JackIn Space: Designing a Seamless Transition Between First and Third Person View for Effective Telepresence Collaborations

Ryohei Komiyama
The University of Tokyo
Tokyo, Japan
komikomi.mikomiko
@gmail.com

Takashi Miyaki
The University of Tokyo
Tokyo, Japan
miyaki@acm.org

Jun Rekimoto
The University of Tokyo / Sony
CSL
Tokyo, Japan
rekimoto@acm.org

ABSTRACT

Traditional telepresence systems only supported first person view and users had difficulty in recognizing the surrounding situation of their remote workspace. *JackIn Space* is a telepresence system that solves this problem by seamlessly integrating first person view with third person view. With a head-mounted first person camera and multiple depth sensors installed in the environment, the surrogate user's first person view smoothly changes to the out-of-body third person view, and the user connected to the surrogate user can virtually look around the environment. The user can also dive into other surrogate users to gain different perspectives. Our evaluation supports that the concept and the function of *JackIn Space* was quite well accepted and our prototype system provides a more natural viewpoint selection. Overall, *JackIn Space* supports better remote collaborations.

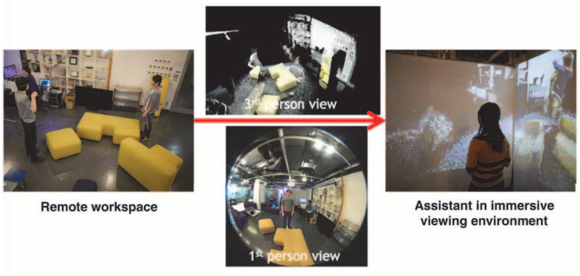

Figure 1: JackIn Space: A telepresence system that supports first and third person perspectives. It also supports seamless viewpoint transition between them.

CCS Concepts

•**Human-centered computing** → **Collaborative interaction;** *Mixed / augmented reality; Virtual reality;*

Keywords

telepresence; remote collaboration; augmented reality; first person view; 3D modeling; Kinect

1. INTRODUCTION

Telepresence is a method of telecommunication where user immersively connects to the remote entity, typically a surrogate robot [22]. The robot's perception is transmitted to the user, and the user's motion controls the robot which enables the user to feel as if the user were in the remote place as a robot. This concept eliminates the limit concerning physical location, and can be used for various remote applications such as remote inspection and manipulation.

Recently this concept has been augmented and includes human-human telepresence [16]. In this case, the first person view of the person wearing a head-mounted camera is transmitted to another person in the remote workspace enabling better mutual communication. For instance, a worker in the field can be remotely supported by an expert by sharing the context. In this case, the human is the surrogate for the other human.

In both cases, current telepresence systems mainly supplies the view from the surrogate(robot/human)'s perspective, which we also called the first person view. However, there are several shortcomings when the first person view is the only viewing method.

First, it causes motion sickness. This is particularly noticeable in the human-human telepresence. When the user in the workspace moves their head to looks around, the first person view tends to be very shaky and is not very comfortable for the remote user to watch. This problem can be solved by using the head mounted omnidirectional (360-degree) camera, and applying image stabilization [23, 29].

The second problem is spatial awareness. In a real co-located collaboration, we can easily grasp positional relationship between humans, nearby objects, and the surrounding environment. For instance, when people are collaboratively doing the maintanace of the machine, they can easily understand the situation by walking around. If they do this through telepresence collaboration, on the other hand, the remote user can only understand the situation from the viewpoint of the surrogate (worker), and the remote user is not able to walk around to gain better understanding. If there are two or more surrogates(robots, humans, drones, etc.) in the workplace, it should be very convenient to "jump around" these viewpoints to understand the situation. However, if the system suddenly switches the viewpoint from one viewpoint to another, the user would lose their spatial context.

To address these issues, this paper proposes a method that en-

AH '17, March 16-18, 2017, Mountain View, CA, USA
© 2017 ACM. ISBN 978-1-4503-4835-5/17/03. . . $15.00
DOI: http://dx.doi.org/10.1145/3041164.3041183

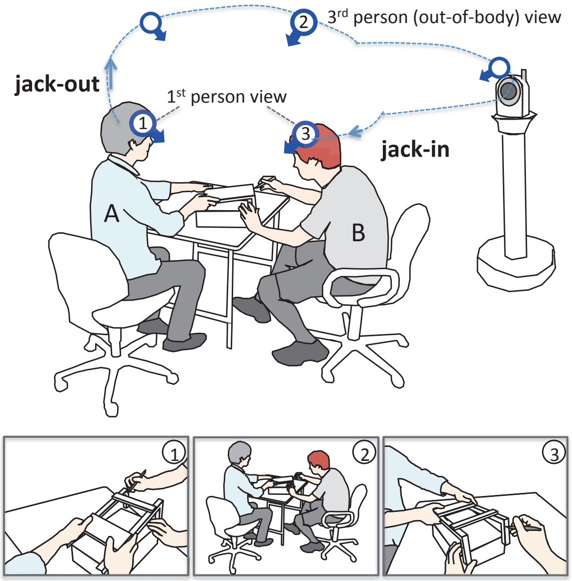

Figure 2: The JackIn Space telepresence concept: (i) A remote ghost user watches through the body user A's first person view 1. (ii) After issuing a jack out command, the remote ghost user's viewpoint smoothly moves out from the ghost user A to become the third person perspective (view 2). (iii) The ghost user can freely change his/her view position, and alternatively jack in to other telepresence robots. (iv) The ghost user locks on the next body user B and issues a jack in command. (v) The ghost user's viewpoint moves into the body user B (view 3).

ables seamless transitions between first person view and third person view, and the seamless transitions between multiple viewpoints in the workspace (Figure 1). To achieve this concept, multiple depth cameras (Kinect 2 sensors) are used to capture and reconstruct three dimensional data of the workspace and person. Once the vision is channeled to the third person view, user can freely change their viewpoint to look at the entire scence and can also go back to the previous first person view or any other first person/robot view.

2. JACKIN SPACE

JackIn Space is our framework that supports the seamless and natural transition between first and third person view (Figure 2). In this framework, the term *jack in* refers to connecting to the remote entity (either humans/robots), and *jack out* refers to disconnecting from the *jack in* state. These words were originally used in a 1984 science fiction "Neuromancer" by William Gibson which is famous as the origin of Cyberpunk [11]. While the original *jack in* mainly means connection to the cyberspace, we slightly extended the meaning to be used in the telepresence context[1].

In addition, we also use the term *body* to refer to a surrogate person/robot who performs tasks in the physical environment, and the term *ghost* to refer to a person who remotely connects (*jack in*) to the body user over the network.

Traditional telepresence systems and past *JackIn* studies [16, 23] did not explicitly consider the *jack out* operation because these systems normally dealt with just one remote surrogate. On the other hand, we consider allowing *jack in* and *jack out* operations would add several advantages to the telepresence concept.

First, it helps a remote (*ghost*) user to understand the situation more easily. While first person view is useful for the user to feel as if he/she is in the environment, there needs to be another method to grasp the entire situation. In this case, exiting from first person perspective is effective. In addition, when two or more surrogate entities (*body* users, robots, or installed cameras) are available, seamlessly transition from one viewpoint to another would also be effective for understanding the situation. For example, in a disaster area inspection, an expert in the remote area can *jack in* multiple workers in the environment, and *jack out* to get third person perspective. At this point, continuous transition of viewpoints specifies from whom and to whom the view transition was implemented.

To enable this framework, our current system install multiple Kinect 2 sensors in the environment for obtaining third person view, and each participant (*body* user) wears the head-mounted camera to provide first person view. As the head position of *body* users are tracked, we can change views seamlessly between first and third person perspectives, i.e., real space and virtual space.

3. RELATED WORK

3.1 Realworld Capturing and Reconstruction

The research of capturing and reconstructing real world by using multiple sensors has a long history since the *virtualized reality* idea was proposed [25]. The Office of the Future is another example of telecommunication using virtual reality, connecting local and remote workspace as integrated 3D [26]. These ideas appear frequently in many science fictions. For example, A.C. Clarke described a three-dimensional replication of the far future city called *Diaspar* in his science fiction "The city and the stars" [3]. With this replicated city model, one can browse from any positions at any timepoint. Similar ideas are shown in the movie " "Déjà-Vu" where the entire real city is captured and archived in realtime, so the user of the system can have a four-dimensional perspective to the city [21]. Gelernter proposed a concept of "Mirror Worlds" where the replication of the real world becomes our primary way of interacting with information space[10]. By combining recent sensing technologies and computational power, these "real world mirroring" ideas become feasible, and telepresence should not be limitted within the perspective of remote entities.

Recently, virtualizing the real world by using multiple Kinects has become popular because of its low-cost RGB-D (depth) sensor [8]. Especially, Kinect2 is suitable for these purpose as they implement time-of-flight sensing which reduces interference between several of those devices, therefore, it enables us to use mulitple devices to gain stable distance image from wide area.

3.2 First and Third Person Telepresence

The first person human-human telepresence is gaining attention as an alternative to traditional human-robot telepresence [16, 23, 29, 30]. This is a system where a person in the field (a *body* in our terminology) wears a head-mounted camera to capture and transmit view from the user's head position. Video stabilization is also used to compensate unwanted video motion caused by the user's head rotation. In this case, the *body* user's head rotation is decoupled from the viewer's (a *ghost* user's) head orientation. Polly [18] has slightly different features, the position of which is on the shoulder and its stabilization and rotation are realized mechanically, however, the concept of which is relatively close to the above researches. In *JackIn Space*, the *ghost* user's view position

[1] Gibson used the term "simstim" for describing human-human perception connection.

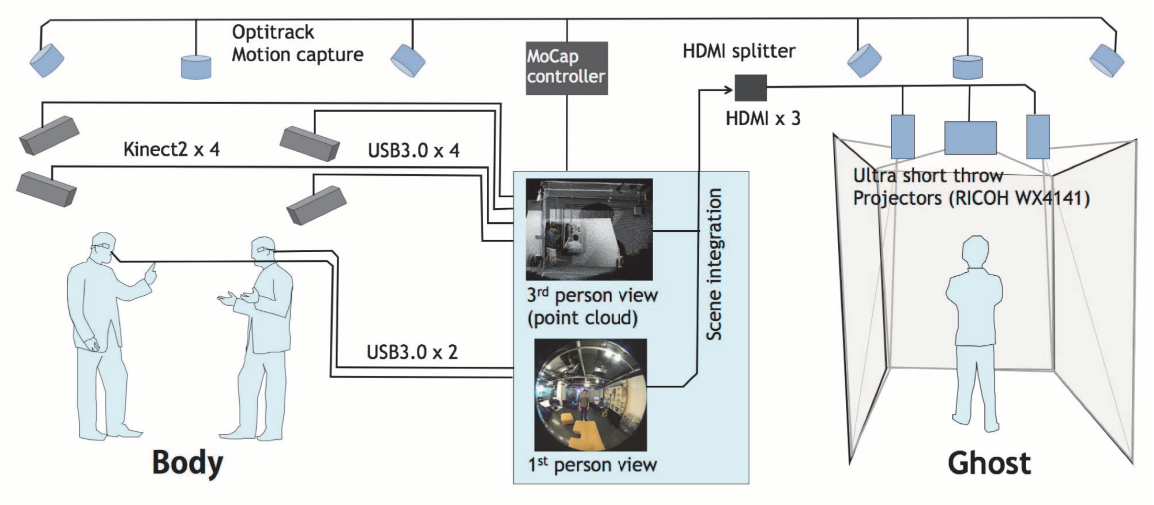

Figure 3: The JackIn Space System Configuration

is also independent from the *body* user's. This means it can be considered that *JackIn Space* extends the degree of decoupling.

The Time Follower's Vision [31] supports an operator who controls a remote robot by supplying a virtual third person view. This view is created using a mixed-reality technology. By overlaying the image of the robot's view on the previous image, an operator control the robot with this virtually synthesized third person view, and thus, there is no need to install real cameras to the environment for third person view.

Recently, Some human-human telepresence systems also offer third person view by creating a virtual 3D space, using monocular vision-based real-time simultaneous localization and mapping (SLAM) [6] or preparing 3D model in advance [9, 24]. However, these systems are not able to update the virtual 3D space in real-time except for the area where the remote user points a camera. Therefore, though they are appropriate for one-to-one telepresence because the created virtual 3D space is clear and the assistant user can check more precise area, not for *JackIn Space* which expects multiple *body* users in remote workspace and enables a view transition between them. In order to realize the *JackIn Space* concept, it needs larger, at least room-scale, virtual 3D space all of which is updated in real-time.

In sports training community, the importance of having the third person view is widely recognized [4, 7, 27, 28]. Several systems were created to enable out-of-body view to capture the human body during the training session by using cameras attached to a stick or to a drone [14, 13]. This out-of-body view can be transmitted to the trainee or to a remote coach. In both cases, third person view is an efficient method for understanding the movement of the trainee. We expect that *JackIn Space* can allow the coach to *jack in* a trainee, and occasionally *jack out* to get the third person view to enhace the quality of training session.

3.3 View Transition

In computer games, smooth camera transition technology, including the transition between first person view and third person view, is established [12]. Many first person computer games often also support the third person mode to show the position of the player (i.e., avatar) in the surrounding environment. *JackIn Space* incorporates these ideas to the area of remote collaboration.

Tatzgern et al. combined augmented reality with virtual reality, and realized a natural transition between the AR view and virtual viewpoints [33]. Sukan et al. also allow users to store snapshots of a scene and revisit them virtually at a later time [32]. However, in these system, the end point of a view transition is chosen when a view transition starts, and never chages after that. *JackIn Space* must consider the additional case in which the destination of a view transition is equal to *body* user's viewpoint which moves around during a view transition.

Rhizomatiks research demonstrated the integration of normal video stream with 3D space in video music performance at The South by Southwest (SXSW'15) [34]. In this case, the performer's movement is recorded beforehand, and the live video camera's position and orientation are also tracked. Then, the video transition effect that seamlessly connects the live video and the 3D model becomes possible, and it gives a strong impression of the integration between real and virtual world. *JackIn Space* also provides smooth transition between the live camera (first person view) and the point cloud (third person view), without requiring measurement of the environment beforehand.

4. JACKIN SPACE SYSTEM CONFIGURATION

To evaluate the concept of *JackIn Space* described in previous sections, we designed a system that contains two simultaneous *body* users and one *ghost* user. Configuration of this system is shown in Figure 3.

4.1 Equipment and Environment for Body Users

The *body* user wears a headset as shown in Figure 5. The headset is composed of a fish-eye camera (e-consystems See3Cam USB3.0 color camera with an M12-mount 185 degree field-of-view lens) and a set of infrared reflective markers for tracking its position and direction. This fish-eye camera provides the first person view of the body user.

The size of the workspace is about $4m \times 4m$. On the ceiling of this space, we installed four Kinect2 RGB-D (depth) sensors (Kinect for Xbox One) and an Optitrack motion capture system. These

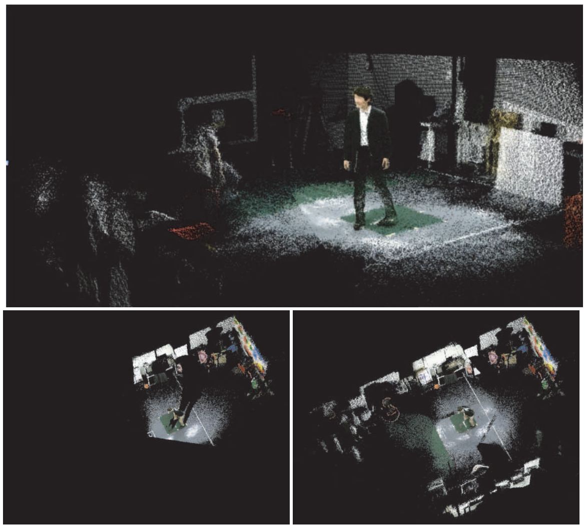

Figure 4: A point cloud scene of the working environment generated by multiple Kinect sensors: (bottom left) A point cloud created by 1 Kinect sensor; (bottom right) created by 4 Kinect sensors.

four Kinect sensors are arranged to look down the floor and used for getting a point cloud for the third person view. The motion capture sensors and Kinect sensors are precalibrated, and coordinate systems used for all of these sensors are identical. Also, slight difference in actual locations between infrared reflective markers and fish-eye camera was dealt with by pre-calibration method.

Since the Kinect 2 sensor, as opposed to Kinect 1, dose not interfere with each other, multiple sensing concept become feasible. We consider it was enabled by the fact that Kinect 2 utilizes the method of time-of-flight sensing and duration of infrared lighting became very short than before.

The RGB-D (depth) information from four Kinect sensors are integrated into a single *point-cloud* space where each point has a three-dimensional position (Figure 4). There will always be a “shadow” of sensing when using a single Kinect, however, it can be elmininated by using multiple Kinects. This point-cloud represents a work environment around the *body* users.

The image obtained from a fish-eye camera mounted on a *body* user’s haed is mapped into a half cube texture in a point-cloud world. When the *ghost* user is in the first person mode, the *ghost* user can actually see this half cube texture as a panoramic view of the *body* user. Turning into the third person mode, this half cube texture gradually becomes transparent, and the *ghost* user can see the environment represented as a point-cloud space.

4.2 Stabilization of First Person View

Since the first person view from a camera on the head of each *body* user can be swayed significantly and wobbled from any movement, this might cause a motion sickness to the *ghost* user. To resolve this issue, motion stabilization method using image processing was applied as it was done in previous systems [23]. This algorithm measures optical flow from the first person images, then estimates to which extent the camera was rotated by aggregating quaternions calculated from each optical flow. After eliminating outliers, we take the average of quaternions as a rotation of the camera. By inversely applying its quaternion to the first person view, this rotation effect of the image can be successfully suppressed.

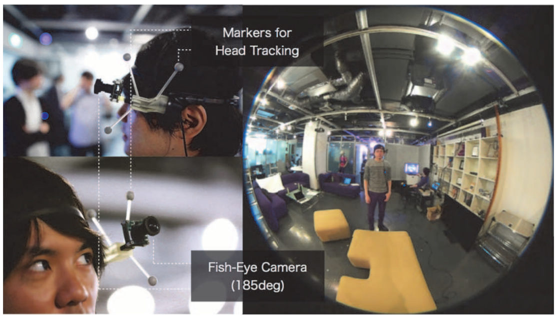

Figure 5: A headset configuration of the body user: (top left) Markers for head tracking; (bottom left) Fish-Eye camera; (right) image captured by fish-eye camera.

Because this rotation angle gradually reaches zero, the rapid head motion can be eliminated while the *ghost* user’s view direction is gradually aligned to that of the *body*’s.

4.3 Equipment and Environment for a Ghost User

The *ghost* user’s visual environment is composed of the CAVE-like projection screen environment [5] with three screens, three projectors (RICOH WX4141), a HDMI splitter, a motion capture system, and 3D glasses.

These three projectors are capable of the frame-sequential stereo projection. As all the projectors are connected to a HDMI splitter and frame-update timings become identical, the *ghost* user with LCD 3D glasses on is able to see stereoscopic scene across the boundary through multiple projectors.

Even though a head mounted display (HMD) was also considered as an option for the visual environment, we decided to adopt this CAVE-like configuration in that we wanted to prioritize the presence of the *body* user the most when the *ghost* user *jack out* from the *body* user. In addition, we concluded that HMD is not suitable for the work done in a remote collaboration because it will isolate the *ghost* user from (*ghost* user’s) environment. During these remote collaborations, there could be a case when the *ghost* user wants to see what’s originally around him/her in order to support the *body* user. Though the single-screen configuration with head-tracking, which is a style of virtual reality known as *Fish-Tank VR* [2], can also be considered, it cannot provide as wide as CAVE style in terms of its viewing angle.

During the first person mode, the *ghost* user can freely look around the environment from the *body* user’s point of view through a hemispheric real time video. Also, due to the larger viewing angle of the *body* user’s fish-eye camera compared to that of an ordinary camera, unlike traditional telepresence systems, the *ghost* user’s direction of view does not always need to be set as identical to the *body* user’s view (head) orientation.

All projectors are compatible with the 3D projection; therefore, the *ghost* user can enter the *body* user’s 3D environment when using third person view mode. First person view mode is useful to look at the details of *body* user’s view because it offers more clear images, although it is not compatible with 3D projection. Combining first and third person view, we are now able to experience such concepts including *jack in* one *body* user, looking around after the *jack out* from that user, and *jack in* a different *body* user to confirm the details.

4.4 Sharing Processing System

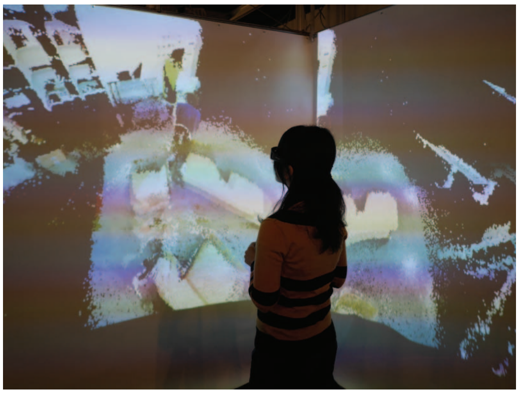

Figure 6: A ghost user navigating in the body's third person perspective.

In the current *JackIn Space* system, one high-end personal computer running Ubuntu Linux (with Intel Core i7-4790K, GeForce GTX TITAN X graphics card and 32GB memory) processes all the information from the *body* user's environment sensors (Kinect and a motion capture system), and generates and shows visual information for a *ghost* user. In order to actualize this, there are two main processes running on the computer, one for the *body* environment and the other for the *ghost* environment.

Since the system currently being mentioned is set as a proof-of-concept prototype, it does not conform to an realistic situation where the *ghost* and the *body* users are separated. Rather it holds a model where all the sensors and displays are connected to the same computer. Particularly, in this paper, we are focusing more on the seamless transition between first and third person view. Regarding the applying a distributed configuration into the system, we will cover more on the design issues in the discussion section.

5. JACK IN / JACK OUT INTERACTIONS

As described in previous sections, a *ghost* user can smoothly switch the mode around among the first person and the third person.

Let us consider a situation where a *ghost* user is initially seeing the world from the perspective of a *body* user's. When that *ghost* user executes a *jack out* command, the viewpoint will automatically move back from the body user's sight (first person view) to the location that is slightly above and behind the *body* user's sight. In this way, the *ghost* user can see the *body* user's head and *body* user after *jack out*. After *jack out*, the initial direction of the third person view is set as it is toward the head of the *body* user, and interpolation is applied to make the transition process smooth (Figure 7, Figure 8).

Once the *ghost* user enters the third person view mode, the *ghost* user now can freely move around the environment presented as a point-cloud space captured from multiple Kinect sensors. This is realized by tracking the *ghost* user's head position and orientation using motion capture system.

During the third person mode, the *ghost* user can also be navigating the environment based on the Point-of-Interest (POI) technique (Figure 9). POI includes positions defined in advance (such as the center of the worktable) and the positions of *body* users. The *ghost* user selects one of these POIs, and approaches toward it with a smooth viewpoint movement, and circles around the POI (circle strafing), as in the POI-based movement in virtual space systems [20]. In order to use these functions, we are currently using a simple hand-held joystick although several other devices are also available.

To *jack in* other *body* user, firslty the *ghost* user looks at the target *body* user (by orienting his/her head toward the target *body* user), or select the target *body* user as one of POIs. Then, the system will show the "lock-on" mark around the head of the target *body* user. When the *ghost* executes the *jack-in* command, the sight of the *ghost* user's perspective will be smoothly moving into that of the target *body* users's (Figure 8). In this case, the view point movement path has a curve, and the final direction of the moving ghost user's sight needs to be in alignment with that of the body user's, so that the *ghost* user can get a feeling of entering the *body* user.

6. EVALUATION

To evaluate the concept of *JackIn Space*, we conducted a pilot study. Each participant took the role of the *ghost* in a CAVE environment, which contained furniture and two *body* users. The participants' task was to figure out the layout of the furniture and sketch it on a paper. Each participant tried the task once with *JackIn Space* function where the participants could *jack out* and *jack in* to either of two *bodies*, and the task only using the first person mode. Each trial was allotted one minute, with the total task time for one participant being about 5 minutes, which included a short explanation session. Participants then answered a questionnaire with four qustions using a 7-point Likert Scale after the task. Figure 10 shows the actual evaluation setup.

Eight participants, all having a computer science background (three of them being computer science department graduate students and others being computer engineers) participated in this experiment. We prepared two sets of furniture layouts. The Latin square order was used to mix the order of methods and the layout of furniture.

The result of the questionnaire is shown in Figure 11.

Overall, the response from the participants was positive. 100% preferred having *jack in / jack out* functionality in the telepresence system (Q4, Median = 6), 87.5% answered they could understand the overall situation by *jack out* (Q3, Meidan = 6.5), and 87.5% answered that the *jack in / jack out* concept was easily understandable (Q1, Median = 7). On the other hand, the answers on view transition smoothness had a slightly lower score (Q2, 12.5% answered negtive, 50% answered positive, Median = 4.5). This indicates the view transition algorithm should be improved in terms of the usability.

As for the correctness of sketches, we could not find a clear difference between *JackIn Space* and normal first person view (Figure 12).

This might be because of the simplicity of the task (we just used three sofas to be detected) and given one minute task completion time, most of them could detect the layout. Nevertheless, participants in the first person mode had to say more direction indication terms to the *body* user such as "turn to the left" and "please look down". On the other hand, participants during the *JackIn Space* mode tended to go up to see the entire environment to grasp the furniture layout.

To summarise the experiment, the concept and function of *JackIn Space* were quite well accepted by the participants, while the view transition between the first and the third person mode were met with more mixed results, indicating need for improvement.

7. DISCUSSIONS

7.1 Self-contained sensing

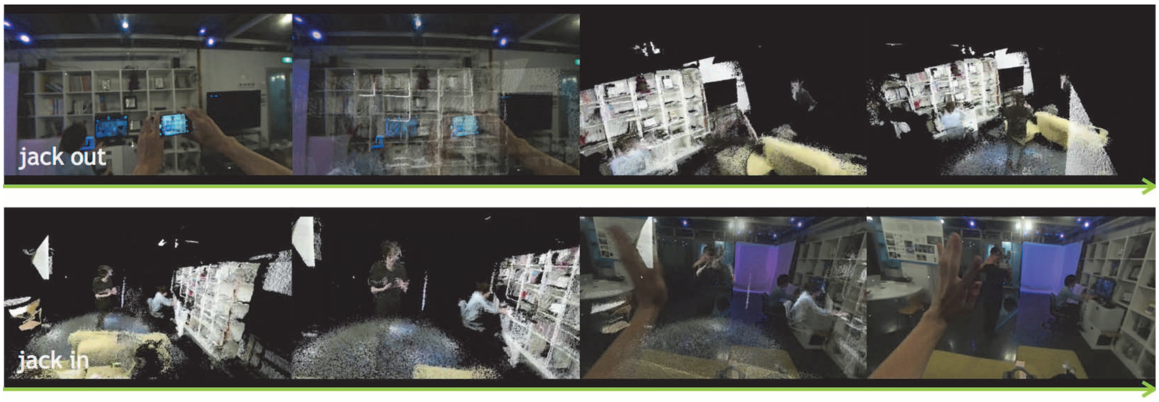

Figure 7: An example of view transition: above: from first person view to third person view (jack out), bottom: from third person view to first person view (jack in)

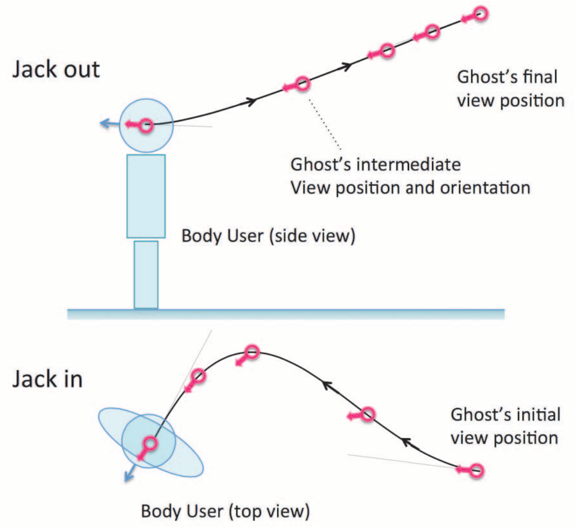
Jack out
Ghost's final view position
Ghost's intermediate View position and orientation
Body User (side view)
Jack in
Ghost's initial view position
Body User (top view)

Figure 8: Jack In and Jack Out viewpoint movement: (above) while switching to the third person mode (jack out), view motion path is created using a spline curve function and the ghost user can smoothly reach to third person viewpoint. (bottom) To move into the first person mode (jack in), third person viewpoint approaches a body user's head position from behind it. During a view transition, the view direction is also smoothly interpolated to preserve continuity.

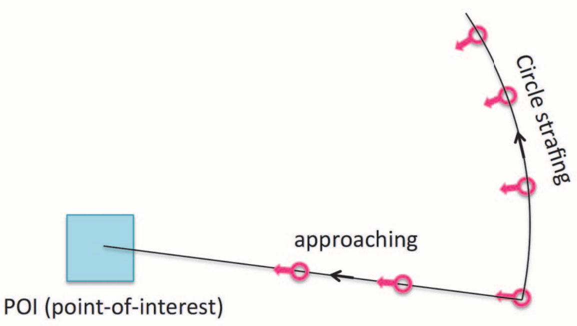

Figure 9: POI based view position movements during the third person mode

Currently, the system suggested in here is based under an assumption that multiple Kinect sensors are installed in the environment to capture 3D space information as well as surrogate user's body shapes. While this configuration might be useful for some applications such as remote surgery and remote laboratory, more solutions are in need for meeting the demands for other applications. The next step of this research would use wearable depth sensors for the surrogate users and incrementally construct the 3D model of surrounding environment, as previous research has covered (such as Kinect fusion [15]). The other approach would include using sensing-oriented surrogate robots. These robots will

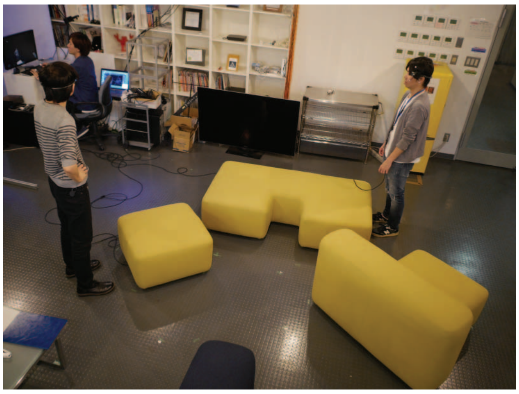

Figure 10: The configuration of the experiment

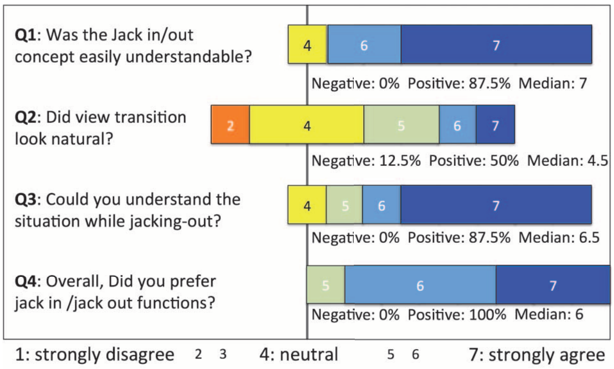

Figure 11: The result of the experiment: 7-point Likart scale questionnaire

follow a certain surrogate user to capture environmental information. These robots would also be useful for providing the *ghost* user with the multiple first person viewpoints so that the *ghost* user can have more freedom in choosing viewpoints.

Ultimately we may not need the first person view, because it is just a part of the third person view from the perspective of the individual. However, qualities of the image obtained by a first person camera and a space sensor (such as Kinect) are still showing a huge difference, and combining these two will be a reasonable decision.

7.2 Distributed configuration

As described earlier in system configuration section, this system assumes both the *body* and the *ghost* sensors are connected with short-range data lines such as USB3.0 or HDMI. To make a distributed system, estimating and reducing the amount of data between the *ghost* and the *body* user environment is significant.

A single Kinect 2 sensor generates information in a format of an RGB-D (depth) image. Its size is about 1.03 Mbytes if uncompressed. Supposed four Kinects are used to capture the environment, the amount of data for transmission at a rate of 30fps would be 124 Mbytes (/sec). If we adopt the point cloud format, the amount of the date becomes even larger because each point has information with a 3D position. However, as the most of the 3D scene information is static, we can eliminate humans from the surrounding environment. In this case, the amount of data to be updated will be much smaller. In actual measurement, the amount of data required for one person (in a form of the point cloud) is around 300 Kbytes, therefore, the amount of data to transmit for four body users will become 18 Mbytes. We also can apply various data compression techniques other than this one to achieve a further reduction in the amount of transmitted data.

7.3 Skill Sharing and Transmission

We expect the *JackIn Space* to be used as a platform for transmission of skills and technology. When teaching skills such as sports, playing instruments and craft, usually teachers demonstrate their performance to learners first and let them mimic. If this process can be transmitted remotely, opportunity to learn something even at a distance will proliferate. In addition, some researchers also mentioned the importance of sharing the first person perspective for skill transmission [17, 1]. Using *JackIn Space*, a learner can observe the teacher's model performance either from third person or firs person view of the teacher. If the teacher's performance is projected on a CAVE environment as a life-size 3D model, a learner is now able to compare teacher's performance with that of one's own. We also consider the possibility of archiving and packaging these model performances as a new interactive textbook for skill transmission.

7.4 Self and Other

We would also like to mention that the *JackIn Space* environment might be a research platform for studies of artificial out-of-body experiences. For instance, Lenggenhager et al. suggested coupling visual and tactile sensation could cause an out-of-body effect (where participant for the experiment will feel as if they are watching themselves from a space outside their bodies) [19]. These researches must be helpful for understanding our notion of self and others. As introduced system can provide a seamless transition between first and third person perspectives, it would also be useful for studying cognitive effects on out-of-body experience.

8. CONCLUSION

In this paper, we described a telecollaboration system *JackIn Space*, which supports seamless transitions between first and third person view. *JackIn Space* also supports multiple first person perspectives from multiple *body* users (remote surrogates), while traditional telepresence systems can only support the perspective of one remote entity. Therefore, the *ghost* user (assistant user) can watch a remote workspace through one body user's first person view, go out from him/her to get a third person perspective, and then enter another *body* user to get a different first person perspective. In third person mode, the *ghost* user can move and look around the remote workspace freely. This *JackIn Space* concept is realized by combining a head-mounted first person camera and multiple depth sensors installed in the environment. Our approach supports both precise work sharing with first person view and workspace understanding with third person view. Our evaluation supports that this configuration provides more natural view position selection, and thus supports better remote collaborations.

9. REFERENCES

[1] J. Amores, X. Benavides, and P. Maes. Showme: A remote collaboration system that supports immersive gestural communication. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*, CHI EA '15, pages 1343–1348, New York, NY, USA, 2015. ACM.

[2] K. W. Arthur, K. S. Booth, and C. Ware. Evaluating 3d task performance for fish tank virtual worlds. *ACM Trans. Inf. Syst.*, 11(3):239–265, July 1993.

[3] A. C. Clarke. *The City and the Stars*. Frederick Muller Ltd, 1956.

[4] A. Covaci, A.-H. Olivier, and F. Multon. Third person view and guidance for more natural motor behaviour in immersive basketball playing. In *Proceedings of the 20th ACM Symposium on Virtual Reality Software and Technology*, VRST '14, pages 55–64, New York, NY, USA, 2014. ACM.

[5] C. Cruz-Neira, D. J. Sandin, and T. A. DeFanti. Surround-screen projection-based virtual reality: The design and implementation of the cave. In *Proceedings of the 20th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '93, pages 135–142, New York, NY, USA, 1993. ACM.

[6] A. J. Davison. Real-time simultaneous localisation and mapping with a single camera. In *Proceedings of the Ninth IEEE International Conference on Computer Vision - Volume*

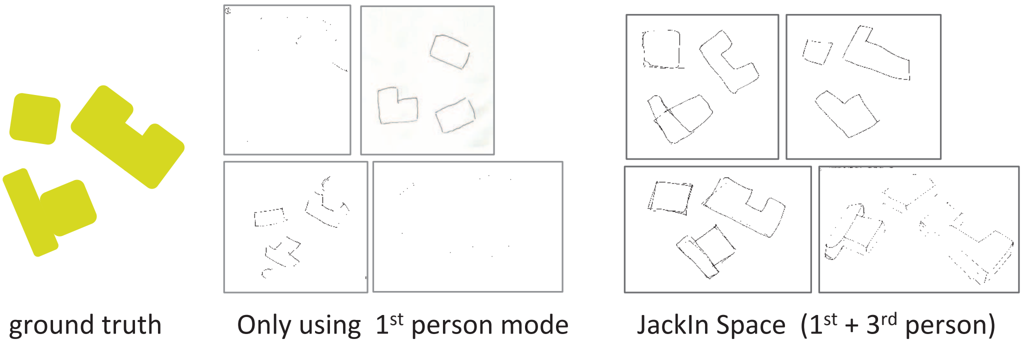

Figure 12: Sketches drawn by evaluation experiment participants

2, ICCV '03, pages 1403–, Washington, DC, USA, 2003. IEEE Computer Society.

[7] H. G. Debarba, E. Molla, B. Herbelin, and R. Boulic. Characterizing embodied interaction in first and third person perspective viewpoints. In *3D User Interfaces (3DUI), 2015 IEEE Symposium on*, pages 67–72, March 2015.

[8] M. Dou and H. Fuchs. Temporally enhanced 3d capture of room-sized dynamic scenes with commodity depth cameras. In *Virtual Reality (VR), 2014 iEEE*, pages 39–44, March 2014.

[9] S. Gauglitz, B. Nuernberger, M. Turk, and T. Höllerer. World-stabilized annotations and virtual scene navigation for remote collaboration. In *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology*, UIST '14, pages 449–459, New York, NY, USA, 2014. ACM.

[10] D. Gelernter. *Mirror Worlds: or the Day Software Puts the Universe in a Shoebox...How It Will Happen and What It Will Mean*. Oxford University Press, 1993.

[11] W. Gibson. *Neuromancer*. Ace Science Fiction, Canada, 1984.

[12] M. Haigh-Hutchinson. *Real-Time Cameras: A Guide for Game Designers and Developers*. CRC Press, 2009.

[13] S. Hasegawa, S. Ishijima, F. Kato, H. Mitake, and M. Sato. Realtime sonification of the center of gravity for skiing. In *Proceedings of the 3rd Augmented Human International Conference*, AH '12, pages 11:1–11:4, New York, NY, USA, 2012. ACM.

[14] K. Higuchi, T. Shimada, and J. Rekimoto. Flying sports assistant: External visual imagery representation for sports training. In *Proceedings of the 2Nd Augmented Human International Conference*, AH '11, pages 7:1–7:4, New York, NY, USA, 2011. ACM.

[15] S. Izadi, D. Kim, O. Hilliges, D. Molyneaux, R. Newcombe, P. Kohli, J. Shotton, S. Hodges, D. Freeman, A. Davison, and A. Fitzgibbon. Kinectfusion: Real-time 3d reconstruction and interaction using a moving depth camera. In *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology*, UIST '11, pages 559–568, New York, NY, USA, 2011. ACM.

[16] S. Kasahara and J. Rekimoto. Jackin: Integrating first-person view with out-of-body vision generation for human-human augmentation. In *Proceedings of the 5th Augmented Human International Conference*, AH '14, pages 46:1–46:8, New York, NY, USA, 2014. ACM.

[17] H. Kawasaki, H. Iizuka, S. Okamoto, H. Ando, and T. Maeda. Collaboration and skill transmission by first-person perspective view sharing system. In *RO-MAN, 2010 IEEE*, pages 125–131, Sept 2010.

[18] D. Kimber, P. Proppe, S. Kratz, J. Vaughan, B. Liew, D. Severns, and W. Su. *Polly: Telepresence from a Guide's Shoulder*, pages 509–523. Springer International Publishing, Cham, 2015.

[19] B. Lenggenhager, T. Tadi, T. Metzinger, and O. Blanke1. Video ergo sum: Manipulating bodily self-consciousness. *Science*, 317(5841):1096–1099, 2007.

[20] J. D. Mackinlay, S. K. Card, and G. G. Robertson. Rapid controlled movement through a virtual 3d workspace. In *Proceedings of the 17th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '90, pages 171–176, New York, NY, USA, 1990. ACM.

[21] B. Marsilii and T. Rossio. DÃĽjÃă vu (2006 film). Buena Vista Pictures, 2006.

[22] M. Minsky. Telepresence. *OMNI*, (July):45–51, 1980.

[23] S. Nagai, S. Kasahara, and J. Rekimoto. Livesphere: Sharing the surrounding visual environment for immersive experience in remote collaboration. In *Proceedings of the Ninth International Conference on Tangible, Embedded, and Embodied Interaction*, TEI '15, pages 113–116, New York, NY, USA, 2015. ACM.

[24] O. Oda, C. Elvezio, M. Sukan, S. Feiner, and B. Tversky. Virtual replicas for remote assistance in virtual and augmented reality. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software &#38; Technology*, UIST '15, pages 405–415, New York, NY, USA, 2015. ACM.

[25] P. Rander, P. J. Narayanan, and T. Kanade. Virtualized reality: Constructing time-varying virtual worlds from real world events. In *Proceedings of the 8th Conference on Visualization '97*, VIS '97, pages 277–ff., Los Alamitos, CA, USA, 1997. IEEE Computer Society Press.

[26] R. Raskar, G. Welch, M. Cutts, A. Lake, L. Stesin, and H. Fuchs. The office of the future: A unified approach to image-based modeling and spatially immersive displays. In *Proceedings of the 25th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '98, pages 179–188, New York, NY, USA, 1998. ACM.

[27] P. Salamin, T. Tadi, O. Blanke, F. Vexo, and D. Thalmann. Quantifying effects of exposure to the third and first-person perspectives in virtual-reality-based training. *IEEE Transactions on Learning Technologies*, 3(3):272–276, July

2010.

[28] P. Salamin, D. Thalmann, and F. Vexo. The benefits of third-person perspective in virtual and augmented reality? In *Proceedings of the ACM Symposium on Virtual Reality Software and Technology*, VRST ’06, pages 27–30, New York, NY, USA, 2006. ACM.

[29] N. Shiroma and E. Oyama. Asynchronous visual information sharing system with image stabilization. In *Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on*, pages 2501–2506, Oct 2010.

[30] N. Shiroma and E. Oyama. Development of virtual viewing direction operation system with image stabilization for asynchronous visual information sharing. In *RO-MAN, 2010 IEEE*, pages 76–81, Sept 2010.

[31] M. Sugimoto, G. Kagotani, H. Nii, N. Shiroma, M. Inami, and F. Matsuno. Time follower’s vision: A teleoperation interface with past images. *IEEE Comput. Graph. Appl.*, 25(1):54–63, Jan. 2005.

[32] M. Sukan, S. Feiner, B. Tversky, and S. Energin. Quick viewpoint switching for manipulating virtual objects in hand-held augmented reality using stored snapshots. In *2012 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 217–226, Nov 2012.

[33] M. Tatzgern, R. Grasset, D. Kalkofen, and D. Schmalstieg. Transitional augmented reality navigation for live captured scenes. In *2014 IEEE Virtual Reality (VR)*, pages 21–26, March 2014.

[34] J. Vincent. This futuristic performance by japanese trio perfume will make you lose your mind. *The Verge*, 2015.