テレイグジスタンスの研究（第 77 報）
―実環境内でのバーチャルな身体を用いた触覚伝送―
Study on Telexistence LXXVII
-Haptic transfer system using virtual body in real world-

〇MHD Yamen SARAIJI, Charith Lasantha FERNANDO, Masahiro FURUKAWA, Kouta MINAMIZAWA and Susumu TACHI

Keio University, {yamen, charith, m.furukawa, kouta, tachi}@tachilab.org

In this paper, we present a method to calculate touch forces in a remote place using low-cost depth camera and present those forces to the user in real-time. Using depth camera's captured data, a 3D virtual scene surface representation is constructed from the remote point of view and used to calculate pushing and stroking forces during user fingers movement. The algorithm computes force vectors for each desired point using point's motion vector and calculating god-points against the scene surface. By using such a method, it is possible to avoid using complex articulated arms and hands in the remote side to measure haptic forces at fingers end points. We show here the process of capturing and constructing the environment, and the method of calculating the interaction forces using depth data. Also we suggest using superimposed virtual hands for natural interaction with the remote environment.

***Key Words*:** Augmented Reality, haptic

1. Introduction

In telexistence applications, the degree of presence that the operator perceives in a remote environment reflects the quality of operation and the sense of presence in that environment [1]. Telexistence robots aim to provide the feeling of presence in a different location by mapping human's senses and actions with a slave robot. TORSO [2] is a telexistence robot providing the operator dexterous head movement and stereovision by mapping it with 6-DOF robot head, but no hands are involved in this representation. In TELESAR V [3] the user is mapped into a full upper body slave representation with a 53 DOF robot allowing the user to operate naturally in a different place. The fingertips of the robot are associated with physical force and tactile sensors. In a previous study, visual and haptic sensory perception of a human works together to compensate each other's biased sensory information misjudgment [4], and using cross-modal perception from both of them, the judgment on object properties (like shape and size) would be improved [5].

With regards to virtual environments, previous works [6][7] addressed the ability to access the virtual space and being able to touch and gain haptic feedback of the virtual objects by designing telexistence robots with mounted haptic sensors and simulating them in the virtual environment. Also, another work [8] was introduced for improving the sense of presence in a remote environment by superimposing user's body into a virtual environment. It showed that it is possible to include your own visual presentation of arms and hands by using kinematics data and a body model matching operator's dimensions. Providing the visual sense of user's arms and hands existence during operating in a different environment gives better experience of natural interaction.

In order to provide haptic feedback for non-armed telexistence robots, we present a depth-based haptic force calculation and extraction from operator's first point of view during interaction with the hand. This method uses a mounted depth sensor aligned with stereo camera vision mapping user's point of view in a remote place, and constructs 3D scene surface using depth-map of the remote place. Using the constructed surface, touch force between user's fingers and the remote environment are calculated and provided to haptic displays mounted on user's fingers. Also we adapt the visual superimposing of realistic virtual arms matching operator's arms posture into remote environment images.

2. Design Considerations

In this work, consumable depth sensor "PrimeSense Carmine 1.08 short range" is used to capture and construct physical environment's geometrical information. Depth sensors such as Kinect and PrimeSense provide 11-bit resolution of depth data, and 320x240 up to 640x480 depth-map generated at about 30Hz which is remarkable quality for many real-time applications. Depth sensors usually use Infra-Red (IR) light beams in specific patterns for depth estimation and work in complete darkness.

However, several challenges remain when using raw depth data from the sensor. Problems such as resulting noise from the sensor, and missing values or 'holes' that no structured light depth reading was possible as shown in Fig. 1. This can be the result of certain materials in the environment doesn't reflect IR light back to the sensor.

To solve such issues, incremental methods create spatial representation of the environment using series of captured depth-maps, such as KinectFusion [9] which uses a single handheld depth camera. Such methods are effective for robotic investigation in unknown environments, and for AR based applications. However such systems are optimized to construct static scenes and non-movable objects such as people passing by the camera which is a common case for telexistence applications. For our purpose, we are only interested in the current state of the scene, thus we calculate the scene surface per-frame with no temporal information.

Fig. 1 Left) Camera images. Right) Depth-map with gaps.

The method is aimed to be used in real-time, thus we avoid using polygonal representation of the 3D scene due to the complexity, content size, and transmission. The calculations are done on the local side, and only a surface map (or normal-map) is generated from the depth-map. Using this representation it is possible to generated not only pushing forces, but also shearing forces along finger's surface.

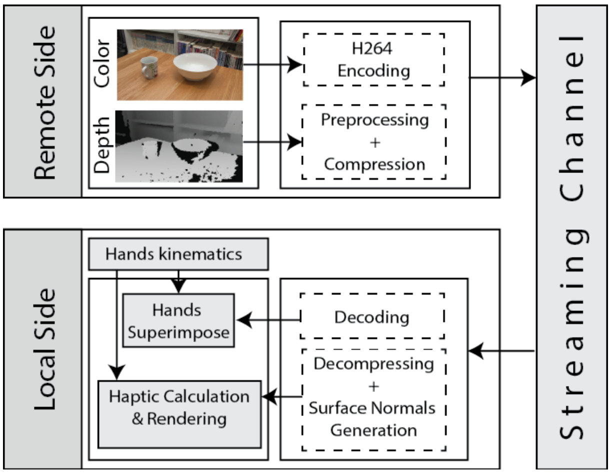

Fig. 2 System data flow between local and remote sides.

The system is divided into two parts: Remote side in which the depth and video cameras are being at and local side, the place that the user is being. Data flow between both sides is shown in Fig. 2. In the following section, more detailed explanation of each step is described.

3. Implementation

3.1 Robot head design

Head design should provide matching between camera images with depth-map. In Fig. 3 a three degrees of freedom head was designed with stereo cameras and a depth sensor.

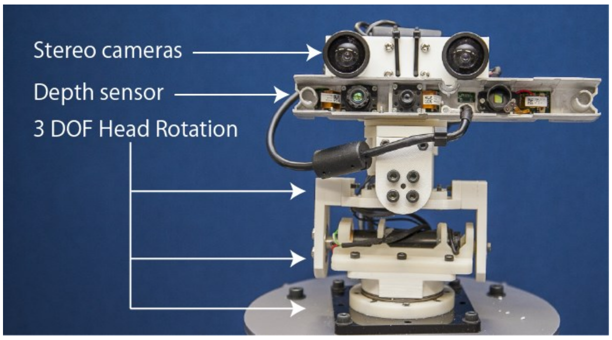

Fig. 3 Remote head design

The viewpoint of the depth sensor matches right eye's horizontal offset. However because the design is not visual conjugated between the depth sensor and the camera, a software offsetting and scaling step for the depth-map is required to correct the mismatched positions and the different field of views.

3.2 Depth preprocessing and surface generation

As mentioned before, using raw depth data from the sensor will produce several issues regarding resulted artifacts and lost information from the scene. Thus it is necessary to process and clean up the data before using it to generate scene surface. To do so, first we propose a filling strategy for the missing depth values, then we apply smoothing filter on the resulted depth-map to remove the presented noise while preserving the hard edges of scene objects.

To recover the gaps of the depth-map, a statistical search model is applied for each missing pixel. The model creates a sliding window and picks the value highest frequent depth in this window. Fig. 4 shows the results of applying this model. However it couldn't fill the entire gaps due to the size of the gaps. To resolve this, temporal methods recover the missing data from previous frames can be used [10].

The processed depth map is then segmented and sent into the operator's side. Since depth-maps are 11-bits, image based lossless compression algorithms will not be applied. Thus it is compressed using ZIP algorithm before sending it.

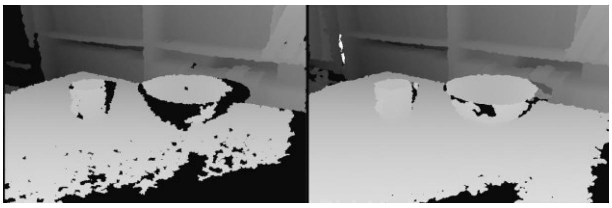

Fig. 4 Recovering missing depth values.
Left) Original depth-map . Right) Recovered map.

Scene surface normal-map is derived from depth-map values. Each pixel of the normal-map is a 3D vector representing surface direction at that point. Normal vectors are calculating using the cross product of the derivatives along X and Y space of the depth values:

$$\overrightarrow{N_{x,y}} = \|\overrightarrow{\partial U} \times \overrightarrow{\partial V}\|$$
$$\overrightarrow{\partial U} = [\partial x, Depth, y]$$
$$\overrightarrow{\partial V} = [x, Depth, \partial y]$$

$\overrightarrow{\partial U}$ and $\overrightarrow{\partial V}$ are derivative vectors for X and Y axis respectively.

However, the noise from the depth sensor become obvious on the normal-map, thus a smoothing filter is applied before calculating the normal. Applying a Gaussian filter on the depth values which works as a low pass filter in the image space. However it doesn't take into consideration the hard edges of the objects in the image. Thus a bilateral filter is used [11] and applied on the depth-map.

Fig. 5 shows the generated normal-map before and after applying the bilateral filter on the depth-map.

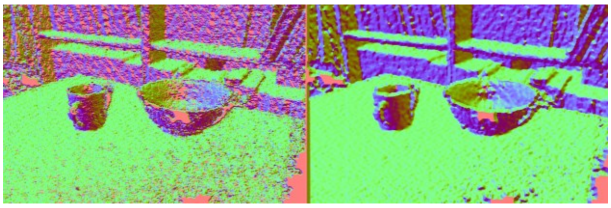

Fig. 5 Generated normal-map. Left) Original noise with no smoothing. Right) Noise reduction using bilateral filter.

3.3 Calculating interaction forces

To calculate the touch forces with the environment, user's finger position should be determined first. To do so, operator body posture is captured via OptiTrack tracking points mounted on the head, shoulders and hands. The kinematics of arms trajectory is calculated using TELESAR V inverse kinematics [3]. Also a 5DT data glove is used to provide bending measurements of each finger. With these information it is possible to determine finger's tips location with respect of user's head and direction in the 3D space. Also these trajectories will be used in the next step for superimposing virtual arms.

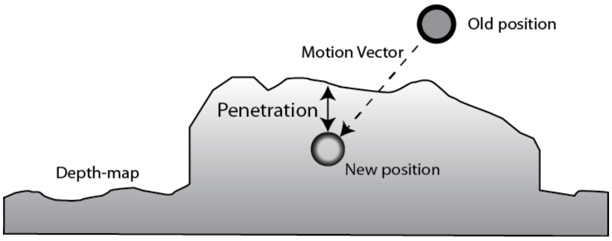

Fig. 6 Touch point intersection with depth-map.

Fig. 6 shows the case of touch with the depth-map. Touch forces are separated into two components: pressure and shearing forces. Pressure is calculated using Hooke's law:

$$F_z = -kx$$

F_z is the resulted force along z axis (pressure axis).

x represents the amount of penetration in depth-map.

To calculate shearing forces, "motion vector" of finger tips is used and projected on the depth-map when penetration happens. This will provide an amount of friction which represents the shearing value along (X, Y) axis.

To render the haptic on finger tips, Gravity Grabber is used [12] which can provide pressure force and one shearing direction force along the finger tip.

3.4 Hands superimposing

The previously calculated and measured trajectories of the body and fingers are provided into Virtual Telesar simulation platform [6] that in turn apply those kinematics into a body model and super impose it into the remote camera images. Fig. 7 shows the result of superimposing the virtual arms into the remote environment, also it shows the mapping of the normal-map and the calculated touch forces between finger tips and the normal-map.

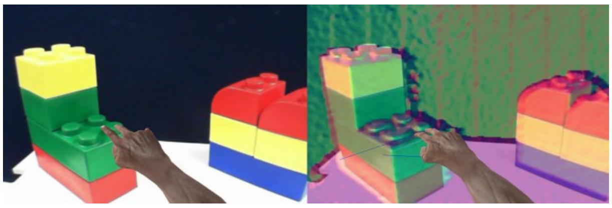

Fig. 7 First point of view interaction using super imposed virtual arms. Left) Interaction with remote scene. Right) Normal-map and touch vectors.

4. Conclusion

In this paper, we presented a real-time haptic interaction system for non-armed telexistence robots. In this method, the user perceive a superimposed virtual representation of his arms, hands and fingers that represents his own body while being operating in a different place. By using depth-map images of the remote environment, scene surfaces are calculated and normal-maps are generated which help to determine when touching occurs with user's fingers. Finally, using normal-maps, force vectors of user's fingers are calculated and passed into mounted haptic displays on user's fingers tips which provide the haptic feedback to the user.

Acknowledgement

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

References

[1] Tachi, S. "Telexistence". World Scientific Publishing Co., Inc.., 2009.

[2] Watanabe, K., Kawabuchi, I., Kawakami, N., Maeda, T., & Tachi, S. "TORSO: Development of a telexistence visual system using a 6-dof robot head". Advanced Robotics, 22(10), 1053-1073, 2008.

[3] Fernando, C.L.; Furukawa, M.; Kurogi, T.; Kamuro, S.; Sato, K.; Minamizawa, K.; Tachi, S., "Design of TELESAR V for transferring bodily consciousness in telexistence," IROS, pp.5112-5118, 2012.

[4] Wu, W. C., Basdogan, C., & Srinivasan, M. A. "Visual, haptic, and bimodal perception of size and stiffness in virtual environments". ASME DYN SYST CONTROL DIV PUBL DSC., 67, 19-26, 1999.

[5] Heller, M. A. "Visual and tactual texture perception: Intersensory cooperation". Perception & psychophysics, 31(4), 339-344, 1982.

[6] Saraiji, M. Y., Fernando, C. L., Furukawa, M., Minamizawa, K., & Tachi, S. "Virtual Telesar-designing and implementation of a modular based immersive virtual telexistence platform". In System Integration (SII) IEEE/SICE International Symposium on (pp. 595-598). IEEE, 2012.

[7] Saraiji, M. Y., Fernando, C. L., Furukawa, M., Minamizawa, K., & Tachi, S. "Haptic Layer Design and Integration for Virtual Telesar". On Robotics and Mechatronics (Robomec 2013), 2013.

[8] Saraiji, M. Y., Fernando, C. L., Furukawa, M., Minarnizawa, K., & Tachi, S. "Real-time egocentric superimposition of operator's own body on telexistence avatar in virtual environment". In Artificial Reality and Telexistence (ICAT), 2013 23rd International Conference on (pp. 35-39). IEEE, 2013.

[9] Newcombe, R. A., Davison, A. J., Izadi, S., Kohli, P., Hilliges, O., Shotton, J., & Fitzgibbon, A. "KinectFusion: Real-time dense surface mapping and tracking". In Mixed and augmented reality (ISMAR), 10th IEEE international symposium on (pp. 127-136). IEEE, 2011.

[10] Camplani, M., & Salgado, L. "Efficient spatio-temporal hole filling strategy for kinect depth-map s". In Proceedings of SPIE (Vol. 8920), 2012.

[11] Tomasi, C., & Manduchi, R. "Bilateral filtering for gray and color images". In Computer Vision, 1998. Sixth International Conference on (pp. 839-846). IEEE, 1998.

[12] Minamizawa, K., Fukamachi, S., Kajimoto, H., Kawakami, N., Tachi, S,. Gravity grabber: wearable haptic display to present virtual mass sensation, ACM SIGGRAPH 2007 emerging technologies, Article 8, 1 page, 2007.