DirActor: Creating Interaction Illustrations by Oneself through Directing and Acting Simultaneously in VR

Seung-Jun Lee
Department of Industrial Design, KAIST
Republic of Korea
seung-jun.lee@kaist.ac.kr

Siripon Sutthiwanna
School of Computing, KAIST
Republic of Korea
siripon.s@kaist.ac.kr

Joon Hyub Lee
Department of Industrial Design, KAIST
Republic of Korea
joonhyub.lee@kaist.ac.kr

Seok-Hyung Bae
Department of Industrial Design, KAIST
Republic of Korea
seokhyung.bae@kaist.ac.kr

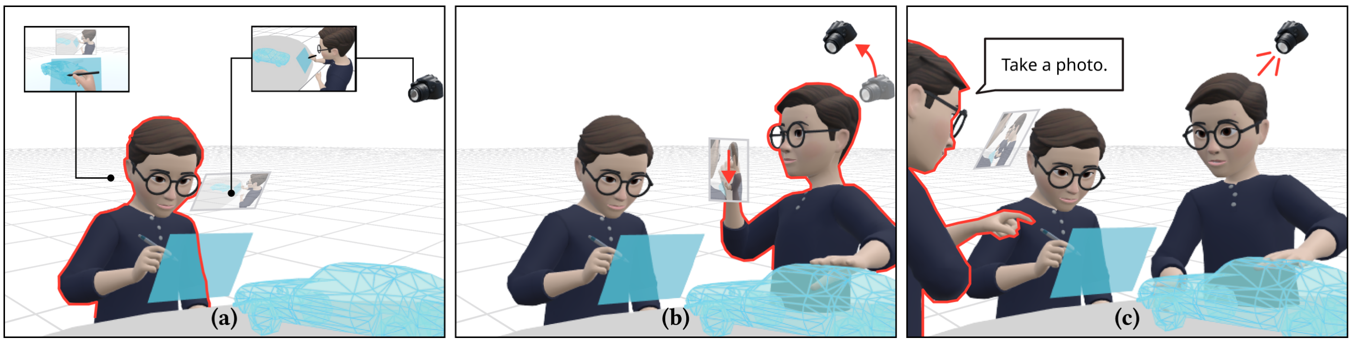

Figure 1: We propose *DirActor*, a novel VR system to help HCI researchers create interaction illustrations featuring multiple human figures in various poses and compositions by oneself. In our system, the user (highlighted in red) can act and create still copies of his avatar. Simultaneously, the user can (a) monitor his appearance on the camera in real-time and (b) control the camera through the *interactive picture frame*, and (c) issue voice cues through *Hold'n'Speak*.

ABSTRACT

In HCI research papers, interaction illustrations are essential to vividly expressing user scenarios arising from novel interactions. However, creating these illustrations through drawing or photography can be challenging, especially when they involve human figures. In this study, we propose the DirActor system that helps researchers create interaction illustrations in VR that can be used as-is or post-processed, by becoming both the director and the actor simultaneously. We reproduced interaction illustrations from past ACM CHI Best and Honorable Mention papers using the proposed system to showcase its usefulness and versatility.

CCS CONCEPTS

• **Human-centered computing** → **Interaction techniques**.

KEYWORDS

interaction illustration, virtual reality, bare hands, voice command

ACM Reference Format:

Seung-Jun Lee, Siripon Sutthiwanna, Joon Hyub Lee, and Seok-Hyung Bae. 2024. DirActor: Creating Interaction Illustrations by Oneself through Directing and Acting Simultaneously in VR. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA '24), May 11–16, 2024, Honolulu, HI, USA*. ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3613905.3650768

CHI EA '24, May 11–16, 2024, Honolulu, HI, USA

ACM ISBN 979-8-4007-0331-7/24/05
https://doi.org/10.1145/3613905.3650768

1 INTRODUCTION

In the research field of human-computer interaction (HCI), drawings or photographs are often necessary to visualize interactions that can be difficult to convey through text alone. However, HCI researchers can face several challenges when creating such interaction illustrations.

Drawing a high-quality illustration requires depicting human figures in the correct anatomy, perspective, and layout, a skill that demands professional training [43]. While researchers can use a photograph as a background for drawing over, a suitable photograph may not be readily available, in which case, researchers must take a new one.

Photography, on the other hand, requires setting up the scene, configuring the camera, and acting, which can be physically demanding to handle by oneself. While working with colleagues can help, they may not always be available in sufficient numbers. Also, it can be tricky to direct actors to pose exactly as wanted when the interaction illustration contains many human figures.

We thus propose DirActor, a novel VR system designed to help researchers quickly and easily create interaction illustrations that contain many human figures by oneself (Figure 1). In our system, by becoming both the director and the actor, the user can demonstrate interactions with the body of his avatar, and at the same time, monitor the shot, control the camera, and give voice cues.

2 RELATED WORK

In this section, we introduce previous works in creating illustrations that contain human figures and provide an overview of the main goals of this study in relation to them.

2.1 Taxonomy of Interaction Illustrations

Antoine et al. [3] suggested a taxonomy of interaction illustrations in HCI research papers. The taxonomy includes five viewpoint types (1^{st} person, 3^{rd} person, over-shoulder ¾, top, and UI only), two user types (single and multiple), and nine interaction types (touch, mid-air, distal, on-body, computer, tangible, controllers, pen, and gaze). In this study, we accommodate depiction of every viewpoint type except the "UI only," which does not involve a human figure, both the single and multiple user types, and every interaction type with the help of pre-imported 3D props.

2.2 Posing 3D Human Figures

While many commercial 3D staging tools [9, 27, 40] support realistic representation of the human figure, manipulating the body to take a desired pose can be difficult and time-consuming. Techniques that extract poses from physical puppets [12, 16, 34] or sketches [6, 13, 25, 29] have been proposed. However, we consider acting and capturing those poses to be more direct and proficient, which can be readily achieved through the hand and body tracking features of commercially available VR headsets.

2.3 AI Assistance

Generative AI models can quickly produce high-quality human illustrations from text prompts [11, 33, 36] and sketches [7, 50]. However, it can still be challenging for the user to attain desired results even after many tries [26]. Instead, we take the approach where the user has full control over the scene, the camera, and the poses. In doing so, the user can issue voice commands, just as the director does to the crew members, which our system accommodates using natural language processing AI.

2.4 Multi-Modal Input

Antoine et al. proposed using the Leap Motion hand tracking sensor to pose the human hand model in creating interaction illustrations [2]. Chi et al. proposed using the Microsoft Kinect body tracking sensor to pose the human body model and voice commands to invoke recording functions while posing in creating step-by-step illustrated instructions [8]. In this study, we incorporate hand and body tracking sensors as well as voice commands in enabling an intuitive and frictionless workflow for creating interaction illustrations in VR.

3 SYSTEM

In this section, we present the DirActor system, which enables the user to monitor the shot, control the camera, and give the cue all the while being an avatar and acting out poses in VR. We designed the *interactive picture frame* and *Hold'n'Speak* interactions inspired by communication between directors and actors in actual photo shooting sessions.

3.1 Monitoring the Shot While Acting

The interactive picture frame is a semi-transparent rectangular panel that displays the real-time view from the camera and follows the user's head so that the user can glance at any time while acting. The user can grab the interactive picture frame and relocate it (Figure 2a), change its aspect ratio (Figure 2b), and resize it (Figure 2c), during which operation it turns opaque.

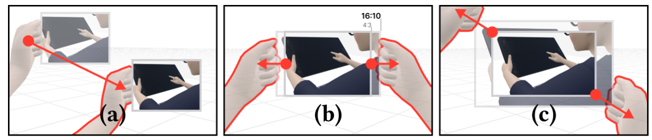

Figure 2: The user can (a) relocate the interactive picture frame by grabbing a border and moving the hand, (b) adjust the aspect ratio by grabbing two parallel borders and moving the hands closer or farther during which it snaps to common aspect ratios (1:1, 4:3, 3:2, 16:10, and 16:9), and (c) resize while maintaining the aspect ratio by grabbing two opposite corners and moving the hands closer or farther.

3.2 Controlling the Camera While Acting

By performing a set of single- and multi-touch gestures with one or two hands on the interactive picture frame, the user can freely manipulate the camera (Figures 3-5) without leaving his position while acting.

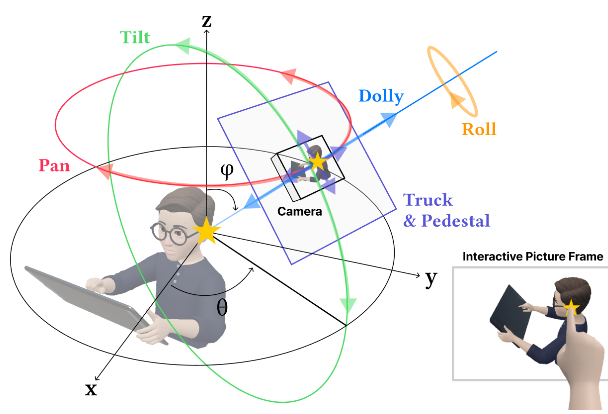

Figure 3: Overview of camera manipulation through the interactive picture frame. When the user touches a point on the interactive picture frame (in this image, the avatar's left ear), a pivot point (⋆) is calculated. The user can pan (red) or tilt (green) around ⋆, dolly in or out (blue) along the axis connecting ⋆ and the camera, roll (yellow) around the axis, and truck & pedestal on the plane parallel to the camera plane (purple). In every camera movement except truck & pedestal, the camera always points to ⋆.

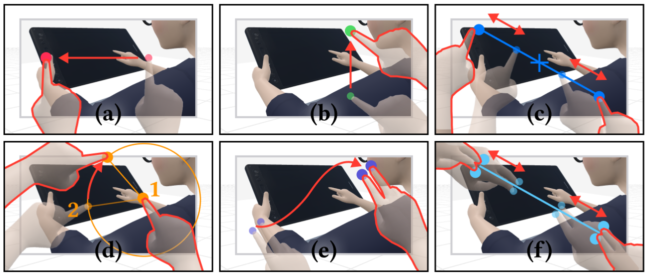

Figure 4: The user can (a) pan by putting one finger on a pivot point and dragging horizontally, (b) tilt by putting one finger on a pivot point and dragging vertically, (c) dolly by putting two fingers to aim the pivot point and moving the hands closer or farther, (d) roll by putting one finger on a pivot point and dragging another along the circle, (e) truck & pedestal by dragging two fingers, and (f) adjust field of view by putting two fingers of both hands at the same time and moving the hands closer or farther.

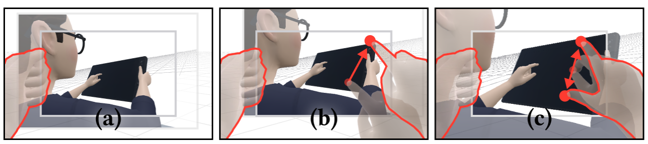

Figure 5: The user can (a) grab a border of the interactive picture frame to reveal the entire image, (b) drag with one finger to translate the crop, and (c) pinch with two fingers to resize the crop.

3.3 Cueing While Acting

Using our Hold'n'Speak interactions, the user can issue voice commands when his hands and body are busy performing a pose. Just as the actor holds his pose until the director gives a cue in actual photo shooting sessions, the user first holds the current pose for a short period (e.g. 0.5 seconds) and then speaks out a command (Table 1), to create (Figure 6), remove (Figure 7), and edit (Figure 8) an avatar and take a photo.

Command	*Example phrase*
Create a new avatar	"Create an avatar.", "New."
Remove the selected avatar	"Remove the avatar.", "Delete."
Edit the selected avatar	"Edit the avatar.", "Re-pose."
Take a photo	"Take a photo.", "Capture."

Table 1: The user can invoke system functions by speaking specific or similar phrases.

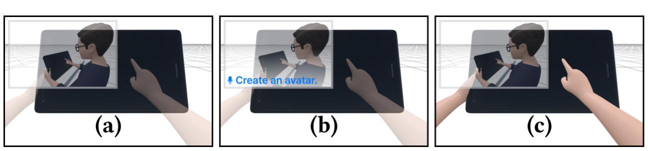

Figure 6: The user (semi-transparent) (a) holds a pose for a short period, and then (b) speaks the avatar creation command to (c) create an avatar (opaque) with the same pose.

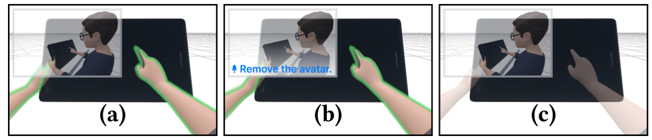

Figure 7: The user (a) approaches an avatar and holds a similar pose for a short period to select the avatar (signaled by green outline), and then (b) speaks the avatar removal command to (c) remove the avatar.

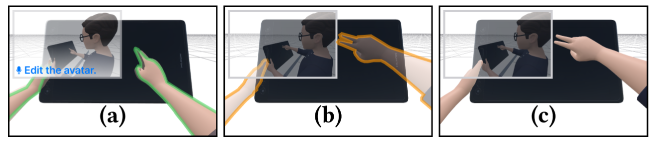

Figure 8: The user (a) selects an avatar and speaks the avatar edit command to (b) unfreeze the avatar (signaled by orange outline) and re-pose. (c) The user holds the new pose for a short period to freeze the avatar.

4 USER SCENARIO

In this section, we present an exemplary user scenario to showcase the workflow using the DirActor system. In this user scenario, a user creates an interaction illustration that envisions a futuristic design tool, which converts a car sketch drawn on a handheld translucent graphics tablet into a 3D hologram, allowing a group of designers to collaboratively review the design. As the interaction illustration involves many human figures engaging with various props and each other, an effective scene layout that highlights each element and interaction is required, which can be challenging to obtain through drawing or photography. The user scenario describes the user's performance before (Figure 9), during (Figures 10-12), and after (Figure 13) the photo shoot in the user's first-person view.

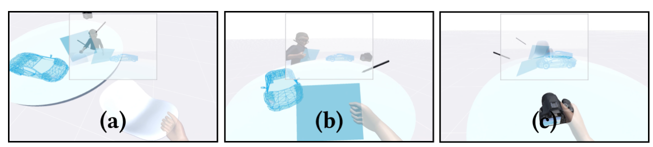

Figure 9: Before the photo shoot, the user enters a VR scene with pre-imported 3D objects. As if setting up a studio, the user grabs and moves (a) large pieces of furniture (in this image, a chair) and (b) small props (in this image, a translucent graphics tablet). Then, the user (c) roughly places the camera which can be finely adjusted in later stages.

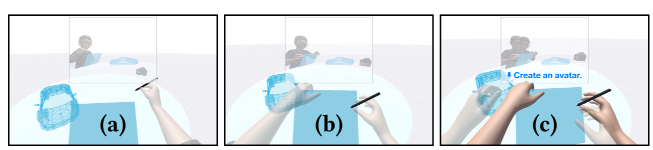

Figure 10: During the photo shoot, the user acts out the first avatar. (a) The user sits in front of the graphics tablet and grabs a pen in the air. Then, the user (b) pretends to sketch while looking at himself through the interactive picture frame and (c) creates an avatar with Hold'n'Speak.

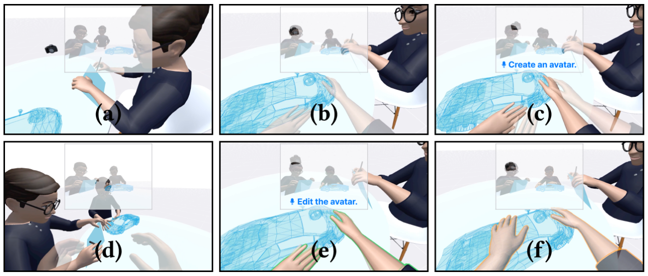

Figure 11: During the photo shoot, the user acts out the second avatar. (a) The user stands up and moves next to the first avatar. (b) The user pretends to touch the car hologram with both hands and (c) creates an avatar with Hold'n'Speak. (d) While heading to create the third avatar, the user decides the second avatar should be touching the car's roof. (e) The user returns and uses Hold'n'Speak to (f) edit the pose.

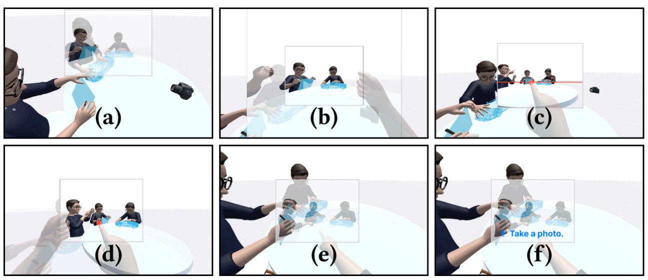

Figure 12: During the photo shoot, the user acts out the third avatar. (a) The user walks to the other side of the table and (b) pulls down the interactive picture frame to see if the avatars and props are in an appropriate composition. Unsatisfied, the user (c) moves the camera and (d) crops the image so that the photo better illustrates each avatar's role in the scene. Then, the user (e) acts out the last avatar by pretending to discuss, and (f) takes a photo with Hold'n'Speak.

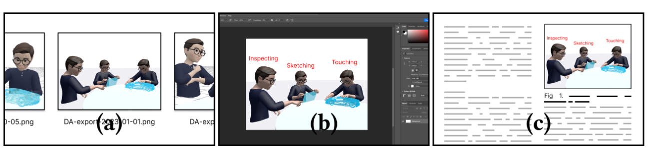

Figure 13: After the photo shoot, (a) the image is saved in PNG format in a local folder. The user (b) post-processes the image with a 2D graphics tool (in this image, Adobe Photoshop) and (c) inserts the image into his paper.

5 IMPLEMENTATION

We implemented our system using the Unity 3D game engine and the hand and body tracking, as well as voice input capabilities of the Meta Quest 2 VR headset. The Meta Avatars SDK currently does not support leg and foot tracking, but we anticipate integrating it to our system when it becomes available in the near future. We process the user's voice command using Wit.ai's speech recognition API [32].

6 SHOWCASE

In this section, we showcase the usefulness of the DirActor system by reproducing interaction illustrations from HCI research papers with it. We chose ACM CHI Best and Honorable Mention papers for their high novelty, representativeness, and presentation quality.

First, out of 705 awarded papers over the past five years, we filtered 192 papers that proposed new interaction techniques or systems. Then, out of 1,838 figures in these papers, we filtered 441 figures that were interaction illustrations.

Next, we selected 29 interaction illustrations out of the 441 based on four criteria. First, the illustration must mainly depict a human figure interacting with the system (e.g. not an illustration with only the tip of a finger). Second, the illustration must not necessitate being an actual photograph (e.g. not an illustration of a mixed-reality system that must distinguish the real and virtual worlds). Third, the illustration must not involve a body part that the VR headset can not currently track (e.g. not an illustration of foot movement). Fourth, we selected only one illustration when a paper included many similar ones.

The 29 illustrations from 29 papers [1, 4, 5, 10, 14, 15, 17–24, 28, 30, 31, 35, 37–39, 41, 42, 44–49] contained 169 human figures, and included every viewpoint type, user type, and interaction type from the taxonomy of interaction illustrations as proposed by Antoine et al. [3] (Table 2).

One of the authors reproduced the illustrations using DirActor by himself (Figures 14-19) and measured the time spent on setting up the virtual scene, posing, and shooting in DirActor. The time spent on finding or creating 3D objects to import and post-processing the captured image were considered outside the scope of this study and not measured. On average, each human figure took 54 seconds to reproduce.

Category	Count
Viewpoint type	
1^{st} person	8
3^{rd} person	20
Over-shoulder 3/4	3
Top	5
User type	
Single	28
Multiple	1
Interaction type	
Touch	10
Mid-air	11
Distal	7
On-body	9
Computer	8
Tangible	6
Controller	3
Pen	3
Gaze	2

Table 2: Counts of interaction illustrations by category. The 29 interaction illustrations selected from 29 papers included every viewpoint type, user type, and interaction type classified by Antoine et al. [3] and targeted by our paper. An interaction illustration could belong to multiple categories.

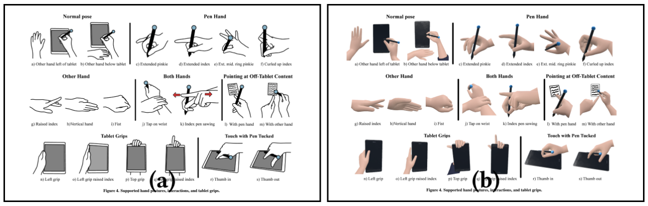

Figure 14: (a) The original interaction illustration of Matulic et al. [30] featuring an interactive tablet pen equipped with a wide-angle camera, which can detect the user's hand gestures (viewpoint types: 1^{st} person, 3^{rd} person, and top; user type: single; interaction types: touch, computer, and pen). (b) Our reproduction (avatar count: 21; time in DirActor: 28m 48s). The permission for the original image was granted by ACM.

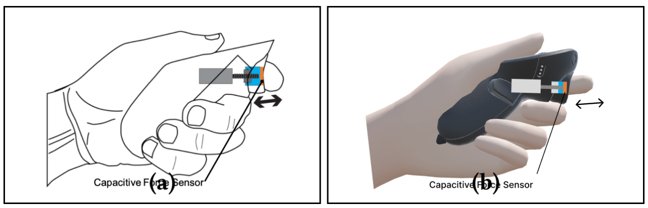

Figure 15: (a) The original interaction illustration of Preechayasomboon et al. [37] featuring a compact linear actuator embedded into a VR controller (viewpoint type: 3^{rd} person; user type: single; interaction type: controller). (b) Our reproduction (avatar count: 1; time in DirActor: 2m 17s). The permission for the original image was granted by ACM.

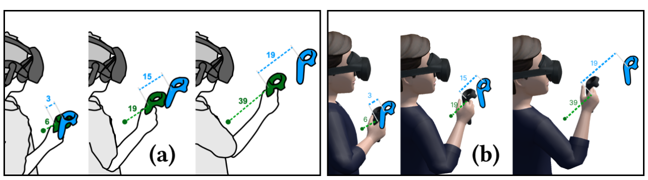

Figure 16: (a) The original interaction illustration of Wentzel et al. [48] featuring a technique that amplifies the distance between a VR controller and the user's body to improve ergonomics (viewpoint type: 3^{rd} person; user type: single; interaction types: mid-air, on-body, and controller). (b) Our reproduction (avatar count: 3; time in DirActor: 4m 18s). The permission for the original image was granted by ACM.

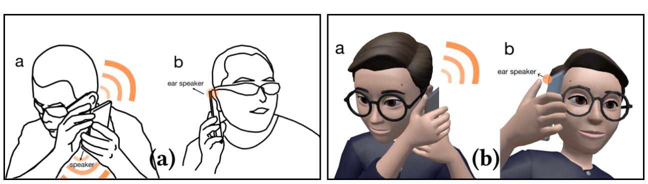

Figure 17: (a) The original interaction illustration of Wang et al. [47] featuring a technique that allows visually-impaired people to use touch screens and hear voice guidance without using a loudspeaker (viewpoint type: 3^{rd} person; user type: single; interaction types: touch and computer). (b) Our reproduction (avatar count: 2; time in DirActor: 8m 45s). The permission for the original image was granted by ACM.

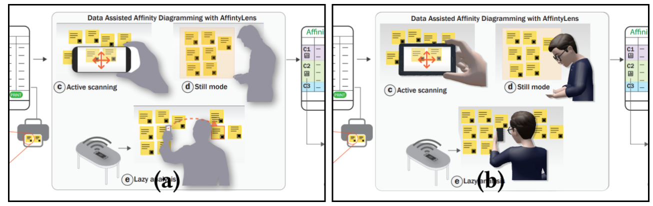

Figure 18: (a) The original interaction illustration of Subramonyam et al. [45] featuring an AR system which can quickly scan post-it notes to create affinity diagrams (viewpoint types: 1^{st} person, 3^{rd} person, and over-shoulder ¾; user type: single; interaction types: touch and computer). (b) Our reproduction (avatar count: 3; time in DirActor: 3m 15s). The permission for the original image was granted by ACM.

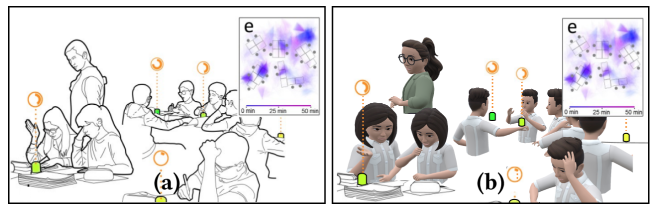

Figure 19: (a) The original interaction illustration of An et al. [1] featuring an interactive lamp that visualizes how much time the teacher has spent with each student (viewpoint type: 3^{rd} person; user type: multiple; interaction type: distal). (b) Our reproduction (avatar count: 9, time in DirActor 7m 45s). The permission for the original image was granted by ACM.

7 CONCLUSION & FUTURE WORK

In this study, we propose DirActor, a VR interaction system to help researchers create interaction illustrations with four viewpoint types, two user types, and nine interaction types. Using the interactive picture frame and Hold'n'Speak interactions, the user can monitor the shot, control the camera, and cue voice commands, all while using his body to pose his avatar. Using DirActor, an author was able to reproduce 169 human figures from 29 interaction illustrations found in ACM CHI Best and Honorable Mention papers, averaging 54 seconds per figure.

In future work, we will evaluate the system's usability through an experiment in which participants reproduce a given interaction illustration as quickly and accurately as possible using baseline and DirActor, respectively. The baseline will be a conventional photo session, where a director and multiple actors will perform the reproduction task in a room where a camera and props are prepared. Then, the director will perform the same reproduction task alone with DirActor in a VR scene where props are pre-imported. We will draw quantitative and qualitative comparisons and publish the findings in a follow-up paper.

ACKNOWLEDGMENTS

This research was supported by the DRB-KAIST SketchTheFuture Research Center, the KAIST Convergence Research Institute Operation Program, and the KAIST Undergraduate Research Participation (URP) Program.

REFERENCES

[1] Pengcheng An, Saskia Bakker, Sara Ordanovski, Ruurd Taconis, Chris L.E. Paffen, and Berry Eggen. 2019. Unobtrusively enhancing reflection-in-action of teachers through spatially distributed ambient information. In *Proc. CHI '19*. Article 91, 14 pages.

[2] Axel Antoine, Sylvain Malacria, Nicolai Marquardt, and Géry Casiez. 2019. Esquisse: using 3D models staging to facilitate the creation of vector-based trace figures. In *Proc. INTERACT '19*. 496–516.

[3] Axel Antoine, Sylvain Malacria, Nicolai Marquardt, and Géry Casiez. 2021. Interaction illustration taxonomy: classification of styles and techniques for visually representing interaction scenarios. In *Proc. CHI '21*. Article 188, 22 pages.

[4] João Marcelo Evangelista Belo, Anna Maria Feit, Tiare Feuchtner, and Kaj Grønbæk. 2021. XRgonomics: facilitating the creation of ergonomic 3D interfaces. In *Proc. CHI '21*. Article 290, 11 pages.

[5] Elise Bonnail, Wen-Jie Tseng, Mark Mcgill, Eric Lecolinet, Samuel Huron, and Jan Gugenheimer. 2023. Memory manipulations in extended reality. In *Proc. CHI '23*. Article 875, 20 pages.

[6] Kirill Brodt and Mikhail Bessmeltsev. 2022. Sketch2Pose: estimating a 3D character pose from a bitmap sketch. *ACM Trans. Graph.* 41, 4, Article 85 (Jul 2022), 15 pages.

[7] Canva. 2023. *Canva*.

[8] Pei-Yu (Peggy) Chi, Daniel Vogel, Mira Dontcheva, Wilmot Li, and Björn Hartmann. 2016. Authoring illustrations of human movements by iterative physical demonstration. In *Proc. UIST '16*. 809–820.

[9] ClipStudio. 2023. *ClipStudio*.

[10] Humphrey Curtis, Zihao You, William Deary, Miruna-Ioana Tudoreanu, and Timothy Neate. 2023. Envisioning the (in)visibility of discreet and wearable AAC devices. In *Proc. CHI '23*. Article 52, 19 pages.

[11] DeepAI. 2023. *DeepAI*.

[12] Oliver Glauser, Wan-Chun Ma, Daniele Panozzo, Alec Jacobson, Otmar Hilliges, and Olga Sorkine-Hornung. 2016. Rig animation with a tangible and modular input device. *ACM Trans. Graph.* 35, 4, Article 144 (Jul 2016), 11 pages.

[13] Martin Guay, Marie-Paule Cani, and Rémi Ronfard. 2013. The line of action: an intuitive interface for expressive character posing. *ACM Trans. Graph.* 32, 6, Article 205 (Nov 2013), 8 pages.

[14] Ollie Hanton, Michael Wessely, Stefanie Mueller, Mike Fraser, and Anne Roudaut. 2022. Enabling hand gesture customization on wrist-worn devices. In *Proc. CHI '20*. 1–13.

[15] Eiji Hayashi, Jaime Lien, Nicholas Gillian, Leonardo Giusti, Dave Weber, Jin Yamanaka, Lauren Bedal, and Ivan Poupyrev. 2021. RadarNet: efficient gesture recognition technique utilizing a miniature radar sensor. In *Proc. CHI '21*. Article 5, 14 pages.

[16] Narukawa Hiroki, Natapon Pantuwong, and Masanori Sugimoto. 2012. A puppet interface for the development of an intuitive computer animation system. In *Proc. ICPR '12*. 3136–3139.

[17] Kai-Chieh Huang, Chen-Kuo Sun, Da-Yuan Huang, Yu-Chun Chen, Ruei-Che Chang, Shuo-wen Hsu, Chih-Yun Yang, and Bing-Yu Chen. 2020. Glissade: generating balance shifting feedback to facilitate auxiliary digital pen input. In *Proc. CHI '20*. 1–13.

[18] Jingun Jung, Sunmin Son, Sangyoon Lee, Yeonsu Kim, and Geehyuk Lee. 2021. ThroughHand: 2D tactile interaction to simultaneously recognize and touch multiple objects. In *Proc. CHI '21*. Article 29, 13 pages.

[19] Bo-Cheng Ke, Min-Han Li, Yu Chen, Chia-Yu Cheng, Chiao-Ju Chang, Yun-Fang Li, Shun-Yu Wang, Chiao Fang, and Mike Y. Chen. 2023. TurnAhead: designing 3-DoF rotational haptic cues to improve first-person viewing (FPV) experiences. In *Proc. CHI '23*. Article 401, 15 pages.

[20] Daehwa Kim, Keunwoo Park, and Geehyuk Lee. 2021. AtaTouch: robust finger pinch detection for a VR controller using RF return loss. In *Proc. CHI '21*. Article 11, 9 pages.

[21] Lawrence H. Kim, Daniel S. Drew, Veronika Domova, and Sean Follmer. 2020. User-defined swarm robot control. In *Proc. CHI '20*. 1–13.

[22] Minju Kim and Jungjin Lee. 2019. PicMe: interactive visual guidance for taking requested photo composition. In *Proc. CHI '19*. Article 395, 12 pages.

[23] Jiasheng Li, Zeyu Yan, Arush Shah, Jonathan Lazar, and Huaishu Peng. 2023. Toucha11y: making inaccessible public touchscreens accessible. In *Proc. CHI '23*. Article 748, 13 pages.

[24] Zisu Li, Chen Liang, Yuntao Wang, Yue Qin, Chun Yu, Yukang Yan, Mingming Fan, and Yuanchun Shi. 2023. Enabling voice-accompanying hand-to-face gesture recognition with cross-device sensing. In *Proc. CHI '23*. Article 313, 17 pages.

[25] Juncong Lin, Takeo Igarashi, Jun Mitani, Minghong Liao, and Ying He. 2012. A sketching interface for sitting pose design in the virtual environment. *IEEE Trans. VCG.* 18, 11 (2012), 1979–1991.

[26] Vivian Liu and Lydia B. Chilton. 2022. Design guidelines for prompt engineering text-to-image generative models. In *Proc. CHI '22*. Article 384, 23 pages.

[27] MagicPoser. 2023. *MagicPoser*.

[28] Karthik Mahadevan, Mauricio Sousa, Anthony Tang, and Tovi Grossman. 2021. "Grip-that-there": an investigation of explicit and implicit task allocation techniques for human-robot collaboration. In *Proc. CHI '21*. Article 215, 14 pages.

[29] C. Mao, S. F. Qin, and D. K. Wright. 2005. A sketch-based gesture interface for rough 3D stick figure animation. In *Proc. SBIM '05*. 175–183.

[30] Fabrice Matulic, Riku Arakawa, Brian Vogel, and Daniel Vogel. 2020. PenSight: enhanced interaction with a pen-top camera. In *Proc. CHI '20*. 1–14.

[31] Andrii Matviienko, Florian Müller, Martin Schmitz, Marco Fendrich, and Max Mühlhäuser. 2022. SkyPort: investigating 3D teleportation methods in virtual environments. In *Proc. CHI '22*. Article 516, 11 pages.

[32] Meta. 2023. *Wit.ai*.

[33] Midjourney. 2023. *Midjourney*.

[34] Naoki Numaguchi, Atsuhi Nakazawa, Takaaki Shiratori, and Jessica K. Hodgins. 2011. A puppet interface for retrieval of motion capture data. In *Proc. SCA '11*. 157–166.

[35] Steve Oney, Rebecca Krosnick, Joel Brandt, and Brad Myers. 2019. Implementing multi-touch gestures with touch groups and cross events. In *Proc. CHI '19*. Article 355, 12 pages.

[36] OpenAI. 2023. *DALL·E 3*.

[37] Pornthep Preechayasomboon, Ali Israr, and Majed Samad. 2020. Chasm: a screw based expressive compact haptic actuator. In *Proc. CHI '20*. 1–13.

[38] Adil Rahman, Md Aashikur Rahman Azim, and Seongkook Heo. 2023. Take My Hand: automated hand-based spatial guidance for the visually impaired. In *Proc. CHI '23*. Article 544, 16 pages.

[39] Hugo Romat, Nathalie Henry Riche, Ken Hinckley, Bongshin Lee, Caroline Appert, Emmanuel Pietriga, and Christopher Collins. 2019. ActiveInk: (th)inking with data. In *Proc. CHI '19*. Article 42, 13 pages.

[40] Roy. 2023. *PoseMyArt*.

[41] Jonas Schjerlund, Kasper Hornbæk, and Joanna Bergström. 2022. OVRlap: perceiving multiple locations simultaneously to improve interaction in VR. In *Proc. CHI '22*. Article 355, 13 pages.

[42] Martin Schmitz, Sebastian Günther, Dominik Schön, and Florian Müller. 2022. Squeezy-Feely: investigating lateral thumb-index pinching as an input modality. In *Proc. CHI '22*. Article 61, 15 pages.

[43] Roselien Steur and Koos Essen. 2011. *Sketching: The Basics*. BIS Publishers.

[44] Zixiong Su, Shitao Fang, and Jun Rekimoto. 2023. LipLearner: customizable silent speech interactions on mobile devices. In *Proc. CHI '23*. Article 696, 21 pages.

[45] Hariharan Subramonyam, Steven M. Drucker, and Eytan Adar. 2019. Affinity Lens: data-assisted affinity diagramming with augmented reality. In *Proc. CHI '19*. Article 398, 13 pages.

[46] Radu-Daniel Vatavu. 2023. IFAD Gestures: understanding users' gesture input performance with index-finger augmentation devices. In *Proc. CHI '23*. Article 576, 17 pages.

[47] Ruolin Wang, Chun Yu, Xing-Dong Yang, Weijie He, and Yuanchun Shi. 2019. EarTouch: facilitating smartphone use for visually impaired people in mobile and public scenarios. In *Proc. CHI '19*. Article 24, 13 pages.

[48] Johann Wentzel, Greg d'Eon, and Daniel Vogel. 2020. Improving virtual reality ergonomics through reach-bounded non-linear input amplification. In *Proc. CHI '20*. 1–12.

[49] Xuhai Xu, Jun Gong, Carolina Brum, Lilian Liang, Bongsoo Suh, Shivam Kumar Gupta, Yash Agarwal, Laurence Lindsey, Runchang Kang, Behrooz Shahsavari, Tu Nguyen, Heriberto Nieto, Scott E Hudson, Charlie Maalouf, Jax Seyed Mousavi, and Gierad Laput. 2022. Enabling hand gesture customization on wrist-worn devices. In *Proc. CHI '22*. Article 496, 19 pages.

[50] Lvmin Zhang, Anyi Rao, and Maneesh Agrawala. 2023. Adding conditional control to text-to-image diffusion models. In *Proc. ICCV '23*. 3836–3847.