https://doi.org/10.2352/ISSN.2470-1173.2018.18.3DIPM-426

Synchronizing 3D point cloud from 3D scene flow estimation with 3D Lidar and RGB camera

Hiroki Usami [1], Hideo Saito [1], Jun Kawai [2] and Noriko Itani [2]; [1]Keio University, Japan; [2]Fujitsu Laboratories LTD, Japan;

Abstract

We present a method for synchronizing three-dimensional (3D) point cloud from 3D scene with estimation using a 3D Lidar and an RGB camera. These 3D points sensed by the 3D Lidar are not captured at the same time, which makes it difficult to measure the correct shape of the object in a dynamic scene. In our method, we generate synchronized 3D points at arbitrary times using linear interpolation in four-dimensional space, time and space. For interpolating the 3D point, we obtain corresponding 3D point matching with the pixel value captured by the RGB camera in a continuous frame. The experimental results demonstrate the effectiveness of the presented method by depicting a synchronized 3D point cloud that is correctly shaped.

1. Introduction

In recent years, free-viewpoint videos [1] have attracted attention in the field of imaging technology. These videos allow viewers to change their viewpoint freely while watching the video. Three dimensional shape recovery based on multiple viewpoint stereo methods is generally performed using multiple RGB cameras, however, the recovered 3D structure is not always captured because the objects lack texture. Other problems such as a limited view angle due to the necessity of a shared field of view and calculation costs are also issues associated with the use of multiple viewpoint stereo methods.

In recent years, various 3D structure sensing devices, such as 3D Lidar, have been available [2]. Using the 3D Lidar device, 3D information can be easily and quickly obtained by Time-of-Flight (TOF) using laser light. The recent development of TOF technologies enable the 3D Lidar to capture a considerably wide angle of view. Moreover, because of its ability to depict the 3D shape of the subject from the captured data, it is expected that the 3D Lidar will have multiple applications in various fields, including automatic scoring of sporting events and recognition of objects by automatic vehicles.

Although 3D Lidar based on TOF technologies provides accurate 3D structures, the 3D Lidar cannot accurately measure a dynamically moving scene. Because the 3D Lidar scans the laser beam and sequentially performs data measurement by exposure, each point of the moving scene cannot be measured at the same time, thus resulting in distortion of the measured three-dimensional shape. In order to suppress such distortion, it is significantly effective to conduct time synchronization of the captured data. Therefore, in this paper, we propose a method to synchronize the point cloud by estimating the motion of the moving object. In this method, we employ an RGB camera without a rolling shutter to estimate the movement of each 3D point by matching the 3D point and the RGB image.

2. Previous works

In this paper, the 3D shape correction is performed by conducting time synchronization of the captured 3D point cloud.

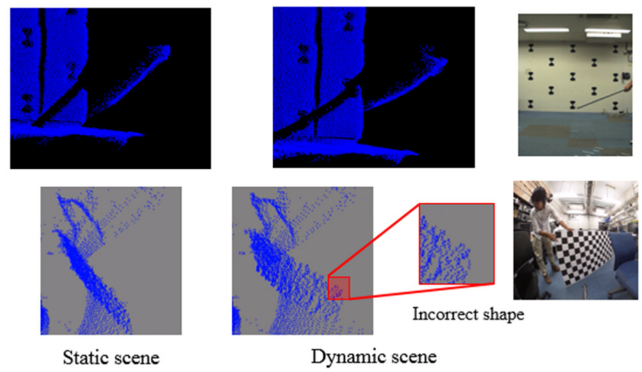


Figure 1. Effect of rolling shutter

When time synchronization is performed, it is necessary to estimate the movement of the 3D point cloud. Several studies have examined such 3D scene flow. One general method involves matching the 3D structure of the scene along the time series. However, this method can only be used when a dense 3D structure can be captured. Vedula et al. proposed a method for estimating 3D scene flow based on two-dimensional (2D) optical flow of an RGB image [3]. This technique extends the optical flow to the 3D scene flow by combining the geometric information of the scene captured in the image. Even with this method, geometrical information is necessary and depends on optical flow accuracy.

Also, in the current situation, it is a problem that objects are distorted due to the rolling shutter, and errors occur in the photographing of moving objects. Several studies have examined the phenomena that occur due to the rolling shutters in RGB cameras [4], but little research has been done on 3D Lidar.

Therefore, as a purpose of the research, correcting the distortion of the 3D shape caused by the measurement of the moving object is considered to be important. In this method, the influence of sequential exposure is reduced by creating a frame in which captured times of each point of the 3D point cloud are synchronized.

3. 3D Lidar

In the future, 3D image technology will be further developed, and the 3D Lidar, which will be capable of easily acquiring 3D images, will attract more attention and become more widespread. The device will be able to capture various scenes. In the field of sports, automatic scoring technology using 3D data is being studied for assigning scores in competitions such as gymnastics.

However, the point cloud obtained from the 3D Lidar is sparse, and the structure of the moving object cannot be acquired accurately by sequential exposure. Most existing 3D scene flow estimation methods deal with things that require a dense 3D point cloud, and treat each point and each pixel in the captured frame as taken at the same time. The fact that the moving object is distorted is not considered.

Therefore, in this paper, we consider the difference in the captured time of each point in the use of the 3D Lidar. In a situation where an accurate 3D structure is not known, we propose a method of finding the 3D scene flow by finding the corresponding 3D point in the frame along the time series. Furthermore, by applying linear interpolation in a four-dimensional (4D) space including time, we propose a method of generating synchronized 3D point clouds to correct the 3D shape. Our method is novel in that each point in a frame is subjected to four-dimensional interpolation by utilizing the fact that it has different captured times due to the influence of the rolling shutter.

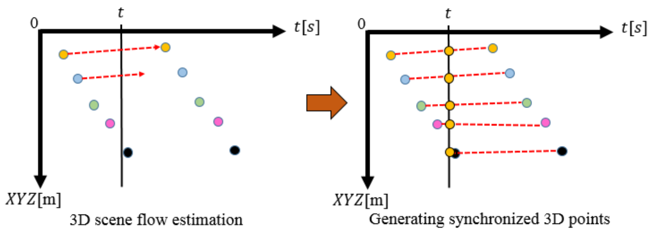


Figure 2. Object in this work.

4. Proposed method

The movement of the point cloud is estimated by obtaining a 3D corresponding point indicating the same region in consecutive frames along the time series acquired by the sensor. In this method, the pixel value of the RGB image is used in order to correspond the 3D points. It is also supposed that analysis is necessary in a scene in which the object to be captured is known beforehand. For this reason, it is considered effective to preliminarily set constraints according to the object to be captured. In this paper, we also propose a method in the case where dynamic objects in a scene can be handled as a plane.

4.1 Proposed 3D scene flow method

In this section, we describe our original 3D scene flow method. Our method contains five processes.

1. First, we distinguish dynamic points and static points from the 3D point cloud obtained by the 3D Lidar. In this method, a dynamic region is obtained by acquiring a background 3D point cloud in advance and performing background subtraction. By preliminarily obtaining a dynamic region, it is possible to apply the processing in this method only to the dynamic region, and it is possible to reduce the processing weight.

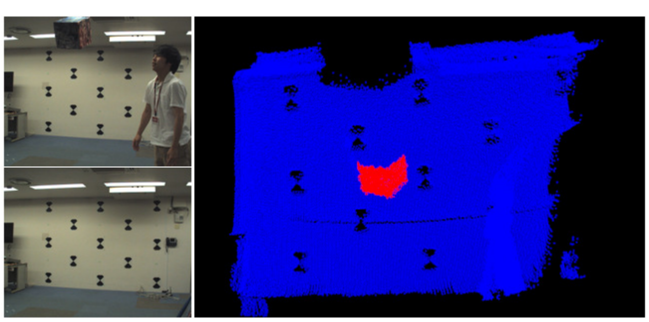


Figure 3. Dynamic points by background subtraction

2. Next, we determine a focused point from the 3D point cloud, and candidates of corresponding point in the next frame are selected from a particular threshold.

3. Next, using the focused point and each of the next frame points obtained in 1, four-dimensional, $X = [x, y, z, t]$, linear interpolation of equation

$$X = X_f + \frac{t-t_f}{t_c-t_f}(X_c - X_f) \quad (1)$$

is applied to calculate a set of 3D coordinates at the time of the adjacent previous time and the time of the RGB image that is captured next. X_f refers to the focused point, and X_c refers to the corresponding point in the next frame.

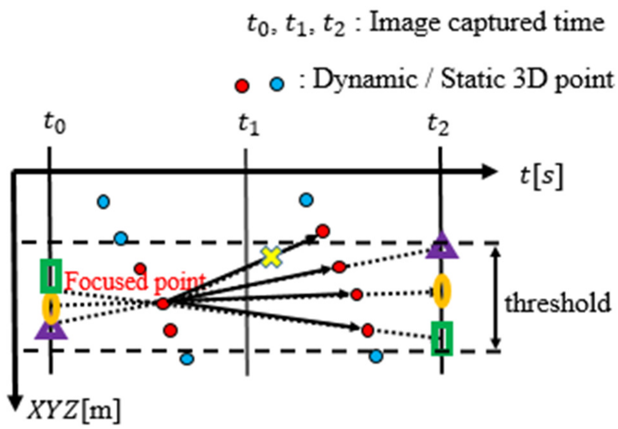


Figure 4. Search for correspond point

4. By projecting the 3D point obtained in 2 to the image coordinates and using the point with the smallest difference between the pixel values at the previous time and the next time as the corresponding point, the focused point can be moved to which three-dimensional point in the next frame or not.

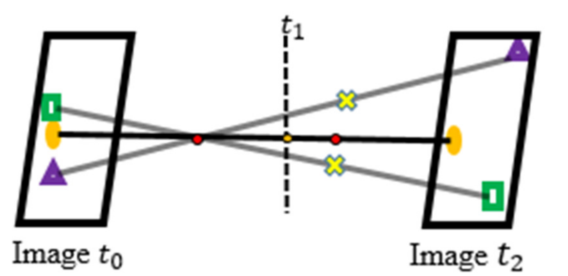


Figure 5. Pixel value matching

5. Finally, we compute the point cloud that is synchronized by calculating the displacement at the specified time of each point from the 3D scene flow of the point cloud.

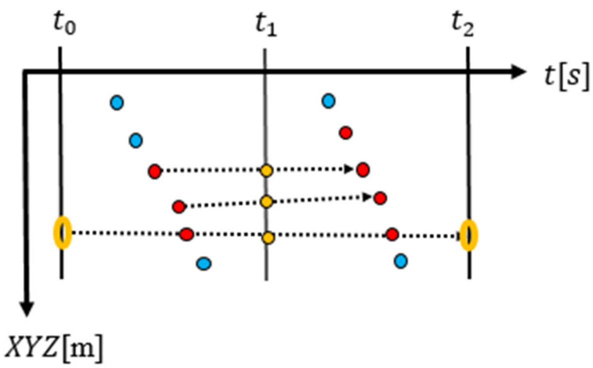


Figure 6. Generating synchronized points

4.2 Up-sampling method with planarity constraint

In this method, it is considered that effective interpolation can be performed when the shape of the object captured can be handled as a plane. Therefore, in this method, planarity is used. In the 3D scene flow estimation method proposed in 4.1, when correcting matching of corresponding points fails, correct interpolation is not performed and the number of point clouds may decrease. Therefore, in this section, we propose a method to increase the 3D points according to the number of RGB image pixels. In this method, super pixel segmentation processing is introduced into RGB images [5]. By introducing this, the region of the moving object that is the subject of processing in this method can be handled as a plurality of segments, and interpolation of the point cloud and the corresponding region of the image can be performed through an easy process. The synchronized3D point cloud generated by the method of 4.1 is projected onto the RGB image and the plane equation is calculated from the 3D point projected in each super pixel. Using this plane equation, pixels on the image are converted to points on the 3D plane.

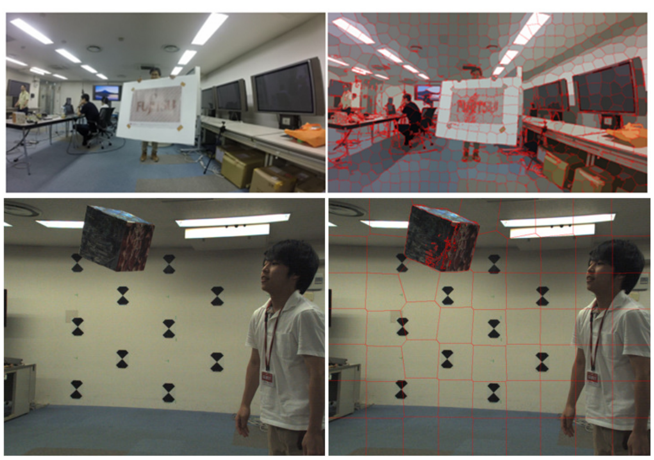


Figure 7. Super pixel segmentation

5. Efficiency of the method

We conducted two experiments in order to confirm the effectiveness of our methods. In our experiments, we used an RGB camera and a 3D Lidar, which were calibrated in advance.

Figure 8. Experiment environment

5.1 Experiment with large plain

In the first experiment, we captured dynamic scenes rotating a large plane using an RGB camera (GoPro HERO3+, 30fps, resolution 1920×1080) and a 3D Lidar (Fujitsu, 30fps, FOV $160° \times 140°$, resolution 300×200). The experiment was conducted in the environment shown in Figure 9.

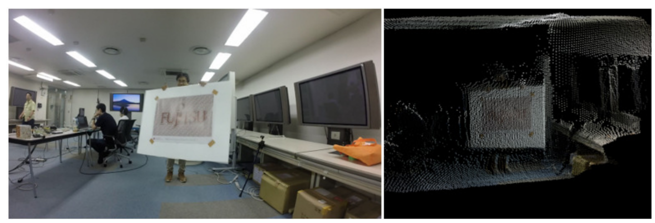


Figure 9. Captured scene

To reduce the effect of the rolling shutter, we generated synchronized 3D points using the proposed method. Figure 10 shows the synchronized 3D point cloud at an arbitrary time obtained by our method. It shows a front view and a side view. The blue points indicate the static region and they are taken as the background. As can be seen from the side view, the plane is also generated from the synchronized point cloud by this method.

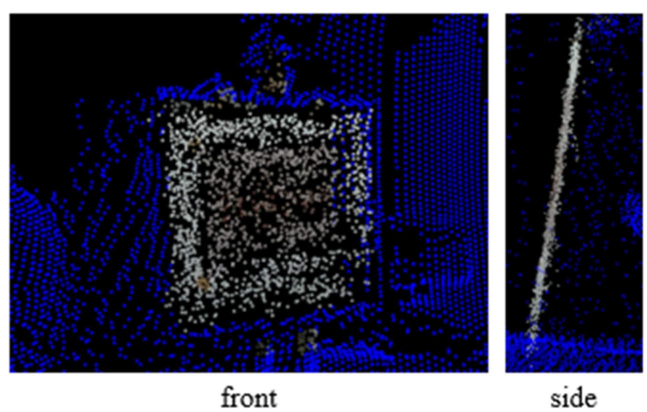


Figure 10. Synchronized 3D point cloud

Figure 11 shows the 3D point cloud generated by our method using the planarity constant.

Figure 11. Up-sampled synchronized 3D point cloud with planarity constraint

As can be seen from comparison with Figure 10, which is not up-sampled, the resolution has been successfully increased. As a result of this method, we can increase the number of 3D points according to the number of pixels the segment contains.

To demonstrate the effect of synchronizing 3D points, Figure 12 shows the comparison of the 3D point cloud captured by the 3D Lidar, colored green and pink, and synchronized by our method. In order to make it easier to compare with the original data, the three point clouds are color-coded and output at the same time.

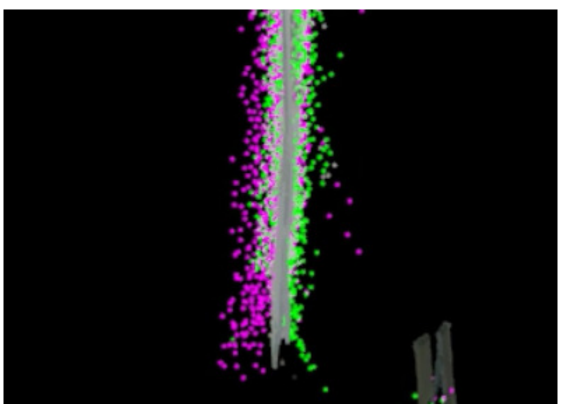


Figure 12. Comparison of the 3D point cloud

Using this method from the point cloud of the previous frame shown in green and that of the next frame shown in pink, a synchronized point cloud indicated the time between them in white, which shows that the interpolation was correct.

5.2 Experiment with box

We also conducted experiments using a box. In the experiment in 5.1, there was only one plane, but experiments were conducted to confirm the effect of this method for more complicated objects. A cube was used as an object composed of a plurality of planes. In this experiment, we used an RGB camera (ImagingSorce DFK33GP1300, 30fps, resolution 1280×1024) and a 3D Lidar (Fujitsu, 30fps, FOV $36° \times 28°$, resolution 300×200). We threw the box into the air and captured the scene.

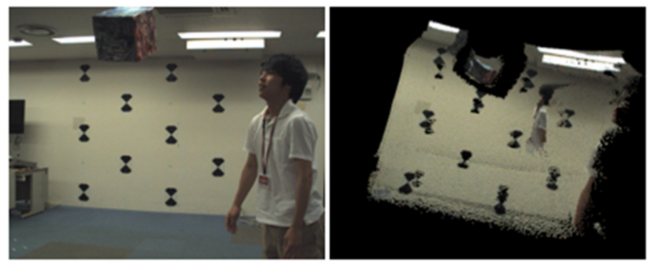
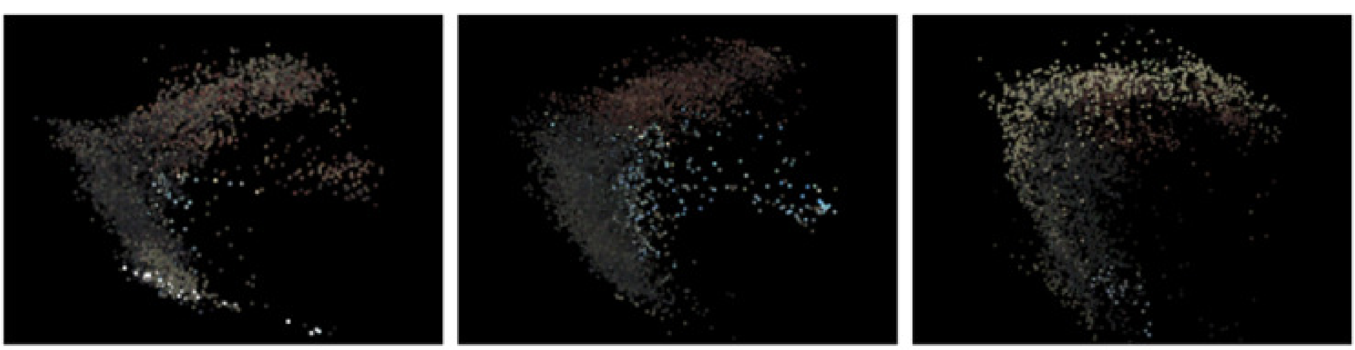


Figure13. Captured scene

The result of generating the synchronized 3D point cloud is shown in Figure 14. The left and right figures are the original data captured by the 3D Lidar and correspond to the previous frame and the next frame used to generate the synchronized point cloud. The middle figure is the synchronized point cloud generated by the proposed method. As the planes that make up the box are generated and the plane intersects perpendicularly, it can be seen that distortion can be reduced.

Also, in order to make it easier to compare with the original data, the three point clouds are color-coded and output at the same time as shown in Figure 15. Using this method from the point cloud of the previous frame shown in green and that of the next frame shown in blue, a synchronized point cloud indicated the time between them in red. In the 3D Lidar, two planes of the box are measured, but the two planes are located between the previous frame and the next frame, which demonstrates that the interpolation is correct.



(a) (b) (c)

Figure 14. Comparison of 3D point cloud: a) Captured 3D point cloud in the previous frame, b) Synchronized 3D point cloud by our method, c) Captured 3D point cloud in the next frame

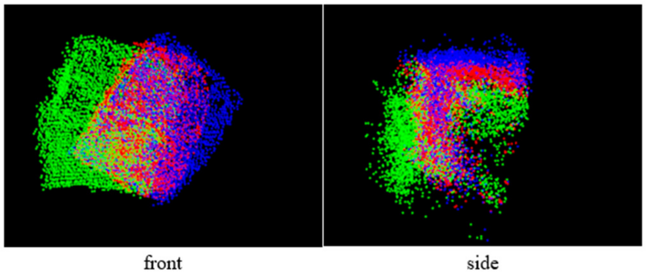


Figure 15. Comparison at the same time

Figure 16 shows the 3D point cloud generated by our method using the planarity constraint.

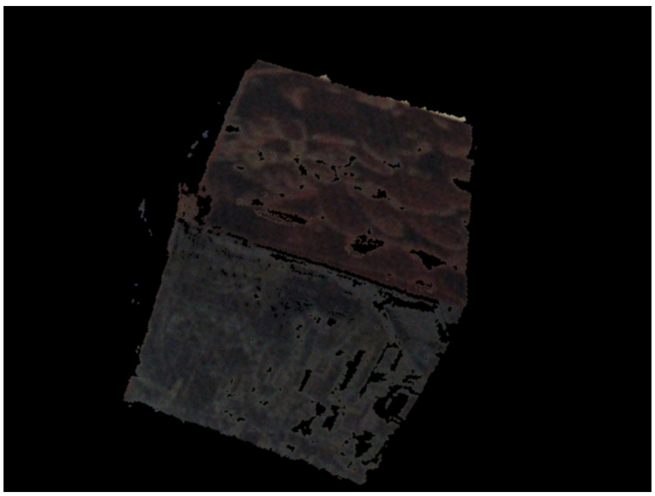


Figure 16. Up-sampled synchronized 3D point cloud

By using this method, even when a box was captured, it was possible to achieve a high resolution of two planes detected by laser. This experiment showed that this method is effective for objects composed of planes. In addition, because point clouds with high resolution are generated on the plane in 3D space, it is possible to interpolate the object without generating distortion.

For demonstrating the effect of the up-sampling method, Figure 17 shows the comparison of the 3D point cloud synchronized using our method, the 3D point cloud that is colored red, and the 3D point cloud that is up-sampled. There was a problem in that the number of synchronized points generated by our method decreased due to matching errors and the accuracy of the original data captured by the 3D Lidar. Due to this influence, the surface of the box indicated by the point cloud generated by our method, shown by the red points in Figure15, is smaller than the original data, the blue or green points, captured by the 3D Lidar. However, by using this up-sampling method, pixels corresponding to the boxes shown in the color image are converted to 3D points, so that it is also possible to generate a region of the surface on which the synchronized point cloud cannot be generated.

It can also be seen from this comparison that the original data of the 3D point cloud is distorted in the outline portion of the surface, but not in the up-sampled point cloud.

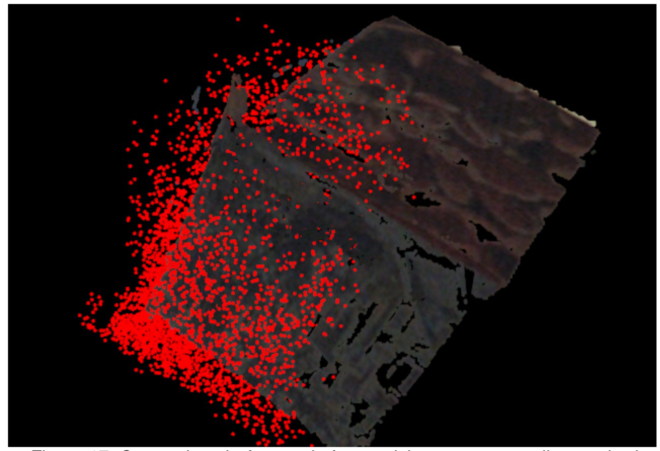


Figure 17. Comparison before and after applying our up-sampling method

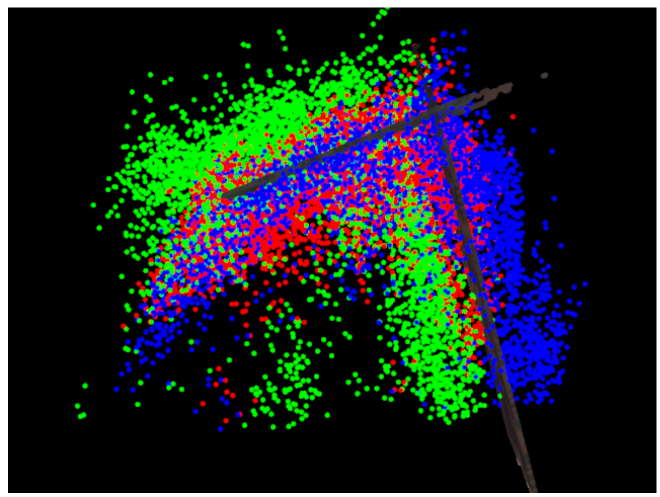


Figure 18. Comparison at the same time.

5.3 Experiment with stick

We also conduct experiments using a stick. These experiments were conducted in the same experimental environment as 5.2. We shook the stick and captured the scene.

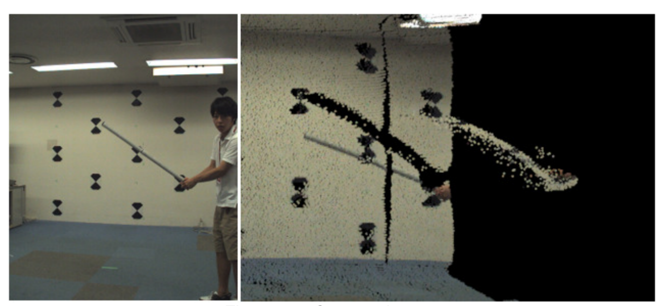


Figure 19. Captured scene

In the input 3D point cloud, the stick is bent, and the RGB values are also not correctly corresponded. This is probably because the shutter speed does not catch up with the movement speed of the stick and because the target area is narrow, the RGB value completely deviates from the area due to small deviation.

The result of generating the synchronized 3D point cloud is shown in Figure 20. The green point cloud is the experimental 3D region in the previous frame, blue is the next frame, and the red point cloud is the synchronized point cloud generated by this method.

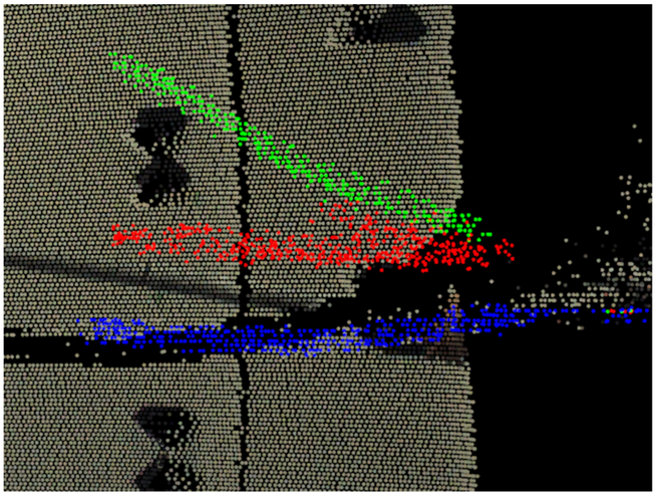


Figure 20. Comparison at the same time

Regarding the depth direction of the laser, although the discussion is not made because the error on the original data is large in principle of the laser, the 3D point cloud is generated at the position between the previous frame and the next frame.

The point cloud projected color image taken at the same time as the generated synchronized point cloud using our method is shown in Figure21.

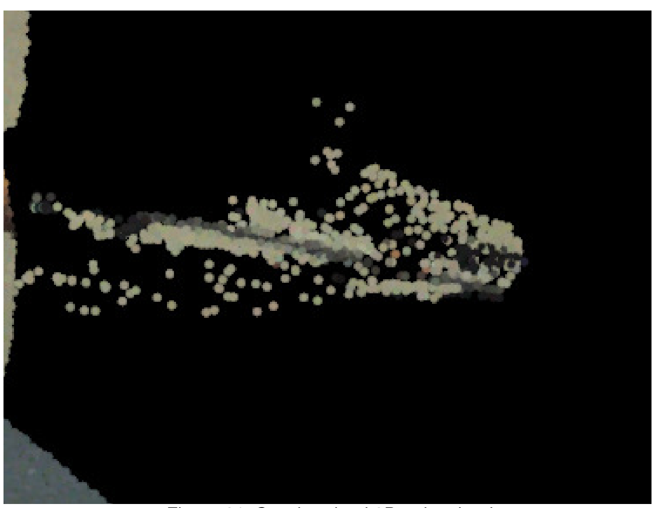


Figure 21. Synchronized 3D point cloud

From this figure it can be seen that the RGB values, gray, on the stick are projected onto the synchronized 3D point cloud. Thus the movement of the point cloud between the adjacent frames can be estimated.

However, because there are many errors in this point cloud, it cannot be confirmed that the distortion of the stick has been taken. This is probably because the point cloud on the stick greatly moved between the frames, so the accuracy of motion estimation by image pixel matching decreased. However, if it is a thin stick, and if several correct interpolation points can be taken, the position of the point cloud can be easily estimated from the shape of the stick.

6. Conclusion

In this research, we proposed an algorithm to reduce the effect of sequential exposure on 3D point cloud data, which is a problem caused by the 3D Lidar structure. We proposed a method for estimating the 3D scene flow of objects by using a 3D Lidar and an RGB camera simultaneously to generate a 3D point cloud with synchronized time.

In the experiment using the large plane, it was confirmed that data with reduced data distortion was created by generating the synchronized 3D point cloud. In addition, we generated an up-sampled 3D point cloud using our method with planarity constraint. Experiments using boxes and experiments on the plane confirmed the effect of reducing distortion and increasing resolution. As a result, we confirmed that this method is effective not only for one plane but also for an object composed of multiple planes.

Also, we conducted experiments using a stick to examine nonplanar objects. In this experiment, we could generate a synchronized 3D point cloud from the movement of the stick, but we also encountered problems such as accuracy. After overcoming this problem, this method should be applied to human body movements such as those of the arms and legs, in future studies.

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

[3] S. Vedula, S. Baker, P. Rander, R. Collins, T. Kanade, Three-Dimensional Scene Flow. Proceedings of the 7th International Conference on Computer Vision, Vol.2, September, pp. 722-729, 1999

[4] S.Baker, E.Bennett, S.B.Kang, R.Szeliski, Removing rolling shutter wobble, Computer Vision and Pattern Recognition (CVPR) pp. 2392-2399, 2010

[5] R.Achanta, A.Shaji, K.Smith, A.Lucchi, P.Fua, Slic superpixels. No. EPFL-REPORT-149300, 2010

Author Biography

Hiroki Usami received his BSc(IT) degree in information and computer science from Keio University, Japan, in 2016. Since 2016, he has been a master's student in science and technology at Keio University, Japan. His research interests include image processing, 3D point cloud processing, and computer vision