Pathfinder Vision: Tele-operation Robot Interface in Consideration of Geometry for Supporting Future Prediction

Naoya Maeda[†1] and Jun Morita[1] and Maki Sugimoto[1]

[1]Keio University

Abstract

Pathfinder Vision is a visual presentation interface for remote robot operation. This interface assists operators by planning the future routes of the robot by presenting images which take into consideration the physical obstructions of the surrounding environment. We also took the occlusion of the 3D environment into account when rendering the CG model of the robot by calculating the current state and 3D geometry of the robot. By visualizing the predicted future with occlusion, our proposed methods makes it possible to prevent collisions while operating tele-operated robots. This system helps operators to find a more secure future trajectory of the robot.

Categories and Subject Descriptors (according to ACM CCS): H.5.2 [Information Interfaces and Presentation]: User Interfaces—Interaction styles

1. Introduction

An environment can become difficult for direct human intervention due to certain types of obstacles, including radiation, poisonous substances, or hard-to-get-around debris. When investigating such environments where the operability takes the priority, the tele-operated robots must be used. Manual operation of these robots is an important approach for critical missions taking places in complex environments and requires educated operators for its best performance. Operation of a vehicle using first person view image makes it arduous to visualize the relative position of the vehicle in the surrounding environment because the vehicle itself is not captured by the camera. This lead to the development of tele-operated robots capable of perceiving both the surrounding environment and its position, as well as an operating interface that uses third person view image generated either by projecting the CG-based vehicle model to the stored sequence of images, or by directly capturing the scene using pole cameras. Using third person view image for the operation lets the operator become aware of the relative position of the vehicle in the surroundings.

However, predicting the near-future event is still dependent on the operator's prior experience and judgement, such as a collision or fall that may or may not result from the current course of vehicular movement, which makes it hard to set the proper route for the vehicle. In order to predict the robot's future state, it is necessary to recognize the shape of the terrain and obstacles of the real environment, and the relative position between the robot and obstacles. Recognizing the shape and position requires the acquisition of three-dimensional information of robot and real environment. In this paper, we propose an informative interface that helps the operator perceive near-future event by generating and presenting images depicting the predicted events while considering the physical interferences and current state of vehicular operation. Our interface allows us to present the current position of the robot clearly while considering the three-dimensional information of the real environment. In our proposed method, the shape model of the real world obstacle is generated based on the data captured by the depth camera in real time. Furthermore, the system predicts where the robot will be after a few seconds based on the current state of the operation. The prediction of the near-future robot state is governed by a physics simulation of the robot model and the shape models of the obstacles. A future prediction image is generated according to this result and is presented to the user. The future predicted image provides a third-person view image taking into account the occlusions in order to present the relative position of the obstacle and the robot in the pre-

---

† nmaeda@imlab.ics.keio.ac.jp

dicted image easily to operators. In this paper, we describe the Pathfinder Vision tele-operational interface and the implementation of our prototype. We introduce the method of depicting the predicted future state, and discuss the method for generating such images. We evaluated the difference in operability by comparing our proposed method with three other methods of images for robot tele-operation: first person view image captured from the robot, third person view image representing the current robot state with CG model, and predicted image without a physics simulation. Figure 1 shows a robot and predicted image.

Figure 1: *Pathfinder Vision: Tele-operate interface assists future prediction.*

2. Related works

There are many challenges in improving operability of tele-operational robots. One remarkable concept is Telexistence which enables the operation of a robot from a remote location by providing natural sensory information from first person view to the operator as if the operator existed in the robot's location.This concept is introduced in TELESAR [Tac98] which shows many advantages for remote robot operation.

On the other hand, it is known that providing third person view image improves operability of vehicles in video games. Time Follower's Vision [SKN*05] provides virtual third person view images by recording camera images with location information and superimposing a wire frame model of a robot according to the current state. An application of this kind of interface is driving vehicles from a remote location for collecting information. Nguyen provides the application of VR interface for exploring the Mars [NBE*01]. The user can operate a rover using virtual third person view image and the information of the surface. Another approach proposed by Nielsen et al in making a virtual third person view image is to capture the 3D geometry around a robot by a range sensor [NGR07]. Based on the captured geometry, a model of the robot is rendered from a birds eye view. Furthermore Okura et al. proposed a virtual free view point robot control system by using a spherical camera and depth cameras [OUSY13]. Other approaches to consider about mapping and robot's position by combining with stored past 3D information [Dav03]. Although these related works were able to make better understanding of the current situation of robots for operators, future states of robots

must be considered to avoid risks of colliding into obstacles. Especially with remote first person view, it is hard to predict future states correctly because the body size of the robot is not visible. Even in the third person view, correctly prtedicting the future state is difficult due to remote camera parameters which does not match with the natural perception of operators. There are even some difficulties when using 3D information. When tracking vehicular position and reconstructing the 3D environment by integrating stored past images, an error can accumulates during mapping.

To solve these issues, there are several visual interfaces which assist trajectory prediction by operators. A popular interface is a car parking assistance system with a rear view camera [TOY]. The system provides rectangles on rear camera images that shows predicted future positions of a car. Another interface is a touch panel based remote robot operation interface named Touch Me [HIII11]. The interface shows a future model of a robot according to touch gestures by an operator.These assistance systems are able to show future models of the control targets. However, these systems are unable to obtain the 3D geometry of the environments around the targets. Therefore, their predicted models of the targets are unable to take into consideration of the collision with their surrounding environment. The predicted image might present the wrong result of simulation, which may result in collision.

One of remarkable possibilities with depth cameras is the ability to detect the 3D geometry of physical obstruction. Mirage Table [BJW12] uses the potential of the depth cameras for real time interaction in a tabletop environment. Also, in the field of interactive systems for designing kinematic art works, there is a prior work which simulate physical obstruction by applying a physics engine with consideration of real time design operation by users [FMIF10]. The prior work successfully visualize future trajectory of a target piece instantly. In this paper, we propose a method which applies future visualization in the third person view for tele-operating robots while taking into consideration physical obstructions by using depth cameras. Moreover, in simulating the collision between the robot and the environment, in order to prevent the accumulations of errors, we apply the simulation on a stored past image.

3. Future prediction interface

3.1. Future predicted image

We introduce three different methods of displaying images. The first method directly shows the captured images from the camera placed on the robot in first person view image. The second method shows the images from a third person viewpoint, either by superimposing a CG model of the robot onto the image scenes captured from the environment or by using specially placed camera for robot observation. In previous work, the third person view image is generated by projecting a CG-based vehicle model to the stored sequence of

images [SKN*05]. By expanding on this concept we propose the third method; the future-predicted image. Future-predicted image refers to a third person view images in which a CG model of the robot visualizes the expected trajectory of the robot if the robot continues the given course of movement. The real-world obstacles are virtually rebuilt within the future-predicted image and physically simulated, therefore making it possible to predict and visualize safe routes for the robot and prevents collision in the presence of obstacles. Figure 2 shows the concept of each method. In this paper, we refer to the image showing the predicted state of the robot with CG model in a third person view image as the "future predicted image."
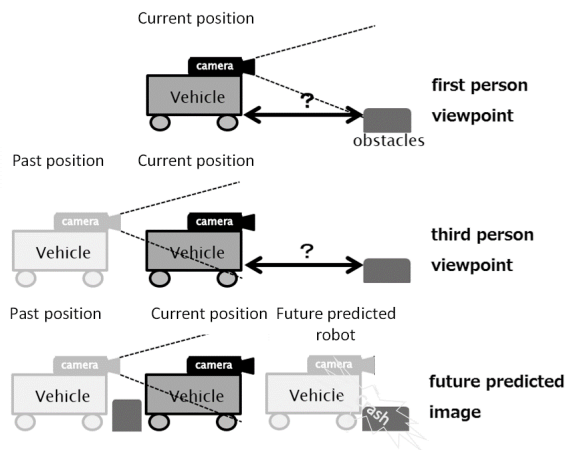

Figure 2: *Concept of each presenting methods.*

3.2. System

The diagram of the system is shown in Figure 3. The main processing PC transmits the commands of the movement of the robot such as movement speed or directional change to the laptop PC that is mounted on the robot via User Datagram Protocol (UDP). At the same time, the laptop PC also sends RGB images, depth images, and the state of the robot to the main processing PC via UDP. The area of obstacles are extracted based on RGB and depth images, and the shape model of the real world obstacle is generated based on the data captured by the depth camera. The predicted vehicle is generated by input command and received images. The predicted vehicle has positional and velocity data, and these are fed into physics simulation for vehicular interaction with the obstacle models. At the same time, based on the positional relationship between the three-dimensional robot and the obstacles, an occlusion can be rendered between the obstacle and the CG of the robot. According to the simulation results and the rendered occlusion, the âĂIJfuture predicted imageâĂİ is generated, and presented to the user. By con-

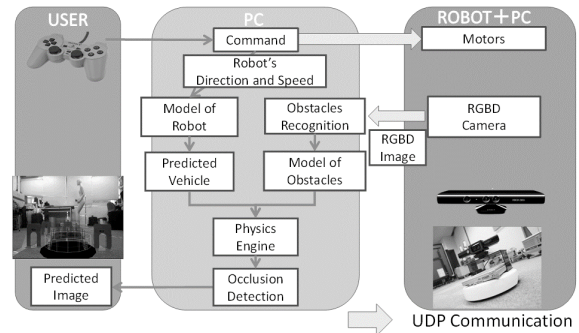sulting this image, the user is able to plan a safe route for the robot.

Figure 3: *The block diagram of the system.*

3.3. Replay of obstacles in real world

It is necessary to be able to know the three-dimensional shape of the robot and obstacles in order to detect and simulate future collisions between the robot and obstacle. We approximate the size of the robot with a cylinder (diameter 34cm, height 30cm). The obstacles in the real environment are portrayed with only the surface information of the obstacles. Our system can predict the collision between the robot and obstacles based on the surface information of the obstacles which can be seen from the depth camera. The obstacles are rendered as sets of point cloud data. Because the processing time depends directly on the size of the point cloud data, we chose to approximate the shape of the obstacles with a collection of spheres (diameter 2cm).Figure 4 shows a three-dimensional shape based on the information from the depth camera and a model that approximates the shape of the obstacles in the image.
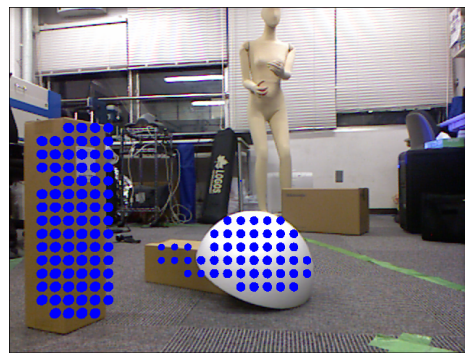

Figure 4: *Three-dimensional shape model of the obstacles and approximates the shape of the obstacles.*

3.4. Simulating collision between robot and environment

Generally if a robot collides with an obstacle while it was moving, the robot may come to a full stop or deviate from

the intended direction. To clearly present to the user the expected movement and position of the robot in each time interval, we use a physics engine. We use a physics engine to present to the user, the movement and the position of the robot in consideration of the passage of time clearly. Our physics simulation runs using Bullet Physics simulation. Figure 5 shows an image before applying physical simulation and an image after applying one.
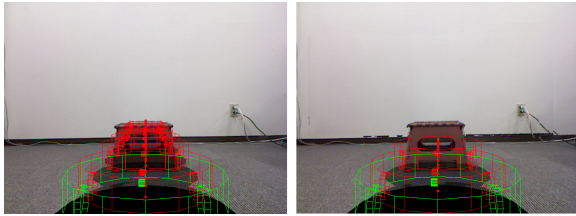

Figure 5: *Applying physical simulation (left: before, right: after).*

4. Implementation

4.1. Tele-operated interface

The depth camera used in our work is incapable of capturing data for objects placed within 50cm from the sensor. The main PC recognizes that there are no obstacles because the depth camera is unable to obtain the 3D data from the RGB-D camera viewpoint less than 50cm. For users to understand the real world, it is necessary to complement the part where three-dimensional shape is not available from the depth camera. Therefore, in our implementation, the CG-based vehicle model is projected on the stored sequence of images to compensate for the shape model of the obstacles. Projecting the state of the vehicle to the past image allows to add the information of the nearby obstacles as RGB color image, and recognize the current state of the robot easily. Figure 6 shows the interface using previous method.
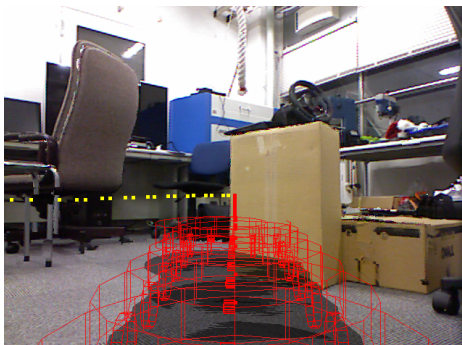

Figure 6: *Future predicted image.*

4.2. Change third-person viewpoint image

It is important for the prediction model and the CG robot that represents the current state of the robot is presented appropriately to the operator. There are some situations where the systems fails at this. For example, if the robot continuously moves a route without having the prediction model update itself, the operator will lose view of the CG robot model within the third person view. In order to appropriately render the CG robot and the prediction model within the prediction image, it is necessary to obtain a different view image of the both the current state of the robot and the prediction model. In our method we used stored past images to solve this problem. In our system the third person view image updates to a new third person view from a new point after a certain length of time. Regarding the updated viewpoint, when we superimpose the CG robot onto the stored past image, we select the optimal stored past image based on the current position of the robot. The system decides the optimal camera view image by the last stored robot's coordinate and amount of the total robot's movement. The optimal image is calculating the difference of the x and y coordinate of the current robot position, and x and y coordinate of the robot position from the camera viewpoint of the stored past image. The CG robot is then projected on the optimal stored past image corresponding to the current position of the robot. Figure 7 shows the selection method of the optimal stored image.
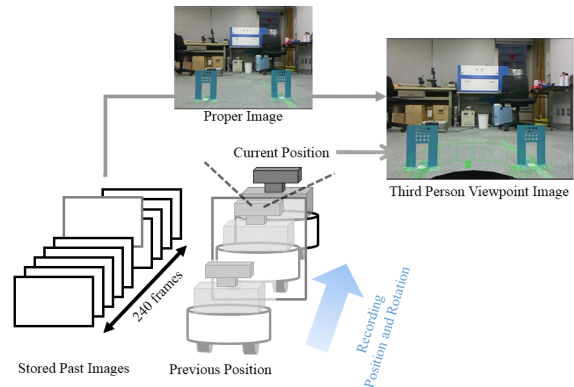

Figure 7: *Selection Method of the optimal image.*

4.3. Occlusion between robot and obstacles

In our research, the three dimensional position information of the obstacles is obtained by the using the depth camera and the three dimensional position information of the robot relative to the obstacles is obtained from the input of the operator. By matching the coordinate system of the real environment and coordinate system of the CG robot, the relative position between the obstacles and the robot is known. Therefore, it is possible to figure out whether the CG robot in the predicted image is in front of the obstacles or not. Figure 8 shows an image without the visual representation consideration of the occlusion and an image with the occlusion.

The predicted image with occlusion clearly presents the relative position between the vehicle and the obstacles.
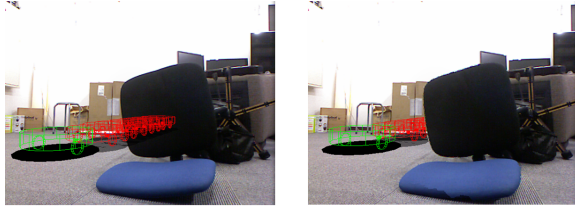

Figure 8: *Applying occlusion (left: before, right: after).*

5. User experiment

We have designed a user experiment to evaluate our proposed method. Table 1 shows the system configuration for the experiment.

Table 1: *System configuration*

Vehicle	iRobot Create
Images resolution	640 × 480 pixel
Wireless standards	Wi-Fi IEEE 802.11a
PC for operating	Desktop PC
CPU	Intel(R) Core(TM) i7-3930K 3.20GHz
GPU	NVIDIA GeForce GTX 680
PC on robot	Laptop PC
CPU	Intel(R) Core(TM) i7-3630QM 2.40GHz
RGB-D camera	Kinect for Xbox360

5.1. Experimental design

We conducted the experiment on courses with obstacles to evaluate the effectiveness of the proposed method. We experimented on the differences of the operability with some obstacles in order to confirm the usefulness of our system. The experimental path is shown on Figure 9. The experimental conditions were the following:

- The subject starts operating the robot and must go through three check points until the robot reaches the goal.
- There are four gates in the path, and the subject must judge if the robot can go through the each gate or not.
- If the subject judges that the robot can go through a gate, the subject must attempt to go through the gate; if the subjects judges that the robot cannot go through the gate, the subject must not attempt to go through the gate and avoid the gate.
- While trying to go through a gate, and if the subject judges the robot cannot go through the gate, the subject may choose to go back and avoids the gate.
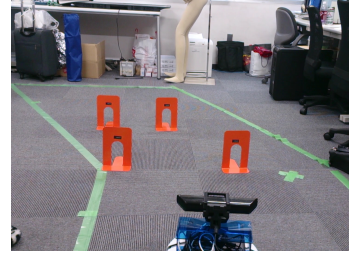- The subjects must avoid as many collisions as possible.

Figure 9: *Experimental path and user experiment.*

We evaluated our proposed method by comparing with three other types of images for tele-operational robots: first person view image, first person view image + third person view image, first person view image + future predicted image without physics simulation, and future predicted image with physics simulation(the proposed method). We compared in the average of the number of collision and the misjudgement rate in deciding if the gates were passable or not. The experimental path had four gates and the widths of these gates were 25cm, 31cm, 37cm, and 43cm each. The width of the robot was 34cm in the experiment. Therefore, two gates (37cm, 43cm) were passable, and the rest were non-passable. The subject used the keyboard or joy pad for operation. 5 subjects (ages 21-24) participated in our experiment and the number of trials were two times per person. The four different methods are shown in Figure 10.
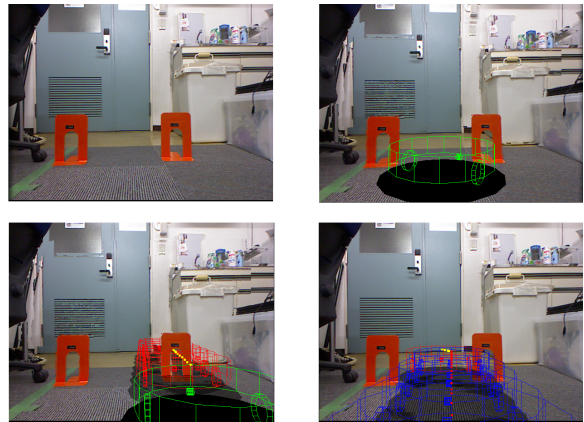

Figure 10: *Four presenting methods (first person view, third person view, future predicted image without physics simulation, future predicted image with physics simulation).*

5.2. Result

5.2.1. Comparing the average of the collision counts in each presenting methods

Figure 11 shows the average of the collision count and standard deviation of the four different methods.
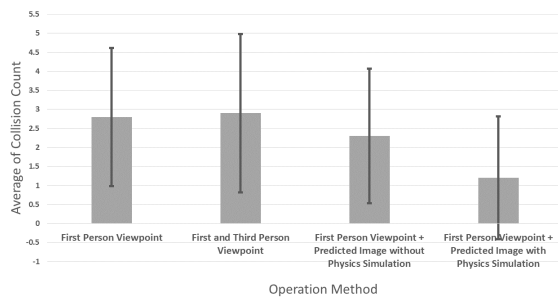
Figure 11: *The average of the collision count and standard deviation.*

In the result, the future predicted image with physics simulation showed less collision count than other methods. All methods except for future predicted image with physics simulation showed no significant difference. Comparing the predicted image without physics simulation and with physics simulation, it is observed that the collision count of with physics simulation is less than half. In terms of the average of collision counts in four methods, it is possible to consider the predicted image with physics simulation reduced collisions against other four methods in this experiment.

5.2.2. Comparing the rate of false recognition in each presenting methods

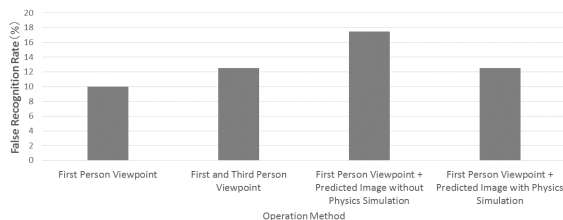Figure 12 shows the false recognition rate in four presenting methods.

Figure 12: *The false recognition rate in three presenting images.*

In this result, comparing the false recognition rate, there was no significant difference in four presenting methods. Therefore, we investigated the detailed the false recognition rate in each width of the gates as shown in Figure 13.

In the result, comparing the false recognition rate in the four widths, the future predicted image improved the false recognition rate at the gates of three widths (25cm, 31cm and 43cm). Especially, the future predicted image with physics simulation shows no false recognition in case of the gates narrower than the robot width (25cm and 31cm). Therefore,
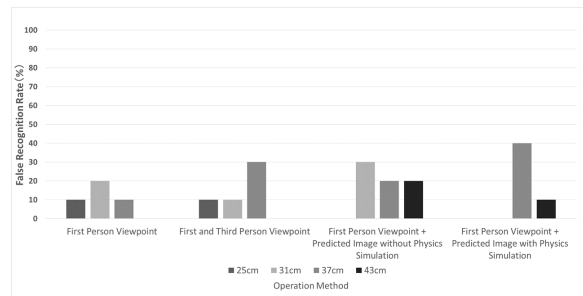

Figure 13: *The false recognition rate in each width.*

we think the future predicted image helped the user to become aware of the future collisions with the surrounding environment, and also assisted in selecting a safe route in those conditions. In contrast, the future predicted image did not reduce the false recognition rate in 37cm. It is wider than the width of the robot. However, future collisions with the passable gates was presented by the physics simulation in some situations due to the approaching angle of the robot and the gate. If the robot approaches to the gate almost orthogonally, the simulator shows no collision. On the other hand, when the setting angle become larger, it showed possible collisions. In consideration of the gate model which were approximated as tiny spheres, in such a situation, the users had tendency to select safer route without changing the approach angle. It was the cause of avoiding the passable gates. It is one of the limitations of the current implementation of future prediction.

6. Limitation and future works

In our current implementation of the system, there are several limitations. One limitation is the inconsistency between the dynamic obstacles and already-captured past images. It is possible to solve this issue by combining past depth images and the current depth image. Another limitation is the drifting of location information. The system assumes that there is no slipping between the floor and the wheels of the robot. Integrating the current frame with the just previous frame might reduce the problem. Furthermore, the system is implemented upon a hypothesis that the robot is on a flat floor to recognize obstacles in the real environment. Improving the algorithm of RGBD images processing for real environment modelling to adapt to complicated 3D geometry remains as a future work. Moreover, for improving our interface more useful, we will investigate the factors for optimal predicted image.

7. Conclusion

In this paper, we proposed a tele-operated robot control interface which assists operators predict future events by visu-

alizing virtual physical collision between models of the real environment and the robot according to real-time information. The environment model was obtained as a point cloud which was approximated as spheres in the physics engine. We conducted a user study to prove the effectiveness of our method during robot operation. The result of the experiment shows that our proposed interface reduced collision count. This implies that our method has potential to improve route planning by operators in real-time operation of tele-operated robots.

References

[BJW12] BENKO H., JOTA R., WILSON A.: Miragetable: free-hand interaction on a projected augmented reality tabletop. In *Proceedings of the SIGCHI conference on human factors in computing systems* (2012), ACM, pp. 199–208. doi:10.1145/2207676.2207704. 2

[Dav03] DAVISON A. J.: Real-time simultaneous localisation and mapping with a single camera. In *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on* (2003), IEEE, pp. 1403–1410. doi:10.1109/ICCV.2003.1238654. 2

[FMIF10] FURUTA Y., MITANI J., IGARASHI T., FUKUI Y.: Kinetic art design system comprising rigid body simulation. *Computer-Aided Design and Applications 7*, 4 (2010), 533–546. doi:10.3722/cadaps.2010.533-546. 2

[HIII11] HASHIMOTO S., ISHIDA A., INAMI M., IGARASHI T.: Touchme: An augmented reality based remote robot manipulation. In *The 21st International Conference on Artificial Reality and Telexistence, Proceedings of ICAT2011* (2011). 2

[NBE*01] NGUYEN L. A., BUALAT M., EDWARDS L. J., FLUECKIGER L., NEVEU C., SCHWEHR K., WAGNER M. D., ZBINDEN E.: Virtual reality interfaces for visualization and control of remote vehicles. *Autonomous Robots 11*, 1 (2001), 59–68. 2

[NGR07] NIELSEN C. W., GOODRICH M. A., RICKS R. W.: Ecological interfaces for improving mobile robot teleoperation. *Robotics, IEEE Transactions on 23*, 5 (2007), 927–941. doi:10.1109/TRO.2007.907479. 2

[OUSY13] OKURA F., UEDA Y., SATO T., YOKOYA N.: Teleoperation of mobile robots by generating augmented free-viewpoint images. In *Intelligent Robots and Systems (IROS), 2013 IEEE/RSJ International Conference on* (2013), IEEE, pp. 665–671. doi:10.1109/IROS.2013.6696422. 2

[SKN*05] SUGIMOTO M., KAGOTANI G., NII H., SHIROMA N., MATSUNO F., INAMI M.: Time follower's vision: a tele-operation interface with past images. *Computer Graphics and Applications, IEEE 25*, 1 (2005), 54–63. doi:10.1109/MCG.2005.23. 2, 3

[Tac98] TACHI S.: Real-time remote robotics-toward networked telexistence. *Computer Graphics and Applications, IEEE 18*, 6 (1998), 6–9. doi:tachi1998real. 2

[TOY] TOYOTA: Back guide monitor. (Proc. Eurographics'98) http://www.toyota.co.jp/jpn/tech/safety/technology/parking/. 2