# AUGMENTED REALITY SYSTEM FOR REMOTE OPERATION

**Young Soo Park*, Jia Luo, Pawel Dworzanski**
Argonne National Laboratory
9700 South Cass Ave., Lemont, IL 60432
ypark@anl.gov

**Joohee Kim**
Illinois Institute of Technology
3301 S. Dearborn St., Chicago. IL 60616
joohee@ece.iit.edu

**Byung Seon Choi**
Korea Atomic Energy Research Institute
111 Daedeok-daero 989 beon-gil, Deogjin-dong, Yuseong-gu, Daejeon, South Korea
bschoi@kaeri.re.kr

## ABSTRACT

This paper presents a new approach to applications of augmented reality techniques for performing telerobotic operations in uncertain and hazardous environments. Particular application examples include remote or partially-automated operation of robotic manipulators in nuclear waste facilities, decommissioning sites, or as a part of incident remediation. Recently advances in virtual reality (VR) and augmented reality (AR) technology presents opportunity for enhancing the performance of such remote operations. VR can provide immersive artificial environment for the operator to interact with. As an advanced concept, AR adds to it the blending or artificial contents with real world - users are able to interact with virtual contents in the real world. This work addresses key technological innovations necessary for adopting the VR/AR technologies for the remote operation of hardware robotic systems in real world tasks. To this end, this work focuses on implementation of 'virtual fixtures', which allows intuitive and precise guidance of remote operation, as well as key enhancements in 3D sensing/reconstruction technology. In addition, the implementation details of the integration of the multi-modal augmented reality system to the telerobotic control system are presented.

*Key Words*: augmented reality, remote operation, 3D sensing and reconstruction

## 1 INTRODUCTION

This paper presents demonstrated applications of augmented reality techniques for performing telerobotic operations in uncertain and hazardous environments, such as nuclear reactors, radioactive waste facilities, hazardous facility decommissioning sites, or as a part of incident remediation. The operations involving robotic manipulators are expected to be performed via remote operation, possibly with partially automated procedures. A key to success in such operations is accurate telepresence - conveying realistic experience of being present in the remote task environment.

Recently significant advances in virtual reality (VR) and augmented reality (AR) technology have been made, with successful applications in data presentation, computer games, and virtual simulation and training. VR can provide immersive artificial environment for the operator to interact with. As an advanced concept, AR adds to it the blending or artificial contents with real world - users are able to interact with virtual contents in the real world. There is a significant potential that such a technology basis can be leveraged to provide effective tools for enhancing the performance of remote operations.

While the potential benefits are great, there are also potential drawbacks. The current VR and AR technologies, whose applications mainly remain in game industry, lack the reliability and accuracy required for robotic manipulation. It has been known that inaccurately presented VR and AR may confuse the operator. Also, in a guided remote operation of a real-world equipment, a small discrepancy between and virtual and real environment may cause serious contact force which may cause damage to the equipment and environment. To this end, this proposal presents a series of innovations which makes VR an AR applicable for application for contact manipulation applications. Such innovations include accurate 3D sensing, reconstruction and dynamic tracking, multiple sensor calibration and coordinate mapping, multi-modal perceptual feedback, as well as robot control system integration.

In this regard, an augmented reality teleoperation testbed is implemented, which include a two-armed robot system for demonstration of physical remote operation, the virtual environment for enhanced telepresence, and sensor-based augmented reality to improve precision and efficiency of in-situ remote operations. Test augmented remote operations are performed for a few simulation cases: remote facility decontamination and decommissioning (D&D), incident remediation, nuclear waste management, and inspection and maintenance of a nuclear facility. The software and hardware integration is performed under a Linux-based distributed open architecture, namely Robot Operating System (ROS).

# 2 SYSTEM OVERVIEW

To develop and demonstrate the utility of augmented-reality technologies for applications in enhancement of teleoperation performance, augmented teleoperation system is composed. Fig. 1 conceptually illustrates the overall composition of the augmented teleoperation system, which incorporates systems of real-time 3D reconstruction, multi-modal augmented reality, and teleoperation testbed.

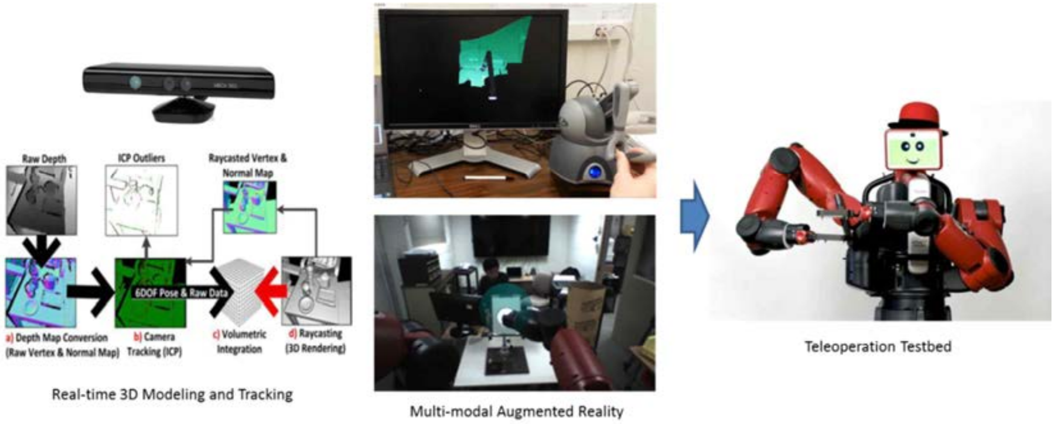


**Figure 1. Concept of the Augmented Teleoperation System**

## 2.1 Real-time 3D Reconstruction System

Nowadays, 3D sensing and reconstruction technology has been established, whose objective is to recreate a real scene, as accurate as possible, within a virtual three-dimensional space using a computer. Recently, low-cost depth cameras such as the Microsoft Kinect [1] have been introduced, which are affordable and capable of generating depth maps at real-time rates. Several software options were available to communicate with the Kinect, for example Microsoft's official Kinect Fusion SDK, or KinFu. KinFu uses OpenNI (Open Natural Interaction) drivers instead of Microsoft's drivers, and it uses PCL (Point Cloud Library) to store and manipulate its data. Our 3D reconstruction is based on KinFu API for 3D reconstruction. Fig. 2 illustrates the processing pipeline of the 3D reconstruction:

1) Depth Map Conversion: It converts the raw depth data captured by a Kinect sensor to 3D points (i.e., vertices and surface normals).

2) Camera Tracking: This stage calculates the location/orientation of current camera and it tracks the pose in every frame by using the ICP algorithm. The ICP algorithm is widely used for geometric alignment of 3D models when an initial estimate of the relative pose is known.

3) Volumetric Fusion: This step is used to convert the depth data into volumetric surface representation rather than point clouds. It processes each frame with a running-average in order to reduce noise.

4) Raycasting: The last stage is to raycast the volume to extract the surface. When using the global pose of the camera, the raycasted view can be considered as a synthetic depth map. Therefore, it is less noisy and better to be used as a reference frame for pose estimation.

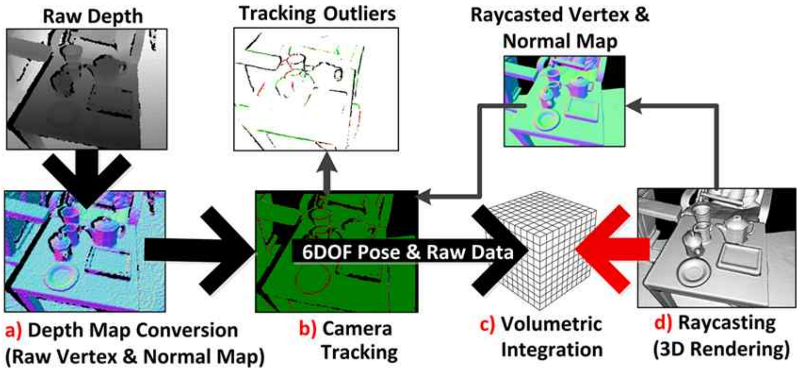
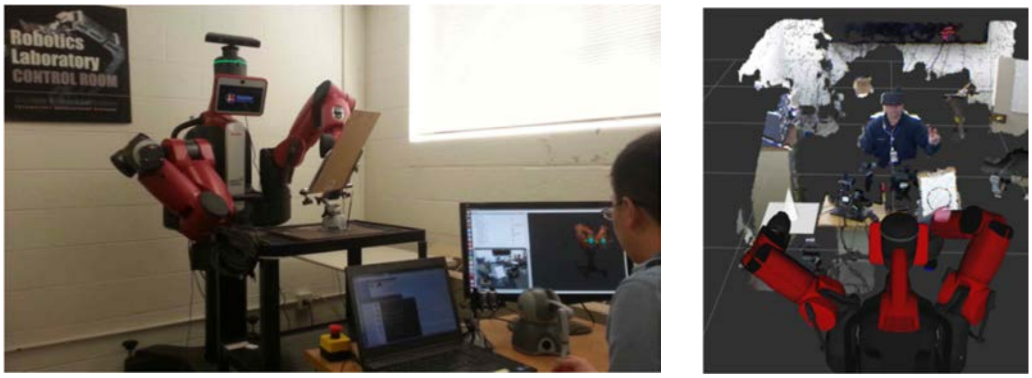


**Figure 2. Illustration of 3D Reconstruction Processing Pipeline**

As a result, a 3D geometric model, a triangle mesh, of the environment is generated. We have developed the capability to import and integrate the mesh into a visual-haptic virtual reality environment.

## 2.2 Teleoperation Testbed

A visual-haptic augmented teleoperation test bed has been constructed (Fig. 3). In this system, a two-arm robot is used as the slave robot, which is remotely controlled by the human operator with a haptic device. A 3D sensor fixed on top of the robot is used to capture the 3D scene in front of the robot. The 3D geometry and well as the camera image of the environment is displayed on an immersive VR headset.



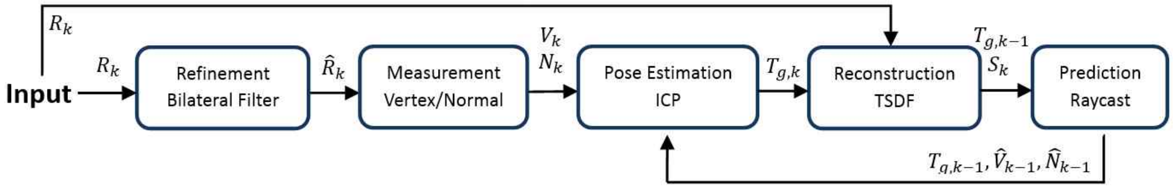
(a) Teleoperation system (b) VR display in RViz

**Figure 3. Augmented Teleoperation Testbed**

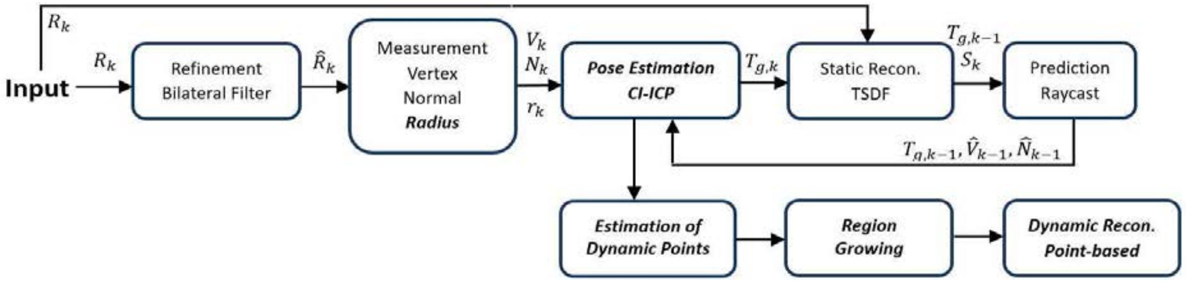
A tele-operation software system is developed using the Robot Operating System (ROS) [2] and the RViz visualization environment [3] for operation of the whole system. ROS allows various utilities for robot control, and RViz allows visual display of sensor information and 3D task environment.

## 3 ENHANCEMENT OF 3D MODEL RECONSTRUCTION

Although the recent advances have made real-time 3D reconstruction possible, the state-of-the-art has some limitations that they are subject to pose estimation error, and are only capable of tracking relatively static objects. Thus modifications are made on the processing pipeline to overcome such limitations. Fig. 4 shows the original processing pipeline of KinFu, explained in section 2-1, and the enhanced pipeline.



(a) Original Processing Pipeline of KinectFusion/KinFu.



(b) Enhanced Processing Pipeline

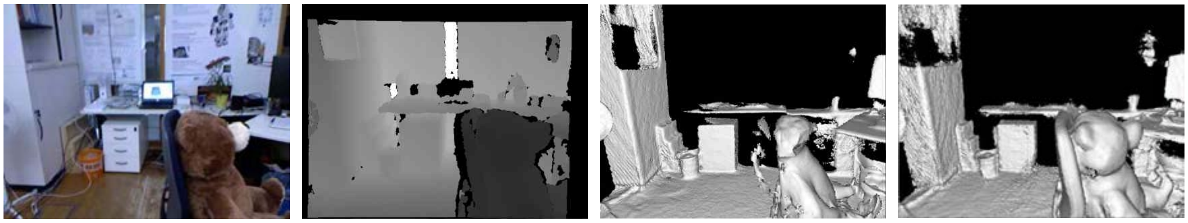
**Figure 4. Pipeline for Enhanced 3D Reconstruction**

### 3.1 Enhancement of Static Reconstruction

Pose estimation is at the heart of 3D reconstruction. The 3D sensors have been shown to have systematic errors in the depth measurement, which are well studied and modelled for Kinect sensors [4]. KinFu uses truncated signed distance function (TSDF) for 3D reconstruction, whereas the systematic error model is incorporated in the TSDF [5]. However, the systematic error models have never been incorporated into the pose estimation. Therefore, we have developed a method to incorporate the systematic error models into pose estimation process in the form of confidence indicator for each depth value.

The raw depth map from the depth sensor is bound to be noisy and usually contain holes. The multi-resolution anisotropic diffusion based depth refinement algorithm proposed in [6] uses both RGB and depth information to achieve hole-filling method in real time. This hole-filling approach was primarily used as depth map refinement stage for view synthesis. The depth value created for the holes have errors proportional to the nearest measured depth value. We designed a method to model this error as a confidence indicator (CI) for each depth value and use it in pose estimation process. Subsequently the Cis are used to weigh both in the pose estimation and reconstruction stages.

We present quantitative and subjective quality results demonstrating the advantages of using confidence indicator for each depth value from systematic errors and filtering errors in pose estimation and 3D reconstruction stage. Fig. 5 illustrates comparison of the 3D reconstruction results. It has been found that the confidence indicator based methods give 59.60% better pose estimate than baseline algorithm in quantitative measures and achieves a significant subjective quality improvement in 3D reconstruction.



(a) (b) (c) (d)

**Figure 5. Comparison of 3D reconstruction of "Teddy" at the 400th frame:** (a) RGB image, (b) depth image, (c) from original KinFu, (d) using WICP with WTSDF.

## 3.2 Dynamic Reconstruction

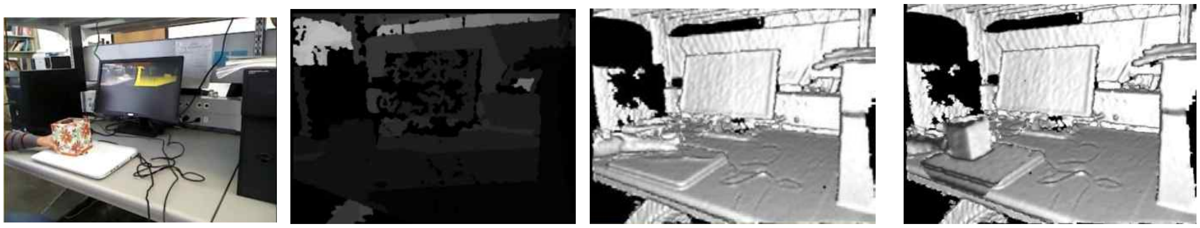
Another drawback of the KinFu pipeline is that if a relatively fast moving object is introduced in the scene, data association fails and there occur tracking errors in the camera pose estimation. To overcome this limitation, we have enhanced the 3D reconstruction for dynamic scene.

The enhanced method is based on the point-based method for dynamic tracking [7]. The approach discussed in [7] is very much similar to that of the conventional 3D reconstruction methods discussed earlier. 1) The data from the depth sensor is pre-processed; 2) the current 6 degree-of-freedom (6DoF) pose of sensor relative to the scene is estimated; and 3) the estimated pose is used to convert depth samples into a unified coordinate space and fuse them into an accumulated global model. The only difference is that this method uses point-based representation throughout the reconstruction process. The point-based fusion method works without the overhead of converting between representations. Along with that, it introduces the use of radius map [8] and confidence counter in the ICP algorithm to detect the dynamic candidates from the non-corresponding points. These dynamic candidates are then segmented using a hierarchical region growing method and used for scene reconstruction. It further uses the surface splatting method [9] for the final 3D reconstruction of the scene.

We have further improved the above point-based reconstruction method to enhance accuracy, reliability and speed. In this regard, the following enhancements are made:

- extended the radius map and confidence indicator to three dimensions,
- combined the point-based fusion and volumetric representation to maintain the level of the reconstruction quality, while reducing the processing time and memory usage,
- added the normal map as a similarity attribute to improve the region growing method,
- implemented it on the GPU to maintain its processing speed,
- the dynamic parts are added into the global model in real-time.

The new algorithm was tested on a reference video set in comparison with the baseline method. The new approach was three times faster, 52% better in absolute tracking accuracy, and resulted in better dynamic tracking as illustrated in Fig. 6.



(a) (b) (c) (d)

**Figure 6. Comparison between the original KinFu and our proposed method for moving_box_1 scene.** (a) RGB map, (b) depth map, (c) results of KinFu and (d) results of the proposed method.

# 4 ENHANCED TELEOPERATION WITH AUGMENTED VIRTUAL FIXTURES

## 4.1 Virtual Fixtures

This section presents implementation of an augmented teleoperation system, which is an enhanced teleoperation concept via use of virtual fixtures. The concept of virtual fixtures - an artificially generated geometric surface overlaid on human perception - was first introduced to enhance teleoperation performance [10, 11]. Such guidance is expected to reduce the operator's mental burden, facilitate precision motion, and improve stability. Also since it is based on local sensory feedback, it is applicable to operation of simple slave robots, i.e. those not requiring complex bilateral system. Fig. 7 conceptually illustrates the architecture of such an augmented reality system.

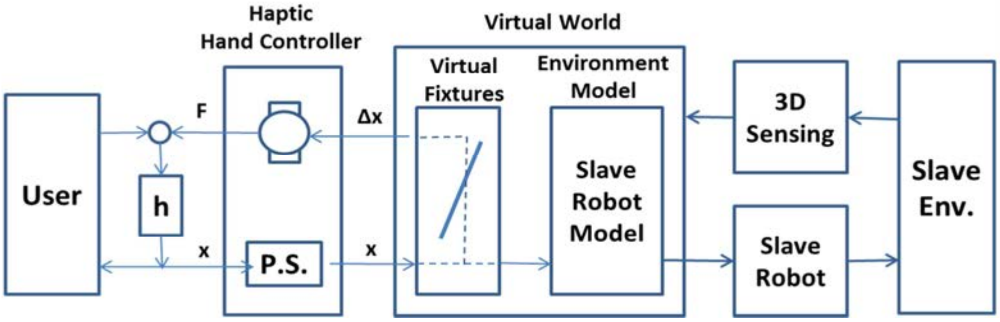


**Figure 7. Architecture of the Augmented Teleoperation System**

The enhanced teleoperation method is implemented on the teleoperation test bed introduced in the previous section. Fig. 8 illustrates the overview of the enhanced teleoperation process.

- A live image captured by the Baxter head camera is displayed to the controller (Fig. 8(a)).
- A virtual fixture, a conical frustum, is introduced to the testbed. The software then automatically detects the circle, and defines the pose and top radius of the virtual fixture (Fig. 8(b)).
- This virtual fixture is a haptic-enabled shape so that it is touchable by the haptic stylus. The haptic cursor stays on top of the surface when in contact with the virtual fixture (Fig. 8(c)).
- After the virtual fixture is placed, the haptic device starts guiding the movement of the Baxter arm. The position of the haptic cursor is tracked by the robotic arm (Fig. 8(d)). As the haptic cursor is constrained by the virtual surface, the movement of robot arm is also bounded by the shape.

The implementation of the above virtual fixture is described in the subsequent sections.


(a) (b) (c) (d)

**Figure 8. Implementation of virtual fixture for guiding in telerobotic operation.** (a) Live image captured by Kinect camera, (b) Virtual fixture placed, (c) Haptic stylus interacting with virtual fixture, (d) Robot arm guided by virtual fixture

## 4.2 Graphics Rendering

The graphics rendering module of the test bed software accomplishes the goals including (i) displaying the actual robot workspace, (ii) rendering a virtual cursor for the haptic stylus, reflecting its position and orientation, and (iii) rendering a set of artificial geometries as virtual fixtures to assist teleoperation. The software uses OpenGL and OpenCV for graphics rendering.

The live image captured by Baxter head camera shows the workspace in front of the robot. This image can be obtained by subscribing the ROS topic “/cameras/head_camera/image”. Referring to the pinhole camera model, a camera projection matrix is used to denote a projective mapping from world coordinates to pixel coordinates. It is derived from the intrinsic and extrinsic parameters of the camera. Fig. 9(a) shows the principle of the camera projection matrix.

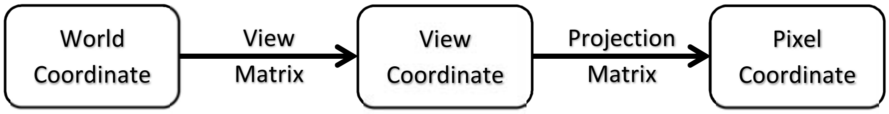
To achieve augmented reality, a virtual scene is established and displayed on top of the actual camera image. It should overlap with the actual scene with sufficient accuracy so that the Baxter arm, which is clearly in the actual scene, can interact with the virtual fixture, which is included in the virtual scene. OpenGL provides a series of matrix stacks to define the mapping from the virtual scene to the display interface, as illustrated in Fig. 9(b).

By comparing Fig. 9(a) with Fig. 9(b), it’s clear that both the virtual and actual scenes will share pixel coordinates, and in turn be merged on display, if

(i) the world coordinates of the virtual scene are the same as that of the camera,
(ii) View matrix equals to camera extrinsic matrix, and
(iii) Projection matrix coincides with camera intrinsic matrix.

The view coordinates of the virtual scene and camera are matched according to the above conditions.

(a) Camera Projection Matrix



(b) Graphics Coordinates Transformation

**Figure 9. View Perspective Transformations for Graphic Rendering**

## 4.3 Haptic Rendering

In the teleoperation test bed, a haptic device is used by the operator to interact with the virtual scene. The haptics rendering module of the software is responsible of (i) map the haptic device to the actual scene, (ii) place the virtual fixture in the scene, and (iii) guide the movement of the robotic arm. The software uses OpenHaptics SDK for haptics rendering.

The haptic workspace is the physical space reachable by the haptic device. In the teleoperation test bed, the haptic workspace should be mapped to the intersection of the range of motion of the robot arm and the field of view of the head camera. In this way, the virtual fixture placed by the haptic device is always visible and reachable. To ensure the haptic feedback reflects the shape of the touchable objects, the mapping should be uniform, i.e. the scale factor should be the same for all dimensions.

OpenHaptics provides a series of matrix stacks to convert the coordinates of haptic workspace to world coordinates, as **Error! Reference source not found.** shown in Fig. 10. World-View matrix defines the transformation to the camera coordinate frame. It is obtained from OpenGL. View-Touch matrix defines the rotation and translation of the haptic workspace relative to view coordinates independent of the workspace mapping. Touch-Workspace matrix defines the mapping of the workspace to view coordinates. The mapping contains a scale to map the workspace and a translation to orient the workspace to the target mapping in view coordinates.

The haptic interaction between the haptic cursor and the virtual fixture is achieved by the proxy-based method implemented in OpenHaptics.

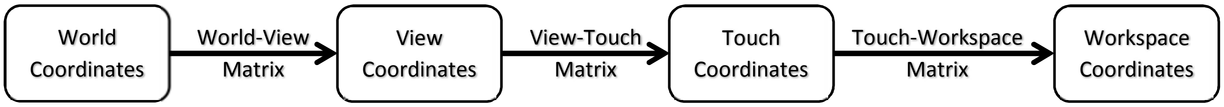


**Figure 10. Haptics Coordinates Transformation**

## 4.4 3D Camera Extrinsic Calibration

The Baxter ROS API by the Rethink Robotics Co. provides the kinematic model of the Baxter robot defined in Unified Robot Description Format (URDF). In our robotic testbed, a Microsoft Kinect is fixed on top of the Baxter robot to capture the 3D scenario in front of the robot. To define the robot motion based on Kinect camera perception, it is necessary to identify the kinematic coordinate transformation between the robot's base coordinate frame and the Kinect camera coordinate frame. This can be obtained by performing extrinsic camera calibration, thus the Kinect-Baxter calibration. One of the robot's hand cameras is used for the calibration.

Calibration is performed using an open source ROS package, baxter_h2r_package, and using a calibration marker.  Fig. 11 some screenshots during the calibration process.

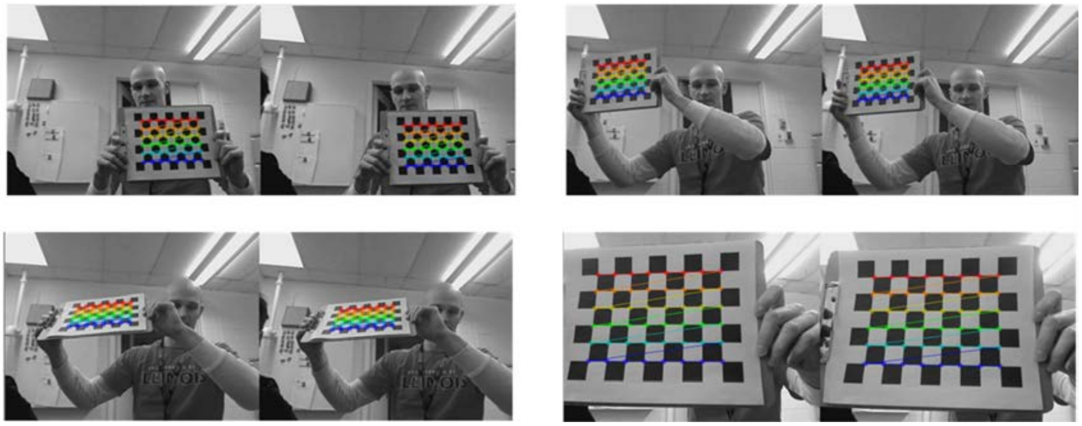


**Figure 11. Screenshots of Calibration Process**

## 4.5 Placement of the Virtual Fixture

Accurate placement of the virtual fixtures is important for the performance of the teleoperation. Based on the 3D sensing, the location of the virtual fixture automatically determined. This process consists of 1) detecting the target geometry, i.e. circle, on the board in front of the robot, 2) identifying the pose of the target geometry, and 3) overlaying the virtual fixture on top of the actual target. This process is illustrated in Fig. 12

(a) original image (b) compute normal (c) shape detection (d) place virtual fixture

**Figure 12. Placement of Virtual Fixture**

## 4.6 Robot Teleoperation

All of the virtual fixture implementation features are integrated into a ROS package, namely *teleop_baxter*. Upon execution of the main program, *main_interface*, the robot operation starts. The program leads through an operation cycle of selecting the target region, identifying the target geometry and location, placing the virtual fixture, and allowing operation of the robot. In addition to the main interface, another ROS node, *teleop*, is running in the background to support the teleoperation of the Baxter robot. It communicates with the robot to get current state information (positions of the joints, etc.) and sensor data, and talks to the controllers on the robot. When the virtual fixture is placed, the position of the haptic cursor will be sent to the *teleop* and trajectory motion is performed.

Fig. 13 shows a snapshot of the test operation with virtual fixture. The task was to teleoperated the robot arm to draw a circle following a circular pattern on a flat panel. During the test operation, the position and force at the robot hand was recorded, which can be obtained from the Baxter ROS package nodes. Fig. 14(a) shows the path of the robot hand, which approximately followed the circular pattern. It was particularly effective in maintaining the depth of the motion on the flat panel. Fig. 14(b) shows the magnitude of the aggregated contact force during the operation. It is noted that the contact force remained at relatively constant level during the kinesthetic interaction of following the circle on the panel.

**Figure 13. Teleoperation Test Setup with a Circular Pattern**

**Figure 14. Plot of the Contact Force at the Robot Hand during the Test Operation.** (a) Positon, (b) Force

## 5 CONCLUSIONS

As an effort to fill the key technology gap required for practical remote system deployment for complex tasks in nuclear applications, a new concept of augmented remote operation method is introduced. This concept aims at augmenting virtual reality-based operator aid to simplify the remote operation task and improve task precision. The development has focus on enhancement of 3D sensing and reconstruction technology, and implementation of a telerobotic testbed based on an open source technology basis. The concept of virtual fixtures has been implemented in the ROS based test bed robot environment. The integrated technology will allow teleoperation of dexterous manipulation tasks using simple and rugged robot system suitable for D&D operations.

## 6 ACKNOWLEDGMENTS

This work was supported by Korea Atomic Energy Research Institute as a part of a larger program by Ministry of Science, ICT, and Future Planning, South Korea.

## 7 REFERENCES

1. J. Smisek, M. Jancosek, and T. Pajdla, “3D with Kinect”, in Proc. IEEE Intl. Conf. on Comp. Vision Workshops pages 1154-1160, Nov. 2011.
2. http://wiki.ros.org
3. http://wiki.ros.org/rviz
4. A. Segal, D. Haehnel, and S. Thrun, “Generalized-ICP,” In Proceedings of Robotics: Science and Systems, Seattle, USA, 2009
5. C. Kerl, J. Sturm, and D. Cremers, “Robust Odometry Estimation for RGB-D Cameras,” In Proc. IEEE Int’l Conf. Robotics and Automation (ICRA), May 2013
6. J. Sturm, N. Engelhard, F. Endres, W. Burgard, and D. Cremers, “A benchmark for the evaluation of RGB-D SLAM systems,” in Proc. Int. Conf. Intell. Robot Syst., 2012, pp. 573–580.
7. M. Keller, D. Lefloch, M. Lambers, S. Izadi, T. Weyrich and A. Kolb. “Real-time 3D Reconstruction in Dynamic Scenes using Point-based Fusion,” In 2013 International Conference on 3D Vision (3DV), pp.1-8, 2013.
8. T. Weise, T. Wismer, B. Leibe and L. Van Gool. “In-hand scanning with online loop closure,” in Proc. IEEE International Conference Computer Vision Workshops, pp. 1630-1637, 2009.
9. M. Zwicker, H. Pfister., J. V. Baar, and M. Gross. “Surface splatting,” in Procs. SIGGRAPH, pp. 371-378, 2001.
10. L.B. Rosenberg, “Virtual fixtures: Perceptual tools for telerobotic manipulation,” Ph.D. Thesis, Dept. of Mechanical Engineering, Stanford University (1994)
11. Y. S. Park, H. Kang, T. F. Ewing, et al, "Enhanced Teleoperation for D&D," Proc. IEEE ICRA (2004)