Two-Level Processing Scheme for 3D-Image Sensing Network

Chongyu Li, Ryoichi Shinkuma, Takehiro Sato, and Eiji Oki
Graduate School of Informatics, Kyoto University, Kyoto, Japan

***Abstract*—This paper proposes a two-level processing scheme for three-dimension-image sensing. The first level processing selects only spatial regions needed for a smart monitoring task to reduce the total volume of data traffic. The second level processing integrates multiple (physical) image sensors into a virtual one to improve the delay and jitter performance in the real-time transmission of data from sensors to the cloud server. We develop a prototype system to implement the proposed scheme. Our demonstration validates that the proposed processing scheme works better than the benchmarks which do not adopt the two-level processing.**

***Index Terms*—smart monitoring, 3D-image sensing network, point cloud selection, merging point clouds**

I. Introduction

Smart monitoring is a key component in smart cities to provide services such as prediction and prevention of traffic accidents on public roads [1], [2]. An image sensing network using three-dimension- (3D-) image sensors, so-called light detection and ranging (LIDAR), is used in a smart monitoring system. The network consists of a local system of devices each equipped with an image sensor and an edge server, which are connected via a local network, and a cloud server connected with the local system via the Internet. Using multiple image sensors is essential to prevent blind area caused by obstacles.

However, conventionally, since data collected by individual image sensors are uploaded to the cloud server directly and independently [3], the total data volume can exceed the network bandwidth. Moreover, since, in the conventional architecture, data from multiple image sensors are aggregated at the cloud server, which experience delay and jitter. The scheme that deals with limited network bandwidth, delay and jitter, is required to achieve the real-time 3D image sensing.

This paper proposes a two-level processing scheme for real-time image sensing. The first level processing performs selection of spatial regions needed for a smart monitoring task. The second level processing integrates multiple (physical) image sensors into a virtual one. The first and second level processing is operated at each device equipped with an image sensor and an edge server, respectively. We develop a prototype system to implement the proposed scheme. Our demonstration verifies that the proposed scheme works effectively for providing image data of sufficient quality against strict bandwidth limitation.

II. Proposed scheme

A. 3D-image sensing network

Fig. 1 illustrates a 3D-image sensing network with the proposed scheme. It consists of multiple image sensors, multiple devices, an edge server, and a cloud server. The devices and the edge server are connected via a local network, while they are connected to the cloud server via the Internet. Each image sensor generates streaming data and sends them to its connected device. The device performs the first level processing and then transmits the streaming data to the edge server. The edge server performs the second level processing and then transmits the streaming data to the cloud server. The cloud server uses the received data for a smart monitoring task such as object detection. We will explain the details of the two-level processing scheme in the next subsections.

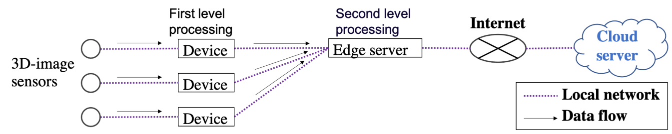


Fig. 1: 3D-image sensing network with proposed scheme

B. First level processing

For the first level processing, each device performs data selection to deal with limited network bandwidth. A field-of-view is decided, which is a part of the 3D space. We divide this field into voxel grids. A *selector set* is predefined by selecting voxels belonging to the selected area. Points inside any one of the voxels in the *selector set* are kept and those outside of all the voxels in the *selector set* are removed. The selected data are transmitted to the edge server in the form of data frames.

C. Second level processing

For the second level processing, the edge server defines virtual frames (frames that merge data frames from multiple devices), and merges data frames from multiple image sensor devices to deal with delay and jitter. Each device embeds a timestamp into each data frame when the device extracts the frame. As illustrated in Fig. 2, the duration of each virtual frame is predefined on the edge server. Which virtual frame that each data frame sent by a sensor device belongs to is determined by the timestamp of each data frame.

III. Prototype development and demonstration

A. Prototype development

We use real sensors and computing devices to implement the prototype system to demonstrate the proposed scheme. For the sensor, we use a LIDAR sensor unit, VLP-16, which is commercialized by Velodyne. Data collected by VLP-16 are in the form of point cloud data. We use Jetson Nano for the device, and Jetson Xavier NX for the edge server, respectively. Both of them are commercialized by NVIDIA. These equipments are connected via a local network.

Annex to ISBN 978-3-903176-39-3 © 2021 IFIP

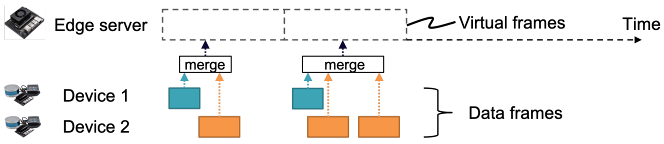


Fig. 2: Calibration in proposed scheme

The experiment field is built indoors as a 1/24 scale model of a parking lot. The parking lot has two perpendicular walls, each of them is 90 cm long and 30 cm high. Three vehicles are parking there. Two VLP-16 LIDAR sensors are put in the parking lot for monitoring. In this experiment, we assume anomaly detection service on the parking vehicles. Therefore, our concerned area is the area where the vehicles are. We divide the space into voxel grids of size 12.5 cm $\times$ 25 cm $\times$ 15 cm such that each vehicle is exactly inside one voxel. Voxels that contain vehicles are set as the *selector set*.

B. Demonstration setup

We demonstrate the performance of our proposed scheme by measuring the quality of point clouds after the two-level processing, under limited network bandwidth. We use the point-to-plane peak signal-to-noise ratio (PSNR), which is defined as the ratio of the point cloud size to the error compared to a reference point cloud [4]. PSNR expresses the quality of point cloud after processing. The larger the PSNR is, the higher the quality is. We use the point cloud collected under ideal network condition without any bandwidth limitation as the reference. The frame time of the reference point cloud is set to 2 seconds, which is long enough so that the point cloud quality should be good enough to be used as comparison reference. The reference point clouds are collected under two situations: 1) without data selection, which means data of the whole parking lot, 2) with data selection, which means only data of the vehicles.

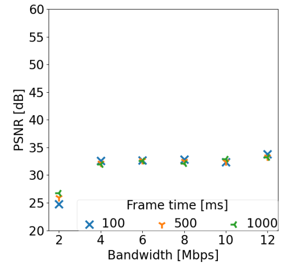
For comparison, we introduce three benckmarks, which do not adopt the two-level processing. The first and second benchmarks collect point clouds from only one LIDAR without data selection. The third one collects point clouds from two LIDARs without data selection. For the three benchmarks, point clouds are collected without data selection, and their comparison reference point clouds are also those without data selection.

For the proposed scheme, point clouds are collected from both two sensor devices, with data selection. Comparison reference point clouds are also those with data selection.

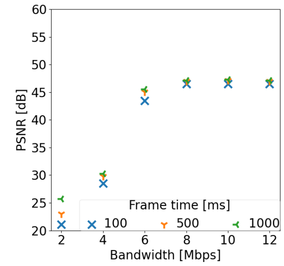
For the three benchmarks and the proposed scheme, we attain the mean PSNR result over about 180 frames, under multiple settings for both virtual frame duration and network bandwidth.

C. Demonstration results

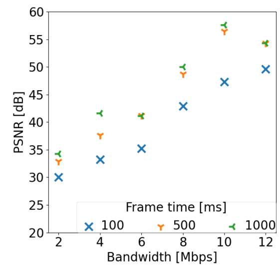
Figs. 3a-3d show the demonstration results, which correspond to the benchmarks and the proposed scheme. The effect of the first level processing is observed by comparing the result of Fig. 3c with that of Fig. 3d. Note that, as we mentioned in Section III-B, the reference point clouds for the PSNR results



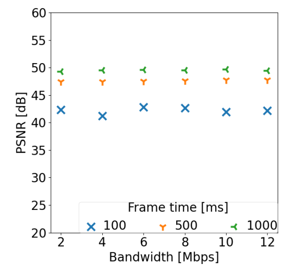
(a) Benchmark 1 (sensor device 1, no data selection)



(b) Benchmark 2 (sensor device 2, no data selection)



(c) Benchmark 3 (two sensor devices, no data selection)



(d) Proposed scheme (two sensor devices with data selection)

Fig. 3: PSNR vs. bandwidth, with virtual frame duration from 100 to 1000 milliseconds.

in Fig. 3d are different from those of Figs. 3a-3c. Therefore, we cannot compare their PSNR results directly. Figs. 3a-3c observe that the stricter the network bandwidth limitation, the worse the PSNR. In comparison, Fig. 3d shows that, with data selection, PSNR does not become worse when bandwidth limitation gets stricter. These results verify that the first level processing improves the quality of collected point clouds under limited network bandwidth.

The effect of the second level processing is observed by comparing the results of Figs. 3a-3c. The PSNR of using both sensor devices is higher than using either sensor device. These results verify that the quality of collected point clouds is improved by adopting the second level processing.

IV. Conclusions

This paper proposed a two-level processing scheme for real-time image sensing. Using our developed prototype system, the demonstration verified that the proposed scheme works better than the benchmarks without the two-level processing.

References

[1] R. Du, P. Santi, M. Xiao, A. V. Vasilakos, and C. Fischione, “The sensable city: A survey on the deployment and management for smart city monitoring,” *IEEE Commun. Surveys Tuts.*, vol. 21, no. 2, pp. 1533–1560, 2018.

[2] S. R. E. Datondji, Y. Dupuis, P. Subirats, and P. Vasseur, “A survey of vision-based traffic monitoring of road intersections,” *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 10, pp. 2681–2698, 2016.

[3] M. A. Alsmirat, Y. Jararweh, I. Obaidat, and B. B. Gupta, “Internet of surveillance: a cloud supported large-scale wireless surveillance system,” *The Journal of Supercomputing*, vol. 73, no. 3, pp. 973–992, 2017.

[4] D. Tian, H. Ochimizu, C. Feng, R. Cohen, and A. Vetro, “Geometric distortion metrics for point cloud compression,” in *2017 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2017, pp. 3460–3464.