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

# Effects of Personalized Avatar Texture Fidelity on Identity Recognition in Virtual Reality

Jerald Thomas[1], Mahdi Azmandian[1], Sonia Grunwald[2], Donna Le[3], David Krum[1], Sin-Hwa Kang[1], Evan Suma Rosenberg[1]

[1]USC Institute for Creative Technologies, Los Angeles, CA
[2]Syracuse University, Syracuse, NY
[3]Pomona College, Claremont, CA

**Abstract**

*Recent advances in 3D scanning, reconstruction, and animation techniques have made it possible to rapidly create photorealistic avatars based on real people. While it is now possible to create personalized avatars automatically with consumer-level technology, their visual fidelity still falls far short of 3D avatars created with professional cameras and manual artist effort. To evaluate the importance of investing resources in the creation of high-quality personalized avatars, we conducted an experiment to investigate the effects of varying their visual texture fidelity, specifically focusing on identity recognition of specific individuals. We designed two virtual reality experimental scenarios: (1) selecting a specific avatar from a virtual lineup and (2) searching for an avatar in a virtual crowd. Our results showed that visual fidelity had a significant impact on participants' abilities to identify specific avatars from a lineup wearing a head-mounted display. We also investigated gender effects for both the participants and the confederates from which the avatars were created.*

**CCS Concepts**

•***Human-centered computing*** → ***Virtual reality;*** *User studies;*

## 1. Introduction

As virtual reality (VR) continues to penetrate the consumer market, the need for social VR applications will grow. One key feature of these applications will be the ability to represent each user virtually as an avatar. With groups or potentially crowds of people interacting with each other in immersive virtual worlds, it will become increasingly important to be able to identify and communicate with specific individuals. In some cases, users may choose digital representations that are fantastical or stylized. However, there are many applications where a user may choose a *personalized avatar* that genuinely replicates his or her real world appearance.

Approaches for creating personalized avatars can be thought of as belonging on a spectrum. On one end, there exist methods that have a low cost, but produce avatars with limited visual and motion fidelity, using commodity equipment and completely automated processes. Solutions at the other extreme may produce avatars with photorealistic quality and highly realistic animations, but may cost tens of thousands of dollars. Furthermore, using professional cameras and artists, such approaches can be an expensive venture in both equipment and effort, and is not a practical option for the typical consumer.

In this work, we investigate the impact of visual texture fidelity on user's abilities to discern the identity of an avatar. We focus on comparing two avatar generation approaches that a typical consumer would arguably have access to in the present or near future. The first method uses a large amount of inexpensive cameras and photogrammetry, aided by small amount of technician effort. This method lies somewhere in the middle of the spectrum. While the required equipment is beyond what can be deployed in a home, these photogrammetric approaches are becoming increasingly commercialized and accessible. In the second approach, the avatar textures are intentionally degraded until they resembled avatars created using a low-cost system using single RGB-D camera.

## 2. Related Work

The ability to quickly and automatically generate a personalized avatar is a process that is both becoming cheaper and more technologically viable. The first of these developments used three fixed location Microsoft Kinects to generate a textured 3D avatar and relies on a motorized turntable on which the user stands [TZL*12]. At the same time there was work using a single fixed position camera solution which did not require a turntable and used a single Microsoft Kinect to generate an untextured 3D avatar [WCM12]. Work from 2014 used a single fixed position Microsoft Kinect to generate a textured 3D avatar without using a turntable [SFW*14]. The method for avatar capture used in this work was first developed in 2015 and presented in [FRS17] and utilizes 100 ultra low-cost cameras in fixed positions around the user. However, even the tran-

sition from [SFW*14] to [FRS17] can represent several thousand dollars of equipment.

Two works from Bailenson et al. looked at recognition of a virtual bust presented on a 2D monitor compared to a photograph and found that the participants were slightly better (~10% error difference) at recognizing somebody using a photograph and from a direct facing angle[BBB03, BBBR04]. A 2014 study showed that an avatar is more recognizable as the confederate if the avatar's gestures are the confederate's and not the gestures of another person [FLM*14]. A 2017 experiment presented in looked at the participant's recognition of gait and motion of both virtual avatars and point light representations of a familiar confederate. The results varied greatly and were not found to be significant [NBF*17].

This work looks at recognition of a virtual avatar in an immersive virtual environment with the entire avatar body presented. All avatar animations were provided by pre-recorded motion capture data so the only information that could be used to identify an avatar is body size, body shape, and the face.

## 3. Methods

*Study Design.* A counter-balanced 2x2x2 factorial study was designed and implemented with two within-subject variables (avatar texture fidelity, confederate gender) and one between-subject variable (participant gender). Each participant was exposed to a male and a female confederate avatar, both in a lineup task and a crowd task, and each with four trials with low fidelity (LF) avatars and four trials with high fidelity (HF) avatars. Avatar fidelity was determined by texture and each mesh had a similar number of faces.

*Participants.* A total of 32 people participated in the study (17 male, 15 female). The mean age of participants was 31.75 (*SD* = 14.55). When asked to rate their experience playing 3D video games, the breakdown was as follows: 11 were inexperienced, 12 were a little experienced, five were experienced, and four were very experienced. Participants were recruited through email and Craigslist online classifieds, and were compensated with $20. They were required to be over the age of 18, able to communicate comfortably in spoken and written English, able to hear spoken words, and have normal or corrected-to-normal vision. We excluded people who were pregnant, had a history of epilepsy or seizures, or were sick with an illness that could be transmitted through contact.

*Avatar Generation.* Avatars were captured using a system consisting of 100 raspberry pis (version b) and raspberry pi cameras as well as supporting hardware following designs described in [FRS17] and [SK14]. AgiSoft Photoscan was used to photogrametrically create the avatar's mesh and texture. An open-source autorigging tool was used to automatically rig the avatar [FCS15]. Once inside the Unity game engine final scale adjustments were made and animations could be applied. The whole process from scan to usable rigged avatar took about 20 minutes. For the LF version of the avatar, the texture was degraded to resemble an avatar created using the method described in [SFW*14] with a circa 2014 commodity RGBD sensor (see Figure 1). We used this method to create a group of avatars which we used as distractor avatars.

*Equipment.* During this experiment, participants were wearing an

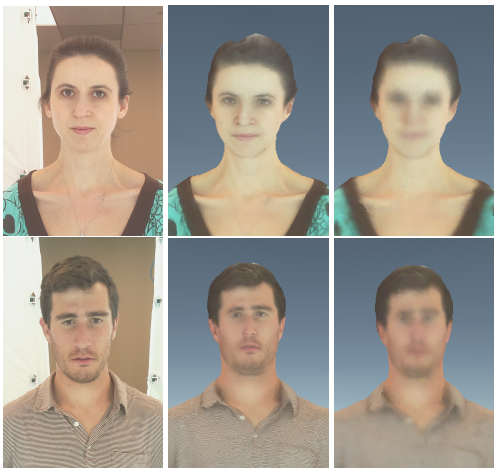

**Figure 1:** *Images taken during the capture process (left). The generated avatars with high (middle) and low (right) texture fidelity.*

Oculus Rift DK2 HMD. The DK2 HMD has a 960x1080 per eye resolution (1080p total resolution), 60Hz refresh rate, 110 degree nominal field of view (FOV), and an internal nine degree of freedom (9DOF) inertial sensor. The participants were tracked in a roughly six meter by ten meter space with a PhaseSpace Impulse X2 motion capture system. To enable 6DOF head tracking by the Impulse X2 system we attached a 5 LED, non-coplanar rig to the DK2. At run-time head orientation was tracked using DK2 internal inertial sensor, while positional information was supplied by the PhaseSpace system. We also performed periodic orientation drift correction between trials using orientation data from the Impulse X2 system. The experiment was run on a Windows 7 64-bit PC with an Intel Core i7-5820k 3.30GHz six-core CPU an NVIDIA GeForce GTX 980 Ti GPU.

*Procedure.* To give the participants a controlled amount of exposure to the confederates we had them watch a short video of each confederate. Before starting the first task with the male confederate the participant was asked to watch a video of the male confederate telling a story. Similarly, before starting the first task with the female confederate the participant was asked to watch a video of the female confederate telling a similar story. There were two stories total, both within five words of each other. The four videos (each confederate was recorded telling both of the stories) were within one second in length of each other. The videos were shot from the same angle with the same background environment and the confederate's whole body was in the shot. These videos were also factored into the counterbalancing in such a way that a participant would not hear the same story told twice. An image of the confederate taken from one of the videos was presented to the participant prior to each trial. Both confederates were scanned eight times, each time with a different outfit, allowing the confederate avatar to be non-distinguishable by clothing for each trial of a task. At the beginning of each task the user was shown a chart similar to a Snellen chart and asked to read the letters. This was to verify that the HMD was fitted and positioned properly so the participant did not experience any blurring of the virtual environment. Participants completed the Kennedy-Lane Simulator Sickness Questionnaire (SSQ) [KLBL93] before and after the experiment. At the end of the study,

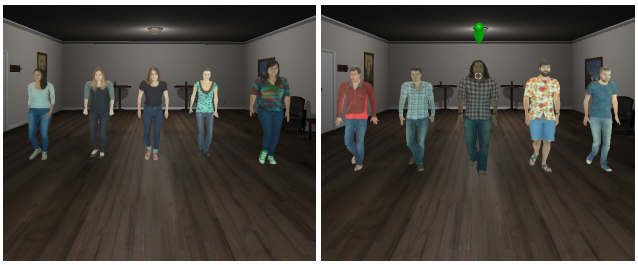

**Figure 2:** *The lineup task (female condition on left, male condition on right) from the participant's perspective. The green object indicates a potential selection.*

**Figure 3:** *The crowd task from the participant's perspective. The green object indicates a potential selection.*

they were also asked to provide basic demographics information, such as age and gender. On average, the entire experiment took approximately 50 minutes, including a five minute break.

*Experiment Tasks.* The experiment was divided into four tasks: the lineup task and the crowd search task, each having to be completed with both a male and a female confederate. These four tasks were counterbalanced in such a way that the two tasks with male confederates were sequential as were the two tasks with female confederates. Each task had eight trials for a total of 32 trials per participant. For each trial the avatars were randomly assigned either a HF texture or a LF texture in such a way that there was a total of four trials with HF avatars and four trials with LF avatars. During each task, the participant was asked to select an avatar by moving the reticle in the center of their vision to the upper torso of an avatar, which would be highlighted by a green icon. The selection would then be completed by pressing a button on a handheld remote.

For the lineup task, the participant was positioned at one end of a rectangular room. A lineup of five avatars, each one meter apart, were arranged facing the participant 10 meters away (see Figure 2). One of the avatars was the confederate's, and the other four randomly chosen distractor avatars of the same gender. When the trial began, the avatars would begin to walk toward the participant until they were within two meters, at which point they would stop walking. During this time, the participant was tasked with finding and selecting the confederate's avatar. Upon selecting any avatar (including an incorrect one) the trial would end. The participant was not given any feedback as to whether the selection was correct, nor was there any limit to the amount of time.

The crowd task was performed in the same rectangular room virtual environment as the lineup task. The room was populated with 23 avatars (one confederate and 22 distractors) in random groupings as to resemble a social gathering (see Figure 3). Once the trial began the participant was tasked with finding and selecting the confederate's avatar, but this time the participant was allowed to walk around. As with the lineup task, the trial was completed after selecting an avatar, and the participant was not given feedback.

*Measures.* The measured variables included time to selection, distance at selection, and accuracy. Time to selection was the amount of time it took the participant from the beginning of the trial until they made a selection, and distance at selection is the virtual distance from the participant to the selected avatar at that point. Accuracy was how often the participant selected the correct avatar.

## 4. Results

Participants' self-reported simulator sickness scores were analyzed with a paired-samples $t$-test, $t(31) = 3.77$, $p < 01$. These results indicated that participants experienced a slight increase in simulator sickness from before the experiment ($M = 16.78$, $SD = 1.16$) compared to afterwards ($M = 18.47$, $SD = 2.72$). All other task performance measures were analyzed using a 2x2x2 mixed analysis of variance (ANOVA) testing the between-subjects effect of participant gender and the within-subjects effects of avatar gender and fidelity (low and high). Unless otherwise noted, all tests cited in this paper used a significance value of $\alpha = .05$. For post-hoc multiple comparison tests, reported significance values have been adjusted using the Holm-Bonferroni method.

### 4.1. Lineup Task

*Time.* Analysis of identification times revealed a significant main effect for avatar fidelity, $F(1,30) = 7.741$, $p < .01$, $\eta_p^2 = .21$. Participants were able to identify HF avatars ($M = 9.30$, $SD = 4.39$) faster than those displayed at LF ($M = 10.65$, $SD = 5.57$). The other main effects and interactions were not significant.

*Distance.* Analysis of identification distances revealed a significant interaction effect between avatar gender and fidelity, $F(1,30) = 15.46$, $p < .01$, $\eta_p^2 = .34$. There were also significant main effects for avatar gender, $F(1,30) = 8.80$, $p < .01$, $\eta_p^2 = .23$, and fidelity, $F(1,30) = 4.81$, $p = .04$, $\eta_p^2 = .14$. To analyze the interaction, we conducted post-hoc paired-samples $t$-tests with a Holm-Bonferroni adjustment. Participants were able to identify female avatars further away when they were displayed at HF ($M = 6.42$, $SD = 2.64$) compared to LF ($M = 5.53$, $SD = 2.64$), $t(31) = 3.71$, $p < .01$. HF female avatars were also identified at greater distances than HF male avatars ($M = 4.88$, $SD = 2.08$), $t(31) = 4.14$, $p < .01$. However, no significant differences were observed between LF male avatars ($M = 5.00$, $SD = 2.21$) and HF avatars, regardless of whether they were male, $p = .50$, or female, $p = .29$.

*Accuracy.* On average, participants were able to accurately identify avatars from the lineup on 83.40% of the trials ($SD = 25.48$). We observed four participants (three male, one female) with very low accuracy scores that were greater than two deviations below the mean, and so we excluded these extreme outliers from the analysis. Analysis of identification accuracy scores revealed a significant interaction effect between avatar gender and fidelity, $F(1,26) = 4.85$, $p = .03$, $\eta_p^2 = .16$. To analyze the interaction, we conducted post-hoc paired-samples $t$-tests; however, none of the multiple compar-

isons tests remained significant after applying a Holm-Bonferroni adjustment. We also found a significant main effect for avatar gender, $F(1,26) = 4.81$, $p = .04$, $\eta_p^2 = .16$, indicating that on average participants were more accurate when identifying female avatars ($M = 86.33\%$, $SD = 30.68$) than male avatars ($M = 80.47\%$, $SD = 27.67$). The other main effects and interactions were not significant.

### 4.2. Crowd Search Task

*Time.* In the crowd search task, the overall average time for participants to identify avatars was 26.61 seconds ($SD = 10.67$). The analysis of identification times did not reveal any significant effects.

*Distance.* Analysis of identification distances revealed a significant main effect for participant gender, $F(1,30) = 7.42$, $p = .01$, $\eta_p^2 = .20$. On average, female participants were closer ($M = 2.00$, $SD = 0.37$) when they identified the avatars compared to male participants ($M = 2.54$, $SD = 0.70$). The other main effects and interactions were not significant.

*Accuracy.* On average, participants were able to accurately identify avatars from the crowd on 67.00% of the trials ($SD = 34.09$). We excluded three extreme outliers (two male, one female) that received accuracy scores of 0% across all trials. Analysis of identification accuracy scores revealed a significant interaction effect between avatar gender and fidelity, $F(1,27) = 9.54$, $p < .01$, $\eta_p^2 = .26$. To analyze the interaction, we conducted post-hoc paired-samples *t*-tests; however, none of the multiple comparisons tests remained significant after applying a Holm-Bonferroni adjustment. The other main effects and interactions were not significant.

## 5. Discussion and Conclusion

We found that in the lineup task, texture fidelity significantly influences how soon and from how far away a person can recognize the confederate avatar in an immersive virtual environment. We also found significant results for recognition distance and time regarding confederate gender, specifically that the female confederate's avatar was easier to recognize than the male confederate's. However, as we only had one female confederate and one male confederate due to resource limitations, we cannot conclude that gender effects were present without a larger sample size.

We did not find any significant results for the crowd search task regarding avatar fidelity and speculate that this is due to two reasons. First, there was much more variability in which the task could be performed compared to the lineup task. Where in the lineup task the confederate avatar always moved towards the participant, the crowd task allowed the participant to move in any direction, including farther away from the confederate avatar. Second, by the nature of the task, the participant can view the confederate from any angle (unlike the lineup task, where the confederate avatar was always presented at a face to face angle). Because the participants were only introduced to the confederate from a face to face angle it may be that the participants did not recognize the confederate avatar's profile or back. This seems to complement the results found in [BBB03] and [BBBR04].

In this study, we found that avatar texture fidelity does seem to have a direct effect on the ability to recognize an avatar as a particular individual. Although this result is not unexpected, we note that it is still valuable to gather empirical data that confirms our intuitions. Additionally, because texture fidelity was expected to have a strong effect, it was a suitable choice to evaluate whether the two presented experimental tasks would be appropriate for future studies of more subtle factors, such as gait, posture, or gestural behavior. Our results suggest that the lineup task is an effective methodology for identity recognition experiments, whereas the crowd task would require substantial redesign to more tightly control for variability.

## 6. Acknowledgements

The work depicted here is sponsored by the U.S. Army Research Laboratory (ARL) under contract number W911NF-14-D-0005 and the National Science Foundation (NSF) grant number CNS-1560426. Statements and opinions expressed and content included do not necessarily reflect the position or the policy of the Government, and no official endorsement should be inferred.

## References

[BBB03] BAILENSON J. N., BEALL A. C., BLASCOVICH J.: Using virtual heads for person identification: An empirical study comparing photographs to photogrammetrically-generated models. *Journal of Forensic Identification 53*, 6 (2003), 722. 2, 4

[BBBR04] BAILENSON J. N., BEALL A. C., BLASCOVICH J., REX C.: Examining virtual busts: Are photogrammetrically generated head models effective for person identification? *Presence: Teleoperators and Virtual Environments 13*, 4 (2004), 416–427. 2, 4

[FCS15] FENG A., CASAS D., SHAPIRO A.: Avatar reshaping and automatic rigging using a deformable model. In *ACM SIGGRAPH Conference on Motion in Games* (2015), ACM, pp. 57–64. 2

[FLM*14] FENG A., LUCAS G., MARSELLA S., SUMA E., CHIU C.-C., CASAS D., SHAPIRO A.: Acting the part: The role of gesture on avatar identity. In *ACM SIGGRAPH Conference on Motion in Games* (2014), ACM, pp. 49–54. 2

[FRS17] FENG A., ROSENBERG E. S., SHAPIRO A.: Just-in-time, viable, 3-d avatars from scans. *Computer Animation and Virtual Worlds 28*, 3-4 (2017). 1, 2

[KLBL93] KENNEDY R. S., LANE N. E., BERBAUM K. S., LILIENTHAL M. G.: Simulator Sickness Questionnaire: An Enhanced Method for Quantifying Simulator Sickness. *The Int. Journal of Aviation Psychology 3*, 3 (1993), 203–220. 2

[NBF*17] NARANG S., BEST A., FENG A., KANG S.-H., MANOCHA D., SHAPIRO A.: Motion recognition of self and others on realistic 3d avatars. *Computer Animation and Virtual Worlds 28*, 3-4 (2017). 2

[SFW*14] SHAPIRO A., FENG A., WANG R., LI H., BOLAS M., MEDIONI G., SUMA E.: Rapid avatar capture and simulation using commodity depth sensors. *Computer Animation and Virtual Worlds 25*, 3-4 (2014), 201–211. 1, 2

[SK14] STRAUB J., KERLIN S.: Development of a large, low-cost, instant 3d scanner. *Technologies 2*, 2 (2014), 76–95. 2

[TZL*12] TONG J., ZHOU J., LIU L., PAN Z., YAN H.: Scanning 3d full human bodies using kinects. *IEEE transactions on visualization and computer graphics 18*, 4 (2012), 643–650. 1

[WCM12] WANG R., CHOI J., MEDIONI G.: Accurate full body scanning from a single fixed 3d camera. In *3D Imaging, Modeling, Processing, Visualization and Transmission* (2012), IEEE, pp. 432–439. 1