ANALYSIS OF THE SPATIOTEMPORAL ACCUMULATION PROCESS OF MAPILLARY DATA AND ITS RELATIONSHIP WITH OSM ROAD DATA: A CASE STUDY IN JAPAN

T. Seto [1] *, Y. Nishimura [2]

[1] Department of Geography, Komazawa University, Japan, tosseto@komazawa-u.ac.jp
[2] Faculty of Letters, Nara Women's University, Japan, nissy_yu@cc.nara-wu.ac.jp

Commission IV, WG IV/4

KEY WORDS: Spatiotemporal analysis, volunteered geographic information, Mapillary, OpenStreetMap, FlatGeoBuf

ABSTRACT:

This paper presents a geospatial analysis of a FlatGeobuf database composed of six years of geographic data on approximately 41.7 million Mapillary photo shooting locations throughout Japan and geospatial data including road data from OpenStreetMap (OSM). Although Mapillary has a shorter track record than OSM, it is a massive data source and its use as a new resource for volunteered geographic information is expected to attract attention in the future. Therefore, we aim to clarify the geographical distribution of Mapillary users and time-series transitions and attempt to analyze the relationships of these road landscape images with OSM road data and the local geographical distribution characteristics. Although the geographical distribution of Mapillary locations is biased, much information is collected for local roads, which may improve the quality of OSM road data and expand the overall road data.

1. INTRODUCTION

OpenStreetMap (OSM: https://www.openstreetmap.org/), a crowdsourcing-based open mapping project, has been active in Japan since 2008. Since OSM was initially started for rudimentary open data sharing, the accuracy of the data is not guaranteed; however, it is now being used to map many applications and web services. There has also been substantial work in free and open-source software for geospatial (FOSS4G) fields to develop tools to support OSM data creation and quality assessment (Senarane et al., 2017; Bégin et al., 2018; Grinberger et al., 2022). For the data collected using OSM since Goodchild proposed the concept of "volunteered geographic information (VGI)" in 2007, research has been conducted regarding its various services and approaches, including data quality, VGI contributors, rationalization of large and multiple data processing and integration, and VGI data sustainability (Yan et al., 2020). It is crucial to examine the relationships between OSM and other VGIs and time series with a focus on OSM's vast data resources.

Another global VGI is Mapillary (https://www.mapillary.com/), which was launched in Sweden in April 2014 as a crowdsourcing-based open landscape photo-sharing service with location information. As of May 2022, approximately 1.5 billion photos have been shared worldwide using Mapillary. According to the Web of Science, only approximately 50 academic studies focus on Mapillary, predominantly after 2016 and focusing on more than the data characteristics and spread of activities (Juhász & Hochmair, 2016; Ma et al., 2020). Mapillary has been compared with similar VGI projects, such as OpenStreetCam (Alvarez Leon & Quinn, 2019; Mahabir et al., 2020) and those involving image recognition and deep learning (Neuhold et al., 2017; Warburg et al., 2020; Kong & Fan, 2020). Mapillary is not only aimed at using open data of landscape images for various purposes but has also focused on developing tools and algorithms such as automatic extraction of geographic features from landscape images to enrich and improve the quality of OSM data from the early stages of its activities. Therefore, although OSM and Mapillary are two separate activities, they share a joint user base and community activities. It is essential to clarify the temporal and spatial differences and characteristics of their accumulated data. There are few large-scale comparative studies of Mapillary analysis at a more detailed spatial level, such as at the country or multi-city levels. It is crucial to compare the spatiotemporal trends of Mapillary data points and their relationship with OSM data to consider the sustainability of both types of VGI activities.

This study aimed to compile a database of Mapillary points of interest (POIs) over six years throughout Japan to investigate regional data differences, considering their spatiotemporal characteristics and the OSM road data coverage. Furthermore, we clarified the relationship between OSM road data coverage and the editing of such road data before and after Mapillary imaging. While previous studies have addressed Mapillary landscape imaging, this study focuses on its geographical context. This paper is organized as follows: Section 2 describes how the database for the analysis was constructed and the spatial characteristics of Mapillary POIs; Section 3 provides an exploratory analysis of the relationship with OSM; and Section 4 summarizes future research issues as a discussion and conclusion.

2. CHARACTERISTICS OF MAPILLARY DATA

2.1 Database Construction

Mapillary POI data from Japan between September 2014 to September 2020 were obtained using the Search Images API (application programming interface) of the Mapillary API (v. 3, 2017) and constructed into a database. The data was output in the open standard geospatial data interchange format GeoJSON after dividing it by one month, using 37 tile indices at a zoom level of

* Corresponding author

https://doi.org/10.5194/isprs-archives-XLVIII-4-W1-2022-403-2022

7 to cover Japan, with 41 765 634 POIs identified. Eight attribute values were obtained from the API including spatial and temporal items.

Mapillary POIs, the subject of the analysis, are extensive point data, whose performance cannot be analyzed using the object-relational spatial database extender PostGIS. Therefore, we used geospatial data abstraction library to convert the data into FlatGeobuf format (.fgb), which can directly binarize and compress the data while ensuring portability, which is one of the advantages of GeoJSON. The results also confirmed that the speed of reading the .fgb format was approximately three times faster than that of PostGIS in the local environment. Spatial merging using the open-source geographic information system QGIS added the administrative names for each city in Japan with Mapillary POIs. The OSM road data attributes, extracted by QGIS (3.2 GB in .fgb format; 8 313 281 records) from the OSM dump file provided by Geofabrik for the whole country of Japan, included the road type, OSM version, last editor's name, and last update timestamp applying to the nearest neighbor of the data point (the maximum search radius was set to 50 m, with attributes outside this radius not being assigned). Therefore, using these procedures, a 17.6-GB point database in .fgb format was constructed. For efficient spatial aggregation, polygon data for each municipality (232.2 MB; 32 attribute values) and for each 1-km mesh (126.5 MB; 20 attribute values) were created by spatially aggregating the point data.

2.2 Overview of Mapillary Data

Table 1 summarizes the data attributes for the 41 765 634 points of interest taken between September 1, 2014, and September 31, 2020. The most common equipment used for Mapillary photography were Apple products, with approximately 15 million images in particular, iPad Pro devices accounted for 1 890 000 images. Additionally, 360-degree cameras (e.g., Ricoh Theta Z1) were used for approximately 5.4 million images, although this value accounted for only approximately 13% of the total.

Property	Description	Type	Range of value
key	Keycode	String	41,765,634
ca	Cardinal points	Real	0 – 360 degrees
camera_make	Device maker	String	91 types
camera_model	Device name	String	488 types
captured_at	Timestamp	Date	from 2014-01-01 to 2020-09-31
pano	Panoramic shooting	Boolean	Yes: 5,413,030 No: 36,352,604
sequence_key	Sequence code	String	41,765,634
user_key	Account code	String	1,513
username	Account name	String	1,513 users

Table 1. Summary of the Mapillary points of interest from September 2014 to September 2020.

The total number of unique users contributing to the Mapillary data collected was 1513. Figure 1 shows the number of users and the timing of the Mapillary shoots per month, with the orange line representing the number of users. In June 2016, the number of users exceeded 50 for the first time and, in July 2017, the number of users exceeded 100. The highest number of users during the period of the aggregate unit was 211 users in June 2019. This is likely attributed to Mapillary being made available for editing in the OSM extensible editors iD Editor and JOSM in June 2016, which significantly increased the awareness of Mapillary among OSM mappers in Japan. Similarly, in June 2017, the iD Editor allowed viewing layers of road signs automatically extracted from Mapillary, coinciding with the subsequent peak and strengthening of the linkage with OSM.

From 2018 to 2019, the Mapillary user community was also active in Japan hosting meetup events. For example, a Mapillary meetup event was held in Nara, Japan, in March 2019, attracting approximately 30 people from across Japan, which may be one of the factors in the subsequent increase in users. As a fundamental statistic, the average number of Mapillary images from any user is 414 421, with a median of 33 353, showing a gap of more than ten times. When data generated by the top 10 contributors, out of the 1513 unique users, were added together, they accounted about 90% of the total data collected (approximately 37 million images), with the top 3 contributors each contributing more than 5 million images, showing that a small number of people made a significant contribution to the number of images. The number of users (mappers) in the Japanese OSM community, which is closely related to the Mapillary community, totaled approximately 35 000 from January 2008 to December 2020 in Japan, according to Seto (2022). The average number of monthly Mapillary users from 2019 to 2020 was approximately 500 to 1000. This trend is approximately half to a quarter of the total number of OSM users.

Figure 2 shows the updated monthly trends of OSM road data. Since activities have been conducted in Japan since 2008, a longer-term situation is evident than that of Mapillary. Additionally, since February 2011, a large scale of data maintained by private companies has been imported for road data only. The data development of major roads had been relatively advanced before Mapillary activities.

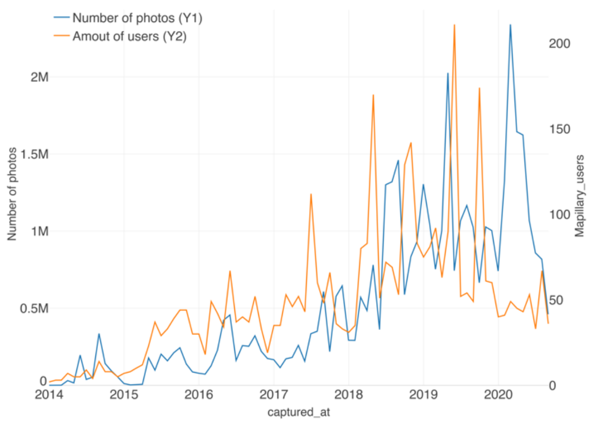

Figure 1. Monthly Mapillary data acquisition dates and number of participating users.

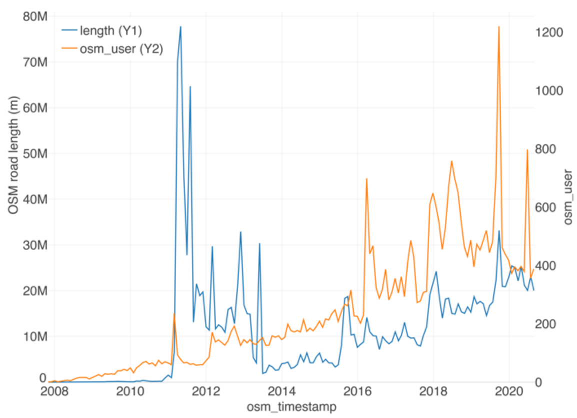

Figure 2. Monthly OpenStreetMap road updates and number of participating users.

https://doi.org/10.5194/isprs-archives-XLVIII-4-W1-2022-403-2022

3. GEOGRAPHIC DISTRIBUTION ANALYSIS OF MAPILLARY DATA IN JAPAN

3.1 Trends in the Distribution of Mapillary Activities

There were a total of 388 762 1-km meshes covering the land area of Japan, of which 71 542, or approximately 19% of the total, had Mapillary contributions. Figure 3 shows the geographic distribution of the Mapillary POIs on each 1-km mesh and four levels of the Jenks Natural Breaks Classification. There are two mesh locations in Japan where more than 100 unique users contributed to Mapillary imaging, both in Tokyo. The distribution of the dark blue areas with a relatively large number of posting users shows that they are distributed on a line along major arterial roads in Japan.

Figure 4 shows the number of Mapillary postings per user, aggregated in units of 1-km mesh and expressed in eight levels, according to the Jenks Natural Breaks Classification. This figure shows the same geographical distribution as Fig. 3; however, the number of postings per user was significantly higher in central Tokyo, in the center of the figure; the Tohoku area, on the right of the figure; and the Kyoto area, on the left of the figure. This distribution covers the area surrounding major roads. The Kyushu area, on the left of the figure, and the Hokkaido area, on the upper right of the figure, also show a linear distribution along major roads, although the number of postings is smaller. These figures show that Mapillary is supported by a small number of contributors in Japan, accumulating data primarily on major roads in Japan and providing areal coverage in some urban areas.

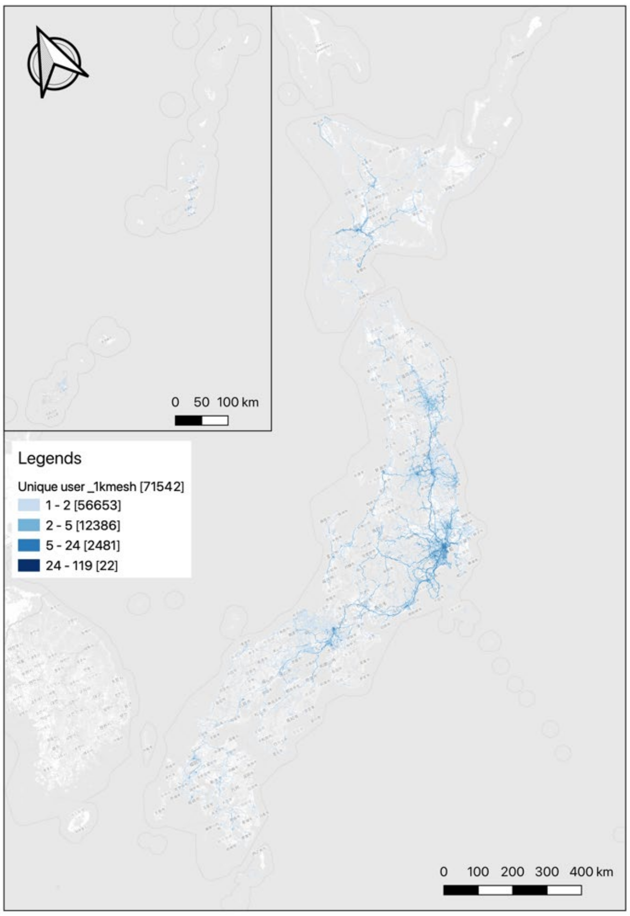

Figure 3. Number of unique Mapillary users per 1-km mesh.

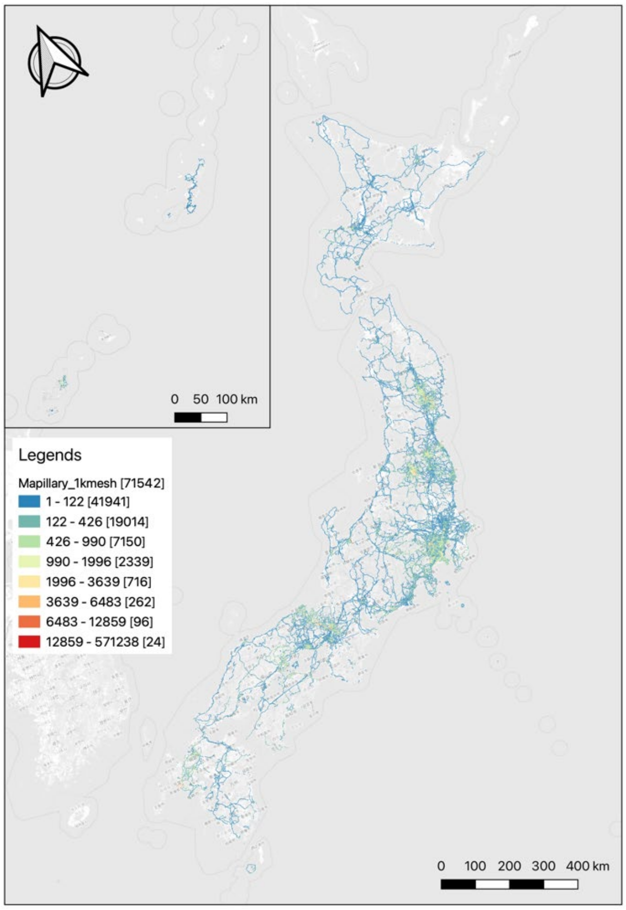

Figure 4. Number of Mapillary postings per user in each 1-km mesh.

3.2 Geographic Distribution between Mapillary and OSM Data

Based on these geographic distribution trends in Mapillary postings, we examined the relationship between Mapillary and OSM road data. The nationwide distribution trend of OSM road data in Japan is shown in Fig. 5, which is characterized by high densities in the three major metropolitan areas (i.e., Tokyo, Nagoya, and Osaka) and comprehensive coverage throughout Japan. Therefore, there is relatively more minor regional bias in OSM data than in Mapillary data.

Figure 6 shows the ratio of the OSM roads with overlapping Mapillary contribution data to the total OSM roads, aggregated onto a 1-km mesh. The results show that the areas covered by existing OSM roads (dark red) show a remarkable trend, with the Tokyo-Shizuoka area, the Tohoku (Fukushima and Iwate) area, and the Kyoto area, accounting for more than 60% of the total, with some urban areas in Hokkaido also being covered.

Figure 7 shows the difference between the last update date of OSM road data and the date when Mapillary data was taken, using the Jenks Natural Breaks Classification, with the number of days as a unit. However, OSM road data was updated after Mapillary data was taken approximately 40% (272 817 roads) of the time, indicating that OSM road data is updated more frequently where Mapillary data is distributed. However, it is necessary to pay attention to more local cases to understand the relationship between road data and OSM update timing.

3.3 Mapillary Data Distribution Trends by Local Area in Japan

While Mapillary-sponsored road data in Japan shows an expansion of linear or areal activities, it relies heavily on coverage in city units. Therefore, this section analyzes Mapillary and OSM data considering municipalities as the spatial units.

Figure 8 shows the coverage rates of Fig. 6, re-calculated by applying them to the city units in Japan. The areas with remarkably high coverage rates were the same as in the previous analysis, and the areas with coverage rates of 60% or higher were limited to 24 cities in Japan. However, when we look at the types of roads where the Mapillary data were distributed, they were not necessarily uniform.

Figure 9 shows a cluster map of the roads with the most Mapillary POIs. In rural Tohoku and Kyoto, Mapillary POIs were distributed predominantly on important roads (OSM highway=motorway, trunk road). Conversely, in urban Tokyo and Osaka, Mapillary POIs were distributed on sidewalks (highway=path, sidewalk, unclassified) and other locations where detailed OSM data were not available, as shown in Fig. 10. In the Tokyo and Osaka metropolitan areas, the data were developed to complement the areas where the OSM data did not provide detailed data, such as sidewalks (highways=path, footway, unclassified).

Next, we explored the localized distribution trends within each municipality, which is a characteristic feature of the OSM data. Figure 10 shows the periphery of Aizuwakamatsu City in Fukushima Prefecture, located in the Tohoku region, which has a vast number of Mapillary POIs (approximately 1.4 million), 34 unique users (within the top 5%), and a high coverage rate (77%) of OSM road data.

Although this is a relatively small city with a population of approximately 120 000, it is home to many contributors who are actively engaged in OSM activities. For example, in 2017, the State of the Map was held in the area, and many global OSM users visited. The region also has a deep knowledge of Mapillary and is trying to train Mapillary users in parallel with OSM mapping activities.

Mapillary POIs are distributed comprehensively in the central part of the city. OSM road data tends to be updated even after the Mapillary image, regardless of the road type. Fukushima Prefecture is one of the areas in Japan where both Mapillary and OSM data generation is high, and this may be partly because the community base tends to be strong as various events are held there. Although there are still only a few such areas in Japan, it helps to indicate the future possibility of data development as the level of activities increases.

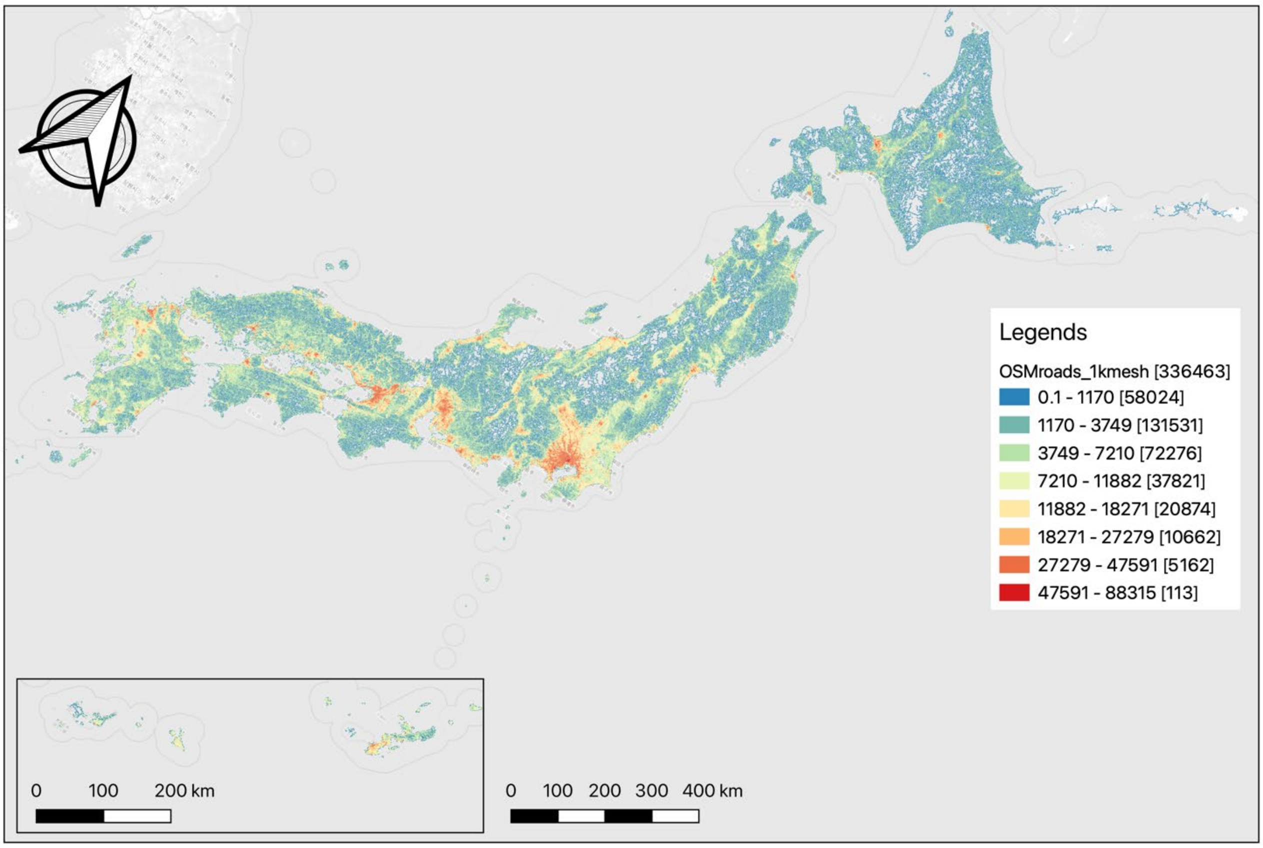

Figure 5. OpenStreetMap road extension in 1-km meshes (unit: m)

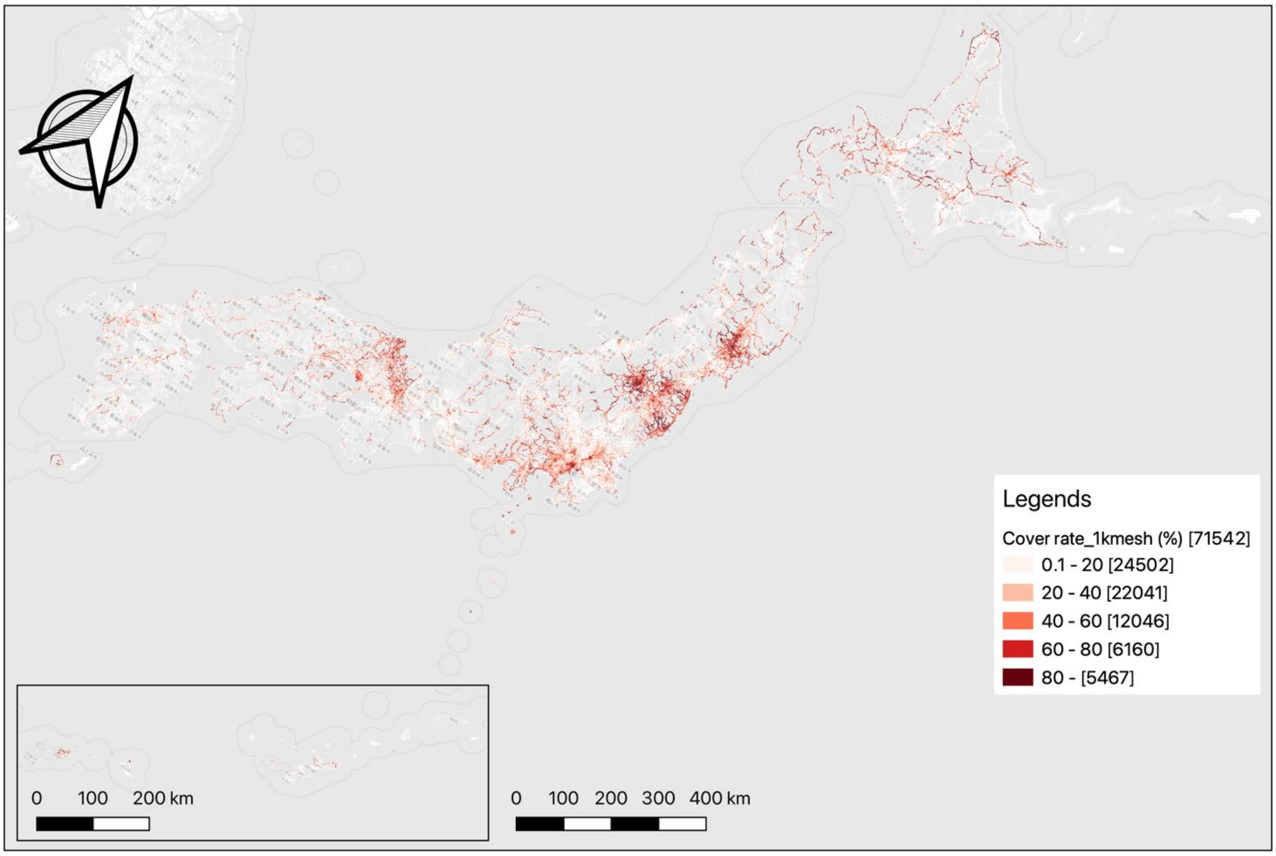

Figure 6. OpenStreetMap-Mapillary coverage in 1-km meshes (unit: %)

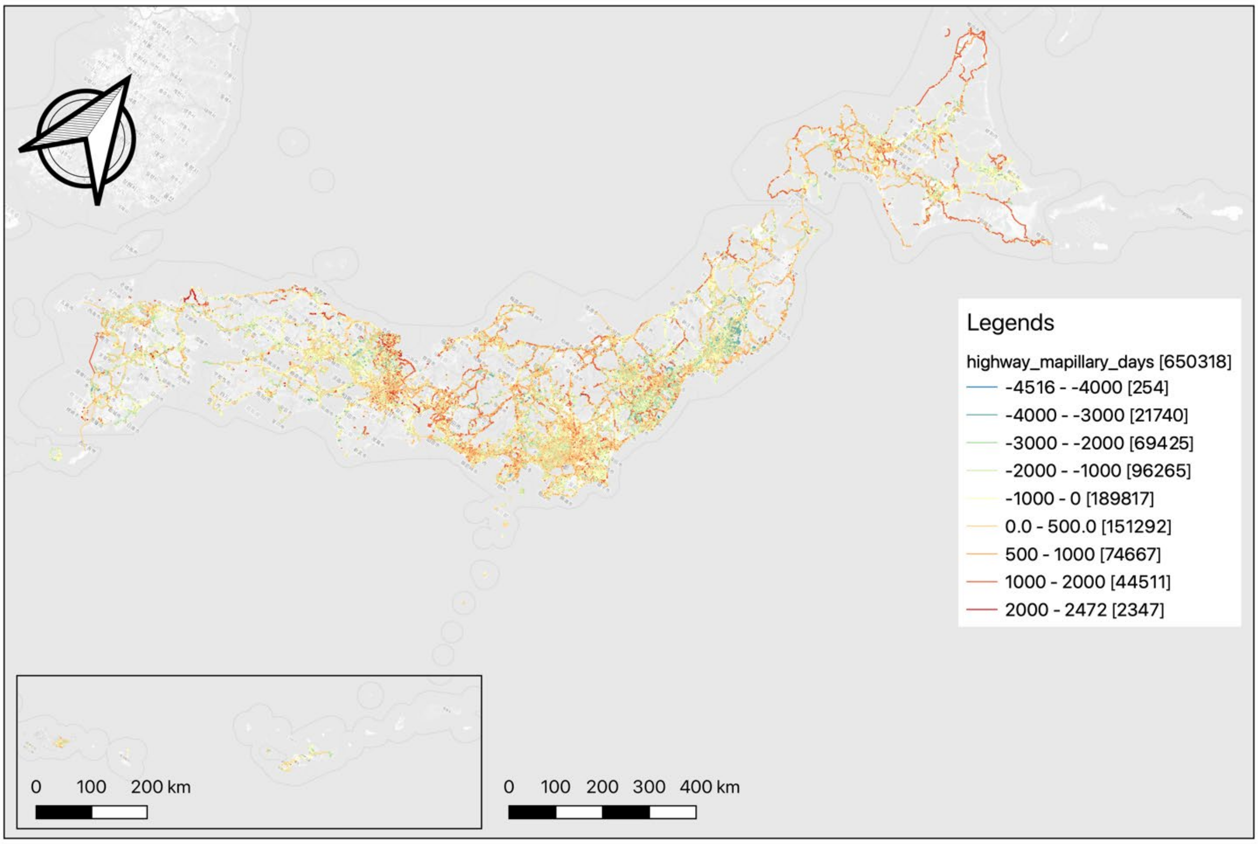

Figure 7. Last update date of OpenStreetMap road before and after Mapillary image taken (unit: days)

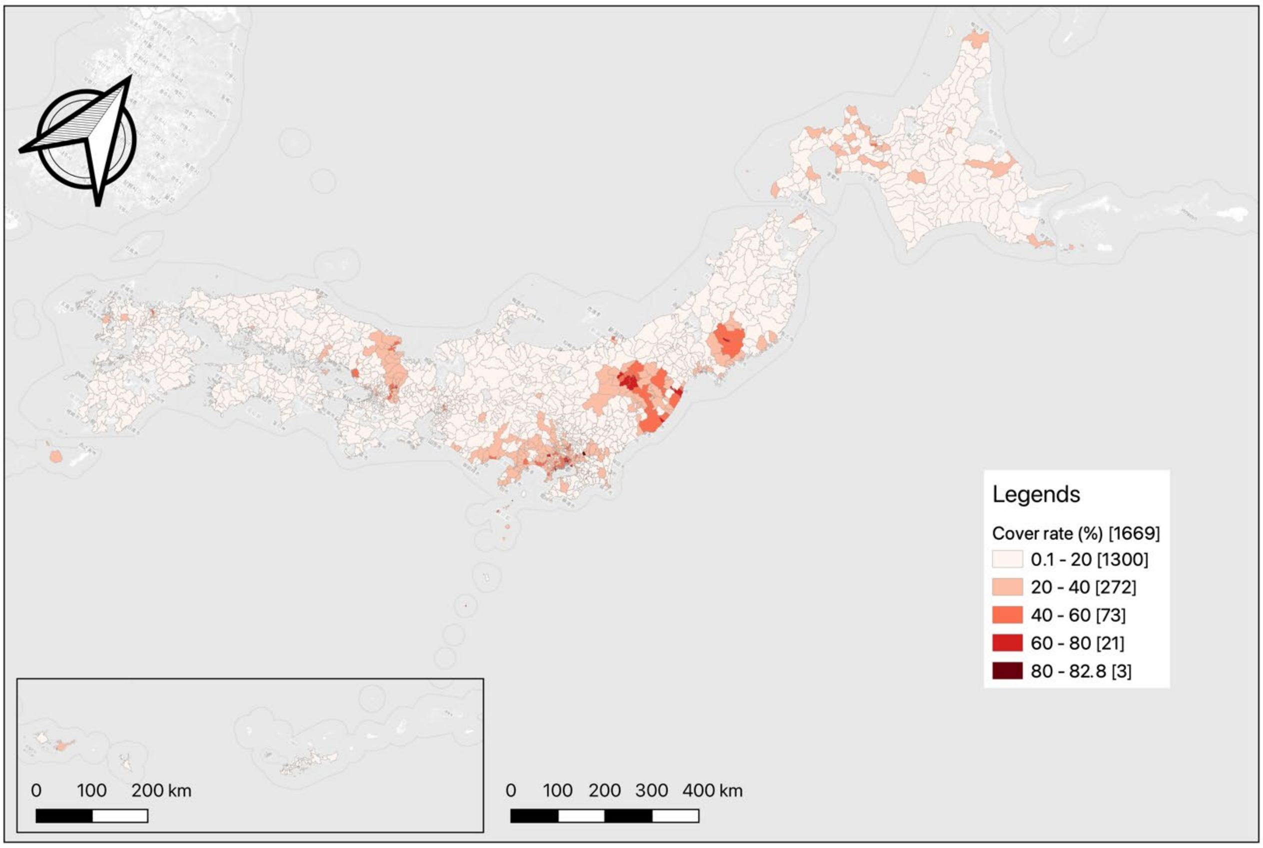

Figure 8. OpenStreetMap-Mapillary coverage by municipality (unit: %)

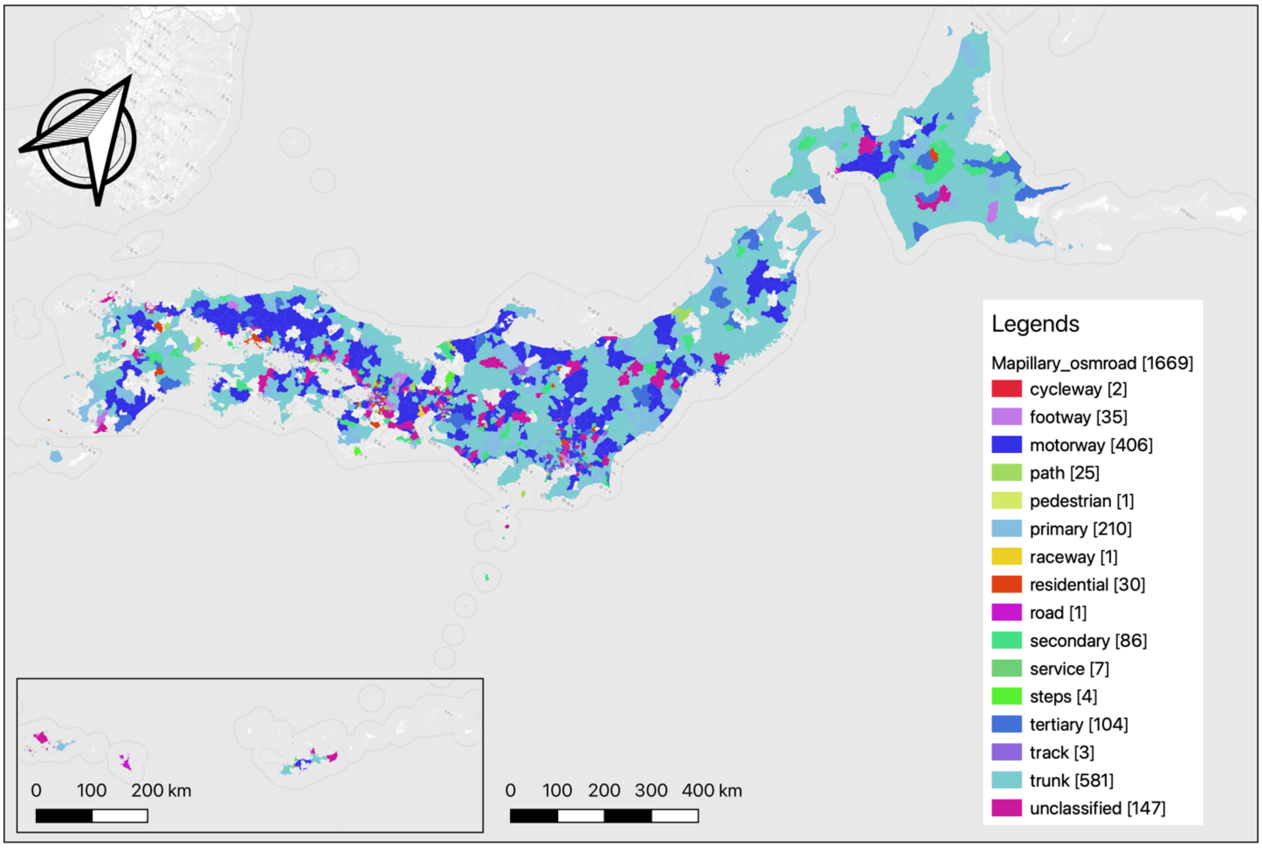

Figure 9. Most common road classifications in terms of Mapillary points of interest (unit: number of cities)

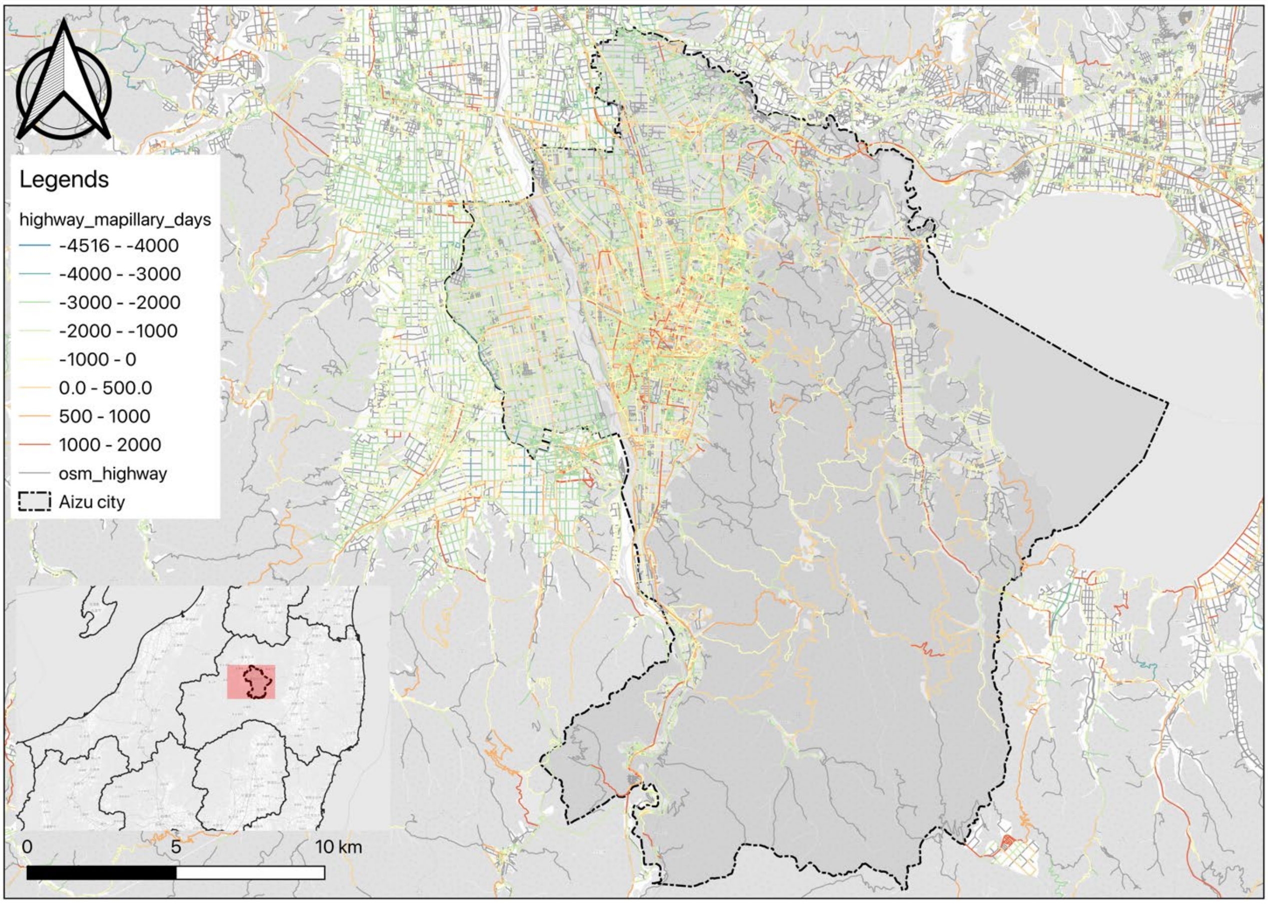

Figure 10. Distribution of Mapillary-OpenStreetMap roads in Aizu City

4. CONCLUSIONS

In this paper, we collected approximately 41 million Mapillary points in Japan, converted them to FlatGeobuf format for easy spatial analysis and visualization, and used QGIS to conduct an exploratory analysis of their detailed geographic distribution nationwide and their relationship with OSM road data. The results showed that Japan had approximately 1500 unique Mapillary contributors during the six years of investigation, with the top 10 users contributing approximately 90% of the images. Geographic distribution analyses of the Mapillary POIs showed that, while most contributions were made along significant roads in major cities, they also tended to be made in surrounding areas such as Fukushima and Kyoto Prefectures. This situation aligns with the growing OSM road data production trend throughout Japan. The proportion of Mapillary data compared to OSM road data has reached about 16% in some areas.

This study also examined the relationship between Mapillary POI and OSM road type and the difference between the date the Mapillary POI was taken and the date the OSM road data was last updated. The survey revealed that not only highways and arterial roads, but also narrower and smaller roads for which there is little official road data, are of particular importance in the urban areas; approximately 40% of all areas have been updated since the Mapillary images were taken, which can help update OSM road data. This case is an excellent example of how the OSM road data can be updated. As a case study, we also mapped the spatiotemporal relationship between Mapillary POIs and OSM road data in Aizuwakamatsu City, Fukushima Prefecture, where Mapillary is active, to show the spatial heterogeneity of the city maintenance.

Limitations of this study include the following. First, Mapillary POIs do not correspond perfectly with the OSM road database; rather, they are estimates between nearest neighbors (roads within 50 meters), making it challenging to explore the fundamental factors that influenced the update of OSM data after Mapillary was taken. Second, identifying data taken at the exact location for multiple periods, which is a feature of Mapillary, is difficult in this approach to spatial analysis. In addition to the spatial density distribution of shooting points and users, it is expected to identify the density distribution considering the time dimension to extract the areas where Mapillary activities have been conducted for an extended period.

Overall, Mapillary can improve the shape and quality of OSM road data and consider time-series trends of the same shooting points and features around the road. Therefore, future studies should evaluate the quality improvement of OSM road data and the presence or absence of mapping of geographical features around the road. Mapillary can also be used as reference data for microgeographic features (e.g., building entrances or road signs) around roads. Analyzing these data in more detail makes it necessary to examine what type of detailed road data can be generated from the accumulated data. It is also essential to spread the activity itself, as it has not reached sufficient coverage in Japan. We hope that VGI research on Mapillary data will be further promoted in Asia to deepen the understanding of the usefulness of the data.

ACKNOWLEDGEMENTS

This study used data accumulated and shared over many years by Mapillary contributors and OpenStreetMap contributors. Mierune Inc. provided technical support for the construction of the first Mapillary data for all of Japan.

REFERENCES

Alvarez Leon, L.F., Quinn, S., 2019: The value of crowdsourced street-level imagery: examining the shifting property regimes of OpenStreetCam and Mapillary. *GeoJournal*, 84, 395–414. doi.org/10.1007/s10708-018-9865-4.

Bégin, D., Devillers, R., Roche, S., 2018: The life cycle of contributors in collaborative online communities: the case of OpenStreetMap. *Int. J. Geogr. Inf. Sci.*, 32, 1611–1630. doi.org/10.1080/13658816.2018.1458312.

Grinberger, A.Y, Schott, M, Raifer, M, Zipf, A., 2021: An analysis of the spatial and temporal distribution of large-scale data production events in OpenStreetMap. *Trans. GIS*, 25, 622–641. doi.org/10.1111/tgis.12746.

Juhász, L., Hochmair, H.H., 2016: User contribution patterns and completeness evaluation of Mapillary, a crowdsourced street level photo service. *Trans. GIS*, 20, 925–947. doi.org/10.1111/tgis.12190.

Kong, G., Fan, H., 2020: Enhanced facade parsing for street-level images using convolutional neural networks. *IEEE Trans. Geosci. Remote Sens.*, 1-13. doi.org/10.1109/TGRS.2020.3035878.

Ma, D., Fan, H., Li, W., Ding, X., 2020: The state of Mapillary: an exploratory analysis. *ISPRS Int. J. Geo-Inf.*, 9, 10. doi.org/10.3390/ijgi9010010.

Mapillary Meetup 2019 in Nara,Kansai, 2019. https://forum.mapillary.com/t/recap-mapillary-meetup-2019-in-nara-kansai/2760/3

Mapillary API Documentation, 2019. https://www.mapillary.com/developer/api-documentation/

Mahabir, R., Schuchard, R., Crooks, A., Croitoru, A., Stefanidis, A., 2020: Crowdsourcing street view imagery: a comparison of Mapillary and OpenStreetCam. *ISPRS Int. J. Geo-Inf.*, 9, 341. doi.org/10.3390/ijgi9060341.

Neuhold, G., Ollmann, T., Bulò, S.R., Kontschieder, P., 2017: The Mapillary vistas dataset for semantic understanding of street scenes, *2017 IEEE International Conference on Computer Vision (ICCV)*, 5000-5009. doi.org/10.1109/ICCV.2017.534.

Senarane, H., Mobasheri, A., Ali, A.L., Capineri, C., Haklay, M., 2017: A review of volunteered geographic information quality assessment methods. *Int. J. Geogr. Inf. Sci.*, 31, 139–167. doi.org/10.1080/13658816.2016.1189556.

Seto, T., 2022: Development of OpenStreetMap data in Japan. Wakabayashi, Y., Morita, T. eds. *Ubiquitous Mapping*, 113-126, Springer Nature, Singapore. doi.org/10.1007/978-981-19-1536-9_7.

Yan, Y., Feng, C.C., Huang, W., Fan, H., Wang, Y.C., Zipf., A., 2020: Volunteered geographic information research in the first decade: a narrative review of selected journal articles in GIScience, *Int. J. Geogr. Inf. Sci.*, 34(9), 1765-1791. doi.org/10.1080/13658816.2020.1730848.

Warburg, F., Hauberg, S., Lopez-Antequera, M., Gargallo, P., Kuang, Y., Civera, J., 2020 Mapillary street-level sequences: a dataset for lifelong place recognition, *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2626-2635. doi.org/10.1109/CVPR42600.2020.00270.

APPENDIX

Spatially aggregated from Mapillary POI data (41 765 634) for all of Japan used in the analysis, we are providing both FlatGeobuf format data per municipal-level (232.2 MB; 32 attribute values) and per 1-km mesh-level (126.5 MB; 20 attribute values), via a Github repository:
https://github.com/tossetolab/mapillary-analysis-japan.