International Conference on Artificial Reality and Telexistence
Eurographics Symposium on Virtual Environments (2017)
R. Lindeman, G. Bruder, and D. Iwai (Editors)

The Effect of User Embodiment in AV Cinematic Experience

Joshua Chen[1], Gun Lee[2,1†], Mark Billinghurst[2,1], Robert W. Lindeman[1], Christoph Bartneck[1]

[1]The Human Interface Technology Laboratory New Zealand, University of Canterbury, New Zealand
[2]School of Information Technology & Mathematical Sciences, University of South Australia, Australia

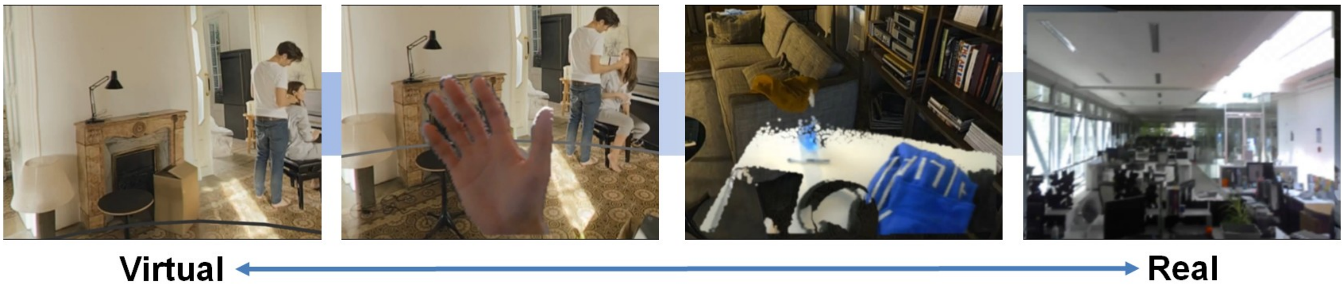

Figure 1: *Transition between the virtual cinematic environment (far left: a 360° panoramic movie the user is watching) and the real world (far right: an office where the user is). The user's hand and the desk in the middle are from the real world, augmented into the movie.*

Abstract

Virtual Reality (VR) is becoming a popular medium for viewing immersive cinematic experiences using 360° panoramic movies and head mounted displays. There are previous research on user embodiment in real-time rendered VR, but not in relation to cinematic VR based on 360 panoramic video. In this paper we explore the effects of introducing the user's real body into cinematic VR experiences. We conducted a study evaluating how the type of movie and user embodiment affects the sense of presence and user engagement. We found that when participants were able to see their own body in the VR movie, there was significant increase in the sense of Presence, yet user engagement was not significantly affected. We discuss on the implications of the results and how it can be expanded in the future.

CCS Concepts

•***Computing methodologies*** → ***Mixed / augmented reality;*** *Virtual reality;*

1. Introduction

The introduction of low cost Virtual Reality (VR) display devices, and devices for capturing 360° panoramic content, are making it possible to explore the use of VR for immersive cinematic experiences. For example, using a mobile phone and Google Cardboard[1] people can immerse themselves into 360° video views from undersea, or historic locations. Devices like the Ricoh Theta-S[2] make it easy to capture immersive content. With the introduction of 360° spherical panoramic videos, exploring the use of VR for immersive cinematic entertainment purposes is becoming more accessible.

VR is capable of "transporting" users completely from their real physical environments into virtual environments (VEs), creating a strong perception of Presence. Previous research has shown that user embodiment can enhance the user's perception of Presence [Sch10]. However, most of the previous work on user embodiment has been conducted in real-time rendered 3D VEs, and not cinematic VEs that use 360° panoramic video as the main content.

Augmented Virtuality (AV) [MK94] is a variation of VR that expands the virtual world to include elements of real life content. One use of AV is to bring a representation of the user's own real body into the virtual scene (see Figure 1). In this paper we describe research on using AV in immersive cinematic experience, where users will view themselves transported into a cinematic VE as a digitized copy. The main contribution of this work is that it is one

† Corresponding author e-mail: gun.lee@unisa.edu.au
[1] https://vr.google.com/cardboard/
[2] https://theta360.com/en/

of the first studies on how user embodiment can affect the sense of Presence and engagement in an immersive cinematic VE.

In the rest of the paper, we first review related work, and then describe the design and implementation of the prototype system. Next we describe the experimental design, report on the results, and discuss the findings from the experiment. We concluded by summarising the paper and offering suggested future work.

2. Related Work

This research extends earlier research in virtual environments, virtual avatars, and user-body segmentation. In this section we review key work in each of these areas and discuss the novelty of our work.

2.1. Cinematic Virtual Environments

Immersive virtual environments are compelling experiences that transport users to a synthetic world [SLF*12]. In contrast to immersive 3D VEs, cinematic VE has a number of unique challenges, such as how far the predetermined nature of the narrative can be relaxed, and how far can the user in a cinematic VR can freely participate in a narrative, rather than acting as a spectator [Ayl99].

Film narrative is made up of a series of edits, sequenced gaps and spatial ambiguities [TT02]. A narrative is therefore understood as the way visual images are framed, positioned and edited; opening up a fictional area where the organisation of visual space contributes significantly to the narrativisation of a film [TT02].

With a 360° visible screen space that viewers look into a cinematic virtual environment, organisation is important. Kwiatek and Woolner [KW10] demonstrated the usage of 360° images and video as the foundation for interactive storytelling. They found that it provided sufficient illusion of the site they want to recreate, made viewers constantly aware of elements of the narrative through usage of spatial sound, and provided options of interactivity in a cylindrical space. An interactive story-graph-based narrative was made and presented on a wrap-around 360° screen to viewers.

2.2. User Embodiment with Augmented Virtuality

As digital bodies, avatars re-embody the user in a virtual environment, promising the affordances of real bodies [Sch10]. Portalés et al. [PVAM10] used an avatar to represent the viewer in a movie that is neither completely virtual nor physical; it is a mixture of both. It involves using spatial registration technology based on AR markers, but these markers have to be constantly worn and changed for every user, and look odd when worn. Günther et al. [GFG15] explains the advantages of AV over virtual avatars in representing the viewer and the saving of complex hardware setups.

Schultze [Sch10] explores the role of the physical body in communication, showing that it plays a crucial role in representing its beholder. It is synonymous with being conscious of one's body existing in, and being distinct from, a pre-designed foreign world, which can be both real and virtual [WW03] [WW06]. Our bodies generally work in the background and are thus taken for granted. However, to understand the digital body, we have to explore the role of the user's body in communication and interactions [Sch10].

In order to combine a representation of the user's real body into the cinematic VR environment, we need to segment out the user's body from the real world background. In our work, we focus especially on the arms and hands since they are known to be our primary sense of touch and control. There are a number of methods for inserting real people into virtual scenes. For example, Portalés et al. [PVAM10] use two projectors for chroma-keying to integrate the image of the user into movie scenes.

The chroma-keying technique has been used very successfully for segmentation [GFG15] [BSRH09]. However, the main disadvantage of chroma-keying is its requirement that the background has to be uniformly coloured with a colour that should not be confused with the skin tones of the user [GFG15]. Therefore, instead of colour-keying, other researchers have explored depth-keying with a depth-capable camera [GKOY03].

In bringing the hands into a VR scene, [PAK*14] present three techniques using GPU shaders to provide realistic hand occlusions, shadows and visual distortion. These use a variety of depth-sensing methods, to render an image of the users real hands inside the VR environment. One disadvantage of this system was the dependency on an additional colour depth camera in the environment for performing hand tracking in the top-down interaction space. It also required a calibration step to precisely align the virtual hands with the real hands. Similarly, the Kinect requires precise calibration of the system and quickly leads to inaccurate segmentations when the user leaves the "sweet spot," [NSC15].

Khattak et al. [KCCH14] used a depth camera attached to an Oculus Rift HMD to perform hand segmentation. For scene reconstruction and occlusion of the user's hands, they capture a depth map of the entire environment to reconstruct in VR, and apply hand occlusion over the reconstructed virtual environment. However the system requires that the real environment behind the user's hands remains static in order to perform accurate segmentation.

Recently, researchers also have looked into incorporating reality into computer generated virtual environments for improving usuability of VR systems [MBMSB15] [BSJ*15]. Applications developed showed how the user could see their own hands [BSJ*15] [GFG15], both their near and far surroundings, and the existence of other people [MBMSB15].

2.3. Summary

Kwiatek [KW10] has shown how projected panoramic environments can be used in immersing multiple spectators into a new narrative world. By personalising this experience and looking through the eyes of an HMD, we similarly used 360° videos to immerse a viewer into a cinematic VR.

We are focused using depth keying to give viewers a virtual body of their own unique self. This is an avatar in the sense that it resembles the viewer but no virtual 3D model of the viewer's body is created and maintained. Rather, the viewer's body visualization is augmented into the cinematic virtual environment. No information about the viewer's real surroundings was needed as we expected that they will not be interacting with it.

Solely focusing on the viewer's avatar combined with the setting of a cinematic virtual environment, we want to observe the sense of Presence and level of user engagement an immersed cinematic user feels within this new medium.

3. System Design and Implementation

In this section, we give an overview of the AV system we designed and developed. The overall goal of the development was to provide an immersive experience for users to feel present in a cinematic VE by applying AV for user embodiment in the movie scene. The basic requirements of the prototype system were to provide:

- an immersive experience of cinematic VE with real-time 360° panoramic movie playback,
- AV visualization through real-time capture of the user's body and surroundings in the real environment, and
- user-input allowing users to control transition between the virtual and real environments.

3.1. Augmented Virtuality Visualisation

The basic function of our prototype is to show 360° spherical panoramic movies on a HMD. The system uses the HMD head orientation tracking capability to let the user look around the panoramic movie scene. The main difference of our system from other HMD based panoramic movie viewers is AV visualisation. So our system has a depth sensing camera attached in the front of the HMD which provides the real world view to the user forming a video see-through configuration. Using the depth image of the real world scene, the system composites the real world scene with the VE (i.e. 360° panoramic movie). The pixels in the real world image that have the depth value of less than a given distance threshold are only shown in the VE. By changing the distance threshold, the system can control how much amount of the real world scene is shown to the user. For example, if the threshold is set to person's average arm length, the user will be able to see his or her body (and other close physical objects, if any) augmented into the movie scene.

The system could also vary the threshold and make transitions between the virtual and real environments, depending on the story, scene, or user's preference. For example, the user can have the full real world view when he or she first puts on the HMD, and the real world can gradually disappear as the movie starts playing, transitioning into the cinematic VE. While the movie is playing, the threshold can be increased in certain scenes to let the user see their own body embedded into the movie scene, or even let the user control the transition between the real and virtual world as they prefer.

To allow the user to control the transition between the virtual and real worlds we designed two different techniques: head shaking and hand gestures. The first method is shaking the head sideways. As a user shakes his or her head, the real world environment will gradually appear staring from the near objects, replacing the objects in the VE. The longer the user shakes his or her head, a larger portion of the real world will become visible, and eventually the user will see the plain video feed of the real world. When the user is not shaking his or her head, the real world view gradually disappears over time starting from the farthest objects until the user will be fully immersed in the VE. The head shaking gesture is easy to implement, as the system already has head orientation tracking.

Figure 2: *Prototype system hardware setup.*

To make the control of the transition more explicit, we introduced a second method of using hand gestures. We used three gestures (*thumbup*, *thumbdown*, and *bigfive*) for increasing, decreasing, and stop changing the amount of the real world shown. When the user makes a *thumbup* gesture, the system starts to increase the threshold of the distance where the real world is shown up to. The distance threshold value keeps increasing until the *bigfive* gesture is made, and the current value is shown at the bottom of the HMD screen as a virtual head-up display (HUD). Similarly, the *thumbdown* gesture triggers the system to decrease the threshold over time. We designed the system to show visual feedback with graphical icons on HUD when the gestures are recognised (see Figure **??**). The HUD also showed an arrow that points up or down when the threshold is increasing or decreasing. The users were allowed to show or hide the HUD by making a *waving* gesture.

3.2. Implementation

Figure 2 shows the hardware setup used in our prototype system. The hardware chosen for our system included (1) a Desktop PC (Intel Core i5-4670 CPU @ 3.40GHz, 8GB RAM, Nvidia GeForce GTX 770), (2) an Oculus Rift DK2 HMD, and (3) a SoftKinetic DepthSense 325 depth sensing camera [3] which can capture colour (up to 720p HD) and depth (up to QVGA) video at 30 fps.

The main software of our system was developed using the Unity game engine [4]. To implement playback of 360° spherical panoramic movies, we used the AV Pro Windows Media plugin [5]. As the panoramic video clips we used were in equi-rectangular projection, the movie texture was mapped onto a reversed sphere, which has the surface normal towards the inside. The virtual camera (the user's view) was placed at the center of the sphere and we used the Oculus SDK[6] Unity plugin to apply VR visualisation with head tracking.

For AV visualisation, we used the SoftKinetic DepthSense SDK

[3] www.softkinetic.com/Products/DepthSenseCameras
[4] https://unity3d.com
[5] http://renderheads.com/product/av-pro-windows-media
[6] http://developer.oculus.com

and developed a Unity plugin that gathers depth and colour video streams from the SoftKinetic DS325 camera. For processing depth images in AV visualisation we used the EmguCV plugin [7], and a point cloud particle system for rendering the processed depth and colour image in the Unity scene. The standard depth-to-colour map provided by the DepthSense SDK had a limitation of not providing values for the pixels further than the camera's depth sensing range (1.5 metres). To overcome this problem, we performed camera calibration between the depth and colour sensors and used this information to back project colour pixels into the depth image coordinates when the standard depth-to-colour map did not provided a proper mapping.

While the alignment between the real world image and the cinematic VE was calibrated, the field of view (FOV) of the depth sensing camera was not wide enough to cover the FOV of the HMD. So we aligned the depth sensing camera to cover the lower part of the FOV on the HMD. This was because the user's hands are mostly placed at the lower part of the user's view.

The Oculus Rift DK2 has inertial sensors to detect rotation and acceleration of the head motion. To detect the head shaking gesture, the software monitors the rotation of the head about the vertical Y axis and identifies fast changes in rotation. For implementing hand gesture recognition, we built a Unity plugin that uses the SoftKinetic *iisu*[8] library which provides hand tracking and gesture recognition from the DS325 depth sensing camera input.

4. User Evaluation

Using the prototype system, we conducted a user study to investigate the effect of AV visualisation on the user's sense of Presence and engagement. We also looked into how much of control do users want to have over transitioning between the virtual and real environments while watching immersive movies.

4.1. Experiment Design

Based on the findings from the pilot study, we designed the main experiment based on three hypotheses:

1. The user embodiment will affect on sense of presence and user engagement.
2. The type of movie will affect on sense of presence and user engagement.
3. There will be a difference in user preference among Manual, Automatic, and Shared control for transitioning between the virtual and real environments.

Three independent variables were identified: User embodiment (user's body present or not), Type of movies in terms of visual realism (Animated or Live action), and Type of control on transitioning between virtual and real environments (Manual, Automatic, or Shared).

[7] www.emgu.com
[8] http://www.softkinetic.com/Products/iisuMiddleware

As dependent variables we defined Sense of Presence, Engagement, and user preference. Sense of Presence is important for VR, and a visual representation of a user in an immersive virtual environment is known to enhance it [BSRH09]. Our bodies are known for making us present in a given place, and sense of presence would affect on user engagement. Thus it would be interesting to investigate how levels of presence and user engagement would be affected by users seeing their own body or not in a cinematic VE.

To test for the sense of presence, we used the Igroup Presence Questionnaire (IPQ) [Sch03]. IPQ consists of three sub-scales, Spatial Presence, Involvement, and Experienced Realism, and a general item which has high influences on all three sub-scales, especially on spatial [Sch03] [SFR99] [SFR01]. To test how each condition affects on user's engagement, we used a questionnaire developed by O'brien and Toms engagement [OT10]. To prevent questionnaires becoming too lengthy, we chose three of the most relevant sub-scales: Focused Attention, Novelty, and Involvement. To investigate what would be the best approach in transitioning between the real and cinematic VE, we asked user's preference using Likert scale rating items and ranking questions.

4.1.1. Procedure

The study started with participants reading the information sheet and signing the consent form. Then they were asked to fill in demographic questionnaire. After giving a briefing of the user study, we let the participant try out the prototype system to get familiarised.

The experiment was split into two sessions both in within-subject design. The first session was in 2x2 factorial design focusing on the first two hypotheses, with two independent variables:; "Body present or not present", and "Animated movie or live-action movie." The second session focused on the third independent variable, consisting of three conditions from the 3 levels of "Type of control" for transitioning between the real and virtual environment.

In session 1, two conditions had a live action movie played, while the other two had an animated movie. As it is hard to control the content of the movie in different types, we carefully designed the study to minimize the results getting biased by the content of the movies. To prevent bias from seeing the content for the first time, the participants were asked to preview the movies used for the experiment on a desktop monitor. Instead of using a different movie for each condition, we used only two movies for each participant to minimize the time spent in preview, and to reduce the chance of the results getting affected by the movie content. In this way participants would likely then no longer rate their overall experience based on the content of the movie, but based on the independent variables.

While restricting the variety of movie clips shown to each participant could improve internal validity, it would reduce the external validity of the study. In order to prevent generalisability of the study getting restricted with a particular content used in the experiment, we sourced variety of movie clips to be shown to different participants. We sourced eight 360 panorama movies in total including four animated and four live action films. This gave us 16 unique animated and live action movie pairs. The pair and order we presented these movies to participants was randomised. Each movie had a running length of 1.5 to 3 minutes.

Once they previewed the movies on a desktop monitor, the participants viewed the movies on the HMD under four conditions. The order of the conditions was counter-balanced with a 4x4 balanced Latin square design. After completing each condition, a per-condition questionnaire was given to rate their experience. The participant also completed a post-session questionnaire at the end of session 1. Participants were allowed take a break in between conditions and sessions as needed.

Session 2 started with the participants watching a video on how to perform the hand gestures to control the real-virtual transition. Then, they were put through a practice trail to get familiarised with the gestures. Once the participant was confident and comfortable with the gesture controls, they watch a 360 movie under three conditions. Participants were given sufficient time to fully experience and utilise transition between real and virtual environments. Therefore, only one randomly chosen movie was played, which had approximately 5 minutes of running length, providing multiple opportunities for participants to use the transitioning between the real and virtual worlds. The order of conditions was counter-balanced with a 3x3 Latin square design, and the movie shown was randomised between participants. A post experiment questionnaire was given to fill in at the end of session 2, and the study conclude with a debriefing. The whole procedure took about 60 to 90 minutes per participant.

5. Results

Twenty-eight participants were recruited for the study (age mean = 21.88, SD = 3.67; 14 female). Twenty of them had never used an HMD before, while 7 had used it a few times and 1 used it weekly.

5.1. Presence and Engagement Questionnaires

We used Repeated measures two-way ANOVA (α = .05) for analysing the results of Presence and Engagement questionnaires from session 1. Table 1 summarises the results.

The results of IPQ (see Figure 3) showed a significant positive main effect of Embodiment on overall sense of presence (F(1,27)=5.036, p=.033). No significant main effect of Type of Movie nor interaction effect between the two independent variables were found. Further analysing the results with each sub-scale in IPQ, the results showed Spatial Presence sub-scale was significantly affected by the Embodiment factor (F(1,27)=6.856, p=.014). While the Realism sub-scale shown similar trend of getting affect by the Embodiment, no statistically significant effect was found (F(1,27)=3.764, p=.063).

While the results of the Engagement questionnaire (see Figure 4) showed similar trend of slightly better engagement with user's body being visualised, no statistically significant effect was found.

5.2. User Preference

At the end of session 1, participants were asked to answer 7-point Likert-scale rating questions on their preference. Four graphs from the top left of Figure 5 summarises these results. We conducted one-sample Wilcoxon Signed Rank tests (α = .05) to check if

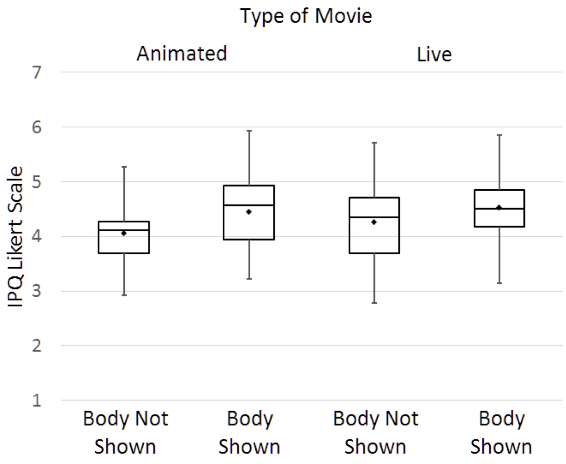

Figure 3: *Participants total IPQ values, higher is better.*

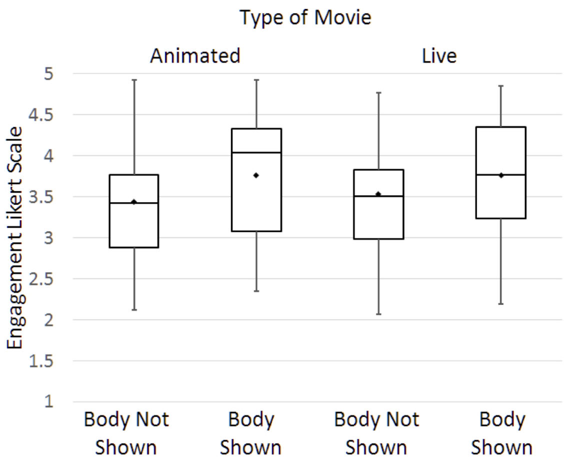

Figure 4: *Participants total Engagement values, higher is better.*

the ratings were significantly different from the neutral value (4). While participants did prefer seeing their body in the VR movies (top right of the figure, Z = -2.449, p = .014), but no significant preference was found on seeing the real world surroundings in the VR movie (bottom left of the figure, Z = -1.669, p = .095). The results show no significant difference in preference over which type of movie the body visualisation should appear (top left of the figure, Z = -1.23, p = .219), while participants did prefer seeing the real world surroundings in Live action film than in Animated film (top middle of the figure, Z = -2.639, p = .008).

At the end of session 2, participants were asked to rate if on a 7-point Likert scale on how participants felt using hand gestures as the interface to control the transition between the real and virtual environments. A one-sample Wilcoxon Signed Ranked test (α = .05) found a statistically significant difference (Z = -2.167, p = .03) between participants' ratings and the neutral value (4) indicating the participants did prefer using hand gesture for controlling

Measure	Type of Movie	Embodimement	Type Of Movie x Embodiment
IPQ (Overall)	F(1,27)=2.283, p=.142	**F(1,27)=5.036, p=.033**	F(1,27)=1.308, p=.263
IPQ - General	F(1,27)=2.309, p=.140	F(1,27)=2.247, p=.146	F(1,27)=1.443, p=.240
IPQ - Spatial Presence	F(1,27)=.849, p=.365	**F(1,27)=6.856, p=.014**	F(1,27)=.801, p=.379
IPQ - Realism	F(1,27)=2.420, p=.131	F(1,27)=3.764, p=.063	F(1,27)=.821, p=.373
IPQ - Involvement	F(1,27)=1.054, p=.314	F(1,27)=.088, p=.769	F(1,27)=.001, p=.975
Engagement (Overall)	F(1,27)=.276, p=.604	F(1,27)=1.183, p=.286	F(1,27)=.118, p=.734
Engagement - Focused Attention	F(1,27)=1.880, p=.182	F(1,27)=.775, p=.386	F(1,27)=.111, p=.742
Engagement - Novelty	F(1,27)=.493, p=.489	F(1,27)=.744, p=.396	F(1,27)=1.284, p=.267
Engagement - Involvement	F(1,27)=.398, p=.533	F(1,27)=1.771, p=.194	F(1,27)=.087, p=.770

Table 1: *Results of IPQ and Engagement questionnaire with two-way repeated measures ANOVA.*

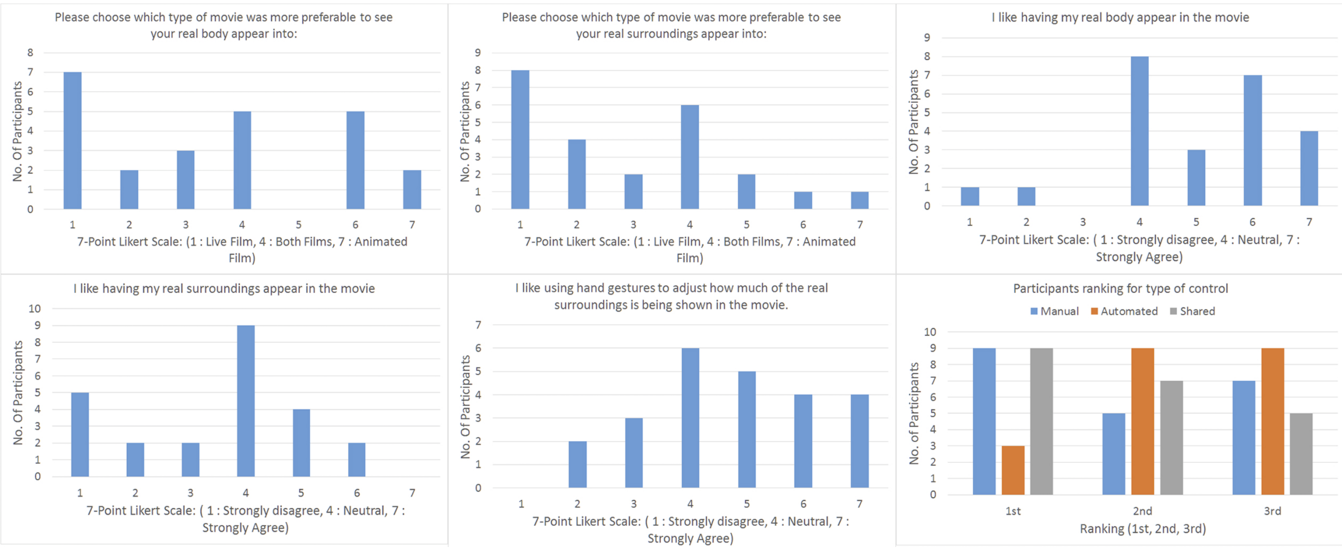

Figure 5: *Results on user preference.*

the transition (see bottom middle of Figure 5). Regarding the results on the question asking participants to rank between the three types of control, Shared control had the leading mean rank (1.81) followed by Manual control (1.90) and lastly Automated control (2.29), however a Friedman test found no statistically significant difference between the conditions ($\chi^2(2) = 2.667$, p = .264).

5.3. Session 1 Qualitative Feedback

We found common themes throughout the subjective feedback and summarised them as follows. The excluded participants commented that if we incorporated purposeful interactions, they would rate their "Body shown" experience as a positive. Quotations shown are feedback from the rest of the participants. These would be important guidelines for future research in user embodiment into immersive cinematic experiences.

Pointless to see surroundings visualised. Many participants mentioned that their surrounding visualisation should play a part in the movie or else it is a distraction to them becoming immersed. This could be used purposely by the movie director to create a special effect for immersing viewers such as an illustration of zapping into space by having the viewer's body and his/her surroundings rapidly blending into the virtual environment. Real items from the physical environment either have to be chosen carefully or visually augmented to fit the theme of the movie. Otherwise, the visualisation would not feel real. For example in the top left of Figure 6, here is an example of sitting on a virtual red chair and the top center of Figure 6 illustrates resting his/her arm on the similar real physical red chair armrest. Top right of Figure 6 demonstrates an idea where one could use their smartphone and the application would visually augment it into a weapon for the user to wield. Most participants agreed that the capability to see their surroundings while still wearing the HMD is helpful for small real-world related tasks such as talking to someone briefly or taking a sip of water without removing the head gear, risking breaking immersion.

"I would prefer if the automatic control was done intentionally by the movie director because the director would want the audience to see from his/her perspective when watching the movie."

We need clothes Participants thought their body fitted, in terms of looks, more in a live-action film than an animated film. Some said that the mismatch between their "live" body and animated movie reminded them about the difference between real and virtual worlds. Some participants suggested modifying their real body visualisation with virtual clothing and accessories to match the theme of the movie. Doing this, they would feel more involved with the movie, allowing user engagement to be natural and easier. For example, imagine in in bottom left of Figure 6 where you would see your body dressed in Hunger Game battle armour, but could still see your unique hands and skin colour.

"If I were to see my body in the film it would be better if I were to see what I am wearing to fit in with the film."

Why do I need a body again? In the current application, "bringing" their real body into the virtual world was great, but pointless if they could do nothing with it. If there were interactions that they had to perform in the movie such as challenges or checkpoints to clear, then it would be more sensible to use see their own hands and body. For example, in the bottom center of Figure 6, one could interact with a cute white rabbit.

"I feel like I didn't need to use my hands and stuff in a movie situation. A more interactive movie would be better."

To see or not to see Participants found that the lack of need for interaction left them satisfied with just turning their heads left and right for a good long while. Doing so caused them to forget their body visualisation because they did not need to look at themselves or raise their hands to perform any actions. If there was interaction they would have to perform, they would naturally first use their hands, reminding themselves, they are "present" there.

"Having to look down to see your body wasn't particularly natural and didn't really improve the experience. It would be better to simply be peripherally aware of your body, as this is what you are naturally used to. A larger field of view could help with this."

5.4. Session 2 Qualitative Feedback

Less control and fast changes if I need to focus on the real world Most participants found the gradual change in blending of the visualisation was too slow if they wanted to take a quick sip of water or food. A state change such as from nothing to just the body and surroundings would be more responsive and quick if they had to attend to minor real world tasks without taking the HMD off. Tasks such as 1) having a drink or eating a snack while watching the movie, 2) briefly talking with friends, or 3) answering a phone call.

More control and precision if I need to focus on the virtual world Interactions with the virtual world require precision, and so participants would like to adjust where they are in the real world. For example if they were sitting at their desk then their body visualisation would be adjusted just enough to "see" their body and not the nearby table. Having this gradual change would give participants more control over what they want to see of the real world while in the cinematic VE, and minimize any distraction as they see fit.

6. Discussion

There was a significant positive main effect of Embodiment on the sense of Presence, yet not on user engagement. From their post-experiment comments, we learnt that users liked to see their body in a movie, but felt there had to be a *reason* for it to be there, and not be shown just because the application could do so. For a convincing immersive "I am really there" experience, it is not enough to provide a visual representation of the real body and surroundings, but there has to be a reason and opportunities for interactions with both the body and surroundings.

Slater [Sla09] mentions that if the user cannot naturally move around and engage with virtual content as if it were real, then the illusion of being in another space may break. It seems that being able to see yourself in a cinematic virtual environment without significant interactions may not be enough for a user to have a strong sense of presence.

With 360° cinematic content being so different from traditional 2D films, the human element is very important for user embodiment. Therefore, there is a lot we do not understand about the effects of placing ourselves in 360° movies. The constantly changing moving viewpoint raises a lot of technical challenges and questions regarding people's reactions to it. There are also questions about the importance of the movie director's influence. For example, if during a scene it is best for viewers to see their body, then the application should have total control and show the body. Viewers do not want to miss or spoil the user embodied cinematic experience that a director has intentionally created.

There are a number of crucial limitations in the research which could be overcome in future work. As shown by the bottom right of Figure 6, the real world visualisation is limited by the field of view of the DS325 camera. This could be overcome by placing two or more of these "windows" side by side by using multiple cameras. It would be interesting to see how much difference in the sense of presence and level of user engagement can be made through increasing the visualisation's field of view. Ideally it should be "large" enough to fake the ability as one participant put it, "to be peripherally aware of the body as one naturally should, without having to physically look down". Other participants' comments are: "Seeing half of your arm was kind of weird and didn't particular help the experience.", and "Dislike the square box of the real body showing in the midst of the virtual world, it reminds me that I am in virtual world when I want to go back to reality."

We admit our current application is limited. A fully developed system would have a movie specially shot and prepared for immersive 360° spherical panoramic viewing with high-quality surround audio. It should also have high quality rendering of the viewer's body capable of wearing virtual clothing. Furthermore, it should support interaction between the viewer and characters and objects.

7. Conclusions and Future Work

In this work we proposed using AV to improve immersive cinematic experiences through user embodiment. Interacting and truly engaging with movie content requires first "being there" and seeing yourself "there." We developed an AV application using inexpensive components to augment the virtual space with real-time 3D visualisation of the hands and body of the user. This allowed them to perceive themselves being present in the cinematic VE. In addition, the proposed system also offered smooth transitioning between the virtual and real environments.

Figure 6: *Augmented Virtuality immersive cinematic experiences with user embodiment.*

From a user experiment using our prototype system, we confirmed Embodiment having a positive effect on sense of presence, yet learnt that to be able to see oneself is insufficient by itself to increase engagement. Activities and interactions must be encouraged for the user experience and movie story to harmonise into an effective experience. Separating them makes it feel as though something is missing, causing levels of immersion to fade.

In the future we plan to explore additional effects such as virtual clothing that matches with the movie which could make the experience more believable while still keeping unique features such as the hands and skin colour. Supporting meaningful interactions in parts of the movie scene is another direction to extend this work.

References

[Ayl99] AYLETT R.: Narrative in virtual environments-towards emergent narrative. In *Proceedings of the AAAI fall symposium on narrative intelligence* (1999), pp. 83–86. 2

[BSJ*15] BUDHIRAJA P., SODHI R., JONES B., KARSCH K., BAILEY B., FORSYTH D.: Where's my drink? enabling peripheral real world interactions while using hmds. *arXiv preprint arXiv:1502.04744* (2015). 2

[BSRH09] BRUDER G., STEINICKE F., ROTHAUS K., HINRICHS K.: Enhancing presence in head-mounted display environments by visual body feedback using head-mounted cameras. In *CyberWorlds, 2009. CW'09. International Conference on* (2009), IEEE, pp. 43–50. 2, 4

[GFG15] GUNTHER T., FRANKE I. S., GROH R.: Aughanded virtuality-the hands in the virtual environment. In *3D User Interfaces (3DUI), 2015 IEEE Symposium on* (2015), IEEE, pp. 157–158. 2

[GKOY03] GVILI R., KAPLAN A., OFEK E., YAHAV G.: Depth keying. In *Electronic Imaging 2003* (2003), International Society for Optics and Photonics, pp. 564–574. 2

[KCCH14] KHATTAK S., COWAN B., CHEPURNA I., HOGUE A.: A real-time reconstructed 3d environment augmented with virtual objects rendered with correct occlusion. In *Games Media Entertainment (GEM), 2014 IEEE* (2014), IEEE, pp. 1–8. 2

[KW10] KWIATEK K., WOOLNER M.: Transporting the viewer into a 360 heritage story: Panoramic interactive narrative presented on a wrap-around screen. In *Virtual Systems and Multimedia (VSMM), 2010 16th International Conference on* (2010), IEEE, pp. 234–241. 2

[MBMSB15] MCGILL M., BOLAND D., MURRAY-SMITH R., BREWSTER S.: A dose of reality: Overcoming usability challenges in vr head-mounted displays. 2143–2152. 2

[MK94] MILGRAM P., KISHINO F.: A taxonomy of mixed reality visual displays. *IEICE TRANSACTIONS on Information and Systems 77*, 12 (1994), 1321–1329. 1

[NSC15] NAHON D., SUBILEAU G., CAPEL B.: "never blind vr" enhancing the virtual reality headset experience with augmented virtuality. In *Virtual Reality (VR), 2015 IEEE* (2015), IEEE, pp. 347–348. 2

[OT10] O'BRIEN H. L., TOMS E. G.: The development and evaluation of a survey to measure user engagement. *Journal of the American Society for Information Science and Technology 61*, 1 (2010), 50–69. 4

[PAK*14] PIUMSOMBOON T., ALTIMIRA D., KIM H., CLARK A., LEE G., BILLINGHURST M.: Grasp-shell vs gesture-speech: A comparison of direct and indirect natural interaction techniques in augmented reality. In *Mixed and Augmented Reality (ISMAR), 2014 IEEE International Symposium on* (2014), IEEE, pp. 73–82. 2

[PVAM10] PORTALÉS C., VIÑALS M. J., ALONSO-MONASTERIO P.: Ar-immersive cinema at the aula natura visitors center. *MultiMedia, IEEE 17*, 4 (2010), 8–15. 2

[Sch03] SCHUBERT T. W.: The sense of presence in virtual environments: A three-component scale measuring spatial presence, involvement, and realness. *Zeitschrift fÃijr Medienpsychologie 15*, 2 (2003), 69–71. 4

[Sch10] SCHULTZE U.: Embodiment and presence in virtual worlds: a review. *Journal of Information Technology 25*, 4 (2010), 434–449. 1, 2

[SFR99] SCHUBERT T., FRIEDMANN F., REGENBRECHT H.: *Embodied presence in virtual environments*. Springer, 1999, pp. 269–278. 4

[SFR01] SCHUBERT T., FRIEDMANN F., REGENBRECHT H.: The experience of presence: Factor analytic insights. *Presence 10*, 3 (2001), 266–281. 4

[Sla09] SLATER M.: Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philosophical Transactions of the Royal Society of London B: Biological Sciences 364*, 1535 (2009), 3549–3557. 7

[SLF*12] SUMA E. A., LIPPS Z., FINKELSTEIN S., KRUM D. M., BOLAS M.: Impossible spaces: Maximizing natural walking in virtual environments with self-overlapping architecture. *Visualization and Computer Graphics, IEEE Transactions on 18*, 4 (2012), 555–564. 2

[TT02] TONG W. L., TAN M. C. C.: Vision and virtuality: The construction of narrative space in film and computer games. *ScreenPlay: cinema/videogames/interfaces* (2002), 98–109. 2

[WW03] WATERWORTH J. A., WATERWORTH E. L.: The meaning of presence. *Presence-Connect 3* (2003). 2

[WW06] WATERWORTH J., WATERWORTH E. L.: Presence as a dimension of communication: context of use and the person. *From Communication to Presence: Cognition, Emotion and Culture Towards the Ultimate Communicative Experience* (2006), 81–96. 2