

Evaluating the Capability of OpenStreetMap for Estimating Vehicle Localization Error

Kelvin Wong, Ehsan Javanmardi, Mahdi Javanmardi, Yanlei Gu and Shunsuke Kamijo

Abstract— Accurate localization is an important part of successful autonomous driving. Recent studies suggest that when using map-based localization methods, the representation and layout of real-world phenomena within the prebuilt map is a source of error. To date, the investigations have been limited to 3D point clouds and normal distribution (ND) maps. This paper explores the potential of using OpenStreetMap (OSM) as a proxy to estimate vehicle localization error. Specifically, the experiment uses random forest regression to estimate mean 3D localization error from map matching using LiDAR scans and ND maps. Six map evaluation factors were defined for 2D geographic information in a vector format. Initial results for a 1.2 km path in Shinjuku, Tokyo, show that vehicle localization error can be estimated with 56.3% model prediction accuracy with two existing OSM data layers only. When OSM data quality issues (inconsistency and completeness) were addressed, the model prediction accuracy was improved to 73.1%.

I. INTRODUCTION

Accurate localization is a crucial requirement of successful autonomous driving, allowing self-driving vehicles to navigate safely. Major sources of positioning information are Global Navigation Satellite Systems (GNSS) such as GPS, GLONASS, Galileo or BeiDou. The positional accuracy using just GNSS, however, can exhibit errors of tens of meters within dense urban environments due to buildings blocking and reflecting the satellite signal [1]. To address this issue, self-driving vehicles use vision and range-based sensors to determine where they are located. One such method utilizes light detection and ranging (LiDAR) sensors for map matching, whereby the input scan is registered against a prebuilt map to obtain a position. Despite the advancement in LiDAR sensor technology, map-based localization can still suffer from error, arising from sources such as the input scan or the matching algorithm.

To improve map-based localization accuracy, research so far has focused on refining the map matching algorithms [2] as well as producing increasingly accurate High Definition maps [3]. In contrast, there is a lack of research on the impact of the map itself, and the *representation of the environment*, as a source of localization error. For example, within a tunnel or urban canyon environment, even with a locally and globally accurate map, the lack of longitudinal features in the map can cause localization error in the moving direction. So far, the author is aware of only one set of studies which have been conducted to evaluate the relation between the representation within the map and localization accuracy [4], [5].

To date, the evaluation of a map's capability for vehicle localization has been limited to the 3D point cloud format.

There are several challenges with this approach. Firstly, localization error can only be estimated for an area if a prebuilt point cloud map is available. If not, data must be collected, which can be costly and time-consuming. These 3D point cloud maps also require substantial effort to maintain and keep up to date. Secondly, the large size of the point cloud data can make it challenging to manage. For example, a 300×300 m area may consist of over 250 million points. Coupled with the unstructured nature of the data, this leads to a large amount of pre-processing, processing, and subsequent management. Considering these challenges, there is a need to explore whether other mapping formats and sources of data, such as geographic information, can offer an alternative approach for estimating vehicle localization error.

This paper aims to investigate whether it is possible to estimate vehicle localization error using OpenStreetMap (OSM) as a proxy. OpenStreetMap is a freely editable 2D map of the world. The main objective is to leverage OSM's public availability, wide coverage, and comparatively 'light' data to enable a more efficient alternative for estimating localization error. This work builds upon Javanmardi et al.'s [4] format-agnostic map evaluation criteria, by defining six new factors specific for 2D geographic information (as used by OpenStreetMap). The resulting work can enable the estimation of localization error without requiring the collection of data to create a prebuilt map. It can help identify and highlight areas with high localization error – either as the result of the map itself, or the real-world environment. Further, the estimated error model can be used in conjunction with other sensor information to improve the accuracy of localization [6].

The remainder of this paper is structured as follows: in Section II, a short overview of vehicle localization and its sources of error is presented, followed by an overview of geographic information and OpenStreetMap. Section III describes the datasets used in this study, as well as the formulation of the six map evaluation factors. The results of the experiment are presented in Section IV. Following a discussion on the capability of OSM, and the advantages and challenges of the approach, the final Section presents the conclusions and future work.

II. BACKGROUND

A. Vehicle localization

Localization is the estimation of where a vehicle is situated in the world. While Global Navigation Satellite Systems can provide a position with meter-level accuracy, this is insufficient for the application of autonomous driving. One

meter can be the difference between driving safely on the road in the correct lane and driving along the pavement. Centimeter-level positional accuracy is therefore required for successful autonomous driving. To achieve this, autonomous vehicles are equipped with a myriad of sensors, including LiDAR, radar, and onboard cameras.

One approach for localization is LiDAR map matching. For this, a map is prebuilt using a high-end mobile mapping system. Subsequently, the LiDAR scan from an autonomous vehicle can then be used to match against the prebuilt map to obtain a position. Despite the importance of accurate vehicle self-localization, there is yet to be an international standard for required accuracy. As guidance, the Japanese government's Cross-ministerial Strategic Innovation Promotion (SIP) Program recommends an accuracy of less than 25 cm. This value is based on satellite image resolution and the physical width of a car tire.

B. Sources of localization error for LiDAR map matching

Sources of localization error for map matching using LiDAR can be divided broadly into four categories: 1) Input scan; 2) Matching algorithm; 3) Dynamic phenomena, and; 4) Prebuilt map.

Firstly, errors from the input scan may be dictated by choice of the laser scanner. For example, the lower-end Velodyne VLP-16 has 16 laser transmitting and receiving sensors with a range of up to 100m. In contrast, the latest model (VLS-128) has 128 sensors and up to 300 m range, allowing the capture of more detail in a wider area, thus enabling more accurate map matching. In addition, distortion can be introduced during the capture phase when in motion – a vehicle traveling at 20 m/s, with a scanner at 10Hz would result in a 2 m distortion as the scanner rotates 360 degrees. Errors can also be introduced in the post-processing where the input scan is required to be downsampled for the matching algorithm. While this reduces noise within the data, it can also remove features and details useful for map matching.

Secondly, the choice of the matching algorithm can be a source of error. As matching is an optimization problem, some algorithms are designed to be more robust to local optimum, while others are more immune to initialization error [7]. In general, Normal Distribution Transform (NDT) is more resilient to sensor noise and sampling resolution (than other algorithms such as Iterative Closest Point) but suffers from slower convergence speed [8].

Thirdly, dynamic phenomena in the environment can contribute error. During the mapping phase, if a large vehicle such as a lorry or a bus blocks the laser scanner, then a large portion of the map will be missing. During the localization phase, some phenomena or features may have moved or shifted. For example, parked cars found in the prebuilt map may have since moved or buildings may have been built or demolished. Furthermore, seasonal changes such as the presence and absence of leaves on trees between the mapping and localization phases can introduce error.

Lastly, the prebuilt map itself is a source of error. As mentioned earlier, having a highly accurate map does not always result in a low localization error. There are instances within the environment, e.g. tunnels, urban canyons, open

spaces whereby map matching is not a suitable localization method. In these cases, it is the physical attributes of the phenomena and its representation on the map which is the source of localization error. This is discussed further in the next section.

C. Quantifying the sources of map-derived errors

To quantify the sources of map-derived errors, Javanmardi et al. [4] defined four criteria to evaluate a map's capability for vehicle localization: 1) feature sufficiency; 2) layout; 3) local similarity, and; 4) representation quality.

Feature sufficiency describes the number of mapped phenomena in the vicinity which can be used for localization. For example, within urban areas, buildings and other built structures provide lots of features to match against, resulting in more accurate localization. An insufficient number of features in the vicinity (such as found in open rural areas) may result in lower localization accuracy.

The layout is also an important consideration, as it is not simply the number of features in the vicinity, but also how they are arranged. Even if there are lots of high-quality features nearby, if they are all concentrated in one direction, the quality of the matching degrades.

To further compound the issue, in certain scenarios, even with sufficient and well-distributed features, accurate localization remains difficult due to local similarity. In these cases, there may be geometrically similar features either side of the vehicle, making it difficult during the optimization process to determine the position in the traveling direction.

Lastly, representation quality describes how well a map feature represents its real-world counterpart. In theory, the closer the map is to reality, the more accurate the prediction. It is important to note, however, that this criterion is slightly different from the level of abstraction or generalization of the map. For example, a flat wall can be highly abstracted as a single line but still have high representation quality.

D. OpenStreetMap

First introduced in 2004, OpenStreetMap is an open source collaborative project providing user-generated street maps of the world. As of March 2019, the project has over 5.2 million contributors to the project who create and edit map data using GPS traces, local knowledge, and aerial imagery. In the past, commercial and governmental mapping organizations have also donated data towards the project.

Features added to OSM are modeled and stored as tagged geometric primitives (points, lines, and polygons). For example, a road may be represented as a line vector geometry with tags such as '*highway = primary; name:en = Ome Kaido; oneway = yes*'. Despite OSM operating on a free-tagging system (and thus allowing a theoretically unlimited number of attributes and tag combinations), there is a set of commonly used tags for primary features which operate as an 'informal standard' [9].

As with any user-generated content, OSM attracts concerns about data quality and credibility. Extensive research has been conducted on assessing the quality of OSM datasets, including evaluating the completeness of the data. For

example, Barrington-Leigh and Millard-Ball [10] estimate the global OSM street layer to be $\sim 83\%$ complete, with more than 40% of countries having fully mapped networks. When comparing the number of OSM tags in Japan to official government counts (TABLE I.), ‘completeness’ varies with different features [11]. While fire departments, police stations and post offices are well represented, temples and shrines are underrepresented. Despite this variance in completeness, OSM still represents a viable and useful source of geospatial data. In fact, in certain areas, OSM is *more* complete and *more* accurate (for both location and semantic information) than corresponding proprietary datasets [12], [13]. As mapping data are abstract representations of the world, it can also be argued that a map can never truly be complete. Instead, mapping data should be collected as fit-for-purpose, and specific for the required application. OSM is successfully used in a wide variety of practical and scientific applications from different domains, demonstrating the usefulness of crowdsourced mapping data.

TABLE I. COVERAGE OF OSM TAGS IN JAPAN 2017 [11]

Tag	# Actual	# in OSM	Coverage
School	36,024	45,568	126.5% ^a
Fire department	5,604	5,028	89.7%
Police station	15,034	13,152	87.5%
Post office	24,052	20,795	86.5%
Traffic lights	191,770	108,498	56.6%
Convenience store	55,176	30,710	55.7%
Bank	13,595	7,077	52.1%
Gas station	32,333	8,944	27.7%
Pharmacy	58,326	7,842	13.4%
Shrine	88,281	10,292	11.7%
Temple	85,045	9,610	11.3%
Postal post	181,523	7,522	4.1%
Vending machine	3,648,600	10,311	0.3%

a. Note that the coverage of schools is higher than 100%. This is the result of schools comprising of multiple buildings and therefore having more than one polygon or point represented within OSM.

E. Related work

There are a several studies on the estimation of vehicle localization error in literature. Akai *et al.* [6] estimate the error of 3D NDT scan matching in a pre-experiment, which was then subsequently used in the localization phase for pose correction. Javanmardi *et al.* [4] proposed ten map evaluation factors to quantify the capability of normal distribution (ND) maps for vehicle self-localization, allowing error to be modelled with an RMSE and R^2 of 0.52 m and 0.788 respectively. The subsequent error model was then used to dynamically determine the NDT map resolution, resulting in a reduction of map size by 32.4% while keeping mean error within 0.141 m. For both studies [4], [6], the estimation of localization error used 3D point cloud and ND maps. As far as the authors are aware, no previous study has investigated the use of OpenStreetMap as a proxy for the estimation of autonomous vehicle localization error.

III. METHOD

A. Study area

The study area is Shinjuku, Tokyo, Japan. The architecture is relatively heterogeneous, with buildings ranging from low-rise shops to multi-story structures. Similarly, roads vary from single narrow lanes to wide multi-lane carriageways. The experiment path is 1.2 km and is shown in Figure 1.

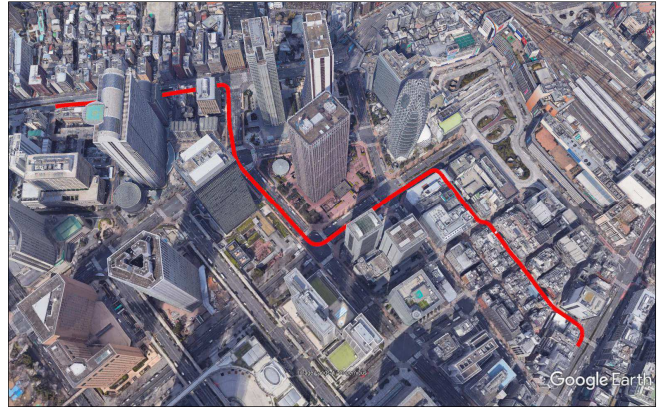


Figure 1. Experiment path in Shinjuku, Tokyo, Japan.

B. Mapping data

Mapping data from OSM was extracted from the *geofabrik.de* server. For the study area, 49 data layers (referencing 23 tags) were available. Note that each tag can be represented by a maximum of three layers, one for each geometry primitive (point, line, polygon). TABLE II. shows a summary of all the tags found in the study area and their feature counts.

TABLE II. OSM TAGS AND FEATURE COUNT IN THE STUDY AREA

OSM tags	Feature count	OSM tags	Feature count
aeroway	12	natural	595
amenity	2,109	office	35
barrier	675	place	4
boundary	16	power	20
building	10,723	public transport	114
craft	2	railway	485
emergency	8	route	100
highway	4,021	shop	580
historic	7	tourism	168
landuse	202	unknown	598
leisure	87	waterway	2
man made	12		

Note that OSM tags can represent geometric or semantic features. For example, features labeled as ‘building’ may represent the physical walls of the structure, whereas a ‘boundary’ feature may represent an administrative region, e.g., ward boundary. For this experiment, the aim is to model what the LiDAR scanner ‘sees’ during the localization phase. As such, only the geometric features from OSM were useful.

To model the surrounding environment, four OSM data layers were used: 1) building_polygon; 2) barrier_polygon; 3) barrier_line, and; 4) barrier_point. The buildings layer was selected as they represent the most stable feature within the urban landscape. Buildings, in general, are least affected by seasonality or time of day (unlike other features such as trees). For vehicle localization, however, the use of buildings alone for map matching may be insufficient. For example, on wide multi-lane roads or within the rural environment, buildings may not be present or visible by the LiDAR scanner. In these scenarios, roadside features (such as central reservations, guard rails, and pole-like features) become increasingly important. Within OSM, these are classified under the ‘barriers’ tag. Areal features, such as hedges, are classed under ‘barrier_polygon’. Linear features, such as guard rails and traffic barriers, are classed as ‘barrier_line’. Lastly, pole-like features (e.g., bollards) are classed as ‘barrier_point’.

The completeness of OSM varies within the study area. For buildings, the data is 74.3% complete when compared to GSI Japan building footprints, using area as a crude measure. For barriers, an authoritative equivalent dataset was not available for reference. In a visual comparison with aerial imagery, over half of the hedges and central reservations (polygon barriers) were missing along the experiment path. There were no line or point barriers mapped within 20 m of the experiment path (although they were present in the wider study area).

To mitigate any issues of inconsistency and incompleteness (as highlighted in Section II.D), an additional set of ‘completed’ OSM layers were created. Supplementary data (aerial imagery and base map) from ATEC and the Geospatial Information Authority of Japan were used to support the manual digitization. By creating these ‘completed’ layers, it allowed the assessment of the capability of OSM, in a scenario where users had correctly and fully mapped all required features. This allowed for a more robust evaluation, without being hindered by data quality issues. In total, 63 polygon-barriers, 31 line-barriers, and 204 point-barriers were manually added along the experiment path. The ‘completed’ OSM layers are presented in Figure 2. TABLE III. presents a summary of all the mapping data used in the study.

TABLE III. SUMMARY OF MAPPING DATA USED

Layer name	Type	Description	Count
<i>Original OpenStreetMap layers</i>			
Building	Polygon	Individual buildings or groups of connected buildings, e.g., apartments, office, cathedral	10,723
Polygon barriers	Polygon	Areal barriers and obstacles, e.g. hedges, central reservation	410
<i>Completed OpenStreetMap layers</i>			
Polygon barriers (completed)	Polygon	Areal barriers and obstacles, manually completed using aerial imagery and external reference maps.	473
Line barriers (completed)	Line	Linear barriers and obstacles, e.g. guard rails. Manually digitized using aerial imagery and external reference maps.	229
Point barriers (completed)	Point	Point barriers and obstacles, e.g. pole-like objects, lamp posts, traffic lights. Manually digitized using aerial imagery and external reference maps	271

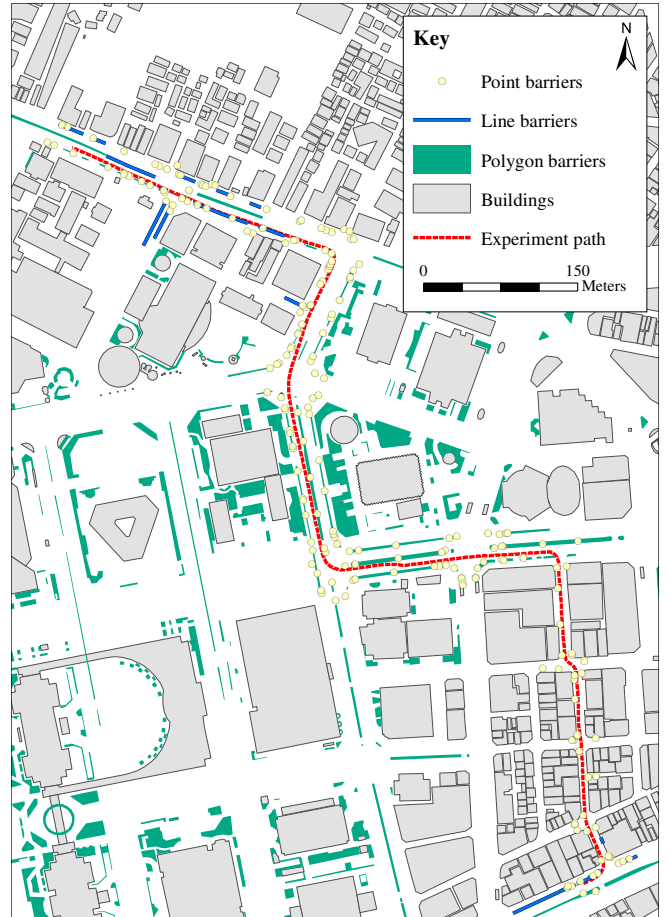


Figure 2. Map showing the ‘completed’ OSM layers and experiment path

C. Localization error data

For localization error, data from a previous experiment was used. Specifically, the mean 3D localization error from map matching using a LiDAR input scan and a prebuilt normal distribution (ND) map with a 2.0 m grid size. The process of obtaining the mean 3D localization error is described below. Firstly, a point cloud map of the area was captured using a Velodyne VLP-16 mounted on the roof of a vehicle at the height of 2.45 m. The vehicle speed was limited to 2 m/s, with a scanner frequency of 20 Hz to mitigate motion distortion. From this point cloud map, an ND map was generated with a 2.0 m grid size. Next, during the localization phase, a second scan was obtained with the laser scan range limited to 20 m. This second scan was then registered to the ND map (NDT map matching) to obtain a location. Thirdly, to evaluate the map factors, sample points along the experiment path at 1.0m intervals were selected. For each sample point, mean 3D localization error was calculated by averaging the localization error from 441 initial guesses at different positions evenly distributed around the sample point, at 0.2 m intervals within a range of 2 m. Further details on the experimental set-up and the creation of the data can be found in [4].

D. Formulating map evaluation factors

Javanmardi et al. [4] proposed ten map evaluation factors to quantify the capability of ND map for vehicle self-localization. Unlike the four map evaluation criteria, the map factors were designed specifically to evaluate 3D ND maps

and cannot be easily transferred to another mapping format without modification. Therefore, six new factors for 2D vector mapping format were defined, based on Javanmardi et al.'s [4] map evaluation criteria. Specifically, the focus was on feature sufficiency of the map and layout of the map. Local similarity and representation quality were not considered in this study, due to the high level of abstraction of OSM and the lack of a higher quality dataset available for comparison.

To model the behavior of the laser scanner and localization process, two auxiliary layers were produced: view lines and view polygons. Firstly, for each sample point, 360 view lines, each 20 m in length, radiating outwards from the center at 1° interval, were generated (Figure 3a). These lines represent the laser scanner's 20 m range. Within the urban environment, the walls of buildings represent a 'solid' and opaque barrier for the laser scanner. As OSM is 2D, it is assumed that buildings are infinitely tall, and view lines were clipped where they intersect with the building layer (Figure 3b). This assumption was based on the 30° vertical field of view (+/- 15° up and down) of the Velodyne VLP-16. Mounted at 2.45 m, and with a laser range of 20 m, the effective vertical view is between 0 to 7.8 m. This is equivalent to an average two-story building. Considering the buildings in the experiment area were, at a minimum, three stories high, this assumption was acceptable. Secondly, from these clipped view lines, a minimum bounding convex hull was generated as an areal representation of what the laser scanner 'sees' and is referred to as a view polygon (Figure 3c).

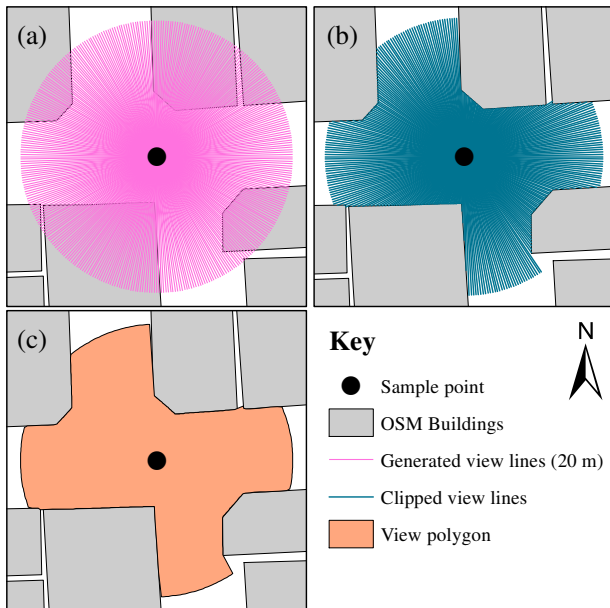


Figure 3. View lines and view polygons.

E. Factors for feature sufficiency criterion

1) Feature count

The first factor is the feature count. It is an adaption of Javanmardi et al.'s [4] factor, providing a simple count of all nearby features. However, unlike 3D ND map formats, there are three geometry primitives for 2D vector mapping: points, lines, and polygons. Further, the 3D nature of the scanner must be abstracted in 2D.

To account for the differences between the geometric representations, feature count was calculated based on the

intersection points between the layer and the view lines or view polygons.

For buildings, feature count was the total number of view lines which intersected with the buildings layer. In this scenario, the intersection points were also the endpoints of the view lines as they were previously clipped (as described in Section III.D). At these end points of the view lines, it is assumed that the sensor cannot 'see' any further past the opaque building walls.

For point barriers, this was a simple count of features which intersected with (and were therefore within) the view polygon.

For line barriers, all intersection points between all view lines and line barriers were counted.

Polygon barriers, unlike buildings, were not considered to be infinitely tall and opaque as the LiDAR sensor can often 'see' over them. Modelling this was not as straight forward, as the intersection between a view line and a polygon barrier is a line. To remedy this, points were generated along the intersection with view lines at 10 cm intervals i.e. the portion of the view line which overlaps with any polygon barrier were converted into a series of points. The value of 10 cm was arbitrary chosen as an optimal balance between resolution and computation time, after trialing multiple other values (1 cm, 5 cm, 25 cm). These generated points were then counted for the feature count. One point to note is that this method results in the generation of multiple points per view line, unlike other layers which may only have single intersection points per feature. This, however, does not affect the factors, as each layer is considered separately during the calculation.

F. Factors for the layout of the map

1) Angular dispersion of features

Angular dispersion is a measure of the arrangement of features around the sample point. Akin to Javanmardi et al.'s [4] Feature DOP measure (which in turn was inspired by GDOP in global positioning systems), this factor describes the uniformity of dispersion and is calculated as follows:

$$Angular\ dispersion = \sqrt{\left(\frac{\sum \sin \alpha}{n}\right)^2 + \left(\frac{\sum \cos \alpha}{n}\right)^2}$$

where α is the azimuth of the features, in radians. The factor returns a value between 0 and 1, with a value of 0 indicating a uniform dispersion and a value of 1 meaning a complete concentration in one direction. For this factor, the dispersion of the intersection points between the view lines or view polygons and the respective layer was calculated.

2) View line mean length

For each sample point, the mean length for every view line which intersected with the building layer was calculated. Non-intersecting view lines were not considered as they would skew the metric.

3) Area of view polygon

This is the area of the view polygon. It represents the theoretical area that the laser scanner can 'see'. A small area suggests that there are lots of building features in the vicinity to localize against.

4) Compactness of view polygon

A compact field of view could suggest that there are a lot of building features nearby to localize against. To measure compactness, the ratio between the area of the view polygon and the area of a circle with the same radius was calculated.

5) The variance of building face direction

Javanmardi et al.'s [4] suggest that if the features in the local vicinity face a greater variety of directions, then the positioning uncertainty decreases. To measure this, at all points of intersection between the view lines and the building layer, the normal of the face was calculated. Subsequently, for each sample point, the variance of all the normal vectors was calculated.

G. Random forest regression

To obtain a prediction of localization error, random forest regression was used. Random forest regression is an ensemble learning method for classification and regression. It uses multiple decision trees and bootstrap aggregation to form a prediction. Random forest regression aims to reduce variance in predictions while maintaining a low bias, thus controlling overfitting. For the implementation, scikit-learn (a machine learning Python library) was used. As random forest regression is a supervised learning method, a training dataset was required. For each model, the datasets were split into two parts: training (75%) and test (25%).

IV. RESULTS

A. General results

Six models were evaluated, using different combinations of data layers from OSM. Model 1 used just the buildings layer directly from OSM. Model 2 considered both the buildings layer and polygon barriers from OSM. From model 3 onwards, completed OSM layers were considered. Model 3 included the completed polygon barriers. Model 4, 5, and 6 then added the line barriers and point barriers individually, and then finally together.

To evaluate the goodness of fit, the R-squared (R^2), root mean square error (RMSE), mean absolute error (MAE), and accuracy are presented. In this scenario, accuracy is defined as the mean average percentage error subtracted from 100%:

$$accuracy = 1 - \frac{1}{n} \sum_{t=1}^n \left| \frac{At - Pt}{At} \right|$$

where At = actual value and Pt = predicted value. The models and their prediction errors are presented in TABLE IV.

B. Factor importance

One of the advantages of random forest regression is the ability to evaluate the importance of different variables used for prediction. TABLE V. shows an overview of the importance of each factor and data layer, with the two most important variables in bold. Note that for models 2 to 5, the feature count of polygon barriers is the first or second most important factor for predicting localization error (ranging from 0.18 to 0.38). With the inclusion of point *barriers* data in

models 5 and 6, the angular dispersion factor has high importance (0.20 and 0.22). Aside from model 1, the importance of factors based on view line and view polygon geometry is below 0.1.

TABLE IV. PREDICTION ACCURACY OF MODELS

Data	Model					
	1	2	3	4	5	6
OSM Buildings	•	•	•	•	•	•
OSM Poly. barriers		•				
Completed Poly. barriers			•	•	•	•
Completed Line barriers				•		•
Completed Point barriers					•	•
Number of factors	6	8	8	10	10	12
R^2	0.182	0.327	0.438	0.692	0.634	0.672
RMSE (cm)	5.0	4.5	4.2	3.1	3.4	3.2
MAE (cm)	3.4	3.0	2.6	2.1	2.2	2.0
Prediction Accuracy (%)	51.2	56.5	62.3	71.9	72.1	73.1

TABLE V. FACTOR IMPORTANCE

Data		Models					
		1	2	3	4	5	6
Feature count	Buildings	0.14	0.09	0.06	0.04	0.05	0.03
	Polygon barriers	-	0.28	0.38	0.18	0.26	0.13
	Line barriers	-	-	-	0.21	-	0.18
	Point barriers	-	-	-	-	0.13	0.08
Angular dispersion	Buildings	0.19	0.10	0.07	0.08	0.06	0.06
	Poly. barriers	-	0.14	0.27	0.15	0.11	0.05
	Line barriers	-	-	-	0.10	-	0.04
	Point barriers	-	-	-	-	0.20	0.22
Other	View line mean length	0.16	0.09	0.04	0.07	0.04	0.05
	Area of view poly.	0.11	0.07	0.03	0.06	0.02	0.04
	Compactness of view poly.	0.11	0.07	0.03	0.06	0.02	0.04
	Variance of building face dir.	0.28	0.17	0.10	0.07	0.10	0.06

V. DISCUSSION

A. The capability of OpenStreetMap for estimating vehicle localization error

Initial results from TABLE IV. show that using both the buildings and polygon barriers layer from OSM, the model can achieve 56.5% prediction accuracy. By improving the completeness of the polygon barriers, the model prediction accuracy is then increased to 62.3%. By adding additional point barriers and line barriers layers, a model prediction accuracy of 73.1% is achieved. Although OpenStreetMap is not specifically designed for feature scale problems such as map matching localization, the work described in this paper show that it is possible to estimate vehicle localization error with reasonable success.

For the application of autonomous vehicles, it is important to note that detecting the points which exceed a threshold error value is perhaps more important than the general accuracy of the prediction. This is because at these locations, the autonomous vehicle is unable to localize itself within its safety threshold. As such, precision and miss rate may be a more suitable assessment of the performance of a prediction model. Using the SIP program’s 25 cm value as guidance, only 20 out of 1096 sample points does localization error exceed the threshold (Figure 4). Of these instances, 18 are ‘detected’ by the model (Model 6). Further work is required to devise a more appropriate accuracy metric which better describes a model’s ability to detect these peaks in localization error, and thus its suitability for the autonomous driving application.

B. Comparison with other studies

TABLE VI. presents a comparison between Models 1 and 6 against Javanmardi et al.’s [5] ND map mean error models. While Model 1 has a relatively low R^2 (0.182), Model 6 can achieve a similar performance with an R^2 of 0.672 and RMSE of 3.2 cm. Although Javanmardi et al.’s [5] 2.0m model has a higher R^2 value (0.77), the RMSE is much higher at 9.0 cm. It is important to bear in mind the difference in methodology – while [5] uses principal component regression, this study uses random forest regression. Further work is therefore required to confirm if OpenStreetMap can truly achieve comparable results with existing approaches.

TABLE VI. COMPARISON TO OTHER MODELS

Author	Model	R^2	RMSE (cm)
Javanmardi et al. [5]	2.0m grid size	0.770	9.0
	5.0m grid size	0.551	20.6
Our models with OSM	Model 1	0.182	5.0
	Model 6	0.672	3.2

C. The importance of different factors and data layers

Of the factors, feature count is the most important for all six models. Feature count of polygon barriers is one of the top two most important factors for models 2 to 5. Comparatively, angular dispersion is less important, except for point barriers for models 5 and 6. Factors based on view line and view polygon geometry were least important. There are several possible explanations for this result. Firstly, the middle half of the experiment route is populated by many polygon barriers (as illustrated in Figure 2). In addition, in this central section

of the experiment path, the combination of wide multi-lane roads, restricted 20m scanner range, and building setback means that buildings cannot be necessarily ‘seen’ by the LiDAR sensor. In these cases, the barrier features become more important for localization. Secondly, the importance of point barriers could be the result of the completeness of the data, as pole-like features can be found all along the experiment path.

Note that these factors have only been applied to a limited experiment path. For example, in other areas where there are no point or polygon barriers, other data layers such as buildings and line barriers could become more important. In addition, only 4 of the 49 layers available from OSM was used in this study. Utilizing additional geometric OSM layers, e.g. road network, may improve the model prediction accuracy.

D. The impact of data quality of OSM

As discussed in Section II.D, OpenStreetMap suffers from data quality issues such as inconsistency and incompleteness. In this study, the data was manually corrected and improved using external datasets, resulting in a 16.6% improvement in model accuracy over using unaltered OSM data. This is encouraging and could suggest that as the quality of OSM improves over time, its capability to estimate localization error also increases. The improved layers in this study can also provide guidance to OSM users on what features to digitize for the application of autonomous vehicles. However, some of the data layers useful for vehicle localization estimation (e.g., pole-like objects) may be difficult to digitize using current data sources for OSM, e.g. satellite imagery.

E. 2D geographic information versus 3D point cloud map

In this study, 2D geographic information in a vector format from OpenStreetMap was used instead of a 3D point cloud map to predict localization error. One of the main differences is that 2D GI is a much more abstracted representation than a 3D point cloud map. From a data handling and computation perspective, it is much ‘lighter’. The vector geometry also provides clear boundaries for delineating features which is not found in point cloud maps. The main drawback of 2D GI is that it lacks the detail found within a 3D point cloud. This means feature scale phenomena such as local similarity cannot be easily evaluated. Nevertheless, even with the inherent data quality issues and lack of detail, the results from this study shows that 2D GI can still be used to assess the feature

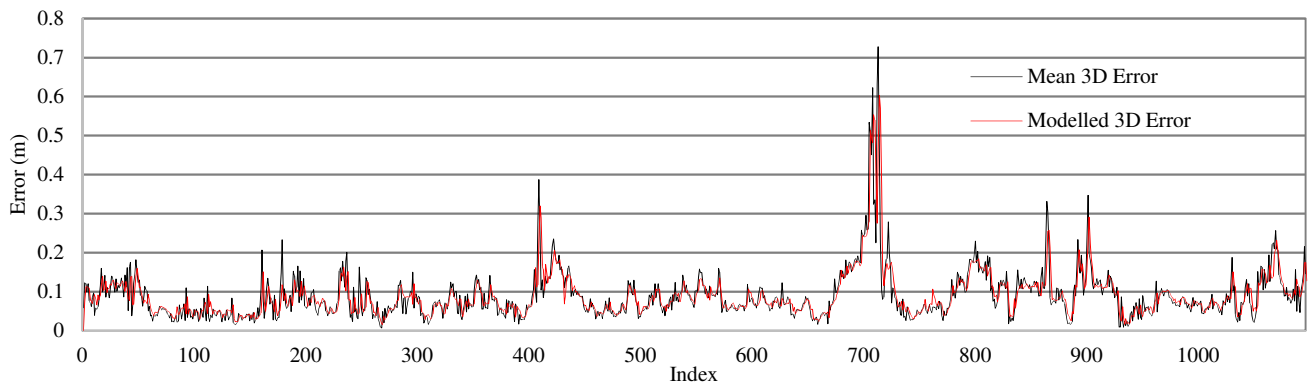


Figure 4. Modelled 3D error (red) compared to actual mean 3D error (black) for the full experiment path (Model 6)

sufficiency and layout of an area. Beyond OpenStreetMap, the same factors could be used with better quality 2D GI. Data from authoritative sources, such as national mapping agencies, could be used to mitigate data quality issues found in open data, and thus achieve higher prediction accuracy.

F. Generalizability of the model

The model is currently only trained on data from a relatively short path, resulting in relatively low generalizability. The addition of more data from multiple paths should reduce any random patterns that may previously have appeared predictive. This would, in turn, enable the use of OSM to predict localization error beyond the modeled area. Coupled with the wide coverage of OSM, it could theoretically be possible to predict localization error for large areas, anywhere in the world, without collecting data to create a prebuilt map as the mapping data is readily available.

G. Alternatives to random forest regression

Random forest regression is only one machine learning approach for prediction. In some cases, alternative approaches such as support vector machine or gradient boosting machines could outperform random forest regression and provide a more generalizable model. These approaches, however, also require additional manual hyperparameter tuning. Further work is required to ascertain which approach is best.

H. Source of localization error

As described in Section II.B, there are many sources of localization error for LiDAR map matching. The precise sources of localization error, however, remain difficult to ascertain. Within the experiment, multiple measures were taken to ensure the reduction of non-map errors, so that any remaining localization error evaluated was directly related to the map as much as possible. Regardless, it is currently difficult to confirm how much of the localization error was derived from the map or the environment. In the future, if the source of localization error is known, map producers could be encouraged to improve mapping for the application of localization. Alternatively, artificial objects could be installed in the environment to improve map matching performance. Furthermore, knowing the specific location of areas of high localization error could inform semi-autonomous vehicle systems of where to disengage autopilot and to prepare for human driver takeover in a timely manner.

VI. CONCLUSION AND FUTURE WORK

This study set out to determine if OpenStreetMap can be used as a proxy to estimate vehicle localization error. Six map evaluation factors were developed for 2D geographic information in a vector format, based on the feature sufficiency and layout of the map. These factors were used to train six different random forest regression models, for an experiment path in Shinjuku, Tokyo. The initial experiment results showed that using just two unaltered OSM data layers, the model can predict localization error with 56.5% accuracy. By improving the data quality (consistency and completeness) as well as using four data layers, the model prediction accuracy increases to 73.1%. This shows that there is potential for OSM to be used for estimating vehicle localization error.

Future work is required in adding more data layers from OSM, as well as other layers which may not currently be available, such as curb information, road markings, and landmarks. In addition, training the model with data from wider areas will help make it more generalizable, enabling prediction beyond the modeled areas. Furthermore, different machine learning models should be investigated to ascertain if it can improve prediction accuracy. The method described in this study can also be adapted to assess other localization approaches, such as camera-based systems.

ACKNOWLEDGMENT

Kelvin Wong is an International Research Fellow of Japan Society for the Promotion of Science (Postdoctoral Fellowships for Research in Japan (Short-term)).

REFERENCES

- [1] C. Ellul, M. Adjrad, and P. Groves, "The Impact of 3D Data quality on Improving GNSS Performance Using city Models Initial Simulations," *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.*, vol. IV-2/W1, pp. 171–178, Oct. 2016.
- [2] J. Levinson and S. Thrun, "Robust vehicle localization in urban environments using probabilistic maps," in *2010 IEEE International Conference on Robotics and Automation*, 2010, pp. 4372–4378.
- [3] M. Javanmardi *et al.*, "Towards High-Definition 3D Urban Mapping: Road Feature-Based Registration of Mobile Mapping Systems and Aerial Imagery," *Remote Sens.*, vol. 9, no. 10, p. 975, Sep. 2017.
- [4] E. Javanmardi, M. Javanmardi, Y. Gu, and S. Kamijo, "Factors to Evaluate Capability of Map for Vehicle Localization," *IEEE Access*, vol. 6, pp. 49850–49867, 2018.
- [5] E. Javanmardi, M. Javanmardi, Y. Gu, and S. Kamijo, "Adaptive Resolution Refinement of NDT Map Based on Localization Error Modeled by Map Factors," in *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, 2018, pp. 2237–2243.
- [6] N. Akai, L. Y. Morales, E. Takeuchi, Y. Yoshihara, and Y. Ninomiya, "Robust localization using 3D NDT scan matching with experimentally determined uncertainty and road marker matching," in *2017 IEEE Intelligent Vehicles Symposium (IV)*, 2017, pp. 1356–1363.
- [7] H. Sobreira *et al.*, "Map-Matching Algorithms for Robot Self-Localization: A Comparison Between Perfect Match, Iterative Closest Point and Normal Distributions Transform," *J. Intell. Robot. Syst.*, vol. 93(3–4), pp. 533–546, 2019.
- [8] M. Magnusson, "The three-dimensional normal-distributions transform: an efficient representation for registration, surface analysis, and loop detection," Ph.D. Thesis, Örebro universitet, Örebro, Sweden, 2009.
- [9] OpenStreetMap, "Map Features - OpenStreetMap Wiki." [Online]. Available: https://wiki.openstreetmap.org/wiki/Map_Features. [Accessed: 25-Mar-2019].
- [10] C. Barrington-Leigh and A. Millard-Ball, "The world's user-generated road map is more than 80% complete," *PLoS One*, vol. 12, no. 8, Aug. 2017.
- [11] OpenStreetMap, "WikiProject Japan/Current coverage - OpenStreetMap Wiki," 2017. [Online]. Available: https://wiki.openstreetmap.org/wiki/WikiProject_Japan/Current_coverage. [Accessed: 20-Mar-2019].
- [12] D. Zielstra and A. Zipf, "A comparative study of proprietary geodata and volunteered geographic information for Germany," in *13th AGILE international conference on geographic information science*, 2010, vol. 2010.
- [13] M. Helbich, C. Amelunxen, P. Neis, and A. Zipf, "Comparative spatial analysis of positional accuracy of OpenStreetMap and proprietary geodata," *Proc. GI_Forum*, pp. 24–33, 2012.