Related Work

2.1. Perceptual Issues in AR

Depth perception in the context of AR refers to the perceived sizes or positions of the virtual objects (e.g., augmentation) as situated in the real environment and in relation to the real objects [8–10]. Among others, the correct depiction of occlusion among objects (real or virtual) is important [16–18], whereas incorrect rendering is known to adversely affect the sense realism [19]. However, acquiring three-dimensional information of the environment and objects in it is not a trivial task [16,20], and as such, virtual objects are often added onto the real by simply overlaying them in the foreground of the screen space [21].

Recent advances in tracking [22,23] and depth estimation techniques [21,24,25] are becoming more amenable even on mobile platforms. However, gracefully handling unknown objects unexpectedly introduced into the scene, such as CIROs, is still a difficult task, especially on the relatively computationally limited handheld AR devices.

A related problem has to do with when the rendered AR scene looks physically implausible [6,19], e.g., objects floating in the air, or overlapping with the real object (i.e., physicality conflict), because of the lack of or erroneous 3D understanding of the environment. An additional provision, such as enabling the virtual objects to react to objects in the real environment in a physically plausible way, can help improve the sense of co-presence with the virtual object [4,6,26].

Another problem, inherent to the relatively small handheld displays, is the dual views problem [15]—due to the limited size of the display, large objects are visible both in the display screen and by the naked eye, but in an inconsistent way because of the difference in perspective parameters between the device's camera and the user's eye [27]. Again, employing additional sensors and adjusting the camera focus [28] can alleviate the problem, but it is practically not feasible for the handheld platforms which strive to be stand-alone, self contained, and inexpensive. In this work, we explore removing (or making it less visible) the source of the problems, the CIRO, in the first place.

2.2. Visual Stimuli Altering Our Perception

It is well known in psychology that visual stimuli, the most dominating human sensing modality, can be manipulated to change one's perception of the world [29,30]. Several studies [31–34] have applied high-visibility visual effects (VFX) augmentation, based on the visual dominance [30], to influence human perception. For example, Weir et al. [31] augmented flame effects on the participants' hands, and despite the awareness that flame effects were not real, participants felt warm. Punpongsanon et al. [35] manipulated softness perception by altering the appearances of objects when touching them.

The field of DR seeks to reduce the visibility of real objects in the environment often for more efficient interaction and focus [36–40]. It can be used not just to hide certain objects but also to affect our perception of the environment or situation in a particular way. For instance, Buchmann et al. [41] and Okumoto et al. [42] used partially transparent (diminished) hands so that objects held by the hand could be seen better. Enomoto et al. [43] used DR to remove the gallery's visiting crowd from blocking the painting so that the user could see it clearly. However, DR is not widely employed due to the heavy computational load (especially on handheld devices), and research explaining how DR might affect user perception are hard to find.

2.3. Dynamic Object Removal

The early typical object removal methodologies segment the dynamic or static objects (e.g., pedestrian) from the image and then replace them with the hidden background by leveraging multi-view capture systems [44,45]. Object segmentation itself is a difficult problem. Traditional approaches include using image subtraction, template matching, and pixel clustering [46], but recently, deep learning-based approaches have shown dramatically superior results [47]. Performances are approaching near real-time with reasonable accuracy as well [48].

The deep learning-based inpainting [49,50] approach is also quite promising compared to simply relying on multi-view images. Most inpainting algorithms operate offline over the video frames, so that they can refer to and make use of as many frames as possible to find the background and fill in the object mask [51]. ViNet [52] is considered one of the state-of-the-art deep learning-based inpainting methods in terms of its computational requirement and accuracy. However, it was not designed for or optimized for small handheld devices. In this work, we used and adapted the work of YOLACT [48] for CIRO segmentation and mask creation, and then used an inpainting algorithm that has been simplified and extended—ViNet—to focus on and inpaint for an object that passed through the scene. We ran it on the mobile platform in real-time.

2.4. Impacts of Diminishing Intensity on User Perception

In our prior work [7], the influence on user perception based on the intensity or degree of visual diminishment in an AR setting was investigated. The AR content featured interaction with a virtual pet via a pedestrian (CIRO) suddenly passing by and intruding into the scene.

Three test conditions of how the intruding pedestrian was treated were compared: Default (DF)—as is; the CIRO is made semi-transparent (STP); and the CIRO is made completely transparent (TP). That is, the intensities of diminishment were 0%, 75%, and 100%, respectively (see Figure 1a–c). This work assumed the ideal ground truth case (e.g., TP) in which the diminishment was perfect without any noticeable artifacts. The experiment was made possible by rendering pre-acquired images rather than using real-time image streams. The results are summarized in Table 1, and they generally show that the diminishment helped reduce one's distraction and the physical implausibility, and improve the sense of presence. Refer to [7] for more details. However, it remains to be seen whether such results will still hold with the deployment of actual DR methods that may exhibit noticeable artifacts, especially on the computationally limited handheld devices.

Table 1. Experimental results of [7] (* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$).

Measure	Tests			
	Cronbach	Friedman		Post-Hoc
Distraction	$\alpha = 0.91$	$\chi^2 = 33.97$,	$p < 0.001$	DF > TP ***, DF > STP ***
Visual Inconsistency	$\alpha = 0.61$	$\chi^2 = 0.63$,	$p > 0.1$	
Object Presence	$\alpha = 0.9$	$\chi^2 = 6.97$,	$p < 0.05$	DF < TP *
Object Implausibility	$\alpha = 0.67$	$\chi^2 = 8.75$,	$p < 0.05$	DF > TP **
Object Realism	$\alpha = 0.87$	$\chi^2 = 9.28$,	$p > 0.001$	DF < TP *, DF < STP *

3. Basic Experimental Set-Up

The main AR content used in the experiments showcased a virtual pet (panda) augmented into the real environment exhibiting various interactive and realistic behaviors. The pet possessed 17 user reactive or autonomous behaviors (e.g., greeting, jumping, falling over, and begging for food) with 13 facial expressions (see Figure 2). The pet was augmented, navigated, and acted on the floor (planar surface) detected using the ARCore SDK [53]. To further strengthen its perceived presence [54], a simple physical–virtual interactive behavior was added where the pet could be commanded to approach and turn on/off an actual IoT driven lamp. The panda could be controlled by voice commands.

While the user in the video interacted with the virtual pet, a pedestrian walked into the space while wandering back and forth perpendicularly to the AR camera's viewing direction. The pedestrian walked over the virtual pet, and this is considered a physicality conflict that causes perceptual issues (see Figure 3e,f).

Figure 2. Snapshots of virtual pet's behavioral animations: (a) idling, (b) greeting, (c) jumping, (d) falling over, and (e) begging for food.

The physical test environment is shown in Figure 3a. A patterned mat was put on the floor for robust positional tracking and 3D understanding of the physical space. As will be explained in more detail, our experiments were not conducted in situ, but online by the subjects watching the captured video of the AR content (with the passerby intruding the scene). Figure 3a also depicts the walking path of the pedestrian (i.e., CIRO). Thus,

two smartphones were used—one as the mobile AR platform providing the AR scene, providing the first person but fixed point of AR view, and the other for recording the entire situation (both the AR scene as shown through the first smartphone, and the larger environment (see Figure 3e–g)).

Figure 3. (a) Experimental set-up [7]. (b–d) The sequence of interactions to command the pet to turn on the lamp [7]. (e–g) The situation wherein a pedestrian walks by and intrudes on the augmented reality (AR) scene.

4. CIRO Diminishing System

System Design

The CIRO diminishing system is mainly composed of two modules for: (1) object segmentation and (2) inpainting, respectively. For detecting and segmenting the CIRO in this experiment, YOLACT [48] was used, setting it to detect and make a binary classification for the human body. The video frames from the mobile AR client are passed to this module running on the server. The result, the segmented image with the binary mask, is passed back to the mobile AR client for filling in hole with the inpainting module. YOLACT uses a fully-convolutional deep learning architecture (and a fast non-maximal suppression technique), and is considered to be faster than any previous competitive approach (reported to be 33.5 fps evaluated on a single Titan Xp [48]). The object segmentation task is broken into two parallel subtasks: (1) generating a set of prototype masks and (2) predicting per-instance mask coefficients. The masks are produced by linearly combining the prototypes with the mask coefficients.

As for inpainting of the CIRO mask, ViNet [52], one of the state-of-the-art of its kind that executes at near real-time, was adapted. ViNet is designed as a recurrent network and internally computes the flow fields from five adjacent frames for the target frame. In addition, it takes in the previously generated frame $\hat{Y}_{t-1}$, and generates the inpainted frame $\hat{Y}_t$ and the flow map $\hat{W}_{t \Rightarrow t-1}$. We employ both the flow sub-networks and mask sub-networks at four scales (1/8, 1/4, 1/2, and 1) to aggregate and synthesize the feature points progressively. Additionally, it uses 3D-2D encoder-decoder networks to complete the missing contents efficiently, and maintains the temporal consistency through recurrent feedback and a memory layer, trained with the flow and the warping loss. Despite being considered one of the best performing inpainting systems, in its original form, it is not suitable for real-time application on the handheld platform, because of the heavy computational load (e.g., it takes 65 ms on our machine equipped with Intel i7-8700K 3.70 GHz CPU and NVIDIA GTX 1080 Ti GPU for an input image size of 512×512 pixels) and requirement for the future image frame.

We created a lightweight version of ViNet to achieve real-time performance on the handheld platform, by removing the four source frame encoder-decoders (which mark the most intensive computational burden in the original architecture). The recurrent feedback and the temporal memory layer (ConvLSTM) were left intact for maintaining the temporal

consistency of the inpainting performance. The pre-trained model was used with no extra training. As a result, the processing time of 32 ms was achieved on the same machine.

Figure 4 shows the overall configuration of the CIRO diminishing system as a server-client system. The video frames captured from the mobile AR client are passed to the server for segmentation of the human body (i.e., CIRO) and creation of the CIRO mask, and then passed back to the AR client for inpainting of it. Considering the communication (TCP) and other (e.g., shared memory among YOLACT and ViNet) overheads, the total latency (i.e., inpainted background update cycle) was measured to be about 60 ms (16fps).

Figure 4. The system architecture of the content-irrelevant real object (CIRO) diminishing algorithm. The video frames from the mobile AR client are passed to the server for segmentation of the human body (i.e., CIRO) and the creation of the CIRO mask, and then passed back to the AR client for inpainting of it.

Inpainting Performance Evaluation

We compared and evaluated the inpainting performance of the original ViNet [52] and our lightweight, real-time version using the structural similarity index (SSIM) and peak signal-to-noise ratio (PSNR) [55,56]. SSIM measures the similarity in structure between the two images, and PSNR, the image's distortion. We prepared two datasets as follows: (1) 17 videos and ground-truth segmentation masks from the DAVIS [57], and (2) 17 videos from DAVIS and segmentation masks generated by the YOLACT [48]. We then generated inpainted images from both datasets using our model and ViNet, and calculated SSIM and PSNR. Our model showed similar or even better performances than those by the original ViNet in both datasets (see Table 2).

Table 2. Performance comparisons between our model and ViNet.

	Ours		ViNet [52]	
	SSIM	PSNR	SSIM	PSNR
DAVIS [57]	0.87	23.88	0.86	23.37
YOLACT [48]	0.84	23.75	0.82	22.89

Despite such a performance, our system has two limitations. First, our segmentation method (or YOLACT) may be vulnerable to fast-moving objects. When pedestrians move too quickly, incorrect segmentation masks could be generated. This in turn may distort the inpainted image. Second, if the dynamic objects occupy large areas in the image or stay for a long time, our system may not be able to restore the hidden background correctly, because it uses mainly the previous frame as the source for inpainting. Note that, however, the CIRO we consider in this paper, a pedestrian, in most cases simply passes by in the background. Additionally, note that obviously, compared to the ground truth (perfect removal and inpainting), any systems, including ours, are bound to exhibit some noticeable artifacts [52,56]. Figure 5 shows the worst-case scenario of visual artifacts

captured in our inpainting system. We can see that part of pedestrian's foot and boundaries are incompletely erased.

Figure 5. Photographs of visual artifacts from the inpaint condition. These snapshots capture a pedestrian walking every 0.3 s. The first row (with red circles) and second row are the sequences of walking from left to right and from right to left, respectively.

5. Experiment: Impacts of the Inpainting Method on User Perception

5.1. Study Design

In our prior study [7], we confirmed the positive effect of removing or diminishing the appearance/presence of the CIRO in terms of the AR user experience. However, the test condition assumed perfect inpainting performance. Here, we conducted an experiment to investigate a more realistic situation, that is, inpainting with possible artifacts, using the basic experimental set-up and system design described in Sections 3 and 4.

The main factor was the specific inpainting method with three values or test conditions:

- **Default/as-is (DF):** The participant sees the pedestrian pass by both through the AR screen and by the naked eye in their real environment. No diminishing is applied (see Figure 1b). This condition serves as the baseline with which much perceptual problems (depth perception and double view) [8] are likely to arise, as also demonstrated in our prior study [7]
- **Transparent (TP):** The CIRO, the pedestrian, is completely and perfectly removed (staged) from the AR scene and filled in, but still visible in the real world by the naked eye (see Figure 1c). The staged imagery is prepared offline using video editing tools. Note that, similarly to the prior study, this experiment also was not conducted in situ, but used offline video review (for more details, see Section 5.2). This condition serves as the ground truth of perfect CIRO removal.
- **Inpainted (IP):** The CIRO, the pedestrian, is removed from the AR scene and filled in using the system implementation described in the previous section, possibly with occasional visual artifacts (e.g., due to fast moving large pedestrians). The pedestrian is still visible in the real world by the naked eye (see Figure 1d).

In summary, the experiment was designed as 1×3 , within subject repeated measure.

5.2. Video-Based Online Survey

Due to the COVID-19 situation, the experiment was conducted as a video based online survey. That is, the three test conditions described in the previous section were recorded as videos (see Section 3 for more details) and accessed for evaluation online through a website (<https://khseob0715.github.io/DBmeSurvey> accessed on 6 April 2021). The

website started with sections for experimental instructions, collecting basic demographic and other background information (e.g., previous exposure to and familiarity with AR). Then, an arbitrary subject visiting the website would be served with the video links presented in a counter-balanced order, with the corresponding questionnaires (see Table 3). Only after completing the evaluation survey, could the subjects can proceed to the next video.

5.3. Subjective Measures

The dependent variables for this experiment measured the AR user experience through the responses to various survey questions in five categories—namely: two constructs related to the CIRO, (1) distraction and (2) visual inconsistency; and related to the perception of the virtual pet, (3) object presence [58], (4) object realism [59], and (5) object implausibility. For details, refer to the Table 3. These questions were answered right after viewing the video for the given test conditions. At the end of all three video viewings and evaluations, the subjects were asked to rank their preferences among the three, and prioritize the distracting objects or factors among the following: (1) the pedestrian outside the AR screen; (2) the appearance and behavior of the virtual pet; (3) the visual inconsistency between inside and outside the AR screen (dual-view); (4) the pedestrian's leg in the AR screen (if visible or its artifact); and (5) the latency of the system. Finally, the subject was asked to count the number of jumps made by the virtual pet as a way to measure their concentration to the main AR content.

Table 3. The questionnaire used to assess the users' perception of visually diminished pedestrian and the virtual contents. Distraction, visual inconsistency, object presence, and object implausibility were addressed with a 7-point Likert scale, and object realism with a 5-point Likert scale.

Distraction	
DS1	I was not able to entirely concentrate on the AR scene because of the person roaming around in the background.
DS2	The passerby's existence bothered me when observing and interacting with the virtual pet.
DS3	To what extent were you aware of the person passing in the AR scene (or real environment)?
DS4	I did not pay attention to the passerby.
Visual Inconsistency	
VI1	The visual mismatch between outside and inside the screen of the passerby was obvious to me.
VI2	The different visual representations of the passerby's leg in the AR scene felt awkward.
VI3	I did not notice the visual inconsistency between the AR scene and the real scene.
VI4	The passerby's leg (or body parts) in the AR scene was not felt awkward at all.
Object Presence	
OP1	I felt like Teddy was a part of the environment.
OP2	I felt like Teddy was actually there in the environment.
OP3	It seemed as though Teddy was present in the environment.
OP4	I felt as though Teddy was physically present in the environment.
Object Implausibility	
OI1	Teddy's movements/behavior in real space looked awkward.
OI2	Teddy's appearance was out of harmony with the background space.
OI3	Teddy seemed to be in a different space than the background.
OI4	I felt Teddy turned on the lamp.
Object Realism; Please rate your impression of the Teddy on these scales.	
OR1	Fake (1) to Natural (5)
OR2	Machine (1) to Animal (5)
OR3	Unconscious (1) to Conscious (5)
OR4	Artificial (1) to Lifelike (5)
OR5	Moving rigidly (1) to Moving elegantly (5)

5.4. Participants and Procedure

We recruited 34 participants from a local community and did not include the participants who carried out our prior work [7]. Amongst them, we omitted data from 8 participants who did not pass our screening criteria; 2 by trick questions and 6 by survey time which was too short or too long. Thus, data from 26 participants (13 males and 13 females, age 18–26, $M = 23.61$, $SD = 2.28$) were used in the statistical analysis. The participants read the instructions as prescribed on the website and followed them through to provide basic and background information. We asked each participant to evaluate their AR familiarity on a 5-point Likert scale, and this yielded a slightly higher level ($M = 3.23$, $SD = 0.86$) compared to our prior work ($M = 2.67$, $SD = 1.28$) [7]. Then our participants watched each test condition video (in the counterbalanced order) and counted the number of jumps by the virtual pet (a way to focus on the video) and answered the evaluation questionnaire.

5.5. Hypotheses

Based on the literature review and in consideration of our experiment conditions, we formulated the primary hypotheses as follows.

Hypothesis 1 (H1). *Subjects are distracted the most by the CIRO among various factors (as seen outside the screen space in the real environment).*

Hypothesis 2 (H2). *The more diminished the CIRO is, the more the subjects will prefer the experience. $TP > IP > DF$.*

Hypothesis 3 (H3). *The more diminished the CIRO is, the less distracted the subjects will feel. $TP > IP > DF$.*

Hypothesis 4 (H4). *Diminishing CIROs can worsen the visual inconsistency of the appearance of the CIRO to make it larger in the AR scene than outside. $TP > IP > DF$.*

Hypothesis 5 (H5). *The overall experience, including the object presence and realism, can be affected in a positive way with the CIRO diminishment, despite some visible artifacts. $TP > IP > DF$.*

5.6. Results

Figures 6 and 7, and Table 4 show the overall results and statistical analysis of our experiment. The non-parametric Friedman tests on the measures were applied at the 5% significance level. For the pairwise comparisons, we used the Wilcoxon signed-rank tests with Bonferroni adjustment.

Figure 6. Results of the distracting factors and preference ranking.

Figure 7. Results of subjective measures in the experiment (*: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$).

Table 4. Experimental and analysis results (* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$).

Measure	Tests			
	Cronbach	Friedman		Post-Hoc
Distraction	$\alpha = 0.94$	$\chi^2 = 14.56$,	$p < 0.001$	TP < IP ***
Visual Inconsistency	$\alpha = 0.78$	$\chi^2 = 35.43$,	$p < 0.001$	TP < IP ***, DF < IP ***
Object Presence	$\alpha = 0.89$	$\chi^2 = 0.58$,	$p > 0.1$	
Object Implausibility	$\alpha = 0.60$	$\chi^2 = 1.95$,	$p > 0.1$	
Object Realism	$\alpha = 0.82$	$\chi^2 = 0.68$,	$p > 0.1$	

6. Discussion

The prior experiment [7] has already shown that the users felt the least degree of distraction and preferred the AR experience of the complete and ideal removal of the CIRO, that is, the TP condition. This experiment assessed the practical situation of where the diminishment is not perfect in the sense that the object might show up unexpectedly in varied degrees of noticeable artifacts (i.e., IP condition). Our main underlying thought was that the quality of the proposed CIRO diminishment algorithm can be judged not just quantitatively, but more importantly by how its perceptual experience fares in regard to that of the perfect case (TP), or the default case (DF). We discuss the raw results presented in the Section 5.6 with regard to the hypotheses, as stated in Section 5.5.

6.1. H1: Subjects Are Most Distracted by the CIRO among Various Factors

Subjects were found to feel most distracted by the CIRO as appearing within the AR scene vs. the remaining part visible outside the screen in the real environment, as shown in Figure 6a. This result is consistent with the prior experiment [7]. Thus, it reasserts the need to remove it from the scene if possible. The removed look could cause an inconsistency with the CIRO visible in the real environment, but the results show subjects did not regard this factor as important as the CIRO's intrusion itself. We also notice that the latency was a relatively significant distractor at 21%. However, the latency must have been caused by the heavy computational load the mobile platform could bear for near real-time performance for the DR processing and rendering. The interactive virtual pet also was listed as a distractor, possibly due to the lack of graphic realism (in terms of blending into the real environment) and natural and physically plausible behaviors. Nevertheless, the results show the importance of having to handle the dynamically intruding objects for the best AR experience.

6.2. H2/H3: The More Diminished the CIRO Is, the Less Distracted the Subjects Will Feel, and This Is Their Preference

What almost naturally follows from H1 is that subjects would rate the diminished AR content to get the best user experience with the least distraction; perhaps a lesser degree of dual view or visual inconsistency improves physical plausibility, thereby affecting the

level of the object's presence (namely, the main character, the virtual pet), and even its realism. Indeed, Figure 6b shows this very result—TP was ranked higher than DF and IP. Table 4 shows that TP was judged to be the least distracting and to have less visual inconsistency than IP (both with statistical significance), and to exhibit the highest object presence, plausibility, and realism. However, no difference was found between DF and IP in the various evaluation criteria. Thus, even though IP showed good quantitative performance in terms of erasing the CIRO, perceptually, it was still insufficient to eliminate the negative factors of the CIRO (see Figure 5). Subjects have reported that this was much more so when the CIRO stayed within the AR view longer (rather than just passing by quickly). This goes to show the imperfect, not completely artifact-free IP condition caused significant visual discomfort almost as much as not erasing it at all. In our prior study, note that the semi-transparently visualized CIRO had the effect of less distraction and overall improved user experience, but no effect on the visual inconsistency (see Table 1). Similar results were found in [60,61]. Thus we can posit that it is not that the CIRO being noticeable was the problem, but that the visual artifact was aggravating. We conclude that H2 and H3 are partially supported.

6.3. H4: Diminishing CIROs May Not Worsen the Visual Inconsistency

As already indicated, AR scenes in open mobile screens can suffer from the dual view problem—where relatively large objects in the scene become broken into two pieces by the screen boundary, possibly not exactly on the same scale. Erasing or diminishing the CIRO can exacerbate this problem—part of the real object visible in the real world is now “diminished” or gone in the AR view—a physically impossible situation [7]. However, the visual inconsistency for TP was unaffected—i.e., no difference from DF. Rather, the IP showed a statistically higher visual inconsistency. Note that in the prior experiment (see Table 1), STP (semi-transparent but artifact-free CIRO representation) had no statistical differences from STP and DF for visual inconsistencies. Thus we conclude that H4 is accepted and IP exhibited high visual inconsistency—again, due to the visual artifact (e.g., occasional incorrectly inpainted rendering) and possibly other performance problems, such as perceptible latency and instability, rather than the diminished representation of the CIRO itself. This result is consistent with the prior experiment as well; see Table 1.

6.4. H5: CIRO Diminishments May Have Positive Effects on User Experience

Being less distracting and with less direct intrusion of the CIRO in the augmented content, we hypothesized that subjects would perceive virtual objects as part of the real space (object presence) and more realistically when the visibility of the CIRO is reduced in the AR scenes. H5 is partially supported for reasons similar to those mentioned previously. In the prior study [7], which used ideally removed methods, the secondary effects we considered on the user experience had significant differences in a positive direction (see Table 1: object presence, object implausibility, and object realism).

Since the passerby, i.e., the CIRO, walked over or nearby the virtual pet, participants could observe incorrect occlusion and thus physically implausible situations. Both are well-known factors that can reduce the presence and realism of virtual objects [11,16]. In this respect, removing the pedestrian in the AR scene also means eliminating the causes of perceptual problems. Thus this could be interpreted as that participants experienced fewer flaws in AR content in the TP condition than the DF condition. Semi-transparency also would have reduced the chances of recognizing such flaws; i.e., some might have been recognized.

Although the performance (see Table 2) of our inpainting method was slightly better than the existing method [52], visual artifacts inherently present in the scene seemed to play a much stronger role in the secondary effects in this experiment. As Blau et al. [62] reported a counter-intuitive phenomenon that distortion and perceptual quality are at odds with each other, we do not fully understand how the artifacts of inpainting methods affect users' perceptions yet. Thus, further investigation in this regard is needed.

6.5. Limitations

The main limitation of our study stems from the fact that the experiment was not conducted in situ and assessed through video recordings (albeit due to the current pandemic) from a fixed view point (no camera movement). Recordings of user controlled and changing AR views could have been used, but it would interfere with the making the test conditions equal, and would have introduced much difficulty in the recording process itself. Camera movement is still regarded an important factor, as it can affect the all the assessment criteria, such as the distraction, visual inconsistency, and object presence/realism.

Our segmentation and inpainting algorithm was tuned to detect only human interference with simply assumed movement types, such as simply passing by and not reappearing. Compromises had to be made to port the diminishing system to run in real time on the handheld platform. With more varied CIRO behaviors, more visual artifacts might occur and again affect the various assessment criteria. However, it still seems clear that the artifacts, rather than the choice of the diminished representation, will play the most important role in the overall user experience. Thus, future work must focus on what kind of artifacts will cause the most serious problems and find an algorithmic approach to eliminating or reducing them further.

7. Conclusions

In this paper, we proposed a deep learning-based inpainting method for real-time diminishment on the mobile AR platform, and evaluated its performance both objectively and subjectively. The qualitative and perceptual user study indicated that the visual diminishment had some positive impacts on perceptual issues (e.g., depth distortion and distraction) in handheld AR. Ideally diminished CIRO could also improve the realism and spatial presence of the virtual objects in real space. However, we also found that the inconsistent artifacts by the diminishment or inpainting can negatively affect the user's experience, despite the reasonable objective performance of the diminishment algorithm. However, it really depends on the type, frequency, and the degree of noticeability of the artifacts. Future mobile implementations will continue to improve, and at some point, should reach a level that is perceived as a perfect case and no different. At this point, given that our implementation is fairly up-to-date with good objective performance as compared to the current state of the art that runs on the server level computers, current DR on the mobile platforms must consider this issue and provide other ways for compensation and deal with the CIROs. In the future, we plan to continue expanding our work for more practical cases where users (points of view) can move freely.

Author Contributions: Conceptualization, H.K. and M.L.; formal analysis, M.L. and H.K.; funding acquisition, G.J.K. and J.-I.H.; investigation, H.K., T.K., and M.L.; methodology, H.K., T.K., and M.L.; software, H.K. and T.K.; supervision, G.J.K. and J.-I.H.; writing—original draft, H.K.; writing—review and editing, all authors. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by a National Research Council of Science and Technology (NST) grant by the Korean Government through the Ministry of Science and ICT (MSIT) (CRC-20-02-KIST) and also by the MSIT under the ITRC (Information Technology Research Center) support program (IITP-2021-2016-0-00312) supervised by the IITP (Institute for Information and communications Technology Planning and Evaluation).

Institutional Review Board Statement: This work involved human subjects. However, our study used an anonymous web survey and not a face-to-face interview. Thus, ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements.

Informed Consent Statement: A consent form was not provided because this experiment was conducted through an anonymous web survey.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Acknowledgments: We express sincere gratitude to the participants in the experiments. We also thank the reviewers for their valuable contributions.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Azuma, R.T. A Survey of Augmented Reality. *Presence Teleoperators Virtual Environ.* **1997**, *6*, 355–385. [\[CrossRef\]](#)
2. Kim, H.; Ali, G.; Pastor, A.; Lee, M.; Kim, G.J.; Hwang, J.I. Silhouettes from Real Objects Enable Realistic Interactions with a Virtual Human in Mobile Augmented Reality. *Appl. Sci.* **2021**, *11*, 2763. [\[CrossRef\]](#)
3. Taylor, A.V.; Matsumoto, A.; Carter, E.J.; Plopski, A.; Admoni, H. Diminished Reality for Close Quarters Robotic Telemanipulation. In Proceedings of the 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Las Vegas, NV, USA, 25–29 October 2020; pp. 11531–11538. [\[CrossRef\]](#)
4. Norouzi, N.; Kim, K.; Lee, M.; Schubert, R.; Erickson, A.; Bailenson, J.; Bruder, G.; Welch, G. Walking your virtual dog: Analysis of awareness and proxemics with simulated support animals in augmented reality. In Proceedings of the 2019 IEEE International Symposium on Mixed and Augmented Reality (ISMAR), Beijing, China, 14–18 October 2019; pp. 157–168.
5. Kang, H.; Lee, G.; Han, J. Obstacle detection and alert system for smartphone ar users. In Proceedings of the 25th ACM Symposium on Virtual Reality Software and Technology, Parramatta, Australia, 12–15 November 2019; pp. 1–11.
6. Kim, H.; Lee, M.; Kim, G.J.; Hwang, J.I. The Impacts of Visual Effects on User Perception with a Virtual Human in Augmented Reality Conflict Situations. *IEEE Access* **2021**, *9*, 35300–35312. [\[CrossRef\]](#)
7. Kim, H.; Kim, T.; Lee, M.; Kim, G.J.; Hwang, J.I. Don't Bother Me: How to Handle Content-Irrelevant Objects in Handheld Augmented Reality. In Proceedings of the 26th ACM Symposium on Virtual Reality Software and Technology, Virtual Event, 1–4 November 2020; ACM: New York, NY, USA, 2020; pp. 1–5. [\[CrossRef\]](#)
8. Kruijff, E.; Swan, J.E.; Feiner, S. Perceptual issues in augmented reality revisited. In Proceedings of the 9th IEEE International Symposium on Mixed and Augmented Reality 2010: Science and Technology, ISMAR 2010, Seoul, Korea, 13–16 October 2010; pp. 3–12. [\[CrossRef\]](#)
9. Berning, M.; Kleinert, D.; Riedel, T.; Beigl, M. A study of depth perception in hand-held augmented reality using autostereoscopic displays. In Proceedings of the 2014 IEEE International Symposium on Mixed and Augmented Reality (ISMAR), Munich, Germany, 10–12 September 2014; pp. 93–98. [\[CrossRef\]](#)
10. Diaz, C.; Walker, M.; Szafir, D.A.; Szafir, D. Designing for depth perceptions in augmented reality. In Proceedings of the 2017 IEEE International Symposium on Mixed and Augmented Reality (ISMAR), Nantes, France, 9–13 October 2017; pp. 111–122.
11. Kim, K.; Bruder, G.; Welch, G. Exploring the effects of observed physicality conflicts on real-virtual human interaction in augmented reality. In Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology—VRST '17, Gothenburg, Sweden, 8–10 November 2017; ACM Press: New York, NY, USA, 2017; pp. 1–7. [\[CrossRef\]](#)
12. Kim, K.; Boelling, L.; Haesler, S.; Bailenson, J.; Bruder, G.; Welch, G.F. Does a digital assistant need a body? The influence of visual embodiment and social behavior on the perception of intelligent virtual agents in AR. In Proceedings of the 2018 IEEE International Symposium on Mixed and Augmented Reality (ISMAR), Munich, Germany, 16–20 October 2018; pp. 105–114.
13. Lee, M.; Norouzi, N.; Bruder, G.; Wisniewski, P.; Welch, G. Mixed Reality Tabletop Gameplay: Social Interaction with a Virtual Human Capable of Physical Influence. *IEEE Trans. Vis. Comput. Graph.* **2019**, *24*. [\[CrossRef\]](#) [\[PubMed\]](#)
14. Mori, S.; Ikeda, S.; Saito, H. A survey of diminished reality: Techniques for visually concealing, eliminating, and seeing through real objects. *IPSJ Trans. Comput. Vis. Appl.* **2017**, *9*, 1–14. [\[CrossRef\]](#)
15. Čopić Pucihar, K.; Coulton, P.; Alexander, J. The use of surrounding visual context in handheld AR: Device vs. user perspective rendering. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Toronto, ON, Canada, 26 April–1 May 2014; pp. 197–206.
16. Sanches, S.R.R.; Tokunaga, D.M.; Silva, V.F.; Sementille, A.C.; Tori, R. Mutual occlusion between real and virtual elements in Augmented Reality based on fiducial markers. In Proceedings of the 2012 IEEE Workshop on the Applications of Computer Vision (WACV), Breckenridge, CO, USA, 9–11 January 2012; pp. 49–54. [\[CrossRef\]](#)
17. Zhou, Y.; Ma, J.T.; Hao, Q.; Wang, H.; Liu, X.P. A novel optical see-through head-mounted display with occlusion and intensity matching support. In Proceedings of the International Conference on Technologies for E-Learning and Digital Entertainment, Hong Kong, China, 11–13 June 2007; pp. 56–62.
18. Cakmakci, O.; Ha, Y.; Rolland, J.P. A compact optical see-through head-worn display with occlusion support. In Proceedings of the Third IEEE and ACM International Symposium on Mixed and Augmented Reality, Arlington, VA, USA, 5 November 2004; pp. 16–25.
19. Kim, K.; Maloney, D.; Bruder, G.; Bailenson, J.N.; Welch, G.F. The effects of virtual human's spatial and behavioral coherence with physical objects on social presence in AR. *Comput. Animat. Virtual Worlds* **2017**, *28*, 1–9. [\[CrossRef\]](#)
20. Klein, G.; Drummond, T. Sensor fusion and occlusion refinement for tablet-based AR. In Proceedings of the Third IEEE and ACM International Symposium on Mixed and Augmented Reality, Arlington, VA, USA, 5 November 2004; pp. 38–47.

21. Tang, X.; Hu, X.; Fu, C.W.; Cohen-Or, D. GrabAR: Occlusion-aware Grabbing Virtual Objects in AR. In Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology, Virtual Event, 20–23 October 2020; pp. 697–708.
22. Holynski, A.; Kopf, J. Fast Depth Densification for Occlusion-Aware Augmented Reality. *ACM Trans. Graph.* **2018**, *37*. [\[CrossRef\]](#)
23. Runz, M.; Buffier, M.; Agapito, L. MaskFusion: Real-Time Recognition, Tracking and Reconstruction of Multiple Moving Objects. In Proceedings of the 2018 IEEE International Symposium on Mixed and Augmented Reality (ISMAR), Munich, Germany, 16–20 October 2018; pp. 10–20. [\[CrossRef\]](#)
24. Godard, C.; Mac Aodha, O.; Firman, M.; Brostow, G.J. Digging into self-supervised monocular depth estimation. In Proceedings of the IEEE International Conference on Computer Vision, Seoul, Korea, 27–28 October 2019; pp. 3828–3838.
25. Luo, X.; Huang, J.B.; Szeliski, R.; Matzen, K.; Kopf, J. Consistent video depth estimation. *ACM Trans. Graph.* **2020**, *39*, 71: 1–71: 13. [\[CrossRef\]](#)
26. Kim, K.; Bruder, G.; Welch, G.F. Blowing in the Wind: Increasing Copresence with a Virtual Human via Airflow Influence in Augmented Reality. In Proceedings of the International Conference on Artificial Reality and Telexistence Eurographics Symposium on Virtual Environments, Limassol, Cyprus, 7–9 November 2018; pp. 183–190. [\[CrossRef\]](#)
27. Mohr, P.; Tatzgern, M.; Grubert, J.; Schmalstieg, D.; Kalkofen, D. Adaptive user perspective rendering for handheld augmented reality. In Proceedings of the 2017 IEEE Symposium on 3D User Interfaces (3DUI), Los Angeles, CA, USA, 18–19 March 2017; pp. 176–181.
28. Baričević, D.; Höllerer, T.; Sen, P.; Turk, M. User-perspective augmented reality magic lens from gradients. In Proceedings of the 20th ACM Symposium on Virtual Reality Software and Technology—VRST '14, Edinburgh, UK, 11–13 November 2014; ACM Press: New York, NY, USA, 2014; pp. 87–96. [\[CrossRef\]](#)
29. Capó-Aponte, J.E.; Temme, L.A.; Task, H.L.; Pinkus, A.R.; Kalich, M.E.; Pantle, A.J.; Rash, C.E.; Russo, M.; Letowski, T.; Schmeisser, E. Visual perception and cognitive performance. In *Helmet-Mounted Displays: Sensation, Perception and Cognitive Issues*; U.S. Army Aeromedical Research Laboratory: Fort Rucker, AL, USA, 2009; pp. 335–390.
30. Spence, C. Explaining the Colavita visual dominance effect. In *Progress in Brain Research*; Elsevier: Amsterdam, The Netherlands, 2009; Volume 176, pp. 245–258. [\[CrossRef\]](#)
31. Weir, P.; Sandor, C.; Swoboda, M.; Nguyen, T.; Eck, U.; Reitmayr, G.; Day, A. Burnar: Involuntary heat sensations in augmented reality. In Proceedings of the 2013 IEEE Virtual Reality (VR), Lake Buena Vista, FL, USA, 18–20 March 2013; pp. 43–46. [\[CrossRef\]](#)
32. Erickson, A.; Bruder, G.; Wisniewski, P.J.; Welch, G.F. Examining Whether Secondary Effects of Temperature-Associated Virtual Stimuli Influence Subjective Perception of Duration. In Proceedings of the 2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), Atlanta, GA, USA, 22–26 March 2020; pp. 493–499. [\[CrossRef\]](#)
33. Blaga, A.D.; Frutos-Pascual, M.; Creed, C.; Williams, I. Too Hot to Handle: An Evaluation of the Effect of Thermal Visual Representation on User Grasping Interaction in Virtual Reality. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, Honolulu, HI, USA, 25–30 April 2020; pp. 1–16. [\[CrossRef\]](#)
34. von Castell, C.; Hecht, H.; Oberfeld, D. Measuring Perceived Ceiling Height in a Visual Comparison Task. *Q. J. Exp. Psychol.* **2017**, *70*, 516–532. [\[CrossRef\]](#) [\[PubMed\]](#)
35. Punpongsanon, P.; Iwai, D.; Sato, K. SoftAR: Visually manipulating haptic softness perception in spatial augmented reality. *IEEE Trans. Vis. Comput. Graph.* **2015**, *21*, 1279–1288. [\[CrossRef\]](#) [\[PubMed\]](#)
36. Kawai, N.; Sato, T.; Nakashima, Y.; Yokoya, N. Augmented reality marker hiding with texture deformation. *IEEE Trans. Vis. Comput. Graph.* **2016**, *23*, 2288–2300. [\[CrossRef\]](#) [\[PubMed\]](#)
37. Korkalo, O.; Aittala, M.; Siltanen, S. Light-weight marker hiding for augmented reality. In Proceedings of the 2010 IEEE International Symposium on Mixed and Augmented Reality, Seoul, Korea, 13–16 October 2010; pp. 247–248.
38. Siltanen, S. Diminished reality for augmented reality interior design. *Vis. Comput.* **2017**, *33*, 193–208. [\[CrossRef\]](#)
39. Guida, J.; Sra, M. Augmented Reality World Editor. In Proceedings of the 26th ACM Symposium on Virtual Reality Software and Technology; Association for Computing Machinery VRST'20, New York, NY, USA, 2–4 November 2020; doi:10.1145/3385956.3422125. [\[CrossRef\]](#)
40. Queguiner, G.; Fradet, M.; Rouhani, M. Towards mobile diminished reality. In Proceedings of the 2018 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct), Munich, Germany, 16–20 October 2018; pp. 226–231.
41. Buchmann, V.; Nilsen, T.; Billingham, M. Interaction with partially transparent hands and objects. In Proceedings of the 6th Australasian Conference on User Interface, Newcastle, Australia, 30 January–3 February 2005; Volume 40, pp. 17–20. [\[CrossRef\]](#)
42. Okumoto, H.; Yoshida, M.; Umemura, K. Realizing Half-Diminished Reality from video stream of manipulating objects. In Proceedings of the 4th IGNITE Conference and 2016 International Conference on Advanced Informatics: Concepts, Theory and Application, ICAICTA 2016, Penang, Malaysia, 16–19 August 2016; doi:10.1109/ICAICTA.2016.7803120. [\[CrossRef\]](#)
43. Enomoto, A.; Saito, H. Diminished reality using multiple handheld cameras. *ACCV'07 Workshop on Multi-dimensional and Multi-view Image Processing. CiteSeer* **2007**, *7*, 130–135.
44. Hasegawa, K.; Saito, H. Diminished reality for hiding a pedestrian using hand-held camera. In Proceedings of the 2015 IEEE International Symposium on Mixed and Augmented Reality Workshops, Fukuoka, Japan, 29 September–3 October 2015; pp. 47–52.
45. Yagi, K.; Hasegawa, K.; Saito, H. Diminished reality for privacy protection by hiding pedestrians in motion image sequences using structure from motion. In Proceedings of the 2017 IEEE International Symposium on Mixed and Augmented Reality (ISMAR-Adjunct), Nantes, France, 9–13 October 2017; pp. 334–337.

46. Comaniciu, D.; Meer, P. Mean shift: A robust approach toward feature space analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* **2002**, *24*, 603–619. [\[CrossRef\]](#)
47. Chen, L.C.; Zhu, Y.; Papandreou, G.; Schroff, F.; Adam, H. Encoder-decoder with atrous separable convolution for semantic image segmentation. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 801–818.
48. Bolya, D.; Zhou, C.; Xiao, F.; Lee, Y.J. YOLACT: Real-time Instance Segmentation. In Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Korea, 27–28 October 2019; pp. 9156–9165. [\[CrossRef\]](#)
49. Huang, J.B.; Kang, S.B.; Ahuja, N.; Kopf, J. Temporally coherent completion of dynamic video. *ACM Trans. Graph. (TOG)* **2016**, *35*, 1–11. [\[CrossRef\]](#)
50. Xu, R.; Li, X.; Zhou, B.; Loy, C.C. Deep flow-guided video inpainting. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 16–20 June 2019; pp. 3718–3727. [\[CrossRef\]](#)
51. Newson, A.; Almansa, A.; Fradet, M.; Gousseau, Y.; Pérez, P. Video inpainting of complex scenes. *Siam J. Imaging Sci.* **2014**, *7*, 1993–2019. [\[CrossRef\]](#)
52. Kim, D.; Woo, S.; Lee, J.Y.; So Kweon, I. Deep Video Inpainting. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 16–20 June 2019; pp. 5792–5801.
53. Google. ARCore. Available online: <https://developers.google.com/ar> (accessed on 6 April 2021).
54. Lee, M.; Norouzi, N.; Bruder, G.; Wisniewski, P.J.; Welch, G.F. The physical-virtual table: Exploring the effects of a virtual human's physical influence on social interaction. In Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology, Tokyo, Japan, 28 November–1 December 2018; pp. 1–11.
55. Wang, Z.; Bovik, A.C.; Sheikh, H.R.; Simoncelli, E.P. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process.* **2004**, *13*, 600–612. [\[CrossRef\]](#) [\[PubMed\]](#)
56. Oh, S.W.; Lee, S.; Lee, J.Y.; Kim, S.J. Onion-peel networks for deep video completion. In Proceedings of the IEEE International Conference on Computer Vision, Seoul, Korea, 27–28 October 2019; pp. 4403–4412.
57. Perazzi, F.; Pont-Tuset, J.; McWilliams, B.; Van Gool, L.; Gross, M.; Sorkine-Hornung, A. A benchmark dataset and evaluation methodology for video object segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 724–732.
58. Vorderer, P.; Wirth, W.; Gouveia, F.R.; Biocca, F.; Saari, T.; Jäncke, F.; Böcking, S.; Schramm, H.; Gysbers, A.; Hartmann, T.; et al. MEC Spatial Presence Questionnaire (MEC-SPQ): Short Documentation and Instructions for Application. Available online: https://www.researchgate.net/publication/318531435_MEC_spatial_presence_questionnaire_MEC-SPQ_Short_documentation_and_instructions_for_application (accessed on 9 April 2021).
59. Bartneck, C.; Kulić, D.; Croft, E.; Zoghbi, S. Measurement Instruments for the Anthropomorphism, Animacy, Likeability, Perceived Intelligence, and Perceived Safety of Robots. *Int. J. Soc. Robot.* **2009**, *1*, 71–81. [\[CrossRef\]](#)
60. Sandor, C.; Cunningham, A.; Dey, A.; Mattila, V.V. An augmented reality X-ray system based on visual saliency. In Proceedings of the 2010 IEEE International Symposium on Mixed and Augmented Reality, Seoul, Korea, 13–16 October 2010; pp. 27–36.
61. Enns, J.T.; Di Lollo, V. What's new in visual masking? *Trends Cogn. Sci.* **2000**, *4*, 345–352. [\[CrossRef\]](#)
62. Blau, Y.; Michaeli, T. The Perception-Distortion Tradeoff. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 6228–6237. [\[CrossRef\]](#)