Proceeding of the IEEE
International Conference on Robotics and Biomimetics
Dali, China, December 2019

Bin-picking of Randomly Piled Shiny Industrial Objects Using Light Transport Matrix Estimation*

Naoya Chiba and Mingyu Li
Graduate School of Information Sciences
Tohoku University
Sendai, Miyagi, Japan
chiba@ic.is.tohoku.ac.jp, li.mingyu.s8@dc.tohoku.ac.jp

Akira Imakura
Faculty of Engineering, Information and Systems
University of Tsukuba
Tsukuba, Ibaraki, Japan
imakura@cs.tsukuba.ac.jp

Koichi Hashimoto
Graduate School of Information Sciences
Tohoku University
Sendai, Miyagi, Japan
koichi@tohoku.ac.jp

***Abstract*— The use of robots in automated factories requires accurate bin-picking to ensure that objects are correctly identified and selected. In the case of objects with multiple reflections from their surfaces, this is a challenging task. We attempted to address this problem by developing a 3D measurement method based on a Light Transport Matrix (LTM), which can be applied to shiny objects or semi-transparent objects. The study presented herein evaluates the accuracy of the proposed method as well as the method for 3D pose estimation we previously reported, by examining a bin-picking task, which is a well-known robot application in factory automation. There is considerable demand for automated bin-picking systems for general objects. However, in the use of cost-effective measurement systems, some objects such as shiny metallic objects continue to prove problematic in terms of bin-picking because of the difficulty to measure their shapes accurately. Our 3D measurement method uses only a projector-camera system; thus, it is cost-effective, and it does not require any special optical system. It is based on fast LTM sparse estimation. We previously demonstrated that this approach can measure the 3D shape of metallic objects and showed that our pose estimation method is applicable to bin-picking. However, we did not verify its accuracy with the application of 3D robot vision. In this study, we integrate these two methods, and demonstrate that our 3D measurement method, in combination with our pose estimation work, can successfully accomplish bin-picking tasks involving shiny metallic industrial objects. Ultimately, we achieved 100 [%] picking success for 5 scenes including 15 pieces. We concluded that our proposed methods are sufficiently accurate to carry out bin-picking tasks in automated factory environments.**

***Index Terms*— Bin-picking, 3D Measurement, 3D Object Detection**

I. Introduction

Random bin-picking for general industrial objects continues to remain a challenging problem even though this technology is a requirement for factory automation. Generally, 3D sensors provide color information, depth information, or both of these. Compared with color information, it is possible to use the depth information to directly recover the full 6 DoF pose of an object because the depth information is invariant to geometric and photometric change. However, 3D measurements of shiny metallic industrial objects continue to be a challenging problem in the field of computer vision.

In this paper, we follow this standard pipeline (3D measurement, pose estimation, and then robot movement); however, we upgrade the 3D measurement stage to our 3D measurement method based on Light Transport Matrix (LTM) estimation[1] which enables the 3D shape of shiny objects and semi-transparent objects to be measured. Achieving this measurement would enable the robot to estimate the position and orientation of an object from among the objects accumulated in the scene with respect to the sensor.

We previously developed a 3D measurement method based on LTM estimation[1, 2, 3, 4] for objects with complex light reflectance, such as metallic or semi-transparent objects. The LTM can be considered to contain a collection of impulse responses between the projector and the camera[5, 6]. Once the LTM is obtained, it is easy to extract the direct reflection components even if the measurement object has complex light reflection. This idea of 3D measurement via LTM was originally proposed in [7]; however, their method requires a high-speed vision system because it is necessary to use a significant number of projections and captured images. Our previous method[1] uses only a common projector-camera system, and we reduce the number of images by using sparse estimation. Although previous studies[8, 9, 10] also applied sparse estimation methods to this problem, obtaining the LTM was time consuming because they mainly focused on the use of computer graphics, which requires a highly accurate LTM. As reported previously [1], we realized that an accurate 3D point cloud can be obtained from a less than accurate LTM. This fact led us to propose 3D measurement based on LTM estimation, which we showed to be capable of measuring the 3D depth of metallic or semi-transparent objects accurately.

Once the 3D depth is accurately obtained, it is necessary to estimate the poses of the objects from the scene. Pose estimation algorithms have been developed to achieve this es-

* This work was supported by ACT-I, JST Grant Numbers JPMJPR16UH, JPMJPR16U6, and JSPS KAKENHI Grant Numbers JP16H06536, 17K12690, JP18J20111.

978-1-7281-6321-5/19//$31.00 ©2019 IEEE

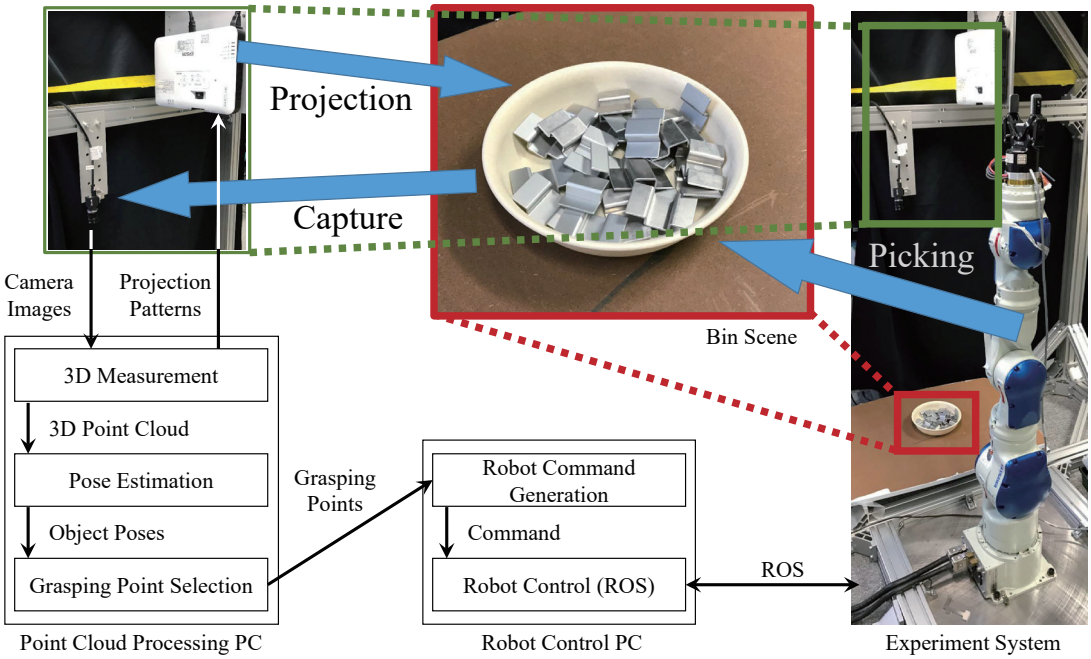


Fig. 1: Experimental system and bin-picking workflow. The scene contains piled metallic industrial objects, which are difficult to measure to determine their pose estimation. Our previous work[1] led to the proposal of a 3D measurement method that can be applied to metallic objects. However, the paper did not evaluate the accuracy of the method to solve real industrial problems. In this paper, we describe the application of the method to a real bin-picking problem to demonstrate that our methods are sufficiently accurate to solve the problem. The system consists of a robot, a projector, and a camera. The bin-picking workflow is as follows: 1) measure the 3D point cloud, 2) estimate the object poses, 3) generate grasping points and select the best one, 4) generate robot commands, 5) control the robot.

timation with 3D depth information. Drost et al.[11] proposed an efficient voting scheme based on the point pair feature. The method could achieve estimation without training data. Another advantage is that it proved capable of processing sparse data, which is crucial for our system because it would save considerable measurement time. Many studies have been carried out based on this method[12, 13, 14, 15, 16]. In this study, we use another method [17], which was proved to be accurate and fast for bin-picking tasks.

The aim of this study is to demonstrate that our 3D measurement method is sufficiently accurate. Although we previously concluded that our method is useful for the bin-picking task [1], we did not evaluate it by using a real application. Therefore, we have since integrated the method into a bin-picking system including an industrial robot. According to experiments, our system performs well for real shiny industrial objects even though the standard workflow is used. In other words, we demonstrate our 3D measurement method can upgrade the performance of bin-picking systems based on 3D vision.

The paper is structured as follows. In Sec. II, we provide an overview of our system. We present a description of the details of 3D measurement, pose estimation, and grasping point selection in Sec. III, Sec. IV, and Sec. V, respectively. Sec. VI shows the experimental results. We compare the performance of our methods with that of existing 3D measurement and pose estimation methods. Sec. VII concludes the paper.

II. OVERVIEW

This section provides an overview of the proposed bin-picking system. This system is designed with the aim of selecting individual metallic objects separately from a bin scene. We follow the bin-picking standard workflow: 1) measure the 3D point cloud of the scene by projecting some images and by capturing light reflected from the scene, 2) use the 3D point cloud to estimate the object poses, 3) generate grasping points for the detected objects and select the best one, 4) generate a robot command to grasp, 5) control the robot to execute the command. The proposed bin-picking system is shown in Fig. 1. It consists of a robot, projector, camera, and bin scene. We assume the scene includes piled metallic objects.

We utilize our 3D measurement method based on LTM estimation[1] with speed-up techniques[18]. This method can

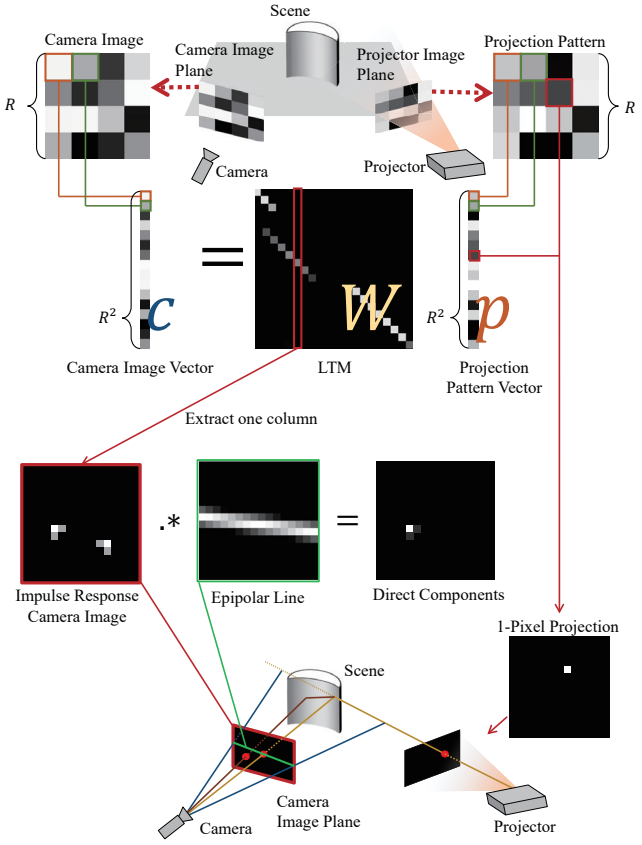


Fig. 2: LTM optical model and extraction of direct components from the LTM [1]. A camera vector $\boldsymbol{c}$ is given by $W\boldsymbol{p}$, which is the result of multiplying LTM W by a projection pattern vector $\boldsymbol{p}$. Each column of the LTM describes an impulse response camera image. Direct components are extracted from the impulse response camera image by masking the epipolar line.

measure 3D shapes generally even for metallic and semi-transparent objects. It needs only a projector-camera system, thereby obviating the need to use an expensive measurement system. We also utilize a pose estimation method[17], which is based on PPF[11] and is faster and more accurate.

III. 3D MEASUREMENT

The 3D measurement method based on LTM estimation[1] was developed as a general 3D measurement method applicable to metallic or semi-transparent objects. LTM estimation-based 3D measurement is briefly introduced here.

The method uses a projector-camera system, and it assumes a linear relation between projector intensities and camera intensities. Suppose $\boldsymbol{p}$ is a projection pattern vector, which is a vectorized projected pattern image, and $\boldsymbol{c}$ is a camera image vector, which is a vectorized captured camera image. The LTM model describes the relation of $\boldsymbol{p}$ and $\boldsymbol{c}$ by

$$\boldsymbol{c} = W\boldsymbol{p}, \tag{1}$$

where W is the LTM(See Fig. 2).

Each column of the LTM consists of impulse responses of a projector-camera system. Here, an impulse response means a camera image captured using only one pixel irradiated by the projector. An impulse response includes a combination of direct and global components of reflection from the measurement scene; however, they can be separated by using epipolar geometry[1, 2, 3, 4, 7]. Once the direct components are obtained separately, a 3D point cloud of the scene is easily reconstructed by classical triangulation even if the scene includes metallic or semi-transparent objects.

The LTM is a very large and sparse matrix; thus, constructing the LTM is highly time consuming[1, 3, 4]. Computer graphics techniques[8, 10] are therefore used to apply compressive sensing to estimate the LTM. However, these techniques are still unacceptably slow for robot vision because their applications, such as scene relighting, require a more accurate LTM. In the study presented in this paper, we utilize multi-scale LTM estimation[1], which is developed for robot vision applications.

We also developed speed-up techniques for LTM estimation. The most important technique is to apply the Sherman-Morrison-Woodbury (SMW) formula[19] to the Alternating Direction Method of Multipliers (ADMM) ℓ_1 minimization [20].

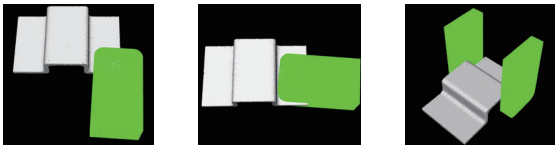
ADMM ℓ_1 minimization includes a computation

$$\left(I + \frac{1}{\mu\lambda}A^\top A\right)^{-1}, \tag{2}$$

where I denotes the Identity matrix and A denotes the observation matrix which consists of the elements of the projection patterns. By applying the SMW-formula, (2) can be computed by

$$I - \frac{1}{\mu\lambda}A^\top\left(I + \frac{1}{\mu\lambda}AA^\top\right)^{-1}A. \tag{3}$$

Note that, the computational cost of (3) is sometimes smaller than (2), depending on the shape of A. By using this, we reduce the total computational cost of ADMM ℓ_1 minimization from $\mathcal{O}(R^6 + R^4N + R^4T)$ to $\mathcal{O}(R^2N^2 + N^3 + R^2NT)$ where R is the resolution of the camera and the projector, N is the number of projection/capture events, and T is the number of iterative cycles in the ADMM process. In this paper, we utilize $R = 16$–256 (controlled automatically by multi-scale LTM estimation[1]), $N = 32$, and $T = 100$. The values of $R = N^{1.25}$ (Maximum value) and $T = N^{1.33}$ can be estimated roughly; therefore, the computational cost is reduced from $\mathcal{O}(N^{7.5})$ to $\mathcal{O}(N^{4.83})$. Finally, we succeeded in reducing the process of LTM estimation to less than 22.76



(a) (b) (c)

Fig. 3: Examples of grasping points of the model. The green blocks represent the gripper, whereas the gray object represents the model.

[sec]. This process is much faster than that in our previous study[1], which required 342 [sec].

IV. POSE ESTIMATION

Generally, the position of the bin is fixed in bin-picking tasks. Before the picking starts, we first capture a cloud of the bin containing no objects. Then for each input scene cloud, the points of the bin are removed because they constitute noise in the estimation process.

Our system uses the pose estimation algorithm of Li and Hashimoto[17], which is based on the point pair feature[11], to estimate the 6 DoF poses of industrial objects. The point pair feature is defined as a four-dimensional vector between two points with normals, containing one distance component and three angle components. We use this method because of its ability to estimate poses with sparse data, which is desirable to improve the efficiency of the 3D measurement.

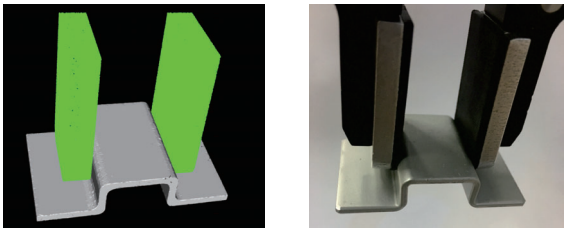
During the offline stage, the features between every point pair in the model are computed and saved in a hash table for efficient search. Then, at the recognition stage, scene pairs are aligned with model pairs with similar feature values, whereby the pose candidates are generated to match the model to the scene. The poses for which it is possible to match relatively more point pairs are saved for the next stage.

A pose is evaluated by transforming the model points into scene space according to the pose and by counting the number of model points for which a scene point could be found nearby, which is the score of the pose. The scene space is divided into small voxels to improve the efficiency of searching neighboring scene points. The non-overlapping poses with top scores are selected as the result poses.

V. GRASPING POINT SELECTION

After estimating the poses of the objects, it is necessary to decide how to lift the objects with our Two-Finger Adaptive Robot Gripper. This section describes the generation of grasping points and selection of the best one.

In our experiments, a human annotated multiple grasping points for the model, as presented in Fig. 3. For each grasping point, we evaluate the grasping stability by measuring the contact area of the grippers and the model, which, in the point cloud, could be expressed as the number of model points sufficiently close to the gripper cloud. The stability of every grasping point is computed and stored offline.



(a) (b)

Fig. 4: (a) Selected picking point (b) Picking point for robot picking.

Sometimes, even though the pose of an object is accurately estimated, it is still difficult to pick it up since the gripper may collide with other objects or the bin. Therefore, during the online stage, given a grasping point, a collision detection method is necessary. If some points of the scene point cloud captured by the camera (including the bin points) are within the gripper point cloud, it means that the gripper would collide when executing the task.

Multiple result poses are estimated for every captured scene cloud, and each of these poses corresponds to multiple grasping points. To select the best among them, the following steps are performed:

1) For each object pose, transform the created grasping points.
2) Rank all the grasping points by the score of their corresponding poses. For grasping points with the same pose score, rank them by their stability.
3) From top to bottom, perform collision detection on the ranked grasping points. The first point that passed the detection is selected as the final grasping point and is transferred to the robot.

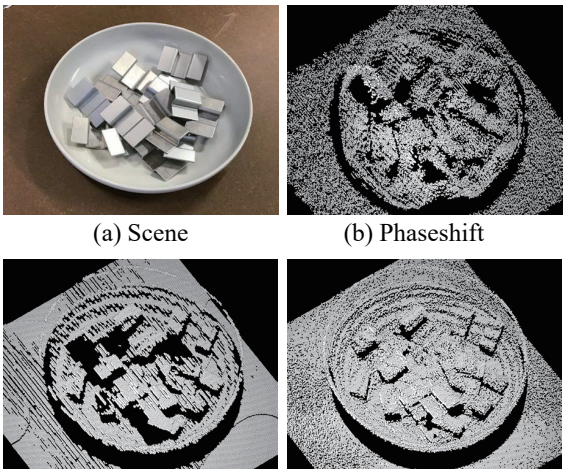
This method enables the robot to pick up the object most likely to be correctly estimated as stably as possible. A picking example is presented in Fig. 4.

VI. EXPERIMENTS

This section presents the results of the 3D measurement and bin-picking experiments. We first describe the measurement of the 3D point cloud by conducting 3D measurement based on LTM estimation for a scene that includes piled metallic industrial parts. Then, we demonstrate a bin-picking task with a real robot system.

A. MEASUREMENT EXPERIMENT

We applied our 3D measurement method based on LTM estimation to the scene containing piled metallic parts. For comparison purposes, we also measured the scene by using the Graycode structured light and Phaseshift methods[3].



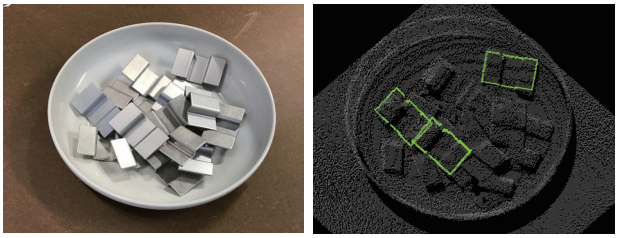
(a) Scene (b) Phaseshift

(c) Graycode (d) LTM estimation

Fig. 5: Measured point clouds. The Phaseshift and Graycode methods have many vanished or inaccurate points compared with LTM estimation. This scene includes specular reflections and inter-reflections, which adversely affect the Phaseshift and Graycode methods.

Both methods are well known, and are often used in real industrial systems. The 3D measurement system consisted of a projector (EB-1795F), and a camera (Point Grey Research, Flea3 FL3-U3-120S3C). We used homographic transformation to change the resolutions of the projector and the camera to 256×256. We calibrated the projector-camera system by using known methods [21] and [22]. Multi-scale LTM estimation was achieved by sequentially utilizing 16×16, 64×64, and 256×256 as the projector resolution. For each resolution, the number of images is 32, and we captured background illumination for LTM estimation preprocessing; thus, the number of projections and captures is 97. Measured point clouds often include noise; thus, we utilized statistical outlier removal[23] as a post-filter, which was also used in our previous study[1] with parameters $K = 20$ and $\alpha = 0.05$ to obtain the results with all the measurement methods. The captured point cloud is shown in Fig. 5. The result obtained with Phaseshift remained noisy even when a denoising method was applied to the area containing the metallic parts. In the result obtained with Graycode, some parts of the point cloud that contain reflected light did not include global components; however, some parts, which seemed to include inter-reflections of specular reflections could not be measured.

We first applied the pose estimation method[17] to the point cloud of the scene Fig. 5 measured by the LTM estimation-based method. The number of points in a measured point cloud is $65,536$ in our experimental setting. We specified the number of objects to be detected as 3. The pose



(a) Scene (b) Estimated Poses

Fig. 6: Pose estimation results for the scene in Fig. 5 measured by the LTM estimation-based method. The number of objects that need to be detected was given as 3.

TABLE I: Number of successfully estimated object poses

Scene ID	Phaseshift	Graycode	LTM Estimation
0	0	0	3
1	1	2	2
2	0	2	2
3	0	2	3
4	1	1	3
5	0	2	3
6	0	1	3
7	1	2	3
8	0	2	3
9	1	3	3
Mean	0.4	1.7	2.8

TABLE II: Time costs for each task (see also Fig. 8)

Task	Mean [sec]	Std. dev. [sec]
Projection/Capture	16.83	0.05
LTM Estimation	22.76	0.99
Reconstruct Point Cloud	2.15	0.05
Pose Estimation	0.58	0.21
Grasping Point Selection	2.99	0.07
Robot Movement	7.50	0.01
Others	0.94	0.16
Total	53.76	1.54

estimation results are shown in Fig. 6, which shows that ours is the only method capable of accurately estimating the poses of the 3 objects.

For the quantitative evaluation, we applied the pose estimation methods to point clouds of 10 scenes, which were measured by each of the 3D measurement methods.We evaluated the number of objects that were correctly detected by our sights because we could not obtain the ground truth object pose from the real scene. The results are shown in Table I. The results in the table show that, with respect to 3D measurement, the LTM estimation-based method performs the best in the comparisons.

B. BIN-PICKING TASK

We finally integrated our 3D measurement and pose estimation methods into a real robotic system. The picking robot was a Yaskawa MOTOMAN SIA-10F, which was

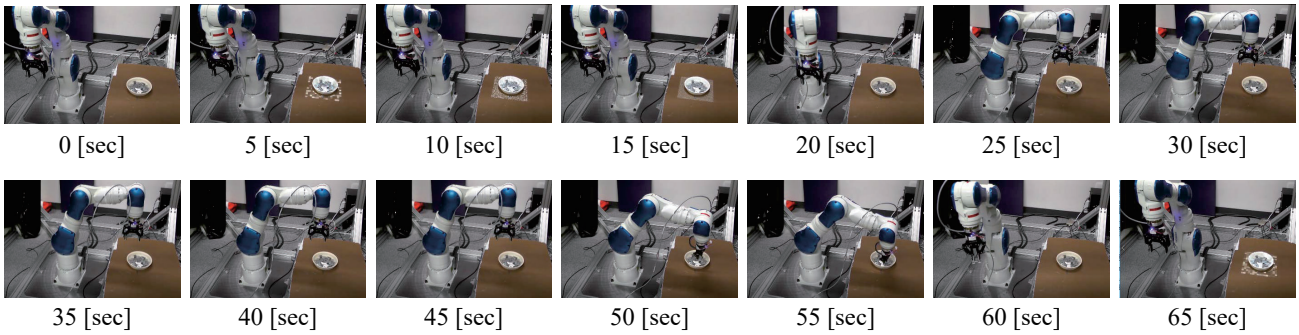


Fig. 7: Robot behavior video sequence for single-object picking. We carried out five trials for each scene including 15 parts. We achieved 100 [%] success for all scenes, including 3 recognition retrials necessitated by collisions. The projection patterns for each scale of the multi-scale LTM estimation are irradiated in the 5, 10, and 15 [sec] frames. While the robot is moving in the 20 and 25 [sec] frames, the PC processing the Point cloud is simultaneously running the sparse estimation calculation. In the frames from 45 [sec] to 60 [sec], the robot picks up the detected object with the selected grasping point.

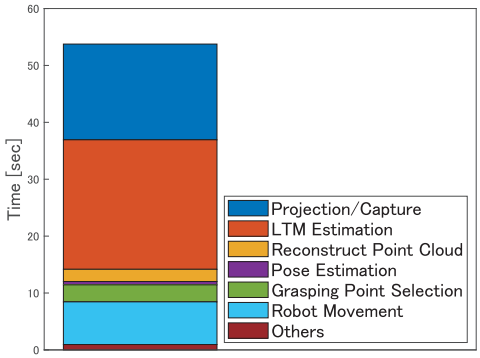


Fig. 8: Breakdown of execution time shown in Table II. LTM estimation takes 22.76 [sec], which is 42 [%] of the entire process. Compared with the Projection/Capture time, which is 31 [%], estimation is not overly time consuming.

equipped with a Robotiq Adaptive Gripper 2F-85 as the end-effector. The 3D measurement system is the same as described in Sec. **??**. The target scene included 15 metallic parts. Sometimes the robot could not reach the detected objects; thus, we conducted the bin-picking experiment until the robot was unable to reach all the detected objects and ran this experiment for 5 different scenes. In cases in which some of the detected poses could not be reached by the robot because of collisions, we retried the recognition. We finally achieved 100 [%] success for all scenes (15 objects in 5 scenes, a total of 75 picking trials).The behavior of the robot is shown in Fig. 7. The execution times are shown in Table II and Fig. 8, and total 53.76 ± 1.54 [sec] for each pick-up motion. LTM Estimation with a speed-up technique[18] is 15 times faster than with the previous method[1], and it requires only 22.76 [sec], which is comparable with the time required by other tasks such as Projection/Capture (16.83 [sec]) and Robot Movement (7.50 [sec]). In other words, when we use our previous implementation[1], the time needed for LTM Estimation becomes 342 [sec] and the execution time becomes 373 [sec]. Additionally, the Pose Estimation task only consumed 0.58 [sec], which is negligible compared with the overall execution time.

VII. CONCLUSION

We demonstrated that the proposed 3D measurement method based on LTM estimation is sufficiently accurate for real bin-picking applications. Our 3D measurement method can measure shiny objects; however, it is computationally costly when a high-resolution result is necessary. Previously[1], we drastically reduced this computational cost, and SMW-ADMM, which was implemented in this study, was shown to accelerate the computation by approximately 15 times, although it is still limited for high-resolution (256×256) intended for online application tasks such as bin-picking. Integration of our method with the projector-camera system and the robot achieved 100 [%] success for 5 scenes containing 15 pieces of metallic objects each. The limitation of our system is: 1) the LTM estimation calculation remains slow compared with the other underlying steps, and 2) we do not provide an automated grasping point generation method. The first limitation has two potential solutions: accelerate the LTM estimation calculation (e.g., by using GPU implementation) and improve the accuracy of the pose estimation under a low-resolution point cloud. We plan to overcome the second limitation by attempting to use existing grasp planning methods (e.g., [24]).

REFERENCES

[1] N. Chiba and K. Hashimoto, "Ultra-Fast Multi-Scale Shape Estimation of Light Transport Matrix for Complex Light Reflection Objects," in *Proc. IEEE Int. Conf. on Robotics and Automation*, 2018, pp. 1763–1768.

[2] N. Chiba, S. Arai, and K. Hashimoto, "Feedback Projection for 3D Measurements Under Complex Lighting Conditions," in *Proc. American Control Conf.*, 2017, pp. 4649–4656.

[3] N. Chiba and K. Hashimoto, "3D Measurement by Estimating Homogeneous Light Transport (HLT) Matrix," in *Proc. IEEE Int. Conf. on Mechatronics and Automation*, 2017, pp. 1763–1768.

[4] ——, "Homogeneous light transport matrix estimation based 3D shape measurement," *Int. J. Mechatronics and Automation*, vol. 6, no. 2–3, pp. 63–70, 2017.

[5] S. M. Seitz, Y. Matsushita, and K. N. Kutulakos, "A Theory of Inverse Light Transport," in *Proc. IEEE Int. Conf. on Computer Vision*, 2005, pp. 1440–1447.

[6] P. Sen, B. Chen, G. Garg, S. R. Marschner, M. Horowitz, M. Levoy, and H. P. A. Lensch, "Dual Photography," *ACM Trans. on Graphics*, vol. 24, no. 3, pp. 745–755, 2005.

[7] M. O'Toole, R. Raskar, and K. N. Kutulakos, "3D Shape and Indirect Appearance by Structured Light Transport," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 38, no. 7, pp. 1298–1312, 2016.

[8] I. Miyagawa, Y. Taniguchi, and T. Kinebuchi, "Radiometric Compensation Using Color-Mixing Matrix Reformed from Light Transport Matrix," *ITE Trans. on Media Technology and Applications*, vol. 4, no. 2, pp. 155–168, 2016.

[9] P. Sen and S. Darabi, "Compressive Dual Photography," *Computer Graphics Forum*, vol. 28, no. 2, pp. 609–618, 2009.

[10] P. Peers, D. K. Mahajan, B. Lamond, A. Ghosh, W. Matusik, R. Ramamoorthi, and P. Debevec, "Compressive Light Transport Sensing," *ACM Trans. on Graphics*, vol. 28, no. 1, pp. pp. 3:1–3:18, 2009.

[11] B. Drost, M. Ulrich, N. Navab, and S. Ilic, "Model globally, match locally: Efficient and robust 3D object recognition," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, 2010, pp. 998–1005.

[12] W. Abbeloos and T. Goedeme, "Point pair feature based object detection for random bin picking," in *Proc. Conf. on Computer and Robot Vision*, 2016, pp. 432–439.

[13] S. Hinterstoisser, V. Lepetit, N. Rajkumar, and K. Konolige, "Going further with point pair features," in *Proc. European Conf. on Computer Vision*, 2016, pp. 834–848.

[14] C. Wu, S. Jiang, and K. Song, "CAD-based pose estimation for random bin-picking of multiple objects using a RGB-D camera," in *Proc. Int. Conf. on Control, Automation and Systems*, 2015, pp. 1645–1649.

[15] T. Birdal and S. Ilic, "Point Pair Features Based Object Detection and Pose Estimation Revisited," in *Proc. Int. Conf. on 3D Vision*, 2015, pp. 527–535.

[16] C. Choi, Y. Taguchi, O. Tuzel, M. Y. Liu, and S. Ramalingam, "Voting-based pose estimation for robotic assembly using a 3D sensor," in *Proc. IEEE Int. Conf. on Robotics and Automation*, 2012, pp. 1724–1731.

[17] M. Li and K. Hashimoto, "Fast and Robust Pose Estimation Algorithm for Bin Picking Using Point Pair Feature," in *Proc. Int. Conf. on Pattern Recognition*, 2018.

[18] N. Chiba, A. Imakura, and K. Hashimoto, "Beyond ultra-fast ltm estimation via smw-admm," in *Prep. for submittion to Int. Conf. Computational Photography*.

[19] G. H. Golub and C. F. van Loan, *Matrix Computations*, 4th ed. JHU Press, 2013.

[20] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers," *Foundation and Trends in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.

[21] D. Moreno and G. Taubin, "Simple, Accurate, and Robust Projector-Camera Calibration," in *Proc. Int. Conf. on 3D Imaging, Modeling, Processing, Visualization Transmission*, 2012, pp. 464–471.

[22] N. Chiba and K. Hashimoto, "Development of Calibration Tool for Projector-Camera System Considering Saturation And Gamma Correction," in *Proc. Robotics and Mechatronics Conf.*, 2018, pp. 2A2–O09.

[23] R. B. Rusu, Z. C. Marton, N. Blodow, M. Dolha, and M. Beetz, "Towards 3D Point cloud based object maps for household environments," *Robotics and Autonomous Systems*, vol. 56, no. 11, pp. 927–941, 2008.

[24] A. T. Miller and P. K. Allen, "Graspit! A versatile simulator for robotic grasping," *IEEE Robotics Automation Magazine*, vol. 11, no. 4, pp. 110–122, 2004.