ShadowHands: High-Fidelity Remote Hand Gesture Visualization using a Hand Tracker

Erroll Wood[1,2] **Jonathan Taylor**[1] **John Fogarty**[1] **Andrew Fitzgibbon**[1] **Jamie Shotton**[1]
[1]Microsoft [2]University of Cambridge

ABSTRACT
This paper presents *ShadowHands* – a novel technique for visualizing a remote user's hand gestures using a single depth sensor and hand tracking system. Previous work has shown that making distributed users better aware of each other's gestures facilitates remote collaboration. These systems presented virtual embodiments as a stream of raw 2D or 3D data – this data is noisy, and requires high bandwidth and favorable camera positions. Instead, our work uses a hand tracker to capture gestures which we visualize with a high-fidelity hand model. Our system is practical, requiring only a single depth sensor placed below the screen, and can be used without per-user calibration. As we use a 3D model rather than raw data, we can augment the hand's appearance to improve saliency and aesthetics. We alpha-blend this visualization over a shared workspace, so the local user perceives the remote user's hand as if they were separated by a transparent display. We conducted an experiment to compare traditional hand embodiments with our new technique, showing a quantitative improvement in selection accuracy and qualitative improvements in feelings of mutual understanding and engagement.

Author Keywords
Remote Collaboration; Hand Gestures; Hand Tracking

ACM Classification Keywords
H.5.3 Group and Organization Interfaces: Computer-supported cooperative work

INTRODUCTION
During remote collaboration, the remote user is often portrayed with a virtual embodiment. Popular systems, e.g. Skype, generally display only a video of the collaborator's face, either adjacent to or super-imposed onto the shared workspace. Improving upon this has been a research goal for some time, and previous work has shown that more advanced embodiments such as "phantom" hand visualizations [9] or whole-body 3D representations [23] can improve cooperation. These allow the communication of important visual cues including hand gestures, eye gaze, and facial expressions. As a result, collaborators are able to understand each other better, and make their actions and intentions clear.

ISS 2016, November 6–9, 2016, Niagara Falls, ON, Canada.

ACM ISBN 978-1-4503-4248-3/16/11 ...$15.00.
http://dx.doi.org/10.1145/2992154.2992169

Figure 1: A remote user guiding a collaborator through an inking task using ShadowHands (white 3d hand model), visualized on top of a shared workspace.

In this work we present *ShadowHands* – a novel technique for visualizing remote hand gestures. Using a hand tracking system, we capture the gestures of a remote user, and render them onto a shared digital workspace. The virtual hand is presented as if the distributed users are separated by a transparent display (Figure 1). Embodiments in previous work used *image-based* techniques – portraying gestures through streams of 2D images or 3D depth data. Despite careful segmentation and post-processing, this data is noisy and requires high bandwidth to transmit. Instead, our approach is *model-based* – we use a state-of-the-art system to track hands in real-time, and reconstruct them by cleanly rendering a posed 3D hand model. These gestures can then be efficiently encoded in a handful of pose parameters, rather than a dense image representation. Furthermore, while previous systems require complex hardware setups including multiple favorably positioned sensors, our approach requires only a single commodity depth camera placed below the screen.

Comparing the effects of different embodiments is an under-researched problem [4]. We conducted an experiment to compare ShadowHands to two alternative embodiments: a point cloud of RGB-D data [23, 6] and a simple lazer pointer. Participants cooperated to complete three tasks: a pointing task, a maze solving task, and a house sketching task. Results showed that our ShadowHands visualization led to fewer selection errors than the point cloud, and was considered easier to use and less distracting. In summary, the contributions of this work are threefold: i) a novel 3D visualization technique that tracks and portrays hand gestures with high-fidelity: they are noise-free, aesthetically pleasing, and visually salient. ii) a

simple and flexible hardware setup that uses a single depth camera, and can work with both small and large displays. iii) a study that compared ShadowHands with previous visualization techniques, demonstrating both the quantitative and qualitative benefits of our approach.

RELATED WORK

Facilitating better awareness between distributed users has been a challenge in computer supported cooperative work (CSCW) research for some time. Gestures enhance communication in two main ways: they are either *expressive* – aiding speech production and interpretation, or *deictic* – referring to objects or a task at hand [18]. So if we don't allow users to produce or observe guestures, collaboration may suffer.

Virtual hand embodiments

Early CSCW research identified the importance of being able to see the remote user. VideoDraw [19] and Clearboard [8] presented remote collaborators as if they were on the other side of a shared interactive surface. These systems streamed videos of a remote user's hands, face, and body under a local workspace, allowing users to collaborate in drawing tasks. However, bulky and impractical hardware setups were required to capture and transmit video, and it was found that simply overlaying digital ink over a user's video feed was distracting [8]. Following work addressed these issues, using modern equipment and computer vision to relax hardware requirements and display improved visualizations.

VideoArms [18], DigiTable [2], and T3 [21] used webcams and simple computer vision techniques to transmit videos of a user's arms as they interacted in video conferencing scenarios. These 2D image-based systems segmented arm images from the background in a video, allowing them to be selectively composited onto a shared digital workspace. For DigiTable this involved a geometric and photometric analysis of the display-facing video feed, segementing out hand pixels that could not be mapped onto the displayed image. C-Slate [9, 1] proposed an alternative low-cost solution that used polarization filters for segmenting the hands for a *phantom presence* visualization. The hand-to-screen distance was also used to modulate the visualization's transparency and bluriness, mitigating occlusion issues and providing an additional channel for communication.

The introduction of commoditity RGB-D sensors [22] and body tracking [16] has allowed researchers to extend these 2D image-based embodiments to 3D, representing a remote user with a stream of 3D point cloud data. 3D-Board [23] is a wall-sized system that presents a 3D image-based embodiment of a whole user, conveying their eye gaze, facial expresssions, and hand gestures. The remote user is rendered locally by fusing together multiple RGB-D streams from two Kinects mounted above the screen. Immserseboard [6] explored novel visualization styles for the whole body, rendering the remote user as if they were writing side-by-side with the local user, or on a shared mirror. Both these systems reported limitations in terms of the 3D image quality. To improve upon this, Zillner et al. [23] proposed improved noise filtering techniques, and

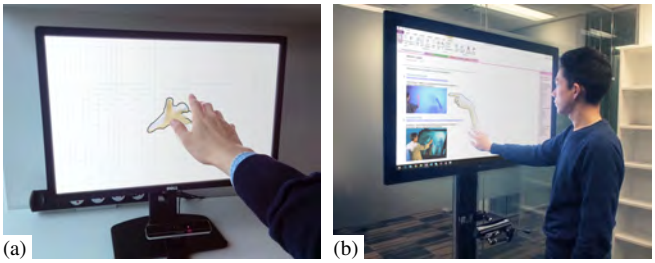

(a) (b)

Figure 2: ShadowHands can be deployed on screens both small, 24-inch (a); and large, 55-inch (b).

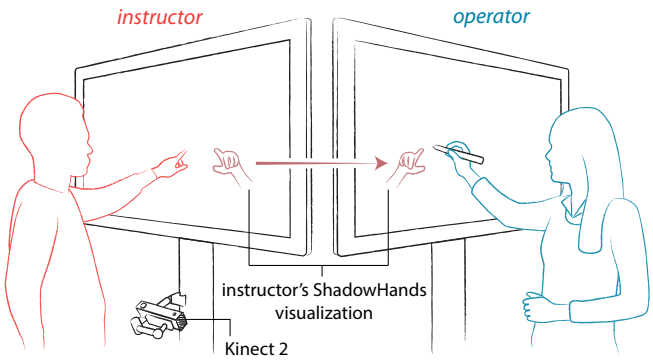


Figure 3: A remote *instructor* performs hand gestures which are tracked using our system and a Kinect V2. The ShadowHands visualization is then blended onto the *operator*'s workspace, who conducts a task using their interactive display.

Higuchi et al. [6] proposed using better quality sensors. Another solution is to cleanly render 3D body models instead of noisy depth data – we examine this approach in our work.

Comparing different embodiments

Though virtual arm embodiments are a common feature of remote collaboration systems in research, there has been little work on directly comparing different types of visualization. [14, 4]. Doucette et al. [4] compared several different hand embodiments: a thin line, 2D shadows, and stretched 2D arm photos. Interestingly, they found that increasing the "level of realism" of the hand embodiment had no effect, though increasing the embodiment's thickness made people better aware of their partner. This suggests that visual prominence is an important factor for collaboration, while realism may not be. However, this study was limited in that it only examined 2D embodiments – it is unclear if these results generalize to the 3D visualizations that are common nowadays.

THE SHADOWHANDS SYSTEM

ShadowHands is a new visualization technique for hand gestures that addresses the limitations of previous work, while relaxing the hardware requirements. We provide a noise-free, visually salient, and aesthetically pleasing hand embodiment, requiring only a single depth camera placed below a screen. To create the impression of looking through a transparent display, we first capture the hand movements of one user using a hand tracker, render these using a 3D model, and finally alpha-blend

Figure 4: The ShadowHands visualization process: Data from the depth sensor (a) is processed by our hand tracker, producing a coarse 3D hand model (b). We then smooth the model using subdivision, shade it with image-based lighting (c), and add black outlines (d). Additionally, our hand pose information allows us to provide alternate stylized visualizations (e).

the rendered image onto both users' screens. In this section we describe the the core components of the ShadowHands system, and the hardware that drives them.

Apparatus

Our hardware setup is practical, requiring only a single depth camera placed below the screen. As can be seen in Figure 2, ShadowHands can be used with both small and large displays. We used an Intel RealSense [7] for close-range desktop operation, or a Kinect V2 [12] for longer-range digital whiteboard use. This paper focusses on the large display scenario: two distributed users collaborating using digital whiteboards.

Figure 3 shows the large screen ShadowHands setup. We use two 55-inch displays: 1) a passive display for a remote instuctor, and 2) an interactive display for a local operator. During our studies, both large displays were situated next to each other, back-to-back. This separated the users so they could not see each other, though they could still hear each other. For convenience, both displays were driven by a single PC which performed both the hand tracking and rendering (3.5GHz CPU, 8 GB memory, K2000 GPU).

Hand tracking

We use a state-of-the-art hand tracker to recover the pose of a user's hand from depth data. It uses machine learning for initialization and recovery; and iterative model-fitting optimization for precise pose and shape alignment. For each input depth frame, the tracker: 1) pre-processes the data to locate the hand, 2) generates a set of initial pose hypotheses, 3) optimizes a smooth fitting energy for each hypothesis, and 4) returns the hand pose that fit best. It can track hands up to several meters away, and is robust against sensor noise. It runs in real time on the CPU only, freeing up the GPU for rendering ShadowHands visualizations in high detail. For more details on the hand tracker, please see previous work [20, 10, 15].

Rendering ShadowHands

We implemented a number of graphics techniques to provide a high-fidelity visualization that was both aesthetically pleasing and visually salient. The output of our hand tracker is a posed mesh of 520 vertices in camera-space (Figure 4b). To create the illusion of viewing the other user's hand in front of them, we first use the camera's intrinsic and extrinsic calibration to re-project this hand mesh as if it were seen from a fixed point 50cm from the centre of the screen.

If visualized directly, this low-resolution mesh would appear unnaturally sharp and blocky, so we instead render a smoothed subdivided mesh (Figure 4c) that better represents the smooth surface of skin. This smoothed mesh is derived by applying two steps of Loop subdivision [11] to the tracked hand mesh. We quickly discovered that users had strong opinions about the appearance of their virtual hands. We therefore experimented with a number of different rendering techniques, including realistic high-resolution skin textures and artistic sketch-style effects. We settled on a clean "toon-style" rendering that users found aesthetically pleasing (Figure 4d), using image-based lighting [3] to illuminate the model. We wanted different hand gestures to be easily identifiable against different workspace backgrounds, so we added black strokes to the outlines of the hand and fingers (Figure 4d). These strokes were added with a post-process: we detected outlines (or edges) as image-space discontinuities in surface normals and depth by filtering with a Laplacian kernel ∇^2 [5].

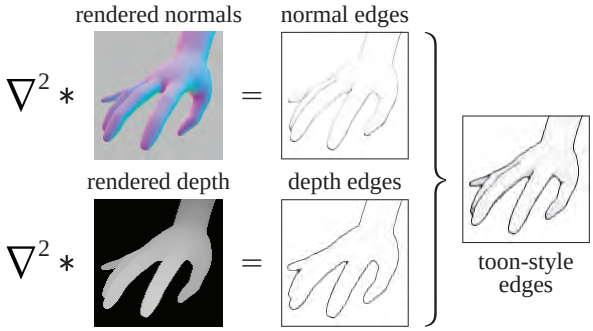


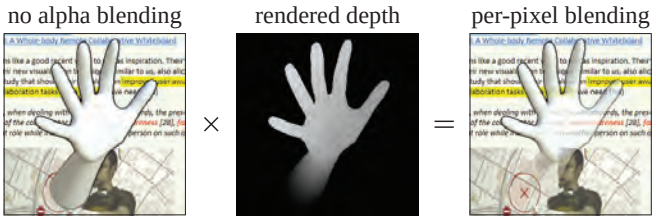
Figure 5 show the improved visual saliency of ShadowHands compared to rendering processed RGB-D sensor data, as done by previous work [23, 6]. This is especially apparent for poses where the sensor cannot see parts of the hand (Figure 5b), or when the RGB sensor is over or under-exposed in difficult lighting conditions (Figure 5c). See the experiment section for details of our sensor data visualizaiton.

Additionally, as our tracker outputs hand pose information in the form of finger-joint angles, we can easily re-pose alternative models to target a user's gestures. This allows for playful communication with non-traditional embodiments e.g. a skeleton hand (Figure 4e) – something that would not be possible without the hand pose data from the hand tracker. All graphics effects were implemented with DirectX.

Blending onto a workspace

The ShadowHands visualization is displayed in a transparent window that appears on top of all other desktop windows.

In this way, it can be used with any other software. While previous work displayed virtual embodiments with a fixed opacity [23], we instead provide a depth-cue by varying the opacity of different parts of the hand model based on their tracked depth. This is done by applying depth-based per-pixel alpha blending in a pixel shader.



We initially render all our 3D geometry with 100% opacity, and then scale the alpha channel for each pixel based on distance to the screen. So a pixel for a finger touching the screen (0cm away) would appear fully opaque ($\alpha = 1.0$), while a forearm pixel 50cm away or more would appear fully transparent ($\alpha = 0.0$). This allows users to control how visible their gestures are: they can draw attention to themselves by gesturing close to the screen; or remove themselves from a discussion by withdrawing away from the screen. Additionally, this prevents the hand visualization from fully obscuring what's underneath it – this can be inconvenient if the remote user is trying to read their screen under the visualization. These transparent window effects were implemented with WINAPI [13].

EXPERIMENT

We conducted an experiment with three tasks to compare our ShadowHands visualization with techniques used in previous work. We focussed on large interactive screen collaboration scenarios, like those between two meeting rooms in two different places. For each experiment we used the same apparatus and compared the same set of virtual hand embodiments.

Participants 14 volunteers (6 male, 8 female) from a universtiy and research lab participated in the experiment, with average age 29.2 ± 4.07[1]. The participants performed all tasks in pairs, with one participant (the remote *instructor*) giving guidance to the other (the local *operator*). The instructors used hand gestures and their voice to give commands, and the operators interpreted these and acted accordingly. 86% of participants reported that they often used video conferencing software for remote collaboration as part of their work. 21% were familiar with large interactive touch screens, and 36% were familiar with hand tracking systems.

Embodiments We compared the effects of using three different virtual hand embodiments:

1. *ShadowHands*: Our high-fidelity tracked hand model.
2. *Point cloud*: This technique was chosen as a comparison to previous work [23, 6]. Similarly to 3D-Board [23], we project the camera's depth image into world space using its intrinsic and extrinsic parameters. Each 3D point is then expanded into a hexagon using a geometry shader, colored using the camera's color image, and rendered from the same virtual camera as ShadowHands. Using tracking information, we only display data from the segmented hand region to reduce background clutter. Furthermore, we median filter the raw RGB-D data to reduce noise.
3. *Laser pointer*: A small red dot in the style of a laser pointer, similar to that used for gesturing in presentation software. This dot was made to follow the user's index finger using the hand tracker.

[1]Results are reported as MEAN $\pm$ STD.DEV.

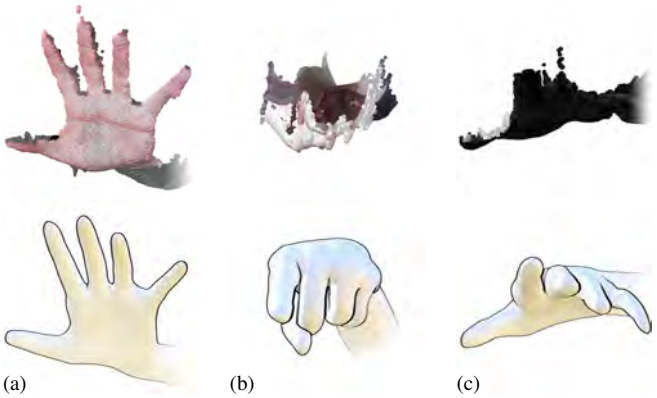


Figure 5: A comparison between rendering RGB-D sensor data (top row), and ShadowHands (bottom row). RGB-D data can be ambiguous for certain hand poses (b) or difficult underexposed lighting conditions (c).

Procedure At the beginning of the session, both participants filled out a pre-study questionnaire on their prior experience with remote collaboration, and the hardware and software used in the experiment. They then performed three tasks: 1) a pointing task, 2) a maze solving task, and 3) a house sketching task. The instructor and operator experienced asymmetric views of each task, so the solution or goal was always visible to the instructor, who had to communicate this to the operator. Before each task, the participants were given as much time as they desired to practice each embodiment as both instructor and operator. The order of embodiments was counterbalanced to reduce learning effects. For the first two tasks, half the participants were operator first, and half the participants were instructor first. After these three tasks were completed, the participants filled out a post-study questionnaire which measured their attitudes towards each different embodiment. The experiment took about an hour to complete in total.

Task 1: Target selection

In the first part of the experiment we investigated the performance characteristics of different embodiments when guiding a user through an abstracted pointing task (Figure 6). This task was chosen to see if ShadowHands provided any improvements for task selection over previous embodiments.

Task design

This task involved the instructor guiding the operator to select circular targets in an 8×8 grid. Before each trail, all targets started off unselected. On beginning the trail, a randomized set of targets in a line chain were highlighted for the instructor

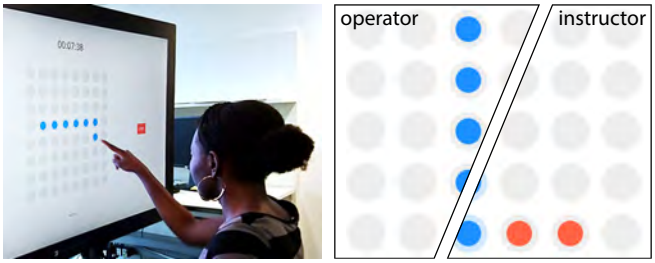


Figure 6: Pointing task: this task involved the operator selecting targets from a grid. The instructor was presented with the candidates for selection, and had to communicate these to the operator using hand gestures only.

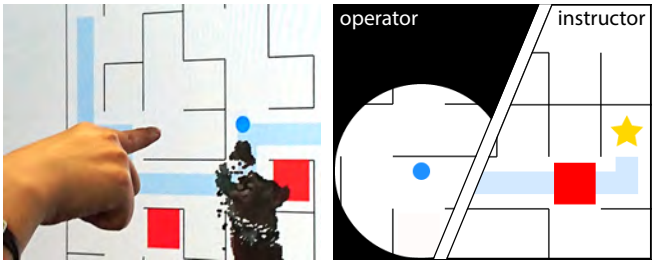


Figure 7: Maze task: the instructor sees the whole maze and must guide the operator through it using guestures and voice. The operator only sees a small region of the maze.

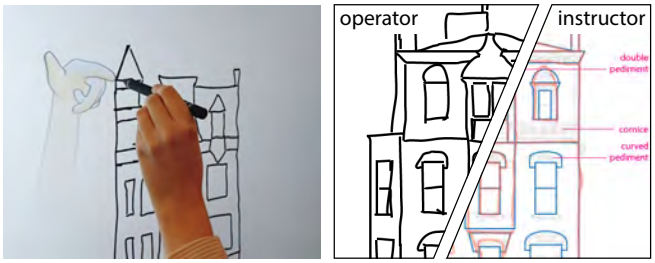


Figure 8: Drawing task: the instructor was presented with a target house design, and the operator was given a plain canvas and a sketching interface. The instructor's goal was to guide the operator so they replicated the target design.

only, marking them as candidates for selection. The instructor then pointed these targets out to the operator, who selected them by touching them. The correct and incorrect selections were visible to the instructor. The participants were told to complete each trial as quickly as possible and were not allowed to talk during this experiment. We measured completion time and selection errors over ten trials for each different embodiment. Once the first operator had completed ten selection trials for each embodiment, the two participants swapped roles and the second operator completed another full set of trials – ten for each embodiment.

Results

Average completion times were $7.05\text{s} \pm 1.91\text{s}$ for ShadowHands, $7.31\text{s} \pm 2.09\text{s}$ for point cloud, and $7.23\text{s} \pm 2.76\text{s}$ for laser pointer. Wilcoxon signed-rank tests showed no significant differences between times for ShadowHands and point cloud ($Z = 4247.5, p = 0.15$), ShadowHands and laser pointer ($Z = 4485.0, p = 0.35$), and point cloud and laser pointer ($Z = 4138.0, p = 0.10$). On average there were 7.10 ± 1.61 targets per trial. Total selection errors were measured as false positives + false negatives. Average errors per trial were 0.41 ± 1.09 for ShadowHands, 0.77 ± 0.12 for point cloud, and 0.21 ± 0.56 for laser pointer. A two-way ANOVA was conducted on the influence of independent variables (embodiment, role) on the number of errors made. The main effect for embodiment type yielded $F(2,414) = 7.26, p < 0.001$, showing significant effect. The main effect for participant role yielded $F(1,414) = 2.18, p = 0.14$, indicating that it was not significant whether participants were operator or instructor first. Wilcoxon signed-ranks tests showed that the point cloud resulted in more errors than ShadowHands ($Z = 463.0, p < 0.05$) and the laser pointer ($Z = 328.0, p < 0.001$). No significant difference in error rate was found between ShadowHands and the laser pointer ($Z = 435.0, p = 0.15$).

Task 2: Maze solving

In the second part of the experiment we examined how hand embodiments were used during a maze solving game (Figure 7). This task was chosen to see if ShadowHands helped instructors communicate spatial and temporal guidance.

Task design

In this task the instructor guided the operator through a randomly generated maze with moving hazard objects. The goal of the operator was to move a blue dot through the maze to touch a gold star, without hitting any red hazards on the way. If the operator collided with a hazard, they were moved backwards in the maze. The instructor and operator experienced asymmetric views: the instructor could see the entire maze while the operator could only see a small region around the blue dot. The participants completed three mazes for each different embodiment. As in the previous task, once the first operator completed three mazes for each embodiment, the participants swapped roles and the second operator completed a further three mazes for each embodiment. We measured maze completion time and hazard collision rate.

Results

Average maze completion times were: $29.9\text{s} \pm 15.4\text{s}$ for ShadowHands, $31.7\text{s} \pm 12.0\text{s}$ for the point cloud, and $28.7 \pm 16.3\text{s}$ for the laser pointer. Wilcoxon signed-rank tests showned no significant differences between times for ShadowHands and laser pointer ($Z = 335.0, p = 0.15$), ShadowHands and point cloud ($Z = 391.0, p = 0.45$), and point cloud and laser pointer ($Z = 287.0, p = 0.05$). Hazard collision rate per trial was measured as the number of collisions / number of hazards. Average collision rates were 0.54 ± 0.56 for ShadowHands, 0.37 ± 0.43 for the point cloud, and 0.53 ± 0.47 for the laser pointer. Wilcoxon signed-rank tests showned no significant differences between collision rates for ShadowHands and laser pointer ($Z = 32.5, p = 0.18$), ShadowHands and point cloud ($Z = 28.0, p = 0.17$), and point cloud and laser pointer ($Z = 25.0, p = 0.78$).

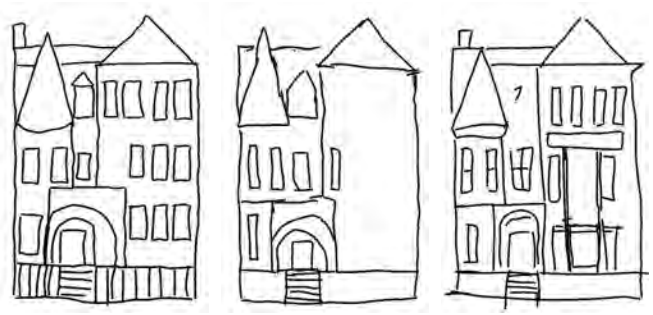


Figure 9: Three sketches of the same house design done by different participants. Note that some participants ran out of time before they could finish a sketch (middle sketch).

Task 3: Sketching a house design

The final task was a collaborative sketching exercise. This task was chosen as we wanted to see how hand embodiments might be used in a more complex and realistic remote collaboration scenario.

Task design

As shown in Figure 8, this task involved sketching a house design. The participants role played as an architect (the operator) being guided through a house design process by a client (the instructor). The instructor was shown an annotated target house design that they communicated to the operator using voice and gestures, while the operator was shown a blank canvas, and was asked to sketch a house design as guided by the instructor. The goal was for the operator's sketch to end up as close as possible to the target design. Four different house designs of the same style [17] were chosen – one to practice on and one for each different embodiment. The participants completed one design per embodiment, and were given a 5 minute time limit for each design. Given the time required, and to prevent fatigue, the participants did not swap roles.

Results

All participants took the maximum time available for each trail. Some examples of the resulting house design sketches can be seen in Figure 9.

Qualitative feedback

Following these tasks, the participants filled out a post-study questionnaire. This included 5-point Likert scale questions to gauge their attitudes towards the different embodiments, and free-form text fields where they could elaborate on what they liked or disliked. The responses to the Likert scale questions can be seen in Figure 10. Significance was measured using the Mann-Whitney U-test.

ShadowHands received positive feedback concerning its appearance – participants found it *"visually appealing"* and appreciated that it *"blended well with the background"*. They reported ShadowHands was *"clear and easy to follow"*, and found it less distracting than the Point Cloud ($U = 158.5, p < 0.005$). The biggest issue with ShadowHands was hand tracking failure – *"when it works, it's super, but I feel like half the time that did not happen"*. 43% of participants reported some level of issue with tracking, ranging from minor glitches (*"sometimes switches from showing only the index finger to two fingers"*) to serious frequent tracking failure (*"hand gesture often wrong"*). Despite this, most participants still found it easy to make themselves understood.

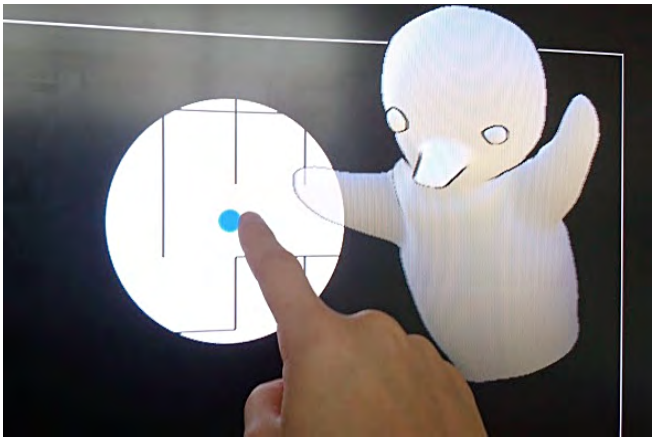


Figure 11: An example of emergent behaviour: some participants role-played using hand puppet visualizations.

Using the point cloud, participants found it harder to make themselves understood compared to ShadowHands ($U = 44.5, p < 0.005$) or the laser pointer ($U = 39.0, p < 0.005$). As instructors, they found themselves awkwardly positioning their hands *"to get into a shape that would be easily understood by the operator"*, for example, positioning their hand *"away from [their] finger for it to be seen"*. Operators reported that it was *"often hard to see where the hand actually is pointing"*, and were confused by the unnatural hand gestures – *"it was hard to tell which was the thumb and/or index finger"*. Furthermore, they complained that the point cloud visualization was distracting – *"I did not like that it was so pixelated and easily dissolved"*. They therefore found it harder to understand point cloud gestures compared to ShadowHands ($U = 28.5, p < 0.001$) or the laser pointer ($U = 30.0, p < 0.001$). Some participants appreciated that the point cloud was *"predictable"* and did not exhibit tracking glitches like ShadowHands.

The laser pointer was a popular technique, with most participants finding it easy to use and understand – *"it was very easy to follow"*. However, some participants felt they missed out on more complex signals – *"there was no way to signal other intentions than an exact point"*. While they liked that the pointer was precise, some participants found it was easy to obscure (*"I kept obscuring the pointer with my finger"*) or lose track of (*"it was hard to locate"*). The biggest complaint was that the laser pointer felt *"impersonal"*, with fewer participants feeling a sense of engagement compared to ShadowHands ($U = 46.0, p < 0.01$).

Emergent behaviour

We were interested in discovering if ShadowHands afforded any communication techniques that we did not expect. The maze and sketching tasks were therefore designed to be fun

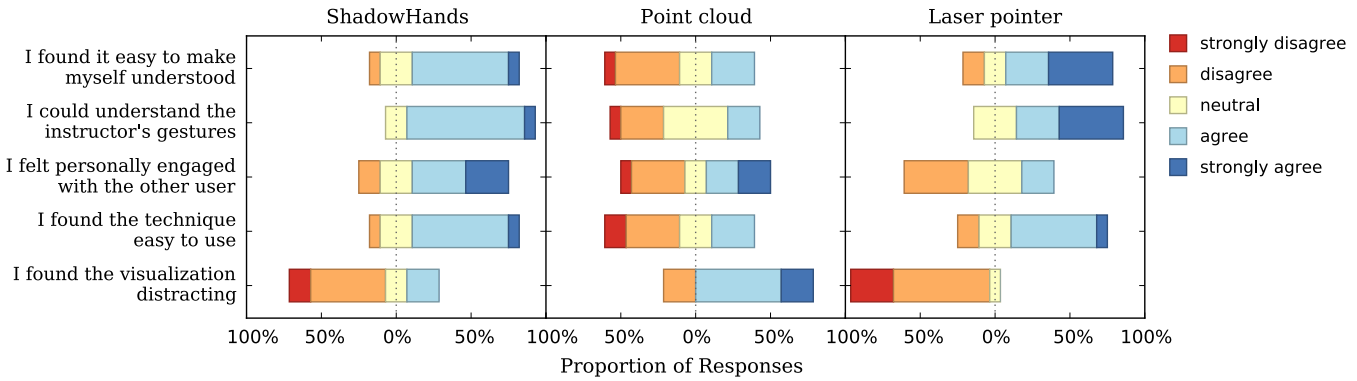


Figure 10: Stacked diverging bar charts of our Likert Scale responses (bars show the number of participants deviating from neutral response). ShadowHands was more popular than Point Cloud in terms of understanding, ease of use, and distraction. Attitudes towards ShadowHands and Laser Pointer were similar, except for feelings of personal engagement.

and complex, to encourage participants to use hand gestures beyond simply pointing at the screen.

We found that ShadowHands were used in several ways beyond pointing. During the maze task, we observed participants holding their palm up to the screen in a "stop" gesture to prevent the operator colliding with nearby hazards. We also observed the ShadowHands being used as a measurement tool during the drawing task: one participant used the width of their fingers to express the width of different architectural features – something tricky to do with only a pointer.

We also observed playful behaviour. During the maze task, a participant accidentally enabled an alternative ShadowHands visualization: a hand-puppet. They then proceeded to role-play as the puppet, gesturing with the puppets arms and head to guide the operator out of the maze[2]. While it is unlikely this provided any measurable benefit to task performance, the participants enjoyed being able to experience the game in this different way.

DISCUSSION

We now summaraize our key findings from the study. In our pointing task we discovered that ShadowHands led to higher accuracy than the point cloud. This was supported by qualitative feedback, with more participants finding ShadowHands easy to use and understand than the point cloud. Current commodity sensors cannot provide good enough RBG-D data for clearly representing hand gestures in our usage scenario. Participants were therefore confused by the ambiguity of the point cloud, preferring the clean ShadowHands visualizations. This finding corroborates previous work suggesting the importance of visual prominence for remote embodiments [4].

As might be expected, the participants felt more personally engaged when using ShadowHands or the point cloud compared to the laser pointer. However, no difference in personal engagement could be found between the point cloud and the laser pointer. This suggests that realism might not be as important for awareness between remote users as previously thought.

[2]Data from these trials were not recorded for the experiment.

If we can render stylised model-based embodiments that faithfully represent a user's gestures, we might improve gesture clarity without affecting awareness.

Though users did not feel personally engaged when they used the laser pointer, this did not lead to any empirical differences in task performance compared to ShadowHands. Furthermore, participant attitutudes towards ShadowHands and the laser pointer were similar, apart from feelings of engagement. This suggests that simple visualization techniques like a pointer are sufficient for many tasks, and future work should include it as a baseline when comparing new visualization techniques.

CONCLUSION

We introduced ShadowHands– a novel technique for visualizing hand gestures to assist remote collaboration. Rather than streaming 2D or 3D data, ShadowHands uses a hand tracker to capture a user's hand gestures, and presents them to a their collaborator. We ensure this visualization is aesthetically pleasing and visually prominent through toon-style shading and per-pixel alpha blending. We ran an experiment to compare ShadowHands to previous visualization techniques. Participants were quantitatively better at selecting targets with ShadowHands (compared with a point cloud), and this translated qualitatively into improved attitudes towards mutual understanding, ease of use, and distraction.

The primary complaint with ShadowHands was that it was unpredictable – some users experienced glitches in hand tracking that they found frustrating. As hand tracking techniques continue to improve, we hope this issue can be resolved in the future. We also discovered that participants enjoyed using playful non-hand visualizations, and would like to investigate these interactions more formally in future work.

ACKNOWLEDGEMENTS

We would like to thank Jeff Han for suggesting we explore this idea. We would also like to thank the reviewers for their detailed and helpful feedback.

REFERENCES

1. Agarwal, A., Izadi, S., Chandraker, M., and Blake, A. High precision multi-touch sensing on surfaces using overhead cameras. In *Workshop on Horizontal Interactive Human-Computer Systems, TABLETOP'07*, IEEE (2007).
2. Coldefy, F., and Louis-dit Picard, S. Digitable: an interactive multiuser table for collocated and remote collaboration enabling remote gesture visualization. In *IEEE CVPR'07*, IEEE (2007).
3. Debevec, P. Image-based lighting. *IEEE Computer Graphics and Applications* (2002).
4. Doucette, A., Gutwin, C., Mandryk, R. L., Nacenta, M., and Sharma, S. Sometimes when we touch: how arm embodiments change reaching and collaboration on digital tables. In *Proc. CSCW'13*, ACM (2013).
5. Gonzalez, R. C., and Woods, R. E. Digital image processing. *Nueva Jersey* (2008).
6. Higuchi, K., Chen, Y., Chou, P. A., Zhang, Z., and Liu, Z. Immerseboard: Immersive telepresence experience using a digital whiteboard. In *Proc. ACM CHI'15* (2015).
7. Intel. Intel realsense camera f200. http://www.intel.com/content/www/us/en/architecture-and-technology/realsense-shortrange.html, 2016. Accessed: 2016-07-07.
8. Ishii, H., and Kobayashi, M. Clearboard: a seamless medium for shared drawing and conversation with eye contact. In *Proc. ACM CHI'92*, ACM (1992).
9. Izadi, S., Agarwal, A., Criminisi, A., Winn, J., Blake, A., and Fitzgibbon, A. C-slate: a multi-touch and object recognition system for remote collaboration using horizontal surfaces. In *Workshop on Horizontal Interactive Human-Computer Systems, TABLETOP'07*, IEEE (2007).
10. Joseph Tan, D., Cashman, T., Taylor, J., Fitzgibbon, A., Tarlow, D., Khamis, S., Izadi, S., and Shotton, J. Fits like a glove: Rapid and reliable hand shape personalization. In *Proc. CVPR'16* (2016).
11. Loop, C. Smooth subdivision surfaces based on triangles.
12. Microsoft. Kinect for windows. https://developer.microsoft.com/en-us/windows/kinect/, 2016. Accessed: 2016-07-07.
13. Microsoft. Windows api index. https://msdn.microsoft.com/en-gb/library/windows/desktop/ff818516(v=vs.85).aspx, 2016. Accessed: 2016-07-07.
14. Pinelle, D., Nacenta, M., Gutwin, C., and Stach, T. The effects of co-present embodiments on awareness and collaboration in tabletop groupware. In *Proceedings of graphics interface 2008*, Canadian Information Processing Society (2008).
15. Sharp, T., Keskin, C., Robertson, D., Taylor, J., Shotton, J., Kim, D., Rhemann, C., Leichter, I., Vinnikov, A., Wei, Y., et al. Accurate, robust, and flexible real-time hand tracking. In *Proc. CHI'15*, ACM (2015).
16. Shotton, J., Sharp, T., Kipman, A., Fitzgibbon, A., Finocchio, M., Blake, A., Cook, M., and Moore, R. Real-time human pose recognition in parts from single depth images. *Communications of the ACM* (2013).
17. Survey, H. A. B. Chapline street row historic district, 2301 to 2319 chapline street. In *Library of Congress* (1933).
18. Tang, A., Neustaedter, C., and Greenberg, S. Videoarms: embodiments for mixed presence groupware. In *People and Computers XX-Engage*. Springer, 2007.
19. Tang, J. C., and Minneman, S. L. Videodraw: a video interface for collaborative drawing. *ACM Transactions on Information Systems (TOIS)* (1991).
20. Taylor, J., Bordeaux, L., Cashman, T., Corish, B., Keskin, C., Soto, E., Sweeney, D., Valentin, J., Luff, B., Topalian, A., Wood, E., Khamis, S., Kohli, P., Sharp, T., Izadi, S., Banks, R., Fitzgibbon, A., and Shotton, J. Efficient and precise interactive hand tracking through joint, continuous optimization of pose and correspondences. In *ACM SIGGRAPH* (2016).
21. Tuddenham, P., and Robinson, P. T3: Rapid prototyping of high-resolution and mixed-presence tabletop applications. In *Workshop on Horizontal Interactive Human-Computer Systems, TABLETOP'07*, IEEE (2007).
22. Zhang, Z. Microsoft kinect sensor and its effect. *IEEE multimedia* (2012).
23. Zillner, J., Rhemann, C., Izadi, S., and Haller, M. 3d-board: a whole-body remote collaborative whiteboard. In *Proc. UIST'14*, ACM (2014).