

Article

Automatic Control of Virtual Cameras for Capturing and Sharing User Focus and Interaction in Collaborative Virtual Reality

Junhyeok Lee ¹, Dongkeun Lee ¹, Seowon Han ², Hyun K. Kim ³ and Kang Hoon Lee ^{1,2,*}

¹ Department of Computer Science, Kwangwoon University, 20, Gwangun-ro, Nowon-gu, Seoul 01897, Republic of Korea; junhyeoklee@kw.ac.kr (J.L.); dongkeun.lee@kw.ac.kr (D.L.)

² School of Software, Kwangwoon University, 20, Gwangun-ro, Nowon-gu, Seoul 01897, Republic of Korea; sewong00@kw.ac.kr

³ School of Information Convergence, Kwangwoon University, 20, Gwangun-ro, Nowon-gu, Seoul 01897, Republic of Korea; hyunkkim@kw.ac.kr

* Correspondence: kang@kw.ac.kr

Abstract: As VR technology advances and network speeds rise, social VR platforms are gaining traction. These platforms enable multiple users to socialize and collaborate within a shared virtual environment using avatars. Virtual reality, with its ability to augment visual information, offers distinct advantages for collaboration over traditional methods. Prior research has shown that merely sharing another person's viewpoint can significantly boost collaborative efficiency. This paper presents an innovative non-verbal communication technique designed to enhance the sharing of visual information. By employing virtual cameras, our method captures where participants are focusing and what they are interacting with, then displays these data above their avatars. The direction of the virtual camera is automatically controlled by considering the user's gaze direction, the position of the object the user is interacting with, and the positions of other objects around that object. The automatic adjustment of these virtual cameras and the display of captured images are symmetrically conducted for all participants engaged in the virtual environment. This approach is especially beneficial in collaborative settings, where multiple users work together on a shared structure of multiple objects. We validated the effectiveness of our proposed technique through an experiment with 20 participants tasked with collaboratively building structures using block assembly.

Keywords: multi-user collaboration; virtual reality; virtual camera



Citation: Lee, J.; Lee, D.; Han, S.; Kim, H.K.; Lee, K.H. Automatic Control of Virtual Cameras for Capturing and Sharing User Focus and Interaction in Collaborative Virtual Reality.

Symmetry **2024**, *16*, 228. <https://doi.org/10.3390/sym16020228>

Academic Editor: Christos Volos

Received: 15 January 2024

Revised: 2 February 2024

Accepted: 10 February 2024

Published: 13 February 2024



Copyright: © 2024 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

The rise in the adoption of virtual reality (VR) equipment, coupled with advancements in computer network speeds, has paved the way for the evolution of social VR. Anchored in these two technological advancements, social VR offers a platform where remote users can converge, communicate, and collaborate within a shared virtual realm, represented by their avatars [1]. One of the commercial examples is VRChat, where users, empowered by VR technology, can craft their own avatars. Then, once inside a virtual 3D space, they have the liberty to maneuver their avatars and engage in interactions with fellow users [2]. Within these social VR platforms, collaboration emerges as a predominant activity. Users can delve into a myriad of collaborative endeavors, ranging from architectural design and music composition to practicing physical exercises [3–5].

Collaborative activities in virtual environments often necessitate multiple users working together to manipulate virtual objects. Take architectural design as an example: multiple participants might need to collaboratively adjust structural elements like columns or beams. Yet, the nature of virtual reality presents challenges. Its communication channels are more constrained; its precision is lower; its expressive range is narrower than that of the real world. This makes collaborative manipulations particularly challenging. A notable difficulty lies in replicating non-verbal cues, such as facial expressions and gaze direction,

which are readily available in the real world, but hard to reproduce accurately in most current social VR systems.

To address this, researchers have explored alternative non-verbal communication methods that might be challenging to implement in the real world, but are more feasible in virtual reality. One such method is the *video mirror* technique, which mirrors the scene captured in one user's field of view for other users [6]. This method provides a wealth of information by rendering scenes from the perspectives of different users. Consequently, users can engage in more-informed and -effective manipulative tasks, having a broader understanding of both the state of their peers and the overarching structure of the virtual environment. Despite the advantages of the *video mirror* technique, it has a limitation in that it solely relies on the users' current field of view and, thus, does not guarantee the provision of visual information closely related to the manipulation. For instance, when a user tries to link one object while holding another, simply sharing that user's perspective would not convey information about the actual objects being manipulated. If such additional information can be shared, it would greatly assist other users in understanding that user's intent of manipulation.

In this paper, we introduce a novel approach that extends beyond merely sharing a user's immediate field of view. Our method captures and disseminates visual information, providing a more-insightful perspective on where each user is directing his/her attention and his/her interactions. To achieve this, we allocated a virtual camera to every user in a symmetrical setting. This camera autonomously navigates to an optimal location and orientation, factoring in the user's gaze direction, the position of the manipulated object, and the proximity of the nearest object, ensuring a holistic view of these elements. To stabilize the visual feed and prevent abrupt fluctuations, we employed the PD control method for the camera. Every frame, rendered in alignment with the virtual camera's perspective, is showcased on a compact 2D rectangular panel positioned above the corresponding avatar's head. To guarantee that each user can thoroughly view the renderings from every other user's camera, we utilized the billboarding technique, adjusting all panels to be perpendicular to the viewer's line of sight. The automatic adjustment of these virtual cameras, along with the presentation of the captured images, is carried out symmetrically for all participants in the virtual environment.

To assess the effectiveness of our proposed method, we established a virtual reality environment where multiple users collaboratively assemble blocks to construct the desired structures. We then conducted an experiment with 20 participants to compare how the efficiency and accuracy of collaboration changed when using our proposed method versus not using it. After comparing the completion time and error rate for two distinct tasks, we found that, for assembling relatively small structures with sufficient visibility, our proposed method yielded higher efficiency and accuracy compared to not using it. However, we also found that, in situations where assembling relatively large structures makes it difficult to check other users' rendered panels, the increased burden of visual information processing can actually lead to lower efficiency and accuracy. These observations suggest that future research should focus on extending the method proposed in this paper to minimize the burden of visual information processing.

2. Related Work

2.1. Three-Dimensional User Interface

Research into 3D user interfaces has been steadily progressing since the 1990s [7]. While there have been advancements with traditional 2D input devices like the mouse and keyboard, there is concurrent exploration into 3D input mechanisms offering six degrees of freedom. Moreover, immersion-enhancing input tools, such as gloves and motion controllers, have been studied, especially in conjunction with Head-Mounted Displays (HMDs) [8].

There have been proposals on various techniques to proficiently select and interact with specific objects in a 3D environment among numerous others [9]. One rudimentary

method showcases a virtual hand that mirrors the movement of a user's actual hand [10]. Much like our interaction in reality, objects in this virtual space are chosen and handled by directly engaging with this virtual representation. This strategy works optimally for objects within the user's grasp, but falters for distant items. To address this, several solutions have been introduced. One such approach speeds up the virtual hand's movement relative to the user's actual hand when it extends a certain distance from the body, facilitating the selection of farther objects [11]. Another strategy employs a ray or beam projected from the virtual hand to select any object it crosses [12].

In the system devised for our research, objects in close proximity can be effortlessly secured by a virtual hand that mirrors the user's actual hand in real-time. For objects that are further away, upon intersection with a ray from the virtual hand, they can be drawn closer for a more-straightforward grasp. This design ensures a fluid and comprehensive object-selection process in the virtual space, irrespective of the object's distance.

2.2. Multi-User Collaboration in VR

In the realm of virtual reality, extensive research has spanned various domains to foster multi-user collaboration for a broad range of purposes. For instance, Du et al. introduced CoVR, a platform that facilitates remote collaboration among team members in a building information modeling project. Through CoVR, participants can converge within a virtual representation of the building, engaging in face-to-face discussions and interactive sessions with the simulated environment. Their findings underscore the potential to bolster communication in construction-related tasks [3].

Galambos et al. presented the Virtual Collaboration Arena (VirCA) platform, leveraging virtual reality as a communication medium. It offers diverse collaborative tools tailored for testing and training within intricate manufacturing systems [13]. Li et al. embarked on an exploration of photo-sharing within social VR settings. Their experiments concluded that the immersion experienced rivals that of traditional face-to-face photo-sharing encounters [14]. Chen et al. crafted a virtual system, allowing individuals to remotely congregate and practice Chinese Taichi. Their work accentuates how immersive environments can expedite the learning trajectory in contrast to non-immersive counterparts [5]. Men and Bryan-Kinns introduced the LeMo system, designed to support duos in musical creation, enhanced by visual aids like free-form 3D annotations. Their analysis delved into the interaction dynamics amongst users within this framework [4]. Mei et al. fashioned CakeVR, a social VR platform that empowers cake customization enthusiasts to remotely design cakes alongside a professional pastry chef. An assessment involving ten experts yielded guidelines for crafting co-design tools in upcoming social VR landscapes [15].

In our study, we ventured to establish a communal virtual space for users to collaboratively assemble blocks. Predicated on this foundation, we initiated tasks where groups constructed edifices, or in certain scenarios, a designated guide spearheaded the assembly, with peers replicating the construction process.

2.3. Augmenting Visual Information

Effective collaboration in virtual reality hinges on seamless verbal and non-verbal interactions among participants. While current social VR platforms offer unhindered verbal communication, the spotlight in research has shifted towards enhancing non-verbal cues like facial expressions, eye movements, body gestures, and visual indicators. This paper introduces a novel approach to non-verbal communication, focusing on the amplification of visual data. This allows collaborators to better understand the intentions and insights of their peers in the virtual environment [16].

Such non-verbal communication techniques, which manipulate or augment visual information, have gained traction. Zaman and colleagues devised the 'nRoom' system, enabling users to share perspectives during spatial design endeavors. Their experiments, involving an expert–novice duo in design collaboration, illustrated the potential of enriched non-verbal cues [17]. In a similar vein, Wang et al. introduced a method for partial

viewpoint sharing, blending one's perspective with another's. This seamless integration of multiple viewpoints facilitated quicker and more-efficient information exchange than traditional view-switching techniques [18]. Building on this, Bovo et al. suggested visualizing a user's focus area as a cone rather than directly sharing views. Such a visualization assists collaborators in discerning where each participant's attention lies. This approach, tested in data analysis collaborations, proved effective at reducing cognitive load and boosting collaboration efficiency [19].

Freiwald et al. explored three methods of sharing viewpoints among collaborators: highlighting objects, using a 3D cone representation for the field of view, and mirroring another's visual feed (*video mirror*). Their experiments showed the video mirror technique to be the most-dependable [6]. Interestingly, these mirroring techniques are not just tools for collaboration; they are also leveraged for intricate 3D manipulations. Ha and Lee proposed a 'virtual mirror' technique that provides users with an additional view by automatically moving the virtual mirror to reflect not only the object being manipulated in the 3D space, but also adjacent objects. This allows users to understand even parts they cannot directly see and precisely control interactions between objects [20]. The method our paper presents is rooted in the *video mirror* concept, but extends beyond mere viewpoint sharing. It provides insights into the objects with which each user interacts, making the process more suited to collaborative tasks that involve manipulation.

3. System Overview

For our usability evaluation, we developed a social VR platform allowing multiple users to concurrently immerse themselves in a shared virtual environment and interact both with the virtual realm and among themselves (see Figure 1). Participants access this environment using the Meta Quest 2 device, wearing a virtual reality HMD and gripping motion controllers in each hand. Upon initial entry, users have the flexibility to personalize their avatars—choosing from an array of facial features, clothing, and accessories.

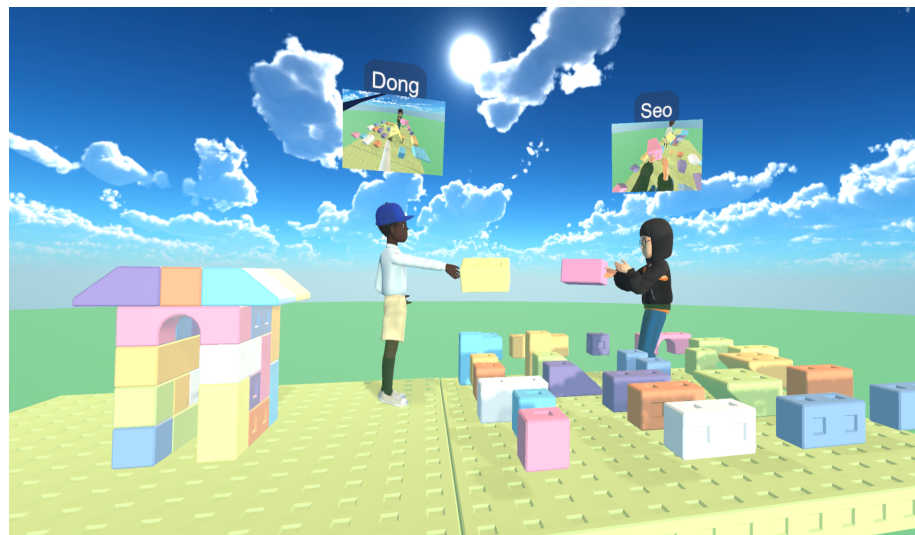


Figure 1. An example screen of the social VR platform developed to demonstrate our method.

After customizing their avatars, users have the choice to either create a new virtual room or enter an already existing one. In this digital realm, every movement users make is naturally reflected by their avatars, offering an intuitive interface. Moreover, when multiple users share a room, their interactions unfold in real-time, maintaining consistency between the avatars and the virtual world. Navigation in this virtual setting is achieved through three methods: walking, thumbstick navigation, and teleportation. For movements in their immediate surroundings, users can walk within the confines defined by their VR equipment. To traverse medium distances, adjusting the thumbstick in the chosen direction gradually changes the user's virtual position. For swifter relocations over greater distances,

users can aim an arched ray to their intended destination and, upon pressing the trigger button, instantly teleport there.

Within the virtual realm, an array of pre-configured assembly blocks is randomly placed. These blocks are designed in various geometric shapes, including cubes, triangular prisms, and cylindrical prisms. When assembled, they can form an expansive variety of intricate structures. Each block features one or more indented sections, functioning similarly to magnets, facilitating connections with adjacent blocks. Unique to these connectors, however, is the absence of polarity, allowing versatile connections in any orientation. To interact with a block, users approach it and press the grip button or direct a hand ray towards the block and activate the grip button. When aligning one block close to another, a visual indication of the impending connection appears. Releasing the grip button at this juncture effectively merges the two blocks. If users wish to separate a previously joined block, they can hold onto the block, pause momentarily, and then, tug it away from its initial placement, effectively disengaging it from the attached structure.

Each avatar will have a panel positioned above its head displaying images captured by a virtual camera. The position of the virtual camera is set at the location of the avatar's head. The direction in which the virtual camera points is automatically determined by considering the direction of the avatar's gaze, the location of the object the avatar is interacting with, and the positions of other objects nearby that object. Each user can continuously check where other users are focusing and what they are interacting with through the display panels placed above their avatars. Providing additional visual information through these display panels allows users to engage in non-verbal communication with each other, enhancing the efficiency and accuracy of collaboration. We will refer to this system of controlling the virtual camera and displaying rendered images on the display panel as the *view-sharing* system in the remainder of this paper. Further details of the view sharing system will be described in the following section.

4. Capturing and Sharing

4.1. Camera Control

The virtual camera in the view-sharing system responds to user movements and changes in perspective, allowing for the smooth and natural sharing of the user's perspective within the virtual space. This system is expected to facilitate effective perspective sharing and interaction among multiple users in a VR environment, thereby enhancing the efficiency and experience of collaboration.

The virtual camera is fixed at the user's head position, but its direction is dynamically changed in response to the user's movements and actions, allowing the camera to actively share a varying perspective. The elements to which the camera responds are divided into three categories: The first is the direction for forward (d_f), representing the user's frontal gaze direction. The second is the direction for interaction (d_i), indicating the direction between the camera and the object with which the user interacts using the motion controller. Lastly, the direction for the surroundings (d_s) is an element representing the direction between the camera and the nearby objects within a certain range around the object the user is interacting with.

The desired direction of the virtual camera is determined as the average direction of these three elements. If the user is not interacting with anything, the camera considers only the first element, d_f (see Figure 2 (top row)). When the user interacts with an object using the controller, the camera's direction is updated to the average of d_f and d_i (see Figure 2 (middle row)). If the object held by the user comes close to other objects, the camera changes its direction to the average considering d_s as well (see Figure 2 (bottom row)). The camera's rotation is not immediately changed to the desired direction, but is made to rotate smoothly by applying physics-based simulation. The PD control method is employed to calculate the torque of the camera by considering the difference between the current direction and the desired direction. This rotational force is applied to the

camera's rigid body, leading to smooth rotation of the camera toward the desired direction. Algorithm 1 describes precisely how our camera control method is implemented.

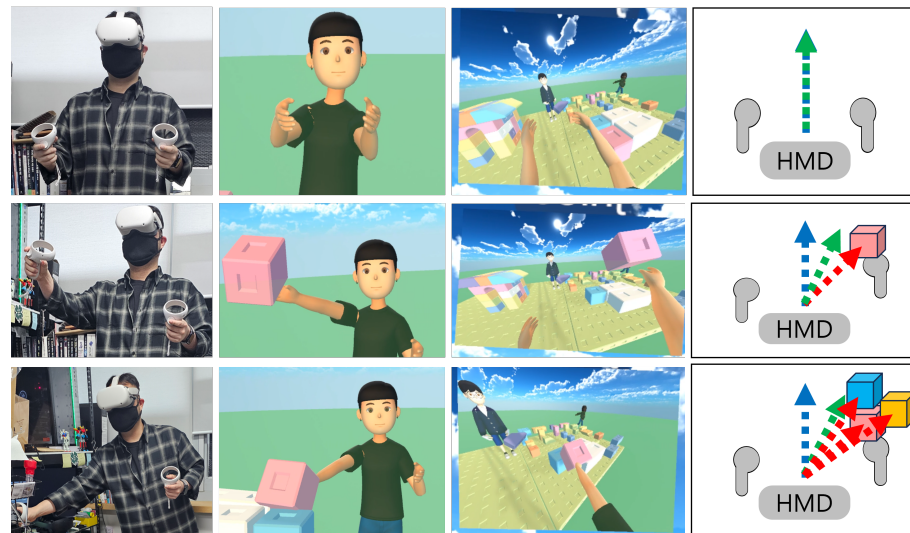


Figure 2. Automatic control of virtual camera. The blue arrow represents the user's gaze direction; the red arrow indicates the direction towards the interacting objects; the green arrow is the desired direction of the camera.

Algorithm 1 Sharing-camera mechanism.

```

1: forward ← GETGAZEDIRECTION()
2: object ← GETOBJECTINHAND()
3: if object is None then
4:   desiredDirection ← forward
5: else
6:   od ← NORMALIZE(object.position − HMD.position)
7:   SOL ← GETSURROUNDOBJECTLIST(object)
8:
9:   if ISEMPY(SOL) then
10:    desiredDirection ← AVERAGE(forward, od)
11:   else
12:    ad ← CALCAVERAGEDIRECTION(object, SOL)
13:    desiredDirection ← AVERAGE(forward, od, ad)
14:
15: desiredQuaternion ← VECTORTOQUATERNION(desiredDirection)
16: torque ← CALCDESIREDTORQUE(desiredQuaternion)
17: ADDTORQUE(torque, Camera)
18:
19: function CALCDESIREDTORQUE(desiredQuaternion)
20:   kp ← 0.5
21:   kd ← 1.0
22:   errorQuat ← desiredQuaternion × INVERSE(currentQuaternion)
23:   TOANGLEAXIS(errorQuat, angle, axis)
24:   if angle > 180 then
25:     angle ← angle − 360
26:   error ← NORMALIZE(axis) × angle × DEG2RAD
27:   derivative ← (error − previousError) / DELTATIME
28:   previousError ← error
29:   return kp × error + kd × derivative

```

4.2. Image Display

In our system, each user should be able to view the image rendered by the virtual camera of every other user. There are several candidate methods to achieve this. One method involves each user performing the rendering of the virtual camera on his/her own computer and, then, transmitting the resulting image to other users via the network. This method has the advantage of distributing rendering computations across multiple computers, thereby reducing the computational cost for each computer. However, it has the downside of potentially causing time delays due to streaming high-resolution image sequences in real-time over a network.

Another method involves each user individually rendering the virtual cameras of all users on his/her own computer. This method has the disadvantage of increasing the rendering computational costs in proportion to the number of participating users. However, it minimizes transmission time delays, as only the positions and orientations of each user's virtual camera need to be transmitted over the network. Since we limited the number of participants to two in our experiment and the computational cost for rendering was not significant, we chose the latter method to reduce network load. The images rendered for each user's virtual camera are mapped as textures on the panel placed above that user's avatar's head. This panel is constantly oriented towards the user, with its direction constantly updated using the billboarding technique from the field of computer graphics.

5. User Study

Experiments were conducted within the environment based on the VR platform introduced in Section 3 and the view-sharing system presented in Section 4. Prior to the experiments, we anticipated the following benefits from this view sharing approach:

- **Enhanced user immersion:** The virtual reality camera system aids users in immersing more deeply into the virtual environment. Through view sharing, users can perceive not only their own avatar, but also those of others, along with their environments, in a more-natural and -realistic manner.
- **Strengthened interaction:** Understanding visually what other users are interested and what they are interacting with helps facilitate the interactions among users. This proves particularly useful in collaborative work environments, allowing users to more intuitively understand each other's focus and interaction.
- **Improved visual feedback:** The system allows each user to gain a richer visual understanding of his/her surroundings through images rendered on other users' display panels. This enhancement can improve task performance, reduce errors, and strengthen the learning process.

The VR system used for the experiment was implemented on a system with specifications including a Ryzen5 5700×, 32 GB of memory, RTX3070, and Windows 10. The Meta Quest 2 was connected via wire to access the virtual environment. The network system was implemented using the Photon Fusion engine, and VR connectivity was achieved through the XR Interaction Toolkit. The functionality of transferring the movements of a user wearing VR equipment directly to an avatar was implemented using Rootmotion's VRIK tool. Our experiment took place in a dedicated space free from external disturbances, with wired Internet connectivity to PC, and VR equipment of identical specifications provided to the participants. Each participant conducted the experiment within a confined space of 1.5 m × 1.5 m. Please refer to Figure 3.

Prior to the experiment, we provided the recruited participants with instructions for the functional experiment of our view-sharing system, which was the purpose of this experiment. We also thoroughly explained the potential safety and motion sickness risks associated with experiments involving VR devices. Participants who agreed to this directly signed the 'Experiment Consent Form' and the 'Personal Information Usage Agreement Form'. The study involved 20 participants, with their ages distributed as follows: early 20 s (12 participants), mid-20 s (6 participants), late 20 s (1 participant), and

early 30 s (1 participant). The gender composition was 11 males and 9 females. Among them, 7 participants reported having prior experience with VR equipment. To assess user proficiency and the potential for motion sickness when using our system, we conducted a brief questionnaire regarding experience with VR devices and video games. Eleven participants had experience with video games, with 5 of them specifically enjoying 3D video games. The remaining 9 participants stated that they did not enjoy playing games. Our system was designed with various methods of movement within the virtual space, including linear motion, teleportation, and rotation, to enhance user convenience. It incorporates complex control mechanisms to facilitate interactions with objects placed within the space. To ensure users are proficient in utilizing these features, we provided them with a learning period ranging from a minimum of 10 min to over 30 min before the experiment. This is why we inquired about participants' gaming proficiency in the preliminary questionnaire.



Figure 3. Experimental scenes: pairs of participants are engaged in collaboratively performing Tasks 1 and 2.

Participants formed pairs and entered the virtual environment. On the left side of this virtual environment, a completed block assembly served as a sample (target), while on the right side, the blocks used in the sample were laid out on the ground (source) (see Figure 4). The objective of each task was for participants to use the source blocks to complete the target assembly. In the first experiment, Task 1, as shown on the left in Figure 4, participants initially created a house located on the left. In the second experiment, Task 2, they created a truck located on the right in Figure 4. Half of the experimental groups had the condition 'with view sharing' in Task 1 and 'without view sharing' in Task 2, while the other half had the opposite condition, 'without view sharing' in Task 1 and 'with view sharing' in Task 2. There were no constraints on the assembly sequence or methods other than this.

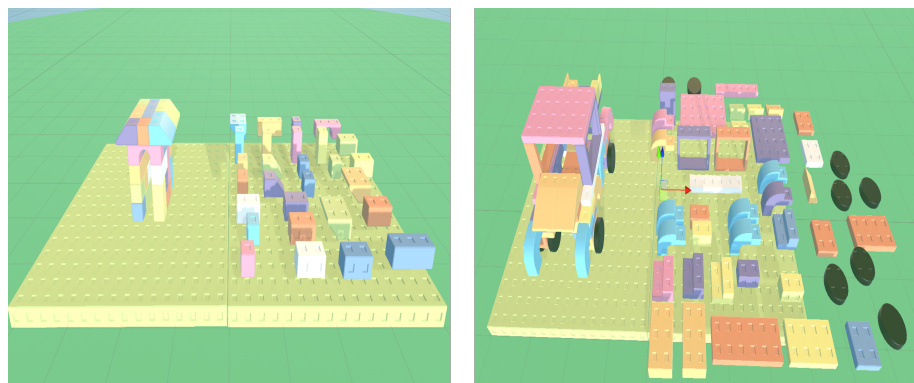


Figure 4. Virtual environments for the experiment: participants are tasked with assembling blocks to exactly match a given sample.

We collected the first set of quantitative data from the moment participants entered the experiment environment until the experiment's completion. Externally, the time taken by participants from accessing the experiment environment to completion was monitored. Participants were not given a time limit. Internally, within the system, we assessed the assembly completeness of the two sets of blocks provided to participants (target and source). This assembly completeness was determined by comparing the local position values of each block to ensure they were placed accurately. Since errors in block assembly typically do not occur unless assembled incorrectly, we assumed that low assembly completeness would be indicative of reduced participant concentration, which could impact the results.

After the first task, participants completed SSQ and NASA-TLX surveys to assess motion sickness and usability, followed by a break. After the second task, another round of SSQ and NASA-TLX surveys were conducted, along with a questionnaire on the view-sharing feature and brief interviews.

6. Result and Discussion

Our study sought to evaluate the effectiveness of a system that enables the sharing of viewpoints among users within a virtual reality environment. This system was designed under the premise that sharing perspectives would enhance the efficiency of collaborative tasks, reduce the time required for task completion, and improve the final quality of work. The quantitative data collected through experimentation served to validate these assumptions, assessing the degree of VR-induced discomfort via the Simulator Sickness Questionnaire (SSQ) and Virtual Reality Sickness Questionnaire (VRSQ) methodologies. Additionally, the NASA Task Load Index (NASA-TLX) was utilized to gather information on the system's overall usability, and interviews provided intuitive feedback from participants.

6.1. Quantitative Evaluation

In our quantitative evaluation, the first point of discussion is the observed task-completion time and the designated accuracy of tasks. The experiment was divided into two main categories: tasks with a shared view and tasks without it. In Task 1 with a shared view, it took 10 min and 12 s to complete, and the assembly accuracy was measured at 99.2%. In contrast, without the shared view, it took 11 min, and the accuracy was 95.2%. For Task 2, the completion time with the shared view was 25 min and 12 s, with an accuracy of 73%, while it took 23 min and 6 s with an accuracy of 79.1% without the shared view. A paired t-test was applied to these results, revealing that the differences in the task-completion times for Task 1 ($p = 0.0086$) and Task 2 ($p = 0.0427$) and assembly accuracy for Task 1 ($p = 0.0216$) were statistically significant, indicating the effectiveness of our view-sharing system, particularly for smaller and simpler structures. However, for larger and more-complex structures, the shared view was less effective, which could be due to the obstruction of the shared view in complex environments, as supported by some users' feedback in post-interviews, stating, "the shared gaze screen often got obscured in complex environments, making it difficult to use". The lack of a significant difference in assembly accuracy for Task 2 ($p = 0.1034$) suggests that other factors may influence the effectiveness of the shared view system in more-complex tasks. These data can be found in Table 1.

Next, we assessed the participants' reactions to view sharing in terms of sickness, as measured by the VRSQ survey (see Table 2). All participants underwent this survey during their rest period after completing Tasks 1 and 2. In Task 1, the sickness response was relatively minor at 8.9% when view sharing was enabled. However, in the absence of view sharing, the response spiked to 37.7%, indicating a significantly higher experience of sickness. This is a very interesting result. Despite providing additional visual information through the display panels, there was a reduction in sickness. This reduction is likely due to the fact that the additional information significantly aided in the task performance. On the other hand, the results were opposite in Task 2; the sickness response was 30.0% with

view sharing and 16.7% without it. This was echoed in the post-experiment interviews, with some participants stating, “Trying to follow the viewer on my partner’s avatar actually increased my sense of sickness”. This suggests that sickness could have been exacerbated by the numerous objects within the field of view, which disrupted gaze and led to discomfort.

Table 1. Whether or not a shared perspective is provided during participant observation makes a difference in the acquisition time and completeness of assembly of a given object.

Task	State	Time (min:sec)	Assembly Accuracy
Task 1	With Sharing	10:12	99.2%
	w/o Sharing	11:00	95.2%
Task 2	With Sharing	25:12	73.0%
	w/o Sharing	23:06	79.1%

Table 2. VRSQ: It shows the total evaluation index for the sickness questionnaire according to each task and the status of visibility.

Task	State	Oculomotor (%)	Disorientation (%)	Average Evaluation (%)
Task 1	With Sharing	10.4	7.5	8.9
	w/o Sharing	43.7	31.6	37.7
Task 2	With Sharing	36.9	23.6	30.3
	w/o Sharing	26.0	7.5	16.7

Upon a detailed analysis of the discomfort scores, it became apparent that there were significant differences between the presence and absence of shared views. In the case of Task 1, the presence of shared views correlated with lower scores in ocular-related symptoms such as general discomfort, fatigue, and eye strain. This was reflected in a substantial p -value, indicating a decrease in discomfort ($p = 0.0000586$). In contrast, for Task 2, even in the presence of shared views, symptoms related to ocular movement and disorientation, such as difficulty focusing and vertigo, were rated higher. This was also reflected in a significant p -value ($p = 0.0000814$). This suggests that discomfort increased when shared views were activated, especially for more-complex tasks. This observation aligns with the responses from the interviews, where a participant stated, “I felt motion sickness because I had to turn my gaze to see the viewer on my partner’s avatar”, indicating that discomfort may have been exacerbated when the field of view was obstructed by a relatively large number of objects in front of them.

The following survey used the NASA-TLX to assess the effectiveness of the work environment provided to the participants during task execution. Similar to other evaluations, it was observed that sharing perspectives in relatively simpler tasks like Task 1 marginally or significantly improved work efficiency. In contrast, for more-complex tasks, the shared perspective did not receive as favorable evaluations. In Figure 5, the horizontal axis summarizes the topics of the survey, while the vertical axis represents the evaluation scores. Depending on the evaluation topic, the scores can indicate positive or negative aspects. For categories like mental demand, physical demand, temporal demand, effort, and frustration, lower scores are more favorable. Conversely, for performance, higher scores indicate greater satisfaction with the outcomes. Continuing, let us examine in detail the specific questions asked in each category of the survey:

- **Mental demand:** “How mentally demanding was performing the given task through this system?” Responses from participants indicated that, in Task 1, the mental demand was slightly lower when the perspective was shared, while in Task 2, a better evaluation was given when the perspective was not shared.

- **Physical demand:** “How physically demanding was performing the given task through this system?” There was a significant difference in physical demand in Task 1 depending on whether the perspective was shared or not. In contrast, Task 2 did not show a significant difference compared to Task 1, which can be attributed to the larger scale of Task 2 and the difficulty in using perspective sharing due to more objects being spread over a wider area.
- **Temporal demand:** “To what extent was time pressure an issue while performing the task through this system?” As no time limit was set for the tasks, this item may not be highly significant in our study.
- **Effort:** “How much effort did you have to put forth to perform the given task through this system?” Similar to other items, Task 1 showed that less effort was needed when the view-sharing camera was enabled. In Task 2, the effort was evaluated to be similar to Task 1 when no perspective was shared.
- **Performance:** “How successful were you in performing the given task through this system?” Performance, unlike other items, reflects user satisfaction, where a higher score indicates a better evaluation. In Task 1, the scores were high and similar regardless of whether the perspective was shared or not.
- **Frustration:** “To what extent did you feel frustrated while performing the task through this system?” Lastly, frustration showed a significant difference in Task 1 based on whether the perspective was shared or not. In Task 2, the sharing of views likely contributed to an increase in participants’ frustration due to the problem of the shared information being obscured by complex structures.

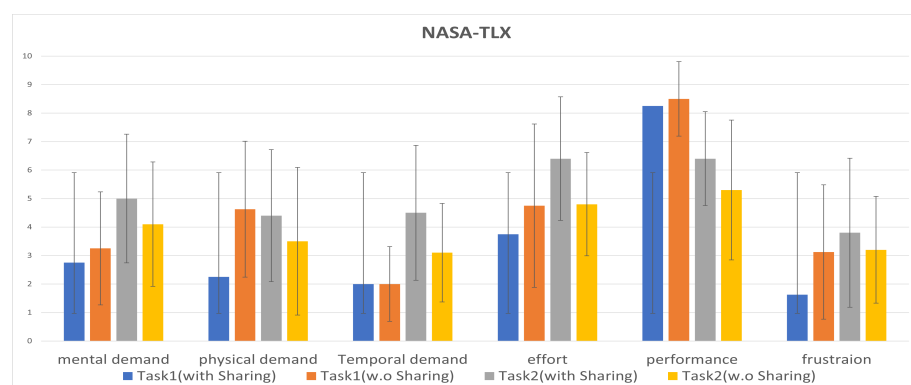


Figure 5. A comprehensive graph of the NASA-TLX survey results. In this graph, the first two vertical columns for each item represent the results of Task 1, while the last two columns represent the results of Task 2.

Just like the time measurements and accuracy evaluations, as well as the VRSQ results, we deduced that the usefulness was better when less obstructive elements were present, as in Task 1, where the structures were smaller and simpler. However, in Task 2, where there were more objects, each larger in size, and the target objects were bigger compared to Task 1, the obstructive elements were more significant, leading to relatively poorer evaluations.

In the final part of our qualitative evaluation, we surveyed all participants regarding their experiences with the view-sharing feature. This survey aimed to determine the overall evaluation of our proposed system in virtual collaboration environments. Each question was designed to address specific aspects of the system’s impact, as detailed below. The results can be referenced in Figure 6. In this graph, the horizontal axis represents the evaluation topics, while the vertical axis indicates the evaluation scores. In this graph, focus, interaction, technical issue, role, performance, satisfaction, and recommended are positively evaluated with higher scores, while information is positively evaluated with lower scores. Detailed explanations are provided in the following:

- **Focus:** This question assessed how much the provision of view sharing helped in concentrating on the task. It scored 4.8, indicating a slightly below-average contribution

to task focus. It seems that view sharing did not play a decisive role in enhancing concentration.

- **Interaction:** This item asked how natural the interactions with other users were due to view sharing. With an average score of 6, it can be concluded that view sharing was helpful.
- **Technical issues:** This question dealt with any technical problems arising from view sharing, receiving a score of 4.25. This slightly below-average score may have been influenced by network errors that occurred during the experiment.
- **Information:** This item inquired about the difficulty in understanding other users' intentions through view sharing. A lower score indicates less difficulty in perceiving information. A relatively low score of 2.75 suggests that it was easy to understand the intentions of other users.
- **Role:** This question asked how clearly participants could perceive their role in collaboration depending on the presence of view sharing. It scored an above-average 6.25, suggesting clarity in role perception.
- **Performance:** This item evaluated how helpful view sharing was in accomplishing the given tasks. It scored a slightly above-average 5.6.
- **Satisfaction:** This question assessed the satisfaction level with view sharing during collaboration. The score of 5.8 indicates a level of satisfaction slightly above average.
- **Recommendation:** The final item asked whether the view sharing experienced in this experiment, or similar features, would be helpful in other virtual collaborative environments. The participants generally rated this positively.

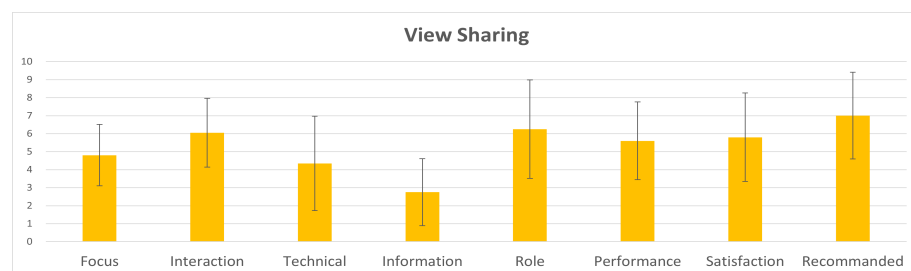


Figure 6. View sharing questionnaire for collaboration by participants group.

6.2. Qualitative Evaluation

In the qualitative evaluation, interviews were conducted after all experiments and surveys were completed. We conducted interviews with the experiment participants to gain insights. The interview consisted of five questions, aiming to understand whether our proposed view-sharing method provided any advantages in the collaborative process. We also sought candid feedback regarding any functional shortcomings. Additionally, we collected opinions on how this technology could be beneficial in the future. The questions and some key responses from the interviews are as follows:

- *“Did using the view sharing feature enhance your overall experience in VR?”* Apart from two participants, the rest responded that it “somewhat improved” their experience.
- *“What was the most helpful aspect of collaborating with the view sharing feature? Were there any inconveniences?”* Sixteen participants mostly responded positively, saying, “It was convenient to see what the other person was working on through their perspective”.
- *“What personal difficulties did you encounter while using the view sharing feature? Did it have a negative impact on the experiment?”* More than half of the participants mentioned, “The location/size of the view sharing panel above the avatar’s head was small and uncomfortable to view” and, “When the blocks were large, it was inconvenient not to be able to see the other person’s view”. Uniquely, some said, “I didn’t pay any attention to the other person’s perspective while assembling objects”.
- *“Was there anything impressive after experiencing the perspective sharing feature?”* Responses included, “I thought view sharing wouldn’t be helpful, but I found it inconve-

nient when it was gone” and, “Having an additional auxiliary perspective beyond my own was helpful in finding and assembling parts”.

- “Do you think the view sharing feature could be applied in other VR settings or in everyday life?” Participants with gaming experience responded, “It could be utilized from a strategic perspective in gameplay”. Others mentioned, “It could be beneficial for people with limited mobility” and, “It might be helpful in medical and nursing procedures in real world”.

The qualitative evaluation revealed that participants were generally positive about the provision of view sharing. However, there were discomforts regarding the location and size of the display panel. In some cases, as view sharing was not mandatory in certain tasks, the utilization of view sharing was not enforced during the tasks, which led to participants not using this feature and focusing solely on the given tasks.

7. Conclusions and Future Work

The proposed system focuses on capturing and sharing users’ attention and interactions, and we implemented a virtual environment system to test it in real collaborative situations. In our view-sharing system, the virtual camera is automatically controlled to consider not only the user’s gaze direction, but also the objects they are interacting with and the surrounding environment. Our goal was to improve communication through non-verbal cues by providing active view sharing, thus facilitating collaboration among multiple users.

Furthermore, we can divide the objectives of our automated view-sharing method into several aspects:

- Interaction-centric automation and efficient information sharing: While previous research [18,19] primarily explored viewpoint-centered automation, we concentrated on automating viewpoint changes based on users’ interactions in the virtual environment to share information about users’ actions during collaboration.
- Enhancing the functionality of mirroring techniques: Previous research has found mirroring techniques [6,20] to be beneficial for expanding the field of view in VR. We aimed to leverage these techniques by automatically sharing views in response to interactions, allowing multiple users to share their perspectives for collaborative purposes.
- Motion sickness reduction: VR experiences using VR equipment can induce discomfort and motion sickness [6]. To alleviate these issues, we expected that actively utilizing view sharing in the virtual environment could reduce user fatigue and contribute to motion sickness relief.

Our proposed system and its goals have versatile applications in various domains, enhancing collaboration in virtual reality environments and finding potential use in educational settings. Additionally, it has the potential to reduce motion sickness when users immerse themselves in virtual environments. The experimental results partially confirmed this: in collaborative tasks involving the assembly of small and simple structures, not only did efficiency and accuracy increase, but discomfort also decreased. However, when the structures were large and complex, the burden of visual information processing increased. More specifically, users reported fatigue and motion sickness in interviews due to excessive movement and rotation required to point at the display panel above the avatar’s head within virtual reality.

The experiment, surveys, and interviews revealed several limitations. First, the two tasks had different levels of complexity. The size and quantity of blocks obstructed the view, negatively impacting the participants. Second, participants frequently cited discomfort with the placement of the shared view screen. To see another user’s viewpoint, one had to turn his/her gaze and search for the other user’s avatar, which often induced motion sickness. Additionally, the experimental scenario was relatively simplistic—“assemble the listed block pieces”. This simplicity meant there were no constraints or roles assigned to participants, occasionally leading them to focus solely on their task (assembling the blocks in front of them) and to completely ignore the shared view screen.

In future research, we will aim to have a nearly equal number of blocks and a similar level of structural complexity for different experimental tasks. Additionally, we will designate specific roles for experiment participants to encourage more-active utilization of the view-sharing camera. For example, one of the two participants will take on the role of a supervisor, not directly involved in assembly, but providing instructions, while the other user will rely on the supervisor's commands and the shared view for task execution. Based on the participant surveys and interviews, we will address the feedback that the display panel above the avatar's head was "uncomfortable to observe" and that "eye movements to locate the viewer induced motion sickness" by offering the display panel in various locations and allowing users to choose, ultimately finding the optimized display position. We expect that these improvements will effectively enhance the use of the display panel and reduce motion sickness.

In conclusion, the proposed method for automated camera control exhibits versatile applicability in various domains. It not only enhances collaboration in virtual reality environments, but also finds potential use in educational settings. For example, it can empower instructors to emphasize specific areas for students or enable remote guidance of practical activities. Furthermore, its application in entertainment, particularly multiplayer gaming, can streamline strategy development and real-time adjustments among players. We aim to explore these potential applications further and adapt the technology for a broader range of collaborative tasks and virtual environments.

Author Contributions: Conceptualization, K.H.L.; methodology, J.L. and K.H.L.; software, J.L., D.L., S.H. and K.H.L.; validation, J.L., D.L., S.H. and K.H.L.; formal analysis, J.L. and H.K.K.; investigation, J.L., H.K.K. and K.H.L.; resources, J.L. and K.H.L.; data curation, J.L.; writing—original draft preparation, J.L. and K.H.L.; writing—review and editing, K.H.L.; visualization, J.L.; supervision, K.H.L.; project administration, K.H.L.; funding acquisition, K.H.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korean government (MSIT) (No. NRF-2021R1F1A1046373).

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki and approved by the Ethics Committee of Kwangwoon University (7001546-202300831-HR(SB)-008-02).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data are available upon request due to privacy or other restrictions.

Acknowledgments: The present research was conducted by the Research Grant of Kwangwoon University in 2023.

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

References

1. Liu, Q.; Steed, A. Social virtual reality platform comparison and evaluation using a guided group walkthrough method. *Front. Virtual Real.* **2021**, *2*, 668181. [[CrossRef](#)]
2. Schäfer, A.; Reis, G.; Stricker, D. A survey on synchronous augmented, virtual, and mixed reality remote collaboration systems. *ACM Comput. Surv.* **2022**, *55*, 116.
3. Du, J.; Shi, Y.; Zou, Z.; Zhao, D. CoVR: Cloud-based multiuser virtual reality headset system for project communication of remote users. *J. Constr. Eng. Manag.* **2018**, *144*, 04017109. [[CrossRef](#)]
4. Men, L.; Bryan-Kinns, N. LeMo: Supporting collaborative music making in virtual reality. In Proceedings of the 2018 IEEE 4th VR Workshop on Sonic Interactions for Virtual Environments, Reutlingen, Germany, 18 March 2018.
5. Chen, X.; Chen, Z.; Li, Y.; He, T.; Hou, J.; Liu, S.; He, Y. ImmerTai: Immersive motion learning in VR environments. *J. Vis. Commun. Image. Represent.* **2019**, *58*, 416–427. [[CrossRef](#)]
6. Freiwald, J.P.; Diedrichsen, L.; Baur, A.; Manka, O.; Jorshery, P.B.; Steinicke, F. Conveying perspective in multi-user virtual reality collaborations. In Proceedings of the Mensch Und Computer 2020, Magdeburg, Germany, 6–9 September 2020; pp. 137–144.
7. Hand, C. A survey of 3D interaction techniques. *Comput. Graph. Forum.* **1997**, *16*, 269–281. [[CrossRef](#)]
8. Kharoub, H.; Lataifeh, M.; Ahmed, N. 3D user interface design and usability for immersive VR. *Appl. Sci.* **2019**, *9*, 4861. [[CrossRef](#)]

9. Argelaguet, F.; Andujar, C. A survey of 3D object selection techniques for virtual environments. *Comput. Graph.* **2013**, *37*, 121–136. [[CrossRef](#)]
10. Poupyrev, I.; Ichikawa, T.; Weghorst, S.; Billinghurst, M. Egocentric object manipulation in virtual environments: Empirical evaluation of interaction techniques. *Comput. Graph. Forum.* **1998**, *17*, 41–52. [[CrossRef](#)]
11. Poupyrev, I.; Billinghurst, M.; Weghorst, S.; Ichikawa, T. The go-go interaction technique: Non-linear mapping for direct manipulation in VR. In Proceedings of the 1996 Symposium on User Interface Software and Technique, Seattle, WA, USA, 6–8 November 1996; pp. 79–80.
12. Bowman, D.A.; Hodges, L.F. An evaluation of techniques for grabbing and manipulating objects in immersive virtual environments. In Proceedings of the 1997 Symposium on Interactive 3D Graphics, Providence, RI, USA, 27–30 April 1997; pp. 35–38.
13. Galambos, P.; Weidig, C.; Baranyi, P.; Aurich, J.C.; Hamann, B.; Kreylos, O. VirCA NET: A case study for collaboration in shared virtual space. In Proceedings of the 2012 IEEE 3rd International Conference on Cognitive Infocommunications, Kosice, Slovakia, 2–5 December 2012; pp. 273–277.
14. Li, J.; Kong, Y.; Röggl, T.; De Simone, F.; Ananthanarayan, S.; de Ridder, H.; El Ali, A.; Cesar, P. Measuring and understanding photo sharing experiences in social virtual reality. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, Glasgow, UK, 4–9 May 2019; pp. 1–14.
15. Mei, Y.; Li, J.; de Ridder, H.; Cesar, P. CakeVR: A social virtual reality (VR) tool for co-designing cakes. In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, Yokohama, Japan, 8–13 May 2021; p. 572.
16. Huyen Nguyen, T.T.; Duval, T. A survey of communication and awareness in collaborative virtual environments. In Proceedings of the 2014 International Workshop on Collaborative Virtual Environments, Minneapolis, MN, USA, 30 March 2014; pp. 1–8.
17. Zaman, C.H.; Yakhina, A.; Casalengno, F. NRoom: An immersive virtual environment for collaborative spatial design. In Proceedings of the 2015 International HCI and UX Conference in Indonesia, Bandung, Indonesia, 8–10 April 2015; pp. 10–17.
18. Wang, L.; Wu, W.; Zhou, Z.; Popescu, V. View splicing for effective VR collaboration. In Proceedings of the 2020 IEEE International Symposium on Mixed and Augmented Reality, Porto de Galinhas, Recife, Brazil, 9–13 November 2020; pp. 509–519.
19. Bovo, R.; Giunchi, D.; Alebri, M.; Steed, A.; Costanza, E.; Heinis, T. Cone of vision as a behavioral cue for VR collaboration. *Proc. ACM Hum.-Comput. Interact.* **2022**, *6*, 502. [[CrossRef](#)]
20. Ha, W.; Choi, M.G.; Lee, K.H. Automatic control of virtual mirrors for precise 3D manipulation in VR. *IEEE. Access.* **2020**, *8*, 156274–156284. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.