
Proceedings Paper:
Jin, Y, Paredes Soto, DA, Rossiter, J orcid.org/0000-0002-1336-0633 et al. (1 more author) (2021) Advanced environment modelling for remote teleoperation to improve operator experience. In: icARTi '21: Proceedings of the International Conference on Artificial Intelligence and its Applications. 2021 International Conference on Artificial Intelligence and its Applications (ICARTI 2021), 09-10 Dec 2021, Bagatelle, Mauritius (Virtual). Association for Computing Machinery . ISBN 9781450385756

https://doi.org/10.1145/3487923.3487939

Advanced Environment Modelling for Remote Teleoperation to Improve Operator Experience

Yixiang Jin
The University of Sheffield
Sheffield, UK
yjin30@sheffield.ac.uk

Daniel Alonso Paredes Soto
The University of Sheffield
Sheffield, UK
d.paredes-soto@sheffield.ac.uk

John Anthony Rossiter
The University of Sheffield
Sheffield, UK
j.a.rossiter.ac.uk

Sandor M Veres
The University of Sheffield
Sheffield, UK
s.veres@sheffield.ac.uk

Abstract

This work presents a novel intelligent robot perception system, including a real-time, high-quality, 3D scanning pipeline for texture-less scenes and a human-supervised grasping system. Comparison is carried out with the state of the art 3D reconstruction systems, and the performance of the proposed system is demonstrated. The scanning methods are applied to a new user interface with object 6D-pose estimation. This work supports human-robot interaction in remote handling operations in hazardous environments by providing a high-quality telepresence. Current teleoperation systems primarily utilise 2D images or point clouds to display the remote workspace to the operator. Operators require extensive training to be able to perceive the spatial relationship between the robot and the target objects by remotely looking at multiple 2D images. Therefore, this paper proposes a new teleoperation system that exploits artificial intelligence to improve the efficiency of operators. The experiments show that the proposed method surpasses state-of-the-art reconstruction systems and successfully complements a simulated nuclear waste handling experiment.

Keywords: human-robot interaction, semi-autonomous system, 3D environment reconstruction, 3D object detection

ACM Reference Format:
Yixiang Jin, Daniel Alonso Paredes Soto, John Anthony Rossiter, and Sandor M Veres. 2021. Advanced Environment Modelling for Remote Teleoperation to Improve Operator Experience. In *Proceedings of ICARTI2021: 2021 The International Conference on Artificial Intelligence and its Applications (ICARTI2021).* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

ICARTI2021, December 09–10, 2021, Bagatelle, Mauritius

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 Introduction

Recently, there has been an increasing number of teleoperated robotic systems deployed in unstructured and hazardous environments, such as the nuclear industry, search and rescue, manufacturing and the international space station, etc. [24]. However, current teleoperations are mostly based on multi-screen displays [15, 19], which means operators need to view the multiple video streams cameras placed in the remote environment and also monitor the robot's state in physical space at any point in time. As a result, operators have to shift their concentration between monitoring the remote camera's video feed and monitoring the robot states constantly, which may lead to fatigue and stress for operators. Apart from this, operator's 3D perception of the remote environment during teleoperation depends on a 2D projection of a 3D environment. This process can be time-consuming, inefficient and less productive.

An advanced 3D visualisation system of the remote environment requires explicit 3D reconstruction scenes. However, the current environment reconstruction technologies cannot meet the demand, especially for the small-scale texture-less scenes. Offline methods often need hours to process, while online approaches have a severe weakness which is relying on object surface texture to compute corresponding relations between different frames.

What is more, an excellent teleoperation system should help operators improving their work efficiency. The majority of state-of-art systems still depend on human's intelligence in excess. With the development of 6D pose estimation technology, an increasing number of works can achieve semantic grasping in the clutter environment. Despite this fact, vision algorithms for 6D pose estimation in teleoperation system is still a challenge. Template-based algorithms (e.g., LineMOD [10], PWP3D[23]) depend on the intact and high-quality 3D object model to extract texture features of the object and compare with scenes. In current industrial applications, all

objects manipulated by robots are made of similar material and cannot easily be differentiated by features such as texture or colour. Deep learning based-methods (e.g., DOPE[28]) can overcome this slightly, but those systems inevitably require long training times and enormous computing resources.

To bridge such contradictory technical challenges, we propose a novel teleoperation system that can significantly improve the operator's experience. A robot can first scan the surrounding environment and generate a high-quality 3D realistic model in this system. Second, the object lists are visualised based on the detection history and operator's guidance from the visual recognition system. Finally, operators can select the target object which needs to be grasped by robots. Through a proposed graphical user interface, the operator can interact with the environment easily.

The main contributions of this paper are listed as follows:

- Proposes a novel 3D visualisation system that presents a high-quality model of a robot's work-space to the operator, thus improving their scene understanding.
- Develops a novel human-robot interface that can significantly reduce the work pressure on the operator.

Note that most teleoperation in the past required visualisation using 2D interfaces [7, 14, 22], 3D point clouds [17] or object surface meshes [29].

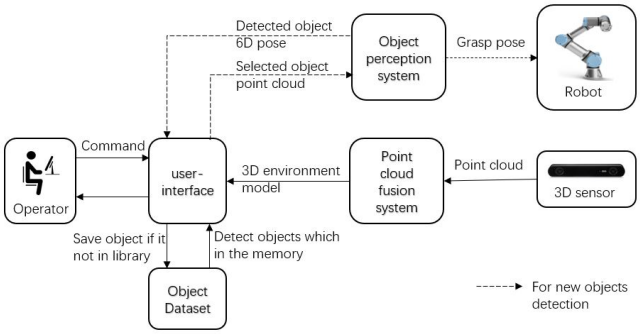
The remainder of this paper is organised as follows: Section 2 gives an overview of the related literature; Section 3 details the pipeline of the 3D environment reconstruction system. Section 4 introduces the human-supervised semi-autonomy system. Section 5 presents the experiments and results. Finally, conclusions are drawn in Section 6.

2 Related Work

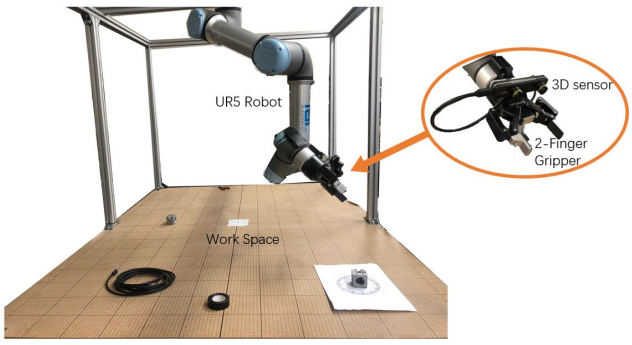
Recently, there has been an increasing amount of research on solving the problem of perceiving 3D task environments in a 2D display. For example, Lipton [14] proposed an innovative homunculus model that created a Virtual Reality Control Room in a head monitor and embedded the operator in the centre. Yew [30] presented a mixed-reality-based maintenance robot by a cyber-physical model, which used a virtual object overlay in the actual target to guide the operator for tasks such as disassembly, cleaning, and repair work. What's more, those improvements can help operators complete a few well-defined tasks more easily but were limited in 3D position perception.

In the field of telepresence vision systems, Kohn *et al.* [12] proposed model-based teleoperation, which can render the figure captured from different angles into a three-dimensional model in real-time. However, they employed a traditional point-to-point ICP method to perform point cloud registration, which proved time-consuming. Similarly, Ni *et al.* [17] provided a new approach to demonstrate the entire view of the robot working space. They used the Kinect sensor to get the raw point cloud and then acquired a full vision point cloud by using a point-cloud registration algorithm. Furthermore, to improve the teleoperability, Kyunghwan *et al.* [5] proposed a visualization system that can help operators recognize target objects in a 3D environment by using a faster region-based convolutional neural network R-CNN. Both [20] and [3] rely on a virtual reality presenting the 3D point cloud scene that allows the operator to perceive the spatial position of the remote environment. To further support the application in industry, several nuclear waste detection and categorization technologies were proposed to speed up the disposal of those wastes [1, 15]. Such approaches, however, fully rely on operating personnel, so that the operator is in the object-handling feedback-control loop; this results in low productivity and operator fatigue.

3 System Overview



(a) Block diagram of the teleoperation system



(b) Experimental configuration showing various components used in this work

Figure 1. System Architecture

The workspace comprises a robotic work cell and a hanging robot system as shown in Fig.1 (a). The robot system is a UR5e series robot arm with two-fingered grippers ROBOTIQ 2F-85. The stroke of the gripper is 85mm while its form-fit grip payload achieves 5 kg. In addition, The size of the robotic work cell is 180 cm × 130 cm × 120 cm, which is suitable for the robot's movement. What's more, this system also includes a stereo camera ZED mini with a maximum depth

distance of 15 meters. In this configuration, the camera will move with the robot end-effector when scanning the scene.

Fig.1 conceptually depicts the architecture of our semi-autonomous teleoperation system. Firstly, a stereo camera, fixed on the robot end-effector, scans the robot work cell over a full 360-degree range. The point cloud captured by the 3D sensor will be fed into the environment reconstruction system. After accumulating point cloud registration, a complete model of robot workspace can be created. The user interface simultaneously displays the robot's surrounding environment model. Those objects that were detected before can be recognized automatically. Regarding items that appear for the first time in the scene, the operator can select them through the interface and type the object's name. Afterwards, the perception system will find out this object instance in the scene. Besides, the point cloud of the selected objects should be stored in the object library automatically. Finally, the grasp list will display the name of all detected object instances, and operators can choose the object that needs to be picked up. The robot will execute and accomplish pick up action *MoveIt!* [4] when it has finished path planning.

4 Model-Based Teleoperation

In this section, a novel pipeline is proposed to implement the reconstruction and detection of unknown objects, which are nuclear waste materials in our applications. As outlined, this method involves three components: *(a)* scenes reconstruction, *(b)* human-supervised object 6D estimation and *(c)* a novel user interface.

4.1 Virtual Environment Reconstruction

Industrial objects are generally texture-less, irregular in shape and even radiant. So it is difficult to obtain 3D models of these materials using regular CAD modelling methods. Most of the online reconstruction approaches are either noisy [6, 21] or completely missed [13] in the fused output. Therefore, we decided to develop an algorithm for small-scale industrial scene reconstruction.

4.1.1 Improving point cloud registration. Fig.2 gives an overview introduction of the 3D scene reconstruction system. The core component of this pipeline is a robust point cloud registration algorithm. Inspired by [21], the colour information is considered along with the geometry information to improve the accuracy of the point cloud alignment.

In [21], the authors pioneered a joint optimization objective that converts both photometric and geometric terms as a continuous function.

$$F(T) = \sigma F_G(T) + (1-\sigma)F_C(T) \tag{1}$$

where F_G and F_C are the nonlinear least-squares function for geometric and photometric items, respectively; $\sigma \in [0,1]$ is the weight of those two items.

F_G is actually a point-to-plane ICP algorithm:

$$F_G(T) = \sum_{(p,q)\in\kappa} ((p - Tq)\cdot n_p)^2 \tag{2}$$

where n_p is the normal of point p and $\kappa = (p,q)$ is the correspondence set from target point cloud P and source point cloud Q.

The photometric term F_C represents a continuous and differentiable colour function:

$$F_C(T) = \sum_{(p,q)\in\kappa} (C_p(q') - C(q))^2 \tag{3}$$

where $C(p)$ is the colour intensity of each point p, $C_p()$ represents a continuous colour defined on the tangent plane of point p function and q' is the projection of point q onto the tangent plane of p.

Despite achieving a good registration result, this algorithm still has many drawbacks affecting the quality of the alignment. Therefore, we attempt the following improvements for this algorithm to improve the accuracy and robustness of the registration.

Firstly, we use gray scale representing the intensity value of each point to calculate the colour gradient.

$$C(p) = 0.299 * R(p) + 0.587 * G(p) + 0.114 * B(p) \tag{4}$$

where $R(), G(), B()$ are point intensity for red, green and blue respectively. Compared with the original method utilizing average intensity value, gray scale can avoid many ambiguities. For example, the average intensity of red (255,0,0) and green (0,255,0) are the same, whereas gray scale can distinguish them easily.

What's more, for objective optimization, we propose a Levenberg-Marquardt Iterative Closest Point (LM-ICP) approach. Specifically, LM-ICP begins with an initial transformation T_0 and executes an optimization iteratively. In each iteration, we calculate the residual r and Jacobian J_r at the T_{k-1} that was estimated in the last iteration, and solve the Levenberg–Marquardt function (5) to obtain increment $\xi = (t_x, t_y, t_z, \alpha, \beta, \gamma)$. In the next iteration, we recompute T_k around T_{k-1} and repeat until convergence.

$$(J_r^T J_r + \mu I)\xi = -J_r^T r \tag{5}$$

where damping parameter μ is adjusted at each iteration, which influences both the magnitude and direction of the step. For the choice of the μ, we refer to [18]. I is a 6×6 identity matrix.

4.1.2 Coarse-to-fine pipeline. To achieve a more precise fusion result, we design the coarse-to-fine pipeline, shown in Fig 2. The original raw point cloud data from 3D sensors are noisy and cluttered owing to the camera's field of view, measurement errors and light reflections; this complicates the point cloud processing. Consequently, before the feature extraction, we need to remove the redundant and outlier points.

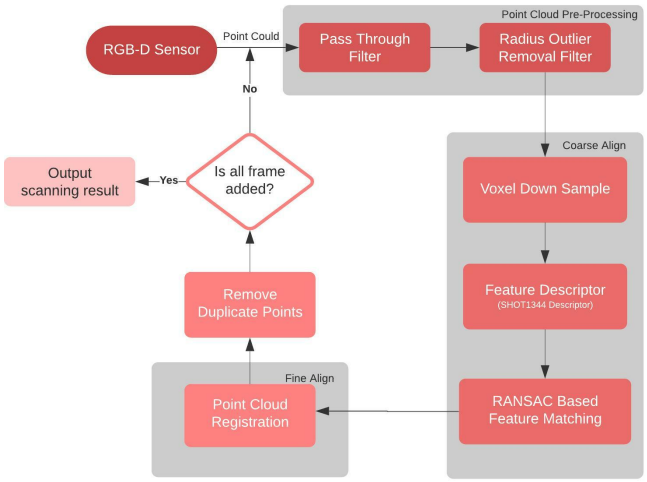


Figure 2. Block diagram of our point cloud fusion system

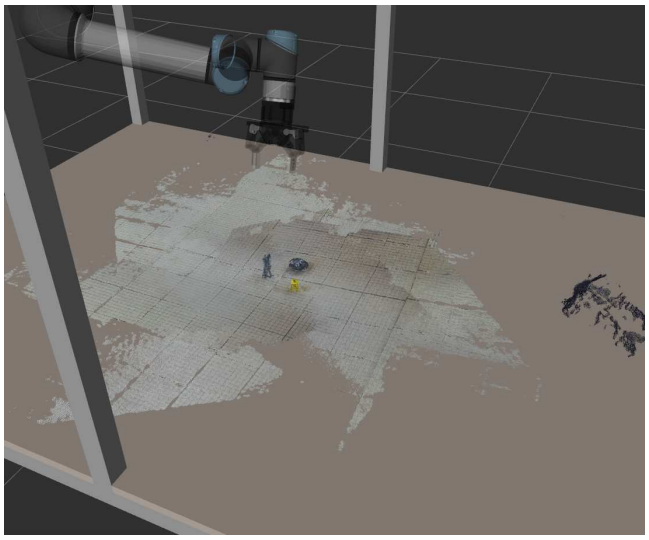


Figure 3. Visualization of the environment modeling

First, the raw point cloud is filtered by a pass-through filter that can delete the points outside the workspace. Secondly, the filtered points will pass to a radius outlier removal filter that can clear outlier points according to the number of neighbours. The point will be seen as an outlier if the number of neighbours within a given radius is less than the threshold.

In the first place, the denoised point cloud will be down-sampled by an approximate voxel filter, which can speed up the alignment process while retaining features detail. Next, we will perform a RANSAC registration [6] to obtain a rough translation between source and target point cloud. Then, we use our improvement point cloud registration algorithm to refine the alignment further.

After alignment, the new point cloud has many overlaps with the target point cloud, and direct merging may cause feature occlusion and increase the computational burden. Consequently, for each point of the source point cloud, we exploit kd-tree data structures to the retrieve neighbours' distance for finding and removing duplicate points. Finally, the new fused point cloud will be defined as the target point cloud for the next iteration until the reconstruction ends. Fig.3 illustrates the reconstruction results of the robot's working environment. To facilitate the operator to perceive the spatial relationship between the robot and the target objects, we render the reconstruction scene and known robot model together in the Rviz Moveit! environment.

4.2 Robot Perception System

The robot can't perceive the surroundings and make decisions autonomously. Thus, we propose a human-supervised robotic perception system that can achieve accurate 3D object detection with the help of humans' intelligence. We obtained a high-quality 3D environment model for unknown scenes through the reconstruction system. Based on it, the operator can provide the complete object geometric model as the template for detection by our interface.

The complete object detection is outlined in the flowchart presented in Fig.4. This algorithm is a point cloud-based method, which presents high accuracy for texture-less objects since a 3D point cloud can provide better geometry information than 2D images. For feature extraction, traditional approaches in the image domain (SIFT [16] and SURF [2]) used at Bundle-Fusion [8] perform poorly in homogeneous surfaces, such as metal, which can be observed from TABLE 1. Point cloud-based methods have greatly improved the extraction of key points. A tested SIFT-3D approach encountered a serious issue regarding the scene keypoints detection due to relying on pixel intensity change. As a result, most of the keypoints detected by SIFT-3D [9] are concentrated in the background environment with large colour variations. By contrast, ISSKeyPoint-3D [31] can extract sufficient feature points for both model and scene, but this can slow down detection speed. Taking the above considerations together, we decide to adopt a Harris-3D [25] keypoints detection method that searches points with large intensity changes in the surface normals' direction to extract interest points. This method can extract enough key points, as well as doesn't influence the subsequent matching process.

Table 1. The number of keypoints detected using five different methods for target model and work scene

	SIFT	SURF	SIFT3D	ISSKeyPoint	Harris3D
Model	16	8	356	845	756
Scene	3648	8790	5387	23594	14986

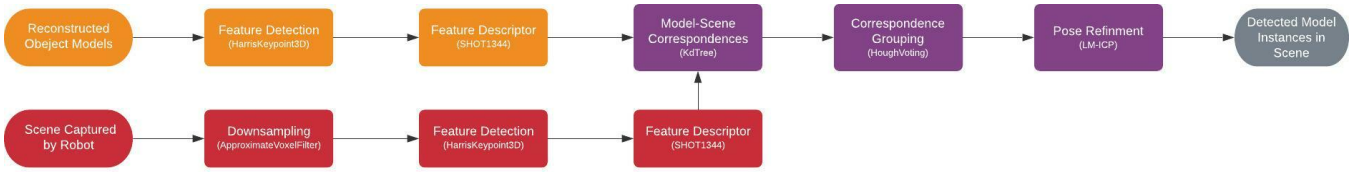
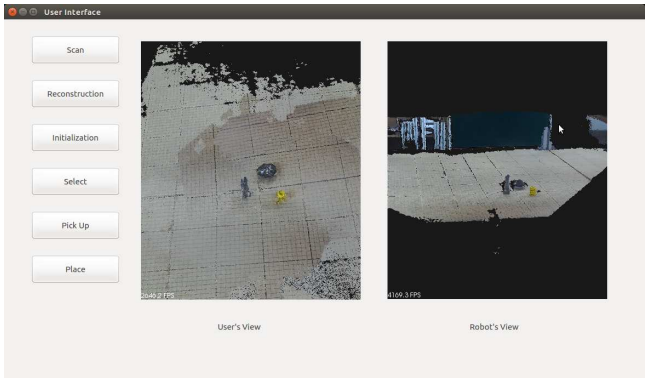


Figure 4. Block diagram of our template-based 3D object detection system

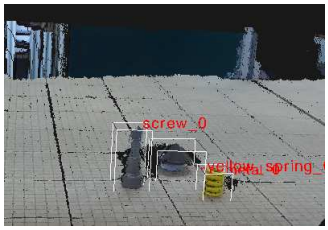
After obtaining sufficient feature points, we use a feature descriptor to describe geometrical patterns based on the information around those points. In the proposed system, a method based on a Signature of Histogram of Orientation (SHOT1344) [27] was considered to describe geometrical features. Then, the correspondence of the extracted features descriptor from model and scene is established if the squared descriptor distance is less than the threshold value (0.45 in our system).

Due to the presence of noise, cluttered background, partial occlusions, and similar features, these correspondence results typically contain several wrong correspondences. Therefore, we add a 3D Hough voting verification to weed out lousy match points by accumulating evidence of the existence of the objects in a 3D Hough space. To be more specific, if there are enough features voting for the existence of the object in the given 3D environment; thus the object is recognized, and its pose is calculated through the computed correspondences.

Accuracy in the object detection process is crucial for high-level manipulation tasks. The position of the object can be detected after the previous procedure, but there is some slight deviation from the orientation of the object. So, we use a, LM-ICP process introduced in the previous section as the post-process to refine the pose detection results. Note that this refining process does not compute for whole scene points instead of the scene points within the 3D minimum bounding rectangle of the instance. Eventually, the operator can view the visualization result of 3D detection through the user interface shown in Fig. 5 (b) and send instructions to the robot by clicking the name of the target illustrated by Fig. 5 (c).



(a) Main window of our user interface



(b) Visulization of detection result

(c) The pick up list of detected object instances

Figure 5. Our user interface with various components.

4.3 User Interface

As mentioned before, the GUI is used to interact with the environment. Therefore, we integrate each function as a button, and the detailed functions are as follows:

Scan: The robot will start to move according to the default scan path.

Initialization: The system will load the object model stored in the database and instantiate object templates into a scene.

Reconstruction: We adapt the Screened Surface Poisson Reconstruction [11] to convert the point cloud model to triangular mesh, which can make the environmental model more realistic.

Select: For a new object, the user can select the target from the left visualization window to provide a model template for the object recognized system.

Pick Up: After selecting the object which it needs to pick up, the system will generate a grasp pose for the object according to the object detected pose. Then, the robot can plan and execute grasping actions.

Place: The user can decide where to place the object by clicking the mouse.

This interface has two visualization windows, the left one is used to interact with the operator, and another is to visualize detection outcomes. The left window displays a reconstruction model, and the operator can select the object model from the reconstruction environment and type the object's property (e.g., name, material and colour). The selected objects information will be stored in the object database simultaneously, and thus this object can be recognized automatically if a new scene contains it. Besides, the user also

designates the place position through this window. The right window reflects the environment perceived by the robot.

Moreover, our interface is designed based on the ROS framework; this can expedite data sharing and communication with MoveIt!, which combines several advanced robotic functions such as motion planning, manipulation, etc.

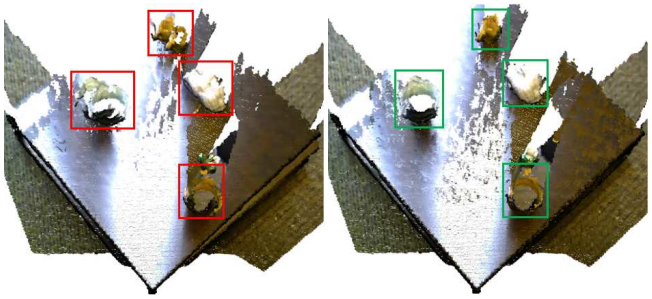
5 Experimental Evaluation

In this section, we present the experimental results obtained for both 3D visualisation system and vision-based semi-autonomous tasks along with a detailed analysis investigating the task performance.

5.1 Evaluation for Point Cloud Fusion Task

We compare against three widely used registration methods for which the function is available in Open3D library [32]. The evaluation is performed on two scenes from the RGB-D Scenes dataset [26].

From Fig 6 (a), we notice that the environment model reconstructed by [21] suffers from misalignment and drifting. In contrast, our method not only accurately aligns the small-scale texture-less model but also keeps geometric information as much as possible. Such high-quality results are beneficial for remote teleportation, point cloud-based deep learning, etc.



(a) The reconstruction result from colored-ICP[21] **(b)** The reconstruction result using our method

Figure 6. Comparison of our approach with colored-ICP [21] in RGB-D Scenes dataset [26]. Red boxes indicate the misaligned part and green boxes show the correctly aligned part.

What's more, we conduct the quantitative evaluation of different registration algorithms using root mean square error (RMSE) and standard deviation (STD) values of relative pose error (RPE) metrics [26]. From TABLE 2 we can notice that our method outperforms all the considered competitors in both translation and rotation terms. Besides, we report the running time of one registration for the different methods in TABLE 2. The running time is measured on a PC with an Inter Core i5-7300HQ CPU and Geforce GTX 1050 GPU. Our method is much faster than point-to-point and point-to-plane ICP, and is comparable to colored-ICP with improved accuracy.

5.2 Evaluation for Robot Perception System

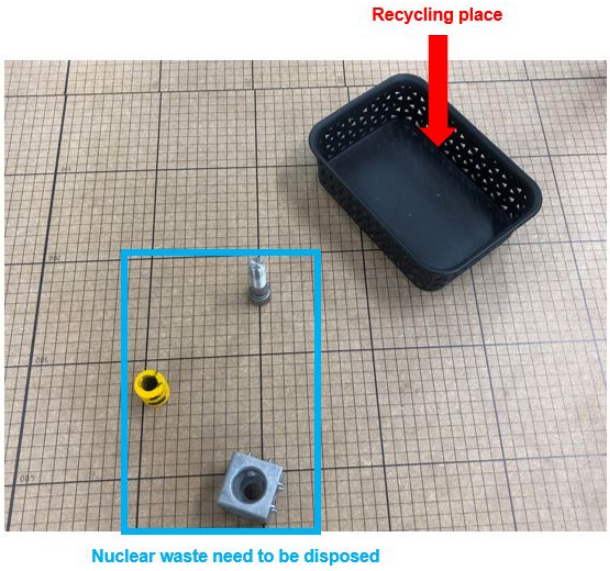


Figure 7. Simulated nuclear waste handling scenario

We applied the proposed system to a simulated nuclear waste handling scenario (shown in Fig 7) to verify the effectiveness of the system. In the experiments we have conducted, our perception system can precisely distinguish those test objects which have similar size and colour. The average handling time for a single object has remained below 30 sec while using a joy controller consumed an average of 65 sec, as demonstrated in the video in https://youtu.be/GQ4yyA6Ooa8. Distinctive from the other complex controller, our mouse-and-keyboard robot manipulation interface is convenient to most of the population. More importantly, the automatic grasping process can free the operator's hands compared with uninterrupted manual teleoperation.

6 Conclusion and Future Work

This paper has presented a novel framework for the implementation of a 3D reconstruction and perception system for remote teleoperation to aid human operators. The experimental results show that the 3D point cloud fusion system performed fairly well when compared against three widely used methods.

Evaluation of the psychological aspects of the 3D models was out of the scope of this paper. We have focused on improved speed and quality of 3D modelling. The visualization system provides a sufficiently precise virtual environment for operators, in contrast to other research in the literature, which still relies on a 2D image or point clouds.

Table 2. The evaluation of point cloud registration on the RGB-D Scenes dataset

	Translation RMSE(m)	Translation STD(m)	Rotation RMSE(degree)	Rotation STD(degree)	Average running time(s)
Point-to-point ICP	0.004027	0.002015	1.2408	0.6172	17.8213
Point-to-plane ICP	0.004373	0.002113	1.7964	0.7925	16.3853
Colored ICP	0.004204	0.002242	1.3262	0.5587	**1.2316**
Our	**0.003367**	**0.001971**	**1.0325**	**0.5145**	1.4679

The proposed visualization system displays the objects perceived by the robot in a virtual scene. An analysis of the modelling experiments revealed that the virtual environment is sufficiently accurate. Thus, this 3D virtual environment can potentially provide intuitive space perception for operators, which can assist operators in reducing work stress and boosting work efficiency.

Moreover, we implement a human-surpervised semi-autonomic system, which is integrated with our concise mouse-and-keyboard user interfaces. This interface allows the operators to select an object quickly and precisely for automated handling. In the second verification experiment, we completed a simulated nuclear waste handling experiment successfully.

Acknowledgments

This work was supported by EPSRC Grant No.EP/R026084/1, Robotics and Artificial Intelligence for Nuclear (RAIN), UK.

References

[1] Jonathan M Aitken, Sandor M Veres, Affan Shaukat, Yang Gao, Elisa Cucco, Louise A Dennis, Michael Fisher, Jeffrey A Kuo, Thomas Robinson, and Paul E Mort. 2018. Autonomous nuclear waste management. *IEEE Intelligent Systems* 33, 6 (2018), 47–55.

[2] Herbert Bay, Tinne Tuytelaars, and Luc Van Gool. 2006. Surf: Speeded up robust features. In *European conference on computer vision*. Springer, 404–417.

[3] Yi Chen, Baohua Zhang, Jun Zhou, and Kai Wang. 2020. Real-time 3D unstructured environment reconstruction utilizing VR and Kinect-based immersive teleoperation for agricultural field robots. *Computers and Electronics in Agriculture* 175 (2020), 105579.

[4] Sachin Chitta, Ioan Sucan, and Steve Cousins. 2012. Moveit![ros topics]. *IEEE Robotics & Automation Magazine* 19, 1 (2012), 18–19.

[5] Kyunghwan Cho, Kwangun Ko, Heereen Shim, and Inhoon Jang. 2017. Development of VR visualization system including deep learning architecture for improving teleoperability. In *2017 14th International Conference on Ubiquitous Robots and Ambient Intelligence (URAI)*. IEEE, 462–464.

[6] Sungjoon Choi, Qian-Yi Zhou, and Vladlen Koltun. 2015. Robust reconstruction of indoor scenes. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 5556–5565.

[7] Matthew Cousins, Chenguang Yang, Junshen Chen, Wei He, and Zhaojie Ju. 2017. Development of a mixed reality based interface for human robot interaciotn. In *2017 International Conference on Machine Learning and Cybernetics (ICMLC)*, Vol. 1. IEEE, 27–34.

[8] Angela Dai, Matthias Nießner, Michael Zollhöfer, Shahram Izadi, and Christian Theobalt. 2017. Bundlefusion: Real-time globally consistent 3d reconstruction using on-the-fly surface reintegration. *ACM Transactions on Graphics (ToG)* 36, 4 (2017), 1.

[9] Gregory T Flitton, Toby P Breckon, and Najla Megherbi Bouallagu. 2010. Object Recognition using 3D SIFT in Complex CT Volumes.. In *BMVC*, Vol. 1. 1–12.

[10] Stefan Hinterstoisser, Vincent Lepetit, Slobodan Ilic, Stefan Holzer, Gary Bradski, Kurt Konolige, and Nassir Navab. 2012. Model based training, detection and pose estimation of texture-less 3d objects in heavily cluttered scenes. In *Asian conference on computer vision*. Springer, 548–562.

[11] Michael Kazhdan and Hugues Hoppe. 2013. Screened poisson surface reconstruction. *ACM Transactions on Graphics (ToG)* 32, 3 (2013), 1–13.

[12] Sebastian Kohn, Andreas Blank, David Puljiz, Lothar Zenkel, Oswald Bieber, Bjorn Hein, and Jorg Franke. 2018. Towards a Real-Time Environment Reconstruction for VR-Based Teleoperation Through Model Segmentation. In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 1–9.

[13] Mathieu Labbé and François Michaud. 2019. RTAB-Map as an open-source lidar and visual simultaneous localization and mapping library for large-scale and long-term online operation. *Journal of Field Robotics* 36, 2 (2019), 416–446.

[14] Jeffrey I Lipton, Aidan J Fay, and Daniela Rus. 2017. Baxter's homunculus: Virtual reality spaces for teleoperation in manufacturing. *IEEE Robotics and Automation Letters* 3, 1 (2017), 179–186.

[15] Naresh Marturi, Alireza Rastegarpanah, Chie Takahashi, Maxime Adjigble, Rustam Stolkin, Sebastian Zurek, Marek Kopicki, Mohammed Talha, Jeffrey A Kuo, and Yasemin Bekiroglu. 2016. Towards advanced robotic manipulation for nuclear decommissioning: A pilot study on tele-operation and autonomy. In *2016 International Conference on Robotics and Automation for Humanitarian Applications (RAHA)*. IEEE, 1–8.

[16] Pauline C Ng and Steven Henikoff. 2003. SIFT: Predicting amino acid changes that affect protein function. *Nucleic acids research* 31, 13 (2003), 3812–3814.

[17] Dejing Ni, AYC Nee, SK Ong, Huijun Li, Chengcheng Zhu, and Aiguo Song. 2018. Point cloud augmented virtual reality environment with haptic constraints for teleoperation. *Transactions of the Institute of Measurement and Control* 40, 15 (2018), 4091–4104.

[18] Hans Bruun Nielsen et al. 1999. *Damping parameter in Marquardt's method*. IMM.

[19] Günter Niemeyer, Carsten Preusche, Stefano Stramigioli, and Dongjun Lee. 2016. Telerobotics. In *Springer handbook of robotics*. Springer, 1085–1108.

[20] Bukeikhan Omarali, Brice Denoun, Kaspar Althoefer, Lorenzo Jamone, Maurizio Valle, and Ildar Farkhatdinov. 2020. Virtual reality based telerobotics framework with depth cameras. In *2020 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 1217–1222.

[21] Jaesik Park, Qian-Yi Zhou, and Vladlen Koltun. 2017. Colored point cloud registration revisited. In *Proceedings of the IEEE international conference on computer vision*. 143–152.

[22] Lorenzo Peppoloni, Filippo Brizzi, Carlo Alberto Avizzano, and Emanuele Ruffaldi. 2015. Immersive ROS-integrated framework for robot teleoperation. In *2015 IEEE Symposium on 3D User Interfaces (3DUI)*. IEEE, 177–178.

[23] Victor A Prisacariu and Ian D Reid. 2012. PWP3D: Real-time segmentation and tracking of 3D objects. *International journal of computer vision* 98, 3 (2012), 335–354.

[24] Thomas B Sheridan. 2016. Human–robot interaction: status and challenges. *Human factors* 58, 4 (2016), 525–532.

[25] Ivan Sipiran and Benjamin Bustos. 2011. Harris 3D: a robust extension of the Harris operator for interest point detection on 3D meshes. *The Visual Computer* 27, 11 (2011), 963–976.

[26] Jürgen Sturm, Nikolas Engelhard, Felix Endres, Wolfram Burgard, and Daniel Cremers. 2012. A benchmark for the evaluation of RGB-D SLAM systems. In *2012 IEEE/RSJ international conference on intelligent robots and systems*. IEEE, 573–580.

[27] Federico Tombari, Samuele Salti, and Luigi Di Stefano. 2010. Unique signatures of histograms for local surface description. In *European conference on computer vision*. Springer, 356–369.

[28] Jonathan Tremblay, Thang To, Balakumar Sundaralingam, Yu Xiang, Dieter Fox, and Stan Birchfield. 2018. Deep object pose estimation for semantic robotic grasping of household objects. *arXiv preprint arXiv:1809.10790* (2018).

[29] Balazs Vagvolgyi, Wenlong Niu, Zihan Chen, Paul Wilkening, and Peter Kazanzides. 2017. Augmented virtuality for model-based teleoperation. In *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 3826–3833.

[30] AWW Yew, SK Ong, and AYC Nee. 2017. Immersive augmented reality environment for the teleoperation of maintenance robots. *Procedia CIRP* 61 (2017), 305–310.

[31] Yu Zhong. 2009. Intrinsic shape signatures: A shape descriptor for 3d object recognition. In *2009 IEEE 12th International Conference on Computer Vision Workshops, ICCV Workshops*. IEEE, 689–696.

[32] Qian-Yi Zhou, Jaesik Park, and Vladlen Koltun. 2018. Open3D: A modern library for 3D data processing. *arXiv preprint arXiv:1801.09847* (2018).