Synthesis of Wide-FOV RGB-D Images by Registration and Upsampling of 3D Lidar with Omnidirectional RGB Camera

Hiroki Usami, **Shogo Miyata**, **Hideo Saito**

Keio University., 3-14-1 Hiyoshi, Kohoku-ku, Yokohama, Kanagawa 223-8522, Japan

Keywords: Free-viewpoint image, Extrinsic calibration, Upsampling

ABSTRACT

We present a method for synthesizing wide-field-of-view (FOV) RGB-D images by combining three-dimensional (3D) Lidar and an omnidirectional RGB camera. In this system, 3D point clouds captured by the Lidar are upsampled and colored by registration with the omnidirectional RGB image. We show free-viewpoint images generated via this method.

1. Introduction

In recent years, free-viewpoint videos [1] have been investigated. This allows viewers to change their viewpoint freely while watching a video, thereby providing immersive observation of the target scene. Depth image-based rendering [2] is a popular technique for synthesizing free-viewpoint images, but has some limitations. One is that a depth camera has limited field-of-view (FOV), which prevents to observe arbitrary area using the free-viewpoint functionality. Another problem is the limited number of points that can be captured by a depth camera, which is usually less than the number of pixels captured by a color camera.

Therefore, the present study proposes three-dimensional (3D) Lidar with wide FOV and an omnidirectional RGB camera to provide wide FOV when capturing a target scene, thereby enabling synthesis of wide-range free-viewpoint imagery. We present a method of extrinsic registration between a 3D Lidar and an omnidirectional camera, based on Pandey's method [3]. In addition to the registration method, we can upsample a 3D point cloud captured by the 3D Lidar according to the registered omnidirectional RGB image.

2. Proposed System

The proposed system involves two processes. The first is extrinsic calibration, which employs a 3D Lidar with wide FOV and an omnidirectional RGB camera. This procedure requires a checker-patterned plane to be observed simultaneously from a 3D Lidar and a camera from several viewpoints. The normal of the plane and 3D points lying on the surface of the plane constrain the relative position and orientation of the 3D Lidar and the camera. The other process involves upsampling the 3D point cloud captured by the 3D Lidar. The 3D point cloud is projected onto the normalized image captured by the omnidirectional camera in order to synthesize a wide-FOV RGB-D image. In this method, the pixel-projected 3D point is sparse. Therefore, our method interpolates depth value to the pixel from the depth value of the peripheral region of the pixel. This allows the creation of a high-resolution free-viewpoint image.

2.1 Calibration method

The calibration technique is similar to that proposed by Pandey (2010). In our method, we employ a checkerboard as a model plane and observe it simultaneously from a 3D Lidar and an omnidirectional camera. At this time, we calculate the normal of the plane and the distance of the model plane from the camera, which we consider as the true value. We also calculate the distance to the plane from the camera via 3D points lying on the surface of the plane. However, to calculate the distance, we need to determine extrinsic parameters, i.e., rotation and translation from the 3D Lidar to the camera. The parameter value is resolved via optimization, utilizing the true value obtained from the camera.

2.1.1 Normal of the plane

In this method, we employ the extrinsic calibration technique proposed by Zhang [4], which utilizes a checkered pattern in several poses. We observe the model plane with the omnidirectional camera to solve the Perspective-n-Point (PnP) problem, to determine an orthonormal rotation matrix that rotates the world frame into the camera frame and translation vector. For convenience, we give the equation of the plane as:

$$Z = 0$$

in the model coordinates. The rotation matrix is expressed as ${}^{c}_{m}R = [r_1, r_2, r_3]$, in which r_3 is considered as rotation related to the Z axis. Thus, r_3 is the unit normal vector of the model plane. Then, as shown in Fig. 1, the orthogonal projection of ${}^{c}t_{cm}$, which is the translation vector from camera to model plane, to r_3, is the distance to the plane from the camera. Therefore, the normal vector of the plane, N_c, is expressed as

$$N_c = (r_3 \cdot {}^{c}t_{cm}) r_3,$$

and distance is expressed as

$$\|N_c\| = r_3 \cdot {}^{c}t_{cm}.$$

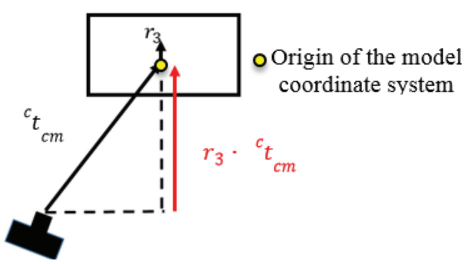


Fig. 1 Normal of the model plane

2.1.2 Registration

In this section, we obtain the relative positions and orientations of the 3D Lidar and omnidirectional camera. At first, we also calculate the distance from the camera to the plane, utilizing 3D points on the surface of the plane captured by the 3D Lidar with the true value (section 2.1.1). Then, by minimizing the error of the two values, we obtain the extrinsic parameter. Let $\{P_l^j; j = 1,2,\cdots,n\}$ be the set of 3D points lying on the plane. These points are known in the 3D Lidar coordinate system. The coordinates of these points in the camera coordinate system are given by

$${}^c_lRP_l^j + {}^ct_{cl}.$$

Here, ${}^c_lR$ and ${}^ct_{cl}$ are the required rotation matrix and translation vector, which transform 3D Lidar coordinates into camera coordinates. Then, the orthogonal projection of ${}^c_lRP_l^j + {}^ct_{cl}$ to the unit normal vector of the model plane is the required distance. We then form an equation from the two values, as:

$$F = \sum_{i=1}^{m}\sum_{j=1}^{n}\left(\frac{N_c^i}{\|N_c^i\|}\cdot\left({}^c_lRP_l^j + {}^ct_{cl}\right) - \|N_c^i\|\right)^2.$$

Therefore, solving this equation as a nonlinear optimization problem, we can determine the required rotation and translation. In our method, we use the Levenberg–Marquardt algorithm [5] to solve this problem.

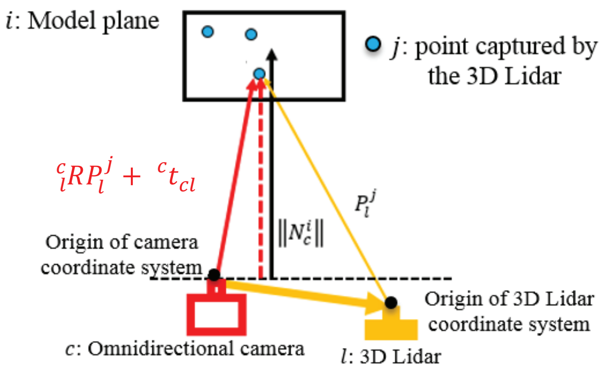


Fig. 2 Registration

2.1.3 Plane fitting

Our system uses 3D points captured by the 3D Lidar, described in section 2.1.1. Therefore, we must distinguish 3D points lying on the surface of the model plane from other points. In our method, we extract target points by taking the difference between the view of the environment installed the model plane and view of environment not installed the model plane. We then perform plane fitting to the target points using the RANSAC algorithm.

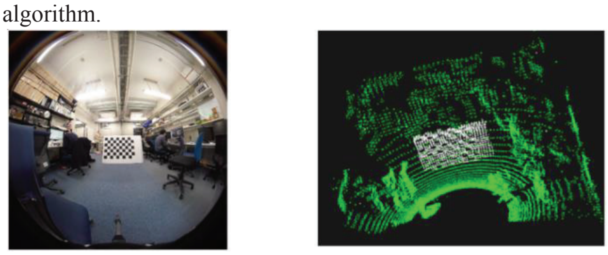


Fig. 3 Captured plane

2.1.4 Number of required planes

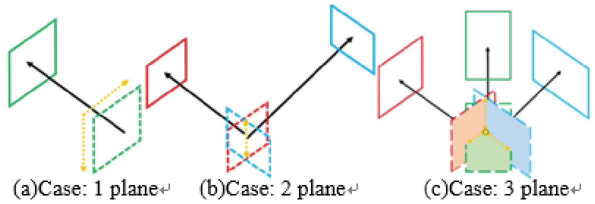


Fig. 4 Number of plane

This section discusses the minimum number of model plane views required. If only one plain is considered, the translation of the 3D Lidar along the model plane and rotation about the axis parallel to the plane normal is not constrained, as shown in Fig. 4(a). Similarly, if two planes are considered, the translation of the 3D Lidar along the line of intersection of the two planes is not constrained, as shown in Fig. 4(b). Therefore, we must set a minimum of three non-coplanar views of the model plane.

2.2 Upsampling method

A 3D Lidar captures fewer 3D points than the number of pixels captured by an omnidirectional camera, and is not sufficient to visualize the scene. Therefore, our system upsamples 3D points according to the number of pixels in the color image.

We first synthesize an RGB-D image by projecting the 3D point cloud captured by the 3D Lidar onto the normalized image captured by the omnidirectional camera. From position of the point, we obtain the depth value, which is the distance from the origin to the 3D point. The image has sparse depth values, as there are few 3D points. Our method then interpolates the pixel depth values. The proposed method focuses on the pixels of the RGB-D image. We provide a filter to the focused pixel. If there is no depth value in focused pixel, i.e., the 3D point is not projected onto the pixel, we interpolate the depth value from those of other pixels within the kernel. The weight of the kernel is determined by the distance of the pixel of the image. Then we obtain the depth value of the focused pixel by calculating the weighted mean with the filter. We thereby create an RGB-D image with a depth map that is sufficiently dense to visualize a free-viewpoint image.

3. Experiment

3.1 Environment

We conducted two experiments to test the proposed method, using 3D Lidar (HOKUYO YVT-X00, FOV 210°×40°) and an omnidirectional camera (Kodak PIXPRO SP360 4K, resolution 1440×1440 pixels, 230°×180°). The 3D Lidar and the omnidirectional camera were combined as shown in Fig. 5(a).

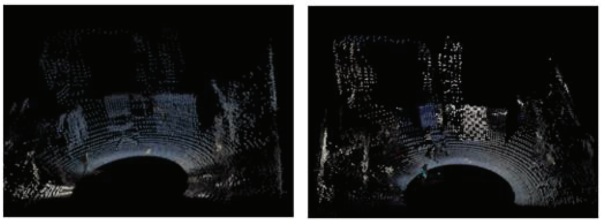
As a first experiment, we calibrated the system using the proposed extrinsic calibration method, obtaining three views of the model plane. In the second experiment, we upsampled the free-viewpoint image via the proposed method. In these experiments, the 3D Lidar captured 10562 points.

(a)Combined system (b)captured environment

Fig. 5 Experiment environment

3.2 Result of calibration method

Following the first experiment, we compare the free-viewpoint image with proposed calibration method to the results without calibration (see Fig. 6). In the calibrated free-viewpoint image, the point cloud successfully colors the chair and checkerboard in the foreground of the input color image. The results confirm that our method obtains correct rotation and translation parameters.



(a)not calibrated (b)calibrated with our method

Fig. 6 Colored 3D point cloud

3.3 Result of upsampling method

The result of the upsampling experiment utilizing the proposed method is shown in Fig. 7 (containing 306386 points). Comparing the number of points in Fig. 6 and Fig. 7 confirms that our method successfully upsampled a free-viewpoint image; and that the upsampled image is sufficient to visualize the scene, as shown in Fig. 8.

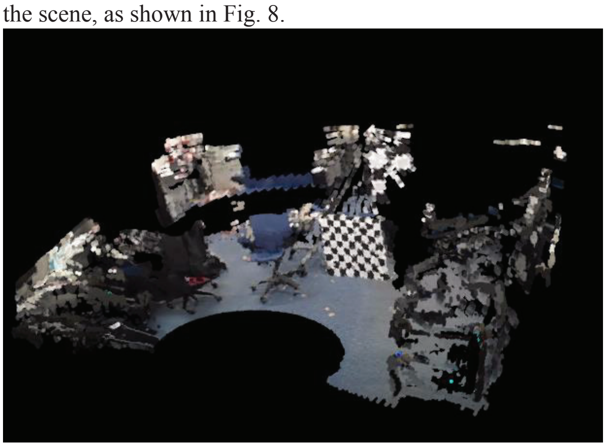


Fig. 7 Upsampled 3D point cloud

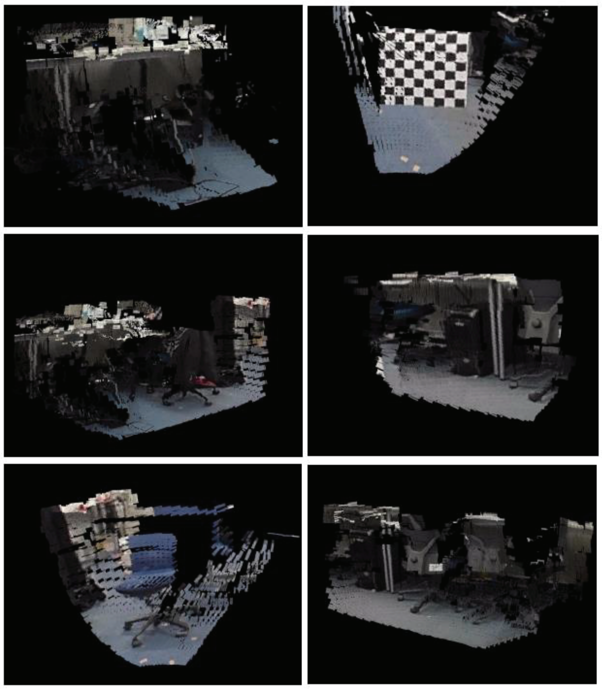


Fig. 8 Free-viewpoint images

4. Conclusion

In this paper, we presented an extrinsic calibration method to estimate rotation and translation from 3D Lidar to an omnidirectional camera; and an upsampling method to create high-resolution RGB-D images from images captured by the omnidirectional RGB camera and the 3D point cloud captured by the 3D Lidar. The two methods were then successfully combined to visualize a high-resolution, accurately colored free-viewpoint image.

REFERENCES

[1] J.Carranza, C.Theobalt, M.A.Magnor, and H.P.Seidel. “Free-viewpoint video of human actors.” *ACM Transactions on Graphics (TOG)*, pp. 569-567, 2003

[2] C.Fehn. “Depth-image-based rendering (DIBR), compression, and transmission for a new approach on 3D-TV.” *Electronic Imaging 2004*. International Society for Optics and Photonics, pp. 93-104, 2004

[3] G.Pandey, J.McBride, S.Savarese, and R.Eustice. Extrinsic calibration of a 3D laser scanner and an omnidirectional camera. *7th IFAC Symposium on Intelligent Autonomous Vehicles*, pp. 336-341, 2010

[4] Z.Zhang. “A flexible new technique for camera calibration.” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1330-1334, 2000

[5] K.Levenberg. “A method for the solution of certain problems in least squares.” *The Quarterly of Applied Mathematics 2*, pp. 164-168, 1944