第 18 回日本バーチャルリアリティ学会大会論文集(2013 年 9 月)

Study on Telexistence LXXV
Realtime self-superimposing in Virtual Environments for Telexistence

テレイグジスタンスのためのバーチャル環境への自己身体の実時間重畳

MHD Yamen SARAIJI [1), Charith Lasantha FERNANDO [2),
古川 正紘 [3), 南澤 孝太 [4), 舘 暲 [5)
MHD Yamen SARAIJI, Charith Lasantha FERNANDO,
Masahiro FURUKAWA, Kouta MINAMIZAWA and Susumu TACHI

慶應義塾大学大学院メディアデザイン研究科
(〒223-8626 横浜市港北区日吉 4-1-1, {yamen, charith, m.furukawa, kouta, tachi}@tachilab.org)

Abstract: On telexistence, visual consistency is important to deliver existing experience to the user. In the current design, the utilization of humanoid robots provides the minimum required degrees of freedom for human bodies. However, it doesn't preserve the visual consistency feedback of user's body. Thus in this paper we provide a new method to present operator self-body visuals and motion into robot's vision during teleoperating. In this method, we utilize the estimated trajectory of user's hands and arms movement to mask his local body via a see-through head mounted display and impose the result over robot's vision. In this paper we describe the design and implementation of the technique.

Keywords: Telexistence, Virtual Reality, Self-superimposing

1. Introduction

During a virtual experience in an immersive environment, we often check for our hands to confirm our existence in that environment. This implies the importance of perceiving an owned body at the expected location to maintain the immersive experience, and as a way of verification of begin in that world.

On telexistence to the virtual environment as well as real environment, it is difficult to show the his/her own with keeping appropriate position and orientation between the two first person's view, masking, fusing and color collection in real time due to some optical constraints, chrome key color errors or depth measuring errors to cut out. So in this paper, we propose essential method to cut out the own appearance and fuse it onto the environment where the user exists as shown in Figure 1.

2. Related work

Such issue can be considered coming from "visual consistency". In other words, the hands and arms which appear in from of you when you put them closer to your face, looks different from your own one, although the appearance of them in the virtual or in the remote place while telexistence is expected to be your own one.

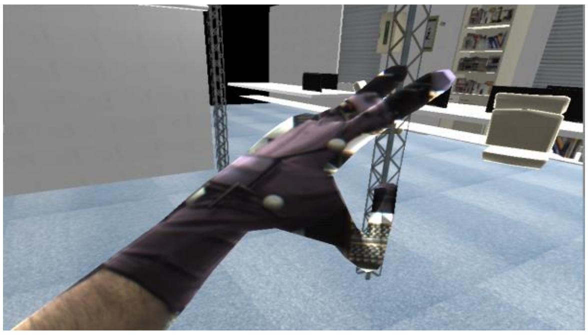

Figure 1: Remained appearance in virtual world

The visual consistency can be categorized in to the two depending on the point of view; first person's view or not. On the telexistence, visual consistency to the audience's perspective is also important. Thus the avatar robot, which represents the remote user but does not look the user, is required an appearance as same as the user in terms of anonymous. This is the visual consistency on the second and third point of view. On "TELEsarPHONE" [1], operator's body is captured and projected on his remote avatar via retro reflective technology [2] to provide mutual communication. However, the user in the avatar could not see his own hand in proper relative position between the remote environment and his/her self.

On the other hand, the first person's view requires the expected trajectory of the hands, which is synchronized with the user's hand in real time as well as relative position between the user's eyes and hands correctly. In other works, bodily kinematics and visual consistency are key points to maintain during the operation. In "TELESAR V" [3] , a 53 DOF dexterous slave robot was designed to maintain the bodily and kinematics consistency between master and slave. In the regard of visual consistency, the aim is to provide seamless operating visual experience to the user which to see his own body rather than a different body – robot. Thus the problem is about how to introduce a physical object, which is operator's body, into another environment such as in teleoperation environment or into a virtual environment.

Noyes and Sheridan [4] proposed a predictive graphic displays where deployed to augment visual feedback and provide a superimposed virtual representation of the slave robot. The slave robot is visible though some obstacles exist between slave 's camera and hands. However, the method does not provide the user's appearance.

In MirageTable [5] , physical objects including user's body are captured via depth and color sensors of Kinect and reproduced in remote environment on a table. This method requires an additional sensor to acquire body outline of the user.

In this paper, we focus on addressing visual consistency for the operator by proposing a new method to superimpose user's body visuals on the slave's vision. The proposed self-projection technique can be integrated with many teleoperation applications as well as in Augmented Reality applications by using mounted camera(s) and body tracking techniques. We describe how the system is organized, and implementation details of this method.

3. System Design

In this technique, we provide a presentation for user's self-body superimposed over his visual feedback from the slave robot. This presentation should override slave's robotic parts using operator's arms and hands which is our focus.

We assume that the operator is using a HMD for visual feedback from the slave robot. We assume also the HMD has a see-through part. This method works by obtaining images from operator's point of view to capture his visual appearance, and we calculate user's active movement and provide it to kinematics solver to construct a virtual representation of his body. This virtual representation allows creating a mask image to identify his body area in the acquired images. Also, we correct light and color information on the resulted masks.

Method's flow can be described briefly in Figure 2, in this diagram "Operator" represents the active user of the system, the operator will generate movements which are transferred to slave robot. "Kinematics Solver" element is responsible to convert operator's movement into slave robot as joint angles values.

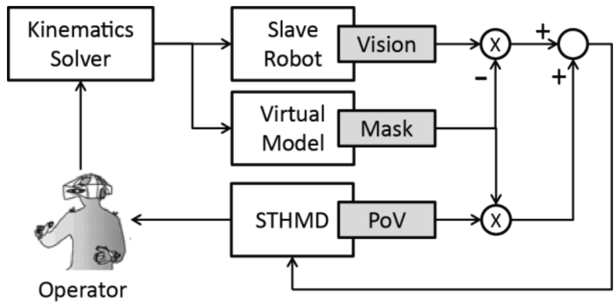

Figure 2: Superimposing flow diagram

The "Slave" represents the remote avatar of the operator which can be a physical robot or virtual environment. It outputs visual images of the environment which is represented as "Vision". "Virtual Model" is a simulated model of slave's body which follows the trajectory of the user. By using this model, we generate a "Mask" image to be used to mask operator body image from the other parts. The operator uses See-Through HMD "STHMD" which is mounted on his head, it has stereo camera system, and outputs user's Point of View "PoV" images.

For masking, we define three basic operations: "(X)" masks out parts of an image by using reference mask layer, "(-)" Inverts the reference mask, and "(+)" which combines two images together into a single one. The resulted image is provided to the user in his HMD. Further details described in the following sections.

4. System Implementation

4.1 Operator's visual acquisition

To acquire operator's appearance in real-time, we utilized a STHMD as shown in Figure 3, which captures stereo video images from the operator's point of view. The HMD, which we designed, provides wide-angle vision of H×V (61◦×40◦), a HD LCD 1280x800px, and two USB cameras, Point Grey Research, Inc. Firefly (FFMV-03M2M), with almost identical field of view lens of the HMD (horizontal 62◦), max resolution of 752x480px@60FPS were installed on the HMD with a horizontal distance of 65mm. Camera resolution was cropped to match LCD resolution.

In Figure 4 a stereo output images from both cameras which show operator's hands. The virtual light path length between the cameras and operator's eyes in this design is 8.5cm, it is important to have a conjugated focal plane: to have the same optical length and center in order to get the similar point of view of the operator.

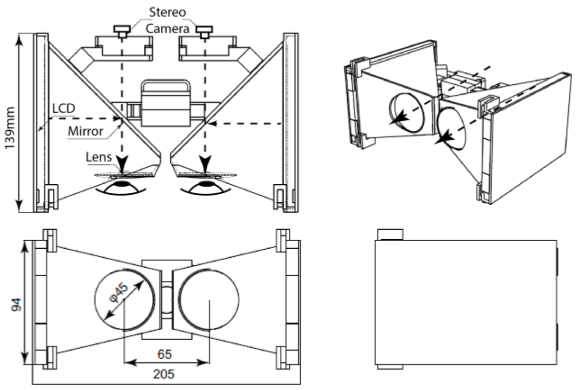

Figure 3: See-Through Head Mounted Display (STHMD)

4.2 Kinematics solver

The Virtual Model and the Slave representation in the remote environment need to be matched. Thus we utilize the same kinematic solver [6] to control the slave robot "TELESAR V" in our virtual environment's slave representation. The solver in this system receives five tracking points from operator's body: both sides of shoulders, both sides of hands, and the head. It map them into 23 DOF body joints. For the hands' fingers, the operator uses data glove to track his fingers and mapping them into 15 DOF per hand. Kinematic solver generates joint angle data to transfer them to Virtual Telesar.

4.3 Masking

To isolate operator's hands and arms from the background pictures, we generate a mask image by using the Virtual Model. The Virtual Model is simulated by using Virtual Telesar Platform [6] , this model receives the same joint angles which were generated by the kinematics solver for the slave robot.

In the simulated environment, the virtual camera is placed with the same displacement between the eye ball and the physical camera of the HMD. This ensures matching the projected size and position between user and Virtual Model.

In Figure 5, a generated black and white image representing the mask of arms and hands. White areas shows the parts which to remain, and black areas are the parts to erase.

Technically speaking, the process of generating the mask was done in the GPU as a post-processing phase. The masks were rendered into a separated GPU Render Target. This allowed us to achieve higher performance than processing it on the CPU.

4.4 Image fusion

After we have acquired the required information to identify an estimated representation of user body via the mask image, we apply masking operation over PoV images. An inverted mask cuts the outer area of the slave's Vision. Figure 6 shows the combined result of both masked images.

As a result of direct fusion of the images, the resulted super imposed hands have flat look and hard edges with the remote environment, which is a result of different lighting conditions between the two environments.

4.5 Lighting and color correction

In order to match lighting conditions of slave's environment, an estimated lighting model is applied on PoV images. Because we have a 3D model representing slave's body, we apply lighting equations on this model.

An estimated light correction layer is generated by applying a Fresnel light equation (1) at each pixel of the simulated model:

$$F = \max(0, 1-(\vec{N} \cdot \vec{V}))^2 \qquad (1)$$

where "F" is the Fresnel value, "N" is Normal vector of the model's pixel, and "V" is the normalized view vector to the pixel. Figure 7 shows the resulted layer. And in Figure 8, the color corrected image shows a consistent lighting appearance between both environments.

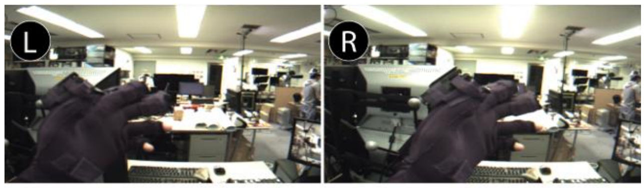

Figure 4: See-Through HMD Cameras images (Left/Right)

Figure 5: Mask images generated by the Virtual Model

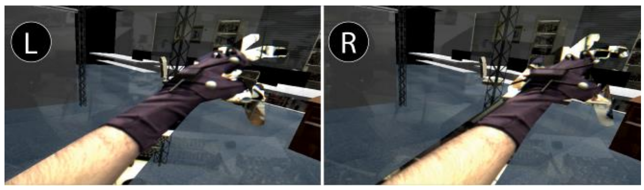

Figure 6: Result after fusing the pictures with the mask

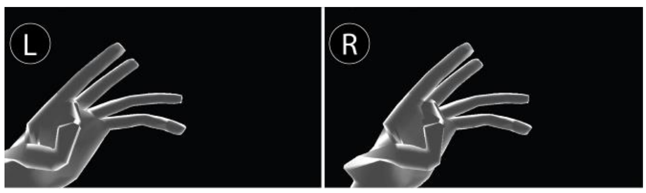

Figure 7: Estimated light correction layer

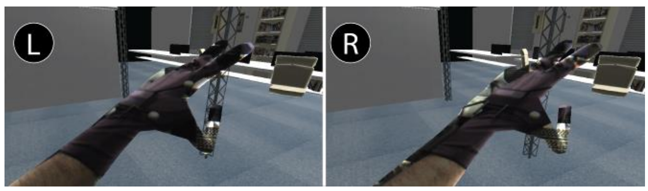

Figure 8: Color correction and light matching

4.6 Experimental Condition

The method was carried on real-time with frame rate exceeding 60FPS using CPU core i7 at 2.30GHz/8GB of memory and NVidia GeForce GT 650M. Image processing and fusion were done on the GPU.

The laboratory room shown in Figure 4 was replicated in Virtual Telesar as shown in Figure 6 for prototyping and testing. Room matched the physical one in terms of scale and objects localization in the scene.

User's active movement was captured by 14 cameras of Natural Point OptiTrack V100:R2. HMD has 5 reflective markers and other 4 points have 3 markers each. The finger movement is measured by 5DT Data Glove 5 Ultra which provides only 5 DOF for each hand, this yielded the lack of mapping to 15 DOF fingers per hand.

4.7 Discussion

It was possible to preserve the hands where we expect it to be. Also the color tone of the imposed images was improved after applying light correction model to user's body images. Light correction model can be changed depending on the situation. Also light placement is possible in the virtual environment to replicate the actual situation for the remote environment.

The appropriate experimental distance for the hands in the current conditions and design of the STHMD is approximately 15 to 40 cm from operator's eyes. The camera position, which is not conjugated with the eye balls resulted the displacement between mask and captured the user's appearance. The issue can be improved using the conjugated optical system with some external mirror.

However, a noticeable shift of the masked area (especially at the fingers) caused a slight outline of the local environment to appear, which is mainly regarded as sensor tracking precision problem. Also hand position shift depending on the distance from the cameras was noticeable. The reason of that shift is due to the optical displacement of the see-through cameras.

In this experiment, the binocular difference was set to 65mm which is considered as an average distance for humans, although it can be changed via HMD slider. This value is reflected into virtual environment camera horizontal placement as well.

Regarding the performance, frame rate was limited to 60 FPS for rendering the previous scene. A slight delay of tracking the movement of the user was noticeable when the he performs sudden moves. A solution for this issue can be performed by providing a temporal tracking to operator's body appearance, and avoiding sudden changes in the new calculated masked color area of the hands.

5. Conclusion

In this paper we proposed self-superimposing system to preserve visual and appearance consistency for teleoperating applications. In order to satisfy that, we presented a technique which utilizes a see-through camera system, a kinematic solver and simulated environment for mask generation.

We will apply the proposed method to the telexistence experience to the real world as well.

Acknowledgement

This work was supported by JSPS Grant-in-Aid for Scientific Research A (23240017).

References

[1] S. Tachi, N. Kawakami, H. Nii, K. Watanabe and K. Minamizawa, TELEsarPHONE: Mutual Telexistence Master Slave Communication System based on Retroreflective Projection Technology, SICE Journal of Control, Measurement, and System Integration, Vol.1, No.5, pp.335-344 (2008.9)

[2] S. Tachi, Telexistence and retro-reflective projection technology (RPT), Proc. 5th Virtual Reality Int. Conf. (IVRIC2003) pp. 69/1–69/9, 2003.

[3] C.L. Fernando, M. Furukawa, T. Kurogi, S. Kamuro, K. Sato, K. Minamizawa, S. Tachi. , "Design of TELESAR V for Transferring Bodily Consciousness in Telexistence" IROS 2012, IEEE/RSJ International Conference on Intelligent Robots and Systems, 2012.

[4] M.V. Noyes and T.B. Sheridan. A novel predictor for telemanipulation through a time delay. In Annual Conference Manual Control, Moffett Field, CA, 1984.

[5] Hrvoje Benko, Ricardo Jota, and Andrew Wilson. MirageTable: freehand interaction on a projected augmented reality tabletop. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12), 199-208, 2012.

[6] M.Y. Saraiji, C.L.Fernando, M. Furukawa, K. Minamizawa, S. Tachi, Virtual Telesar - designing and implementation of a modular based immersive virtual telexistence platform, SICE-SI, 2012.