PROCEEDINGS OF THE INTERNATIONAL DISPLAY WORKSHOPS, VOL.28, 2021

Luggage Object Classification from Range Images

Akinobu Watanabe

akinobu.watanabe.fz@hitachi.com
Hitachi, Ltd.
Keywords: TOF, Object, Classification

ABSTRACT

We developed the luggage object classification technique from range image captured by TOF sensor and confirmed the accuracy of 10 class classification is about 98% and loss is 3% by the 3D-data based model trained by small point cloud data.

1 Introduction

In the analyses of the operation of workers in the manufacturing premise and the customer action in the retail store, the inflection of provided range image data by TOF (Time of Flight) sensor is expected. Many techniques are suggested as a person posture estimate using range image data, but, as for the use case in the point of view looking down, examination does not advance enough.

1.1 Background

A sensor apparatus and an IT system have been developed and become low price. An IoT (Internet of Things) market using Information collected with them from a machine, a vehicle and a building, to use the information for analysis and control is spreading.

With progress of IoT, there is the movement that is going to feed back the result of future prediction by collecting the data of a machine and the facilities which are on-site physically and reappearing as “digital twin” within the cyber world of the IT system, using an information processing technology. Siemens and GE stimulate research and development in conjunction with the digital twin, too and have begun to already send information in a general medium [1][2].

1.2 Purpose

It was aimed for the establishment of the luggage object classification technology suitable for an available privacy sensitive use with the general-purpose 3d sensor including the TOF sensor in order to solve the restriction of RGB camera raised on privacy issues.

1.3 Target

In this study, we intend for processing to classify luggage objects from the range image.

2 Experiment

In this study, we captured range images of luggage, removed background point cloud, created luggage object dataset, trained deep learning model and predict the luggage class.

2.1 Previous method

As shown in the prior publication [3], we developed detection function carrying or wearing something or not with 3D point cloud captured by TOF sensor [4] in real-time on general environment.

2.2 Issue of accuracy and robustness

The previous method can detect human position and existence of luggage near the person. As the next step, we need to detect luggage itself even with no pedestrian carrying it.

Such method as PointNet [5] is known to recognize objects from 3D point clouds. These techniques recognize objects directly from a 3D point cloud. Recognition accuracy is excellent, but processing cost could be high because the number of point cloud is large.

In this study, we decided to consider a method that can process real-time at a lower processing cost because it was intended to be applied to a system with a cheaper and simpler configuration.

Then we decided to develop classification function of luggage with 3D point cloud in real-time on general environment.

2.3 Data creation condition

Table 1 shows the experimental environment.

Table 1 Experimental environment

Object	TOF Sensor	Sensor Height	Sensor Angle
10 class (Suitcase Tool case Shoulder bag etc.)	HLS-LFOM5	2.4[m] from floor (1.8[m] from rotation base)	X=30 degree

In table 1, Object is configured with 10 kinds of luggage such as suitcase, tool case, shoulder bag and so on. 3D LiDAR sensor is one TOF sensor provided by Hitachi LG Data Storage Inc., which is set 2.4[m] high from floor. Sensor direction is 30 degrees raised from the vertical downward axis.

2.4 Point cloud data extraction

To obtain point cloud data effectively, each object is put on rotation base, and turned around slowly on the base to measure and get the point cloud of the object.

To exclude other points than of the object, such as floor or experimenter, the cylinder shape of space is defined to extract point which is usable as the dataset of the object.

Fig. 1 shows the screen shot of data creation tool. The upper left part is top view of 3D point cloud including object and experimenter. The circle shows the boundary surface of the side of cylinder area. The lower left part shows the front view and shows the base of cylinder area as the yellow and magenta lines. The middle white line is the floor level. The upper right part is IR view, and the lower right part is Depth view. Motion extraction method is enabling to remove the still points, with the same manner of the previous study.

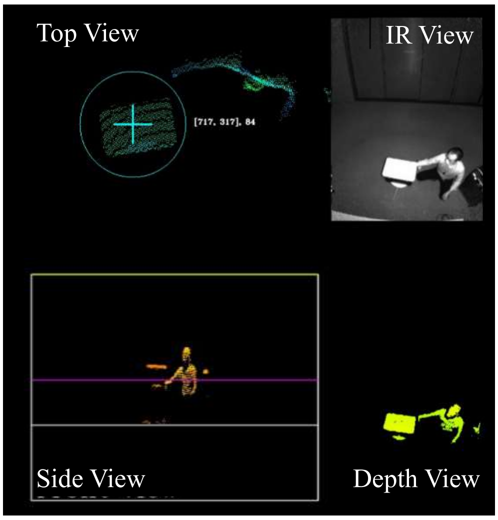

Figure 1 Data Creation Tool

2.5 Dataset

Fig. 2 shows images of 10 objects to be classified in this study. And Fig. 3 shows mapping of dimensions of each object.

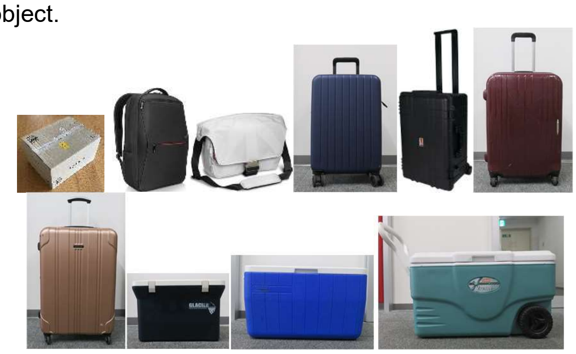

Figure 2 Image of Object (upper row: #1-5, lower row: #6-10)

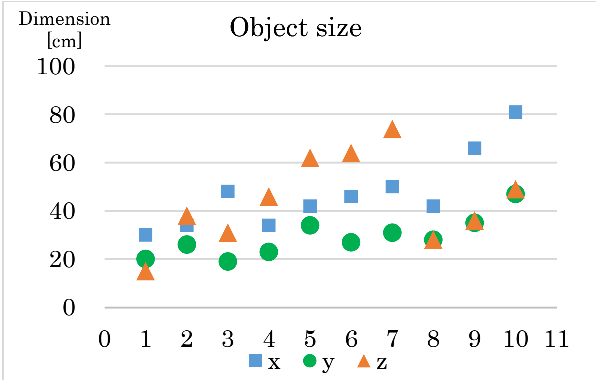

Figure 3 Object size

The data set created for this experiment is shown in Table 2.

Table 2 Dataset of 10 class object

#	Object	Dimension [cm]			Frame num	Point num per frame
1	Small box	30	20	15	84	267
2	Backpack	34	26	38	386	826
3	Shoulder bag	48	19	31	250	1369
4	Suitcase S	34	23	46	414	960
5	Tool case	42	34	62	325	428
6	Suitcase M	46	27	64	356	1076
7	Suitcase L	50	31	74	462	2539
8	Cool box S	42	28	28	135	1222
9	Cool box M	66	35	36	227	2480
10	Cool box L	81	47	49	176	3580

2.6 Modeling

We adopted a model using PointNet network configuration [6].

The final output is 10 class classification. And we used TensorFlow [7] as the framework for deep learning.

2.7 Sampling

To train prediction model, small number of points are selected from each objects' point cloud data. For example, 64 points are randomly selected from 267 points in case of typical #1 “Small box” object. And repeating this selection enables sampling various combinations of point clouds from the same data. This approach helps to increase the number of train data and test data.

In this study, 90% of data is used for train data, and the remaining is for test data.

2.8 Dataset subjunction

To increase dataset size and robustness, subjunctive data are created using measured point cloud data, by shifting the position of each point based on normal distribution.

3 Results

Fig. 4 and Fig. 5 shows the results of classification accuracy and loss for each condition. It shows the average of 16 attempts.

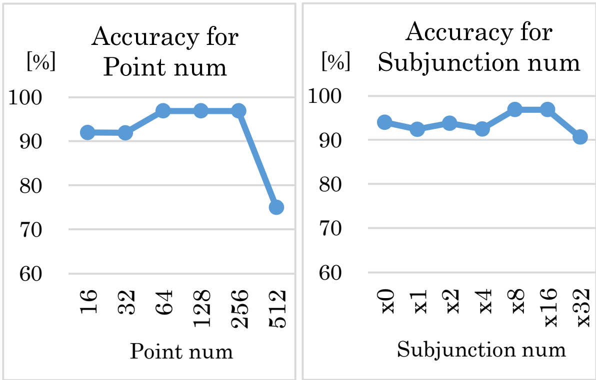

Figure 4 Accuracy for Point num (left) and Subjunction num (right)

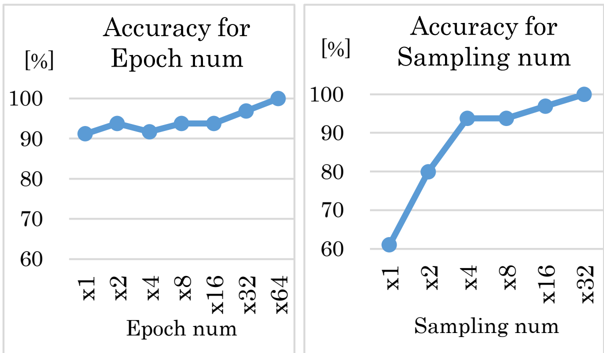

Figure 5 Accuracy for Epoch num (left) and Sampling num (right)

Fig. 6 shows prediction time of previous study and this study.

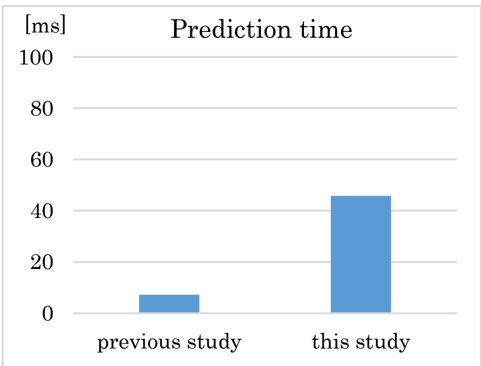

Figure 6 Prediction time

Fig. 6 shows that previous study took 7.27[ms] and this study takes 45.78[ms].

Previous study recognized 3 class objects which are carried by pedestrian, and its network configuration was CNN, 2D based recognition method. On the other hand, this study recognizes 10 class objects with PointNet network, 3D based recognition method.

Table 3 shows a comparison of accuracy and loss between previous study and this study.

Table 3 Z-R Histogram and Object Classification

	Z-R Histogram for 3 class classification	This Study for 10 class classification
Accuracy	98.3	98.0
Loss	6.9	3.2
Input Data Size [byte]	3072	768
Prediction Time [ms]	7.27	45.78

In Table. 3, although dataset and class number are different each other, we confirmed that the accuracy is almost same, and loss of this study is better.

Regarding prediction load, about 6 times long time is needed for single object prediction of this PointNet method.

4 Discussion

The best combination of the numbers of point, subjunction and sampling that provides the good result of accuracy is 64, x8 and x8. The x16 and x32 cases of sampling num are considered as over fitting cases, because the training history shows the validation loss and validation accuracy are not converged.

The classification accuracy and loss are almost same as the previous study, then we can apply both previous method and this method selectively to the real-time luggage recognition system.

Prediction load of this study is about 6 times longer than the previous study. To realize real-time prediction, all transaction such as moving object extraction, segmentation and classification for single frame should be completed in 100ms, because TOF sensor can take 3D point cloud data 10fps.

We need to reduce both of classification time and segmentation time under 100msec.

Regarding this data set, the number of data frame is different each other. To reduce the affects of this aspect, we need to enlarge the number of data frame as the same level.

5 Conclusions

The luggage classification technique from range image captured by TOF sensor achieved the accuracy of 10 class classification is 98% and loss is 3%. On the other hand, the prediction load is 46ms, 6 times longer than previous method.

As the next step, we are developing luggage segmentation technique based on this study to realize real-time luggage recognition system on general environment.

References

[1] Digitalization in machine building - The digital twin (http://www.siemens.com/customer-magazine/en/home/industry/digitalization-in-

machine-building/the-digital-twin.html)
[2] 'Digital Twin' Technology Changed Formula 1 and Online Ads. Planes, Trains and Power Are Next (http://www.gereports.com/digital-twin-technology-changed-formula-1-online-ads-planes-trains-power-next/)
[3] A. Watanabe, 'Carried Objects Recognition from Pedestrians' Range Images ', IDW '20
[4] Hitachi-LG Data Storage 3D LiDAR TOF Sensor (https://hlds.co.jp/product-eng/)
[5] PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation (https://arxiv.org/abs/1612.00593)
[6] Point cloud classification with PointNet (https://keras.io/examples/vision/pointnet/)
[7] TensorFlow (https://github.com/tensorflow/tensorflow)