# 3D Sensing by Light Transport Matrix Estimation and Point Cloud Deep Learning

Naoya CHIBA

## Abstract

With the cost reduction of 3D sensors and the performance improvement of computers, the capture and processing of 3D data have become widely used. Methods for handling 3D data have also recently been developed. This dissertation focuses on robot vision, especially the utilization of 3D data for industrial robots. Industrial robots are expected to be able to handle any object; however, due to the limitations of 3D sensors, the objects that can be handled well are limited. In particular, to measure 3D shapes of metallic or translucent objects are often desired; however, because they have complex lighting specifications, it is not easy to measure the shape in a cost-effective way such as active stereo.

In this dissertation, I propose a new 3D measurement method based on the sparse estimation of the Light Transport Matrix (LTM). The method uses only a calibrated projector-camera system consisting of a single projector and a single camera; therefore, the system is inexpensive. The difficulty in successfully measuring the shape of metallic or semi-transparent objects is due to their non-Lambert reflections. Assuming a projector-camera system is used, the reflected light is captured by the camera when the lights are projected from the projector onto the measurement scene. The reflected light is divided into two components, a direct component and a global component. Classical 3D measurement methods assume that the reflected light contains only the direct component; therefore, if the reflected light contains global components, these methods cannot measure the 3D shape og the scene accurately. When the projector-camera system is utilized to measure the shape for the objects which have the non-Lambertian reflects, they often make global components of the reflected lights; thus, these objects are hard to measure their shape. The light reflection model LTM describes the propagation from the projector pixel to the camera pixel by a linear equation; therefore, this can be considered as the impulse response of the projection camera system. By using epipolar geometry, direct components can be easily extracted from LTM.

Now, the 3D measurement problem can be considered how to obtain LTM efficiently because LTM elements are all pairs of the projector pixels and the camera pixels. I introduce a sparse estimation method to LTM estimation because LTM should be sparse in principle. When the existing sparse estimation method is applied directly, the estimation of LTM is still slow. I propose computational cost-reducing methods for the LTM sparse estimation: the row-wise estimation, the multi-scale estimation, introducing the Sherman-Morrison-Woodbury (SMW) formula for the Alternating Direction Method of Multipliers (ADMM), and multi-row simultaneous estimation. By applying them, I finally measure 3D shape in 41.74 [sec] for $256 \times 256$ projector-camera system in practice. In this method, the 3D surface can be measured even if the scene contains metallic objects or semi-transparent objects.

It is found that the LTM model is not completely fitted to the actual observations. Actual observations of projector-camera systems include saturation or under-exposure problems. These non-linearities affect the accuracy degradation of the LTM estimation. Therefore, I modify the sparse estimation algorithm to be capable of these non-linearities. For LTM sparse estimation, I utilized $\ell_1$ minimization via ADMM; thus, I apply the clipping and the crushing functions for the observation model of the $\ell_1$ minimization problem. The LTM sparse estimation utilized the ADMM $\ell_1$ minimization. For the robustness, clipping function and fracture function were applied to the observation model of $\ell_1$ minimization problem. I give a closed-form update rule for this problem and demonstrate that the proposed method is robust against saturation and under-exposure problems.

Although the proposed 3D measurement method is fast, it is still slow to apply real robot vision tasks. I work on the bin-picking task for metallic rigid objects. In the bin-picking task, measurement results are used for object detection and posture estimation tasks. Existing handcraft feature-based methods require high-density point clouds as an input such as approximately tens of thousands of points to a scene. Meanwhile, the point cloud processing method using the neural network has rapidly developed in these 4 years, and often consumes the point cloud of low density such as 1024 points for the scene. I try to apply the point cloud deep learning architecture to the task: the object detection and the 6 degree-of-freedom (DoF) posture estimation.

Finally, I integrate the 3D measurement method and robot to be the bin-picking system. In the bin-picking experiment via high-density point cloud, I utilize the $256 \times 256$ project camera system and evaluate that the proposed 3D measurement method is sufficiently accurate as a 3D robot vision sensor. In the bin-picking experiment via low-density point cloud, I try to integrate the $128 \times 128$ project-camera system and the proposed object detection and attitude estimation method together to make bin-picking faster. Moreover, I suggest the other application of the LTM, which is material segmentation, in the section.

In conclusion, this dissertation is summarized, and follow-up works are described.