GrabView: A Scalable Street View System for Images Taken from Different Devices

Jiong Huang*, Sheng Hu*, Yun Wang*, Chunhong Zhao*, Guanfeng Wang*, Xudong He*,
Xiaocheng Huang*, Shaolin Zheng*, Tom Galloway*, Yifang Yin†, Roger Zimmermann†
*Grab R&D Centre, Singapore
†National University of Singapore, Singapore
Email: *jiong.huang, sheng.hu, yun.wang, chunhong.zhao, guanfeng.wang,
xudong.he, xiaocheng.huang, shaolin.zheng, tom.galloway@grab.com,
†idsyin@nus.edu.sg, rogerz@comp.nus.edu.sg

***Abstract*—In the last decade, many researchers and applications focused on street view systems' geo-spatial and multimedia content. However, most street view systems have high data collection costs and have out of date content in certain regions. An efficient and effective street view system reflecting real world multimedia content on a weekly basis has been an elusive goal.**

We present GrabView, a street view system that:

- **Uses a data capture during ride sharing service trips model.**
- **Collects up-to-date geo-referenced multimedia content at low cost.**
- **Processes both the multimedia and geo-sensor content.**
- **Serves a navigation and browsing user experience equivalent to that from dedicated mapping vehicles.**

Grab ride hailing service vehicles collect most of GrabView's geo-sensor data and multimedia content. This results in much better up-to-the-minute road network coverage at no extra cost.

***Index Terms*—multimedia content indexing, street view, geo-referenced**

I. Introduction

Google published its initial street view system for multiple U.S. cities in 2007 [1]. Over the next decade, Google increased its coverage across the world. During that period, others entered the field, such as Mapillary [2], OpenStreetCam [3], Bing Streetside [4], CycloMedia [5], Yandex [6], etc..

There has also been academic research in the street view field, including, but not limited to:

- Image-based localization based on street view visual contents [7]
- Semantic segmentation for street view images [8]
- Urban greenery assessment based on street view images [9]
- City 3D modeling from street view panoramic sequences [10]

A street view system's ease of access to rich multimedia content both:

- Significantly helps people understand the street level physical world.
- Serves as a building block for several different research areas.

However, street view progress in Southeast Asia is far behind what you might expect. Image coverage from different systems is very poor, and most multimedia content is obsolete. Global operating companies and organizations usually don't have Southeast Asia as their highest priority. Conversely, Grab, a leading car hailing app that only operates in Southeast Asia, has an obvious interest in pushing the region's street view standards to the next level.

Unlike the traditional making a map from images collected by dedicated cars, Grab's driver fleet records geo-referenced multimedia content while going about their business. Advantages to this approach are the low collection cost and timeliness of the multimedia content. Our drivers cover most of the roads in Grab-serviced cities every week.

On the other hand, without dedicated vehicles and their recording hardware, our system has the challenges of using different devices and various collection methods. The next section discusses how we designed the system and addressed those issues.

II. System Architecture

In this section we describe *GrabView*'s overall system architecture and its end-to-end data processing. As shown in Figure 1, we first collect two data sets:

- Multimedia content, such as images and videos.
- Geo-sensor data, such as GPS and gyroscope readings from heterogeneous devices.

We then apply separate multi-purposed processing modules on both data sets. Finally we merge them into geo-referenced image frames, with an efficient indexing system for the rendered map-based web interface.

A. Data collection from different devices

We use different cameras such as *GoPro Fusion*, *Insta360* and *Waylens* to capture street-level images. We mount cameras on the roofs of regular cars and on bikers' helmets. These non-fixed helmet-mounted cameras are challenging for following multimedia content and geo-sensor processing. We must guarantee that all images have the same angle in both the Transverse and Longitudinal axis.

978-1-7281-5527-2/19/$31.00 ©2019 IEEE
DOI 10.1109/BigMM.2019.00056

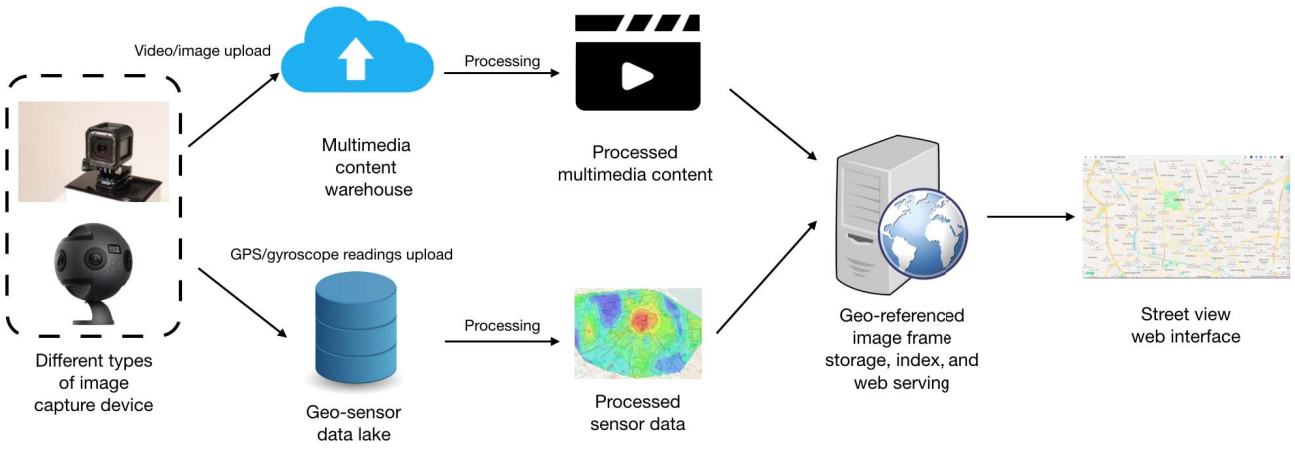

Fig. 1. Overview of the GrabView system.

B. Geo-data lake and multimedia warehouse

Multimedia contents are uploaded to a centralized AWS S3 storage. We organize files by uploader indices and cities. Geo-sensor data is stored in a separate S3 storage and indexed in ElasticSearch clusters. For efficient access, we put query engines above both storage locations.

C. Multimedia content processing

Images and videos are cached in cameras. They later upload to a Grab server when a high-throughput low-cost network is available.

GrabView does the following processing (via separate modules) on its multimedia contents when needed:

- Privacy-based image removal such as car license plates and faces via dockerized APIs
- Quality evaluation and filtering
- Image stitching

Image/video quality varies significantly due to our use of many different camera types and circumstances when pictures were taken. Quality evaluation and filtering is a key processing pipeline component.

D. Geo-sensor data processing

Geo-sensor data, including GPS, gyroscope, and accelerometer readings and multimedia content, is recorded simultaneously. After the data is uploaded to the GrabView service, the system extracts it from the data lake and does:

- Trajectory cleaning
- Map matching
- Direction adjustment and smoothing (to solve the non-fixed helmet-mounted camera issue)
- Topology indexing

We'd like to highlight the importance of direction adjustment. Motorbike drivers turn their heads and bend and stretch for multiple reasons. So when we collect data from motorbike helmet cameras, the pictures are not all from the same fixed direction as are those from car-mounted cameras. It's essential for a good viewing experience that we adjust images and video frames into the same shooting direction, based on the geo-sensor data analysis .

E. Geo-referenced image frame merging, indexing, and web serving

When GrabView is loaded in a web interface, it sends a bounding box of the displayed area (effectively, a map) to the server. The server queries all multimedia content in the displayed area using their ElasticSearch-stored geo-sensor data. This sends a list of image paths to the browser. The web page then indicates the system's street view pictures by displaying pins on the map at the pictures' locations. When a user clicks on a pin, GrabView serves an image retrieval request. A panorama viewer then renders the returned street view picture of that location.

III. DEMONSTRATION

For this demo, we collected videos and photos from a *GoPro* fusion camera and an *Insta360* camera.

After a user uploads their data, we:

- Transcode the multimedia content.
- Compress the geo-spatial sensor data.
- Host the above separately in a managed AWS S3-based file system. We built a query engine on top of the file system for fast access and retrieval.

The GrabView web-based user interface visualizes 360 degree panoramic images. Users can navigate adjacent images along the road network on an *OpenMapTiles* canvas.

Blue dots, called pins, represent all indexed and filtered images on a 2D map. When a user clicks on or touches a pin, a map-overlay image viewer launches. The 360 degree panoramic frame is rendered from the designated starting location. During user navigation, stitched and orientation-corrected images are seamlessly rendered on the same interface.

REFERENCES

[1] D. Anguelov, C. Dulong, D. Filip, C. Frueh, S. Lafon, R. Lyon, A. Ogale, L. Vincent, and J. Weaver, "Google street view: Capturing the world at street level," *Computer*, vol. 43, no. 6, pp. 32–38, 2010.

[2] "Mapillary." https://www.mapillary.com/. [Online; accessed May-2019].

[3] "Openstreetcam." https://www.openstreetcam.org/. [Online; accessed 19-May-2019].

[4] "Bing streetside." https://www.microsoft.com/en-us/maps/streetside. [Online; accessed May-2019].

[5] "Cyclomedia." http://www.cyclomedia.com/. [Online; accessed May-2019].

[6] "Yandex panoramas." https://yandex.com/maps/. [Online; accessed May-2019].

[7] A. R. Zamir and M. Shah, "Accurate image localization based on google maps street view," in *European Conference on Computer Vision*, pp. 255–268, Springer, 2010.

[8] J. Xiao and L. Quan, "Multiple view semantic segmentation for street view images," in *2009 IEEE 12th international conference on computer vision*, pp. 686–693, IEEE, 2009.

[9] X. Li, C. Zhang, W. Li, R. Ricard, Q. Meng, and W. Zhang, "Assessing street-level urban greenery using google street view and a modified green view index," *Urban Forestry & Urban Greening*, vol. 14, no. 3, pp. 675–685, 2015.

[10] B. Micusik and J. Kosecka, "Piecewise planar city 3d modeling from street view panoramic sequences," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2906–2912, IEEE, 2009.