Mitigating Autonomous Vehicle GPS Spoofing Attacks through Scene Text Observations

Erald Troja, Joan DeBello, Nikhil Yadav, L. Truong, J. Worden
St. John's University
trojae, debelloj, yadavn, truongl, wordenj@stjohns.edu

Mehrdad Aliasgari
California State University
mehrdad.aliasgari@csulb.edu

Abstract

This paper investigates both from an empirical and a systems-based perspective, how surrounding textual information can be leveraged towards the mitigation of Autonomous Vehicle (AV) and self-driving cars Global Positioning System (GPS) signal spoofing attacks. The paper presents and proposes methods of how AVs and self-driving cars can extract, as they travel along a trajectory, surrounding textual information through machine-learning based Scene Text Recognition (STR). The paper researches and proposes geospatial models which can be applied to the extracted textual information in order to build a text-based geolocation system for the purposes of validating the received GPS signal. The ultimate contribution of the paper is to lay the groundwork towards enhancing the Cybersecurity of the current and future Autonomous Vehicle and self-driving car ecosystem by addressing its Achilles heel, namely insecure and inaccurate geolocation due to GPS spoofing attacks.

Keywords: GPS spoofing attacks, Scene Text Recognition, Geolocation, geospatial query

1. Introduction

Global Positioning Signal (GPS) geolocation is the cornerstone of a myriad of location-based services and applications. GPS geolocation is used as the first step towards providing location based services such as navigating through a city or issuing location-dependent queries to a location based service (LBS) database i.e., issuing a k nearest neighbor (kNN) query which is expected to return the k points of interest with respect to a given coordinate [Jiang et al., 2021, Ignacio, 2021, Hort et al., 2021]. Furthermore, the ubiquity of self-driving cars and Autonomous Vehicles (AV) [Perrine et al., 2020, Bansal and Kockelman, 2018, Simoni et al., 2019] is growing as their applications varying from last-mile delivery [Feng et al., 2021b], to accurate concrete wall cracking detection [Bohari et al., 2021], and water quality monitoring/water surface cleaning [Chang et al., 2021] is ever expanding. Hence, the growing need for secure and accurate geolocation is of an utmost societal importance.

1.1. Motivation

Unfortunately, currently deployed civilian mode GPS technology is identical to the way it was deployed in 1970's and it is not secure. Civilian mode GPS infrastructure assumes a honest threat model without the presence of any malicious threat actors. Hence, GPS signal can be falsified to introduce various malicious payloads and the infrastructure does not have any built-in countermeasures. For example, GPS signal can be pre-recorded at one location and can be replayed by malicious threat agents at another location to mislead any receiving devices into believing that they are located at the coordinates where the GPS signal was pre-recorded [Seco-Granados et al., 2021]. Alternatively, it has been shown that adversaries can build a portable spoofer with low costs (about $225), which can be utilized to generate spoofed GPS signals to mislead any receiving devices into believing that they are located at alternate coordinates [Zeng et al., 2018]. Therefore, research which proposes mitigation solutions towards secure and accurate GPS geolocation is of high importance and timely relevance.

1.2. Contribution

In this paper we address GPS signal spoofing by researching and investigating the problem, both from an empirical and systems-based perspective, of how surrounding textual information can be meaningfully leveraged towards secure and accurate geolocation. Our proposed mitigation solution aims to overcome the above challenges and since it requires no modifications of the current GPS infrastructure, it has the potential

URI: https://hdl.handle.net/10125/103445
978-0-9981331-6-4

for a higher societal and industrial adoption rate. The contributions of this paper can be summarized as follows. First, the authors define the GPS spoofing adversarial threat model with respect to Autonomous Vehicles and self-driving cars. Second, the authors research and investigate methods of how Autonomous Vehicles can extract, through machine-learning based Scene Text Recognition (STR), textual information from surrounding scenes as they travel along a trajectory T. Third, similar to how a human driver/operator leverages surrounding textual information in order to validate the coordinates/location shown in GPS navigation system, the authors research and propose geospatial and machine-learning models which leverage the extracted textual information in order to build a text-based geolocation system to validate the received GPS signal. Fourth, the authors evaluate the proposed methods empirically in a laboratory setting. Fifth, the authors propose a pathway towards a systems-based real-world setting deployment.

The ultimate contribution of this paper is to lay the groundwork on how one can leverage surrounding scene textual information in order to verify the GPS location of self-driving/Autonomous Vehicles. Section. 2.1 provides the system model and Section. 2.2 describes the assumed adversarial threat model. Section. 3 provides details of the state-of-the art approaches in STR and provides details of our GPS signal spoof mitigation techniques through STR including preliminary results. Lastly, Section 4 provides the pathway towards investigating, comparing and contrasting, from a systems-based deployment, in a real-world setting, the degree of accuracy and efficiency of the proposed text-based geolocation models towards the mitigation of the GPS spoofing attacks.

2. Model

In this section we introduce the system model and the assumed adversarial threat model. We then proceed to elaborate on the research steps towards the ultimate goal of enhancing the Cybersecurity of the current and future Autonomous Vehicle ecosystem by mitigating GPS signal spoofing attacks through surrounding textual information.

2.1. System Model

Our assumed system model is based on the economy of mechanism principle [CISA, 2013] which takes into account the cost, deployment overhead, effectiveness, and robustness of the proposed system. A critical aspect of our contribution is that unlike other mitigation countermeasures which require severe alterations of either the GPS satellites, GPS receivers, and ground infrastructure [Kuhn, 2004, Yan et al., 2008, Wesson et al., 2012, Jansen et al., 2018, Moser et al., 2016, Nielsen et al., 2011, Wesson et al., 2011], *our assumed system model does not require modifications of the currently deployed GPS system*, hence it is more likely to be adopted by the Autonomous Vehicle (AV) and self-driving car industry. Other navigation systems such as inertial navigation [Woodman, 2007] are orthogonal to our research since they are mainly used for aircraft and ships navigation, along with tactical and space missions. Hence, while other GPS spoofing countermeasures might be feasible as stated in [Zeng et al., 2018], the methods provided in this paper are focused on providing countermeasures based on analysis of the surrounding scene textual information.

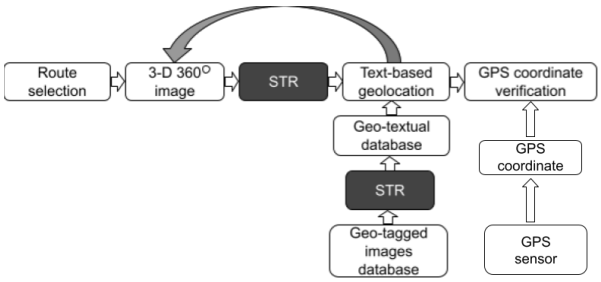

Figure 1. Overview of our proposed system model.

Our proposed methods, based on our proposed system model (Fig. 1), are contingent to the following assumptions (i) we assume that in the future all autonomous vehicles would be equipped with an overhead-mounted 360°panoramic camera similar to how Google StreetView cars and (ii) a mobile internet connection is available in order to issue geotextual queries and receive replies (Section 3.2.3), and a modern end-user computing system consisting ample CPU and RAM would be available as on-boarding computing facility in the autonomous vehicle. Our optimistic assumptions are based in the fact that the cost of 360°panoramic cameras is projected to fall while their accuracy is expected to increase in the near future [Barazzetti et al., 2018, Sun and Zhang, 2019]. Self-cleaning camera technology has been explored in [Lee et al., 2017] which can lead to excellent image capturing capabilities. Furthermore, ample on-board computing capabilities already exist in self-driving cars [Chakaravarthy et al., 2021].

Lastly, our proposed system model assumes the existence of an honest-and-trusted party (industry partner) which has the ability to accurately represent virtually the physical surroundings of the real world.

Specifically we assume that the honest-and-trusted party has the ability to collect, align, and collate imagery taken from the physical world and represent it into a database of geo-tagged images [Inc., 2021b] which can be publicly queried. In our model we assume that the honest-and-trusted party is Google and the virtual representation of the physical surroundings is represented through Google StreetView [Inc., 2021c]. The StreetView geo-tagged images database is utilized both in our empirical evaluations as well as in our real-world evaluations. Specifically, AVs will utilize the real-time 3-D 360°images to extract textual information and use it to geolocate themselves in the region by matching their real-time location's surroundings textual information to the textual information of the StreetView database.

The proposed model is meant at detecting GPS spoofing on autonomous vehicles operating in environments with ample scene signage such as Manhattan, NYC. As hinted on in Section 2.1, the accuracy of the proposed model increases in environments with more surrounding scene signage. It is one of our future research goals to analyze the accuracy of the system based on the amount of surrounding scene signage in order to derive a baseline of the system's accuracy.

2.2. Adversarial Threat Model

We now provide a brief overview of the current GPS technology, then define the capabilities, goals and limitations of the adversary.

2.2.1. Global Positioning System (GPS) GPS is a space-based radio navigation system composed of (i) **space segment** consisting of 31 atomic clock synchronized satellites broadcasting the GPS signal while orbiting in medium Earth [Petropoulos and Srivastava, 2021, Hofmann-Wellenhof et al., 2012] (ii) **control segment** consisting of worldwide monitoring and control stations that operate and maintain the proper orbit of the satellites in medium Earth (iii) **user segment** which receives the broadcast GPS signal and calculates its position through a triangulation of 3 orbiting satellites. GPS signal is provided to civilian and military concurrently through the same satellites, yet *only the military signal is encrypted in order to provide authenticity of the signal and avoid GPS spoofing* [Maps., 2021].

2.2.2. Adversarial Capabilities We assume that the adversary has access to a portable GPS spoofer which can be put together from many off-the-shelf components such as a HackRF One-based front-end, a Raspberry Pi, a portable power source and an antenna. The financial cost to purchase all the required components is approximately $225 [Zeng et al., 2018]. We assume that the adversary can (i) either attach the GPS spoofer to the Autonomous Vehicle (AV) or (ii) has the ability to maintain close proximity to the victim AV by tailgating.

2.2.3. Adversarial Goal(s) and Limitations In this paper we assume that the ultimate adversarial goal is to purposely deviate the AV in order to launch the payload of the attack i.e., ambush, rob, steal the AV and its contents. However, the targeted deviation attack is not limited to those types of payloads. For example, a GPS spoofing attack can be feasibly launched onto a busy section of a city in the hopes of luring end-users as they travel in their self-driving car and who are querying, via their self-driving car's smart-dashboard, a location-based system (LBS) via specific keywords such as "restaurants nearby" or "gas stations nearby". The payload in this case would be to increase customer foot traffic and financial revenue at the geographical location/region specified by the spoofed GPS signal since a k nearest (kNN) query would be presented to the LBS based on the spoofed GPS coordinates. Specifically, in the near future, self-driving and Autonomous Vehicles (AVs) are envisioned to be programmed to take tourists, who are oblivious to their exact coordinates/location, on individualized and customized tours of major tourist destinations [Ribeiro et al., 2021, Webster and Ivanov, 2019]. The attack in this case consists of spoofed GPS signals, targeting such tourist-carrying AVs, who might be searching, through their AVs smart-dashboard, for things such as "food near me", "bike rentals", "souvenir shop". The payload of such an attack is to deviate the AVs in order to increase tourist foot traffic onto an "attacking radius".

A sample deviating attack is shown in Fig.2 and Fig.3. Fig.2 shows the original intended route with the source being "Merrick Bl/115 Ave" and the destination being "St. John's University Queens Campus".

In Fig. 3 the large red **X** identifies the location where the GPS spoofed signal is introduced, and the attacking radius represents the geographical region where the malicious payload (ambush, robbery, theft, financial gain) is executed. Upon receiving the spoofed GPS signal, the victim is oblivious to the fact that it is being re-routed inside the attacking radius. Throughout this paper we assume that the threat agents do not have the capability to either (i) jam the communication

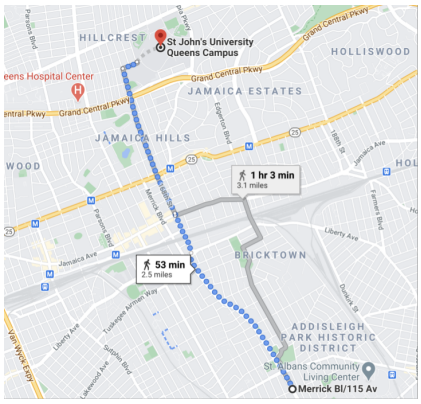

Figure 2. Original route.

systems of the AVs and (ii) destroy/vandalize neither the overhead-mounted 360°panoramic camera nor the AV's computing capabilities (iii) significantly modify/destroy the textual signs in the surrounding target area. Destruction/modification of the overhead camera will reduce the attack to a trivial one hence we do not address such in the paper. Furthermore, one of the assumptions we make is that the geotagged-images database shown in Fig 1 is kept up to date frequently i.e., any changes in the physical environment such as new street signs or awnings etc, are reflected in the geotagged-images database within days.

Figure 3. Attacked route.

3. Proposed Research

3.1. Textual Information Extraction through STR

The ultimate goal of the paper is to leverage surrounding textual information and utilize it as a mitigative countermeasure towards GPS spoofing attacks. Therefore, in a similar fashion to how a human driver/operator utilizes surrounding textual information in order to validate the coordinates/location shown in a GPS navigation system, the goal here is to utilize STR toolkits as a black box component in order to accurately detect and extract textual information from surrounding 360°panoramic imagery. The extracted textual information would then be utilized as input to our text-based geolocation models. For example, a human operator who relies on his eyesight will sporadically cross-reference and verify his GPS navigation system location by simply looking around in his immediate surrounding. However, prior to any coordinate/location verification, the human operator must consciously extract surrounding textual information. Hence, as our first step, our goal is to (i)evaluate the black box text-extraction models which will be utilized in our mitigative research approach (ii) select the highest accurate models.

3.1.1. Technical Challenges and State of the Art

The main challenge is related to the accuracy of the selected black box textual extraction tool since street signs, awnings etc., come in various fluid shapes, size, colors, complex backgrounds, various fonts, imperfect imaging conditions, and viewing angles. For example,

(a) Street sign (b) St. John's University text reference

Figure 4. Sample imagery taken at coordinates 40.721406, -73.790383 yielding texts KILDARE RD, UTOPIA PKWY, ST. JOHN'S UNIVERSITY

in Fig. 4, the adopted black box textual extraction method is expected to output the same text as what a human operator is able to extract and infer via his eyesight under optimal weather conditions. Therefore, the first challenge is to identify a black box textual extraction tool which has a very high accuracy rate under any weather and/or driving condition. While we assume that the Google StreetView 360°panoramic imagery is taken in optimal weather and lighting conditions, in a real world setting, the overhead 360°panoramic overhead-mounted camera would be operating under all sorts of lighting and weather conditions.

Text recognition in natural scenes has drawn the attention of researchers and practitioners, as indicated

by the 2021 ICDAR Robust Reading Competitions [Competition., 2021] which focus on scene text recognition. An end-to-end STR system includes many components such as text detection, text localization and verification, text segmentation, text recognition and post-processing steps [Li et al., 2017, Du et al., 2021, Wang et al., 2021, Feng et al., 2021a]. The ultimate goal of a STR is to transcribe all text regions from a target image into a target string sequence. The two most heavily used state-of-the-art end-to-end STR platforms are Google Cloud Vision (GCV) [Inc., 2021a] and the open-source Tesseract [Inc., 2021d]. In our case, from an applied systems perspective, we are concerned with the *transcription accuracy* and *transcription speed* of each platform when utilized against randomly selected images drawn from StreetView and/or real-world settings. Ultimately, our goal is to mitigate, in real-time, the GPS signal spoofing for self-driving and Autonomous Vehicles (AVs) traveling at potentially high speeds.

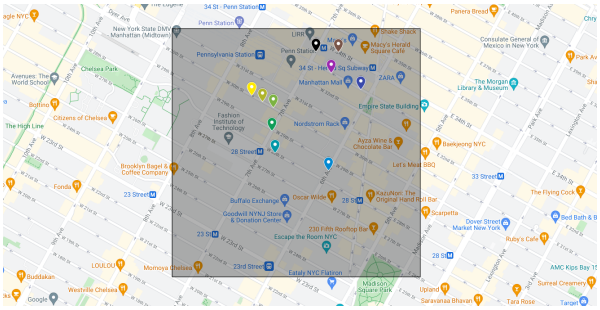

Figure 5. Sample Manhattan 1km×1km target area.

3.1.2. Approach: Evaluate Empirically and select a black box STR model As the first step, we were interested into evaluating empirically the transcription accuracy and speed of Google Cloud Vision vs Tesseract in order to choose the proper black box end-to-end STR. The input to both Google Cloud Vision and Tesseract will come from randomly selected images drawn from StreetView database by querying for images along a 1km×1km area of Manhattan as shown in Fig. 5. The transcribed textual output will allow us to compare the transcription accuracy of both Google Cloud Vision and Tesseract. We will also measure the transcription speed of both candidate black box STRs as the images are being processed and transcribed.

3.1.3. Preliminary Work We have identified our target region towards the evaluation of the two candidate end-to-end STRs. The shaded square in Fig. 5 shows the target 1km×1km Manhattan region from where we randomly sampled the StreetView database geotagged images. The lat/lon of the bounding box are (top left) 40.751,-73.997 and (bottom right) 40.742,-73.985.

Figure 6. Sample image captured at lat:40.749092, lon:-73.993929.

We have selected a few preliminary data points consisting of geotagged images, and have started experimenting with Google Cloud Vision (GCV). On average, the transcription accuracy is 96.5%. However, one of the issues we are running into is to figure out how to extract textual information in a human manner through GCV. For example Fig. 6 shows a portion of the panoramic image taken at lat:40.749092, lon:-73.993929. This corresponds to the yellow pin of Fig. 5. While a human should be able to extract the textual information as "Kaufman Furs", "HIMA & PRODI" in our preliminary work, we are getting tokenized versions of the textual information such as "Kaufman", "Furs", "HIMA", "&","PRODI". Hence, while the GCV confidence level remains very high (96%) in this step, one research venue would be to modify Tesseract by employing methods of [Du et al., 2021], [Wang et al., 2021], and [Feng et al., 2021a] through more meaningful NLP approaches. One of the challenges presented in this phase is the side effect of utilizing machine-learning textual extraction toolkits related to their inability to discern spurious/temporary text which is not part of the permanent environment. For example, in Figure 4(a), a human is able to discern spurious/temporary text "CLEANERS" which is part of a temporary, passing white vehicle (van). One possible solution is to utilize time lapsed images of the same location in order to identify spurious/temporary text. For example, if a second or third panoramic image is used from location lat:40.749092, lon:-73.993929 such that the images differ in a few hours/weeks worth of time, it might be possible to identify spurious/temporary textual

references since the permanent textual references would appear over and over despite the time lapse.

3.2. Text-based Geolocation

In this part of the paper we are interested into building a text-based geolocation system where queries are identified via multiple STR-extracted keywords and the returned value is a location ID.

3.2.1. Technical Challenges and State of the Art

Assuming that the STR black box component yields a high transcription accuracy and speed, the most challenging part is the organization of the textual information in such a way that it can be utilized to effectively answer text-based geolocation queries. For example, as shown in Fig. 7, assuming that the only surrounding textual information extracted and provided by the black box STR is "7-eleven", when querying a geotagged image database via the extracted keywords, the vast results might prohibit an accurate geolocation.

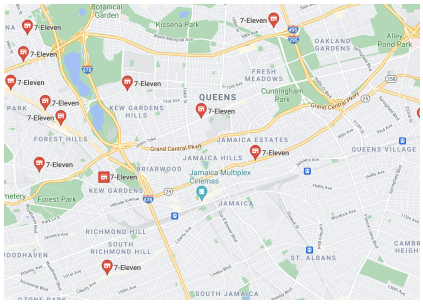

Figure 7. Geolocation via "7-eleven" keyword.

However, as more textual information is utilized as part of the set of the querying keywords, the accuracy of the text-based method increases. Fig. 8 shows the text-based geolocation attempt when two textual references are used in this case "7-eleven" and "cheeper peepers". The research challenge here is to find an efficient way to index the images based on their textual information such that the query, consisting of the set of keywords, returns the image ID that contains the most keywords.

Most studies [Bouros et al., 2012] [Fan et al., 2012] [Hu et al., 2015] [Rao et al., 2014] focus on the use of indexing in order to efficiently find object pairs that are spatially close and textually similar. For example, [Fan et al., 2012] designs a spatial signature and a textual signature for each object and utilize them to prune dissimilar object pairs. [Rao et al., 2014] develops two spatial-first and two text-first indexing schemes. [Hu et al., 2015]

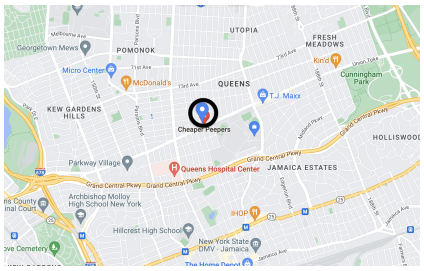

Figure 8. Geolocation via "7-eleven" and "cheeper peepers" keywords.

generates a spatio-textual signature set for each object and leverage these sets to prune dissimilar object pairs. [Bouros et al., 2012] propose different spatial-index-based algorithms. [Pat and Kanza, 2017] focuses towards the problem of geosocial search over geotagged posts. [Pat and Kanza, 2017] introduces a novel two-step search process of (i) quickly finding relevant areas by using an arbitrarily indexed partition of the space, and (ii) applying clustering to the geotagged posts in the discovered areas, to present more accurate results.

3.2.2. Approach: Develop a baseline text-based geolocation proof-of-concept model prototype As the first step towards a text-based geolocation method, we will test the feasibility and accuracy of querying the geotagged images database via the surrounding STR keywords by building a proof-of-concept prototype. Specifically in this phase we will utilize the 56k 360°panoramic images drawn from the StreetLearn dataset [DeepMind and Inc., 2021] [Hermann et al., 2020] and built a relational model such that the keys represent the STR keywords, and the values represent the StreetLearn image IDs where the text keyword has been observed and extracted.

table Images		
keywords	imageID	coordinates
7-eleven	492232	40.695, -74.023
7-eleven	992334	40.123,-74.044
7-eleven	910032	40.123,-74.023
cheeper peepers	492232	40.4695,-74.023

Figure 9. Relational structure depicting the proof-of-concept text keywords and Image ID relationship.

The research goal of our approach is to know if we can safely rely on textual surrounding information for accurate geolocation purposes. Note that efficiency evaluations are not the main focus here. For example,

Fig. 7 shows that text-based geolocation is not feasible when only one keyword is observed in the surrounding scene. However, when two or more keywords are observed, as shown in Fig. 8 accurate geolocation can occur. Here we will model the surrounding textual information geolocation by (i) selecting all the image IDs via the observed and extracted surrounding keywords, and (ii) performing an inner joint on the matching image IDs. Since each image in the StreetLearn database has been captured 5 to 10 meters apart from each other, extracted keywords overlap in several images. This gives us the textual geolocation proof-of-concept that we seek. A partial relational structure corresponding to the text-based geolocation method of Figs. 7 and 8 is shown in Fig. 9. Figs 7 and 8 are respectively generated via the following 2 queries, assuming the existence of the relational table *Images* with three columns titled *keywords, imageID, coordinates*.

```
Q1: select coordinates from Images
where keywords="7-eleven";

Q2: select i1.coordinates from
( select * from Images
where keywords="7-eleven")
as i1 inner join
( select * from Images
where title="cheeper peepers")
as i2 on i1.keywords=i2.keywords;
```

3.2.3. Approach: Model text-based geolocation through spatial keyword query processing In this section we model the text-based geolocation as a spatial keyword query processing problem. Specifically, we define a geo-textual object to consist of the geographical location where each 360°panoramic photo has been captured along with the STR extracted textual information. We can apply the text extraction models selected previously and pre-process each of 56K 360°panoramic images from the StreetLearn database to form a database D of geo-textual objects. Formally, each spatial object $o \in D$ is defined as the pair $(o.c, o.k)$ where $o.c$ is the 2-dimensional GPS coordinates of the 360°panoramic image, and $o.k$ is the set of STR extracted textual information associated with coordinate $o.c$. We approach the text-based geolocation problem, as a modified boolean range query (BRQ). Specifically, given a query $q = \langle k, r \rangle$ where k is the set of keywords, and r is a query spatial region, the results of a boolean range query $q(D)$, is the subset of D containing all of the objects such that $\forall o \in q(D)(o.c \in q.r \wedge q.k \subseteq o.k)$. One observation is that a modified BRQ can be utilized towards textual-based geolocation under two settings. In the first setting, the anonymous vehicle (AV) attempts to perform initial geolocation based on surrounding textual imagery, similar to what a GPS unit performs during the first fix GPS geolocation [University., 2020]. In the second setting, the anonymous vehicle (AV) is aware of the initial source location and requires geolocation services as it travels along a priori trajectory T. In the first setting, one immediate observation is that we can utilize BRQ by setting the bounding region $q.r$ to a very large region i.e., entire Manhattan, and $q.k$ can be specified according to the actual surrounding textual information. However, one immediate issue with BRQ is that the results are not ranked based on the text relevance. In the second setting, when the the AV requires continuous geolocation services, it would be beneficial to retrieve contiguous records in order to minimize the amount of queries sent. Here we will explore to modify and extend the BRQ query in order to retrieve ranked results i.e., according to the relevance of surrounding keywords $o.k$ and utilize the modified version of BRQ to support the verification of the initial first-fix GPS geolocation based on surrounding textual information.

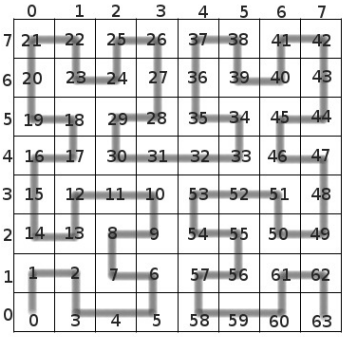

Figure 10. Sample level 3 Hilbert SFC.

With respect to the continuous GPS signal verification, one immediate observation is that the query should ideally retrieve, not only the immediate coordinates, but also surrounding ones. The intuition is to organize the spatial database in such a way as to retrieve contiguous spatial objects through one query. For this setting we will explore the application of two crucial geospatial structures namely (i) space filling curves (SFCs) (ii) Voronoi diagram. SFCs have the ability to map space from 2-D into 1-D hence nearby points in the SFC line will be guaranteed to be spatially close in their 2-D representation. A popular SFC is the Hilbert SFC [Kamel and Faloutsos, 1993] shown in Fig. 10. Therein, one can notice 64 records, which are indexed according to their Hilbert SFC ID. We first studied this problem in [Troja and Bakiras, 2015], in the context of mobile clients in a database-driven DSA deployment and we introduced two processing

Table 1. Sample DB segmentation with 4 segments

D segment 0	D segment 1	D segment 2	D segment 3
0–7	8–15	16–23	24–31
32–39	40–47	48–55	56–63
64–71	72–79	80–87	88–95
96–103	104–111	112–119	120–127
128–135	136–143	144–151	152–159
160–167	168–175	176–183	184–191
192–199	200–207	208–215	216–223
224–231	232–239	240–247	248–255

improvements. First we experimented with various ways of how to index the spatial database D. We experimented with different space filling curves which allowed query retrievals to process spatially consecutive locations of D. Furthermore, we introduced a spatial retrieval method where the database D was split into multiple, k disjoint segments. This method which allowed us to retrieve k times the number of spatially close records can be applied towards retrieval of the required records in order to significantly reduce the overall amount of issued queries during the traversal of the trajectory T.

[Christoforaki et al., 2011] proposed several evaluations of hybrid geo-textual retrieval schemes, which combine SFCs with inverted index files. Among them SFC-QUAD is shown to perform best according to the experimental results in [Christoforaki et al., 2011]. Specifically, SFC-QUAD index is built of an inverted file where all of the compressed geo-textual objects in the inverted list is assigned and ordered through their spatial position in the SFC. Since SFC-QUAD is an optimized geospatial data structure, we will utilize it with the goal of improving the baseline proof-of-concept accuracy and efficiency. By using the SFC-QUAD data structure, another possible research approach is to investigate the retrieval optimization based on partitioning of the database in multiple segments. For example, a toy geo-textual database D, consisting of 256 geo-textual objects which are indexed through their SFC ID, can be partitioned into 4 segments according to Table 1. As shown in Fig. 11, retrieval of consecutive geo-textual objects can be performed by retrieving the entire first row containing all of the 4 geospatially close segments. Ultimately the optimization research question we are solving in here is: What is the optimal database segmentation which minimizes the number of queries with respect to a priori trajectory T?

Another interesting venue, with respect to the geo-textual localization, is to create a spatial index based on the Voronoi diagram [Du et al., 1999]. In its most basic state, a Voronoi diagram (VD) has the property to partition space into disjoint polygons based on a set of arbitrary generators G and to associate all locations in the plane to its respective closest generator $g \in G$. Therefore, the nearest neighbor of any query point which is located inside a Voronoi polygon is the generator $g \in G$. A sample space partitioning as well as the nearest neighbor (NN) query through Voronoi diagram (VD) indexing is shown in Fig. 12.

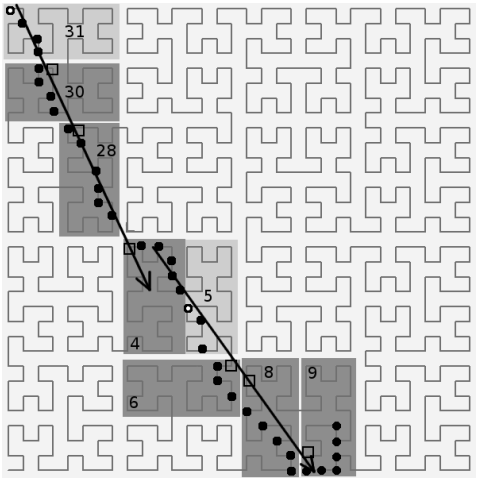

Figure 11. 4 segments partitioning.

A research pursuit in here would be to build a VD index, with the generators g being the set of all spatial objects o in the database D. Continuous Autonomous Vehicle (AV) geo-textual localization would then utilize not only the immediate surrounding textual information, but also the immediate nearest spatial objects. For example, an AV who is located at point p_1 will utilize its surrounding textual keywords $q.k$ to geolocate itself at point p_1. However, through a VD index, it can seamlessly retrieve all of the nearest neighbors (NN) by selecting all of the adjoining Voronoi cells. Here an Autonomous Vehicle (AV), incapable of accurately calculating its first-fix at point p_1 (through textual geolocation), can utilize the textual information contained in the nearest neighbor (NN) objects to approximate its hypothesized first-fix textual geolocation coordinates.

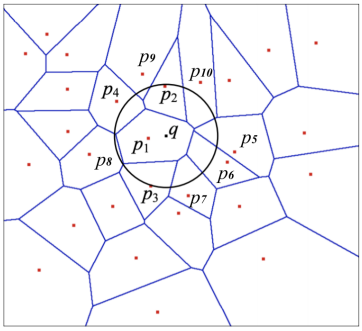

Figure 12. NN query through Voronoi diagram.

4. Conclusion and Future Work

The paper presents an approach to GPS signal spoofing mitigation by researching and investigating the problem, both from an empirical and systems-based perspective, of how surrounding textual information can be meaningfully leveraged towards secure and accurate geolocation. The paper presents a mitigation solution that aims to overcome the above challenges and since it requires no modifications of the current GPS infrastructure, it has the potential for a higher societal and industrial adoption rate. The contributions of this paper can be summarized as follows. The paper elaborates and emphasizes the GPS spoofing adversarial threat model with respect to Autonomous Vehicles and self-driving cars. The paper researches and investigates methods of how Autonomous Vehicles can extract, through machine-learning based STR, textual information from surrounding scenes as they travel along a trajectory T. In a similar fashion to how a human driver/operator leverages surrounding textual information in order to validate the coordinates/location shown in GPS navigation system, the paper researches and proposes geospatial and machine-learning models which leverage the extracted textual information in order to build a text-based geolocation system to validate the received GPS signal. The paper's future research directions are to evaluate the proposed methods empirically in a laboratory setting and provide preliminary results in order to build a baseline of system accuracy. Another research direction includes a pathway towards a systems-based real-world setting deployment. The ultimate research goal extension of the paper is to build a real-time physical systems model, where 360°images which are captured in real-time from an overhead mounted panoramic camera, are utilized as explained in the the scene text imagery extraction method of Section 3.

References

[Bansal and Kockelman, 2018] Bansal, P. and Kockelman, K. M. (2018). Are we ready to embrace connected and self-driving vehicles? a case study of texans. *Transportation*, 45(2):641–675.

[Barazzetti et al., 2018] Barazzetti, L., Previtali, M., and Roncoroni, F. (2018). Can we use low-cost 360 degree cameras to create accurate 3d models? *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*, 42(2).

[Bohari et al., 2021] Bohari, S. N., Amran, A. U., Zaki, N. A. M., Suhaimi, M. S., and Rasam, A. R. A. (2021). Accuracy assessment of detecting cracks on concrete wall at different distances using unmanned autonomous vehicle (uav) images. In *IOP Conference Series: Earth and Environmental Science*, volume 620, page 012005. IOP Publishing.

[Bouros et al., 2012] Bouros, P., Ge, S., and Mamoulis, N. (2012). Spatio-textual similarity joins. *Proceedings of the VLDB Endowment*, 6(1):1–12.

[Chakaravarthy et al., 2021] Chakaravarthy, R. V., Kwon, H., and Jiang, H. (2021). Vision control unit in fully self driving vehicles using xilinx mpsoc and opensource stack. In *2021 26th Asia and South Pacific Design Automation Conference (ASP-DAC)*, pages 311–317. IEEE.

[Chang et al., 2021] Chang, H.-C., Hsu, Y.-L., Hung, S.-S., Ou, G.-R., Wu, J.-R., and Hsu, C. (2021). Autonomous water quality monitoring and water surface cleaning for unmanned surface vehicle. *Sensors*, 21(4):1102.

[Christoforaki et al., 2011] Christoforaki, M., He, J., Dimopoulos, C., Markowetz, A., and Suel, T. (2011). Text vs. space: efficient geo-search query processing. In *Proceedings of the 20th ACM international conference on Information and knowledge management*, pages 423–432.

[CISA, 2013] CISA (2013). *Economy of Mechanism Principle*. https://us-cert.cisa.gov/bsi/articles/knowledge/principles/economy-of-mechanism.

[Competition., 2021] Competition., R. R. (2021). *ICDAR 2021 competition on Document Visual Question Answering.* https://rrc.cvc.uab.es/.

[DeepMind and Inc., 2021] DeepMind and Inc., G. (2021). *StreetLearn Dataset.* https://sites.google.com/view/streetlearn/dataset?authuser=0.

[Du et al., 2021] Du, C., Wang, Y., Wang, C., Xiao, B., and Shi, C. (2021). Unconstrained end-to-end text reading with feature rectification. *Pattern Recognition Letters*.

[Du et al., 1999] Du, Q., Faber, V., and Gunzburger, M. (1999). Centroidal voronoi tessellations: Applications and algorithms. *SIAM review*, 41(4):637–676.

[Fan et al., 2012] Fan, J., Li, G., Zhou, L., Chen, S., and Hu, J. (2012). Seal: Spatio-textual similarity search. *arXiv preprint arXiv:1205.6694*.

[Feng et al., 2021a] Feng, W., Yin, F., Zhang, X.-Y., He, W., and Liu, C.-L. (2021a). Residual dual scale scene text spotting by fusing bottom-up and top-down processing. *International Journal of Computer Vision*, 129(3):619–637.

[Feng et al., 2021b] Feng, X. et al. (2021b). Time and cost efficiency of autonomous vehicles in the last-mile delivery: A uk case. *International Business Research*, 14(3):1–26.

[Hermann et al., 2020] Hermann, K. M., Malinowski, M., Mirowski, P., Banki-Horvath, A., Anderson, K., and Hadsell, R. (2020). Learning to follow directions in street view. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 11773–11781.

[Hofmann-Wellenhof et al., 2012] Hofmann-Wellenhof, B., Lichtenegger, H., and Collins, J. (2012). *Global positioning system: theory and practice*. Springer Science & Business Media.

[Hort et al., 2021] Hort, M., Kechagia, M., Sarro, F., and Harman, M. (2021). A survey of performance optimization for mobile applications. *IEEE Transactions on Software Engineering*.

[Hu et al., 2015] Hu, H., Li, G., Bao, Z., Feng, J., Wu, Y., Gong, Z., and Xu, Y. (2015). Top-k spatio-textual similarity join. *IEEE Transactions on Knowledge and Data Engineering*, 28(2):551–565.

[Ignacio, 2021] Ignacio, A. E. (2021). Implementation of an android mobile location-based service application for general auto repair shops. *International Journal of Multidisciplinary: Applied Business and Education Research*, 2(1):49–62.

[Inc., 2021a] Inc., G. (2021a). *Detect text in images through GCV.* https://cloud.google.com/vision/docs/ocr.

[Inc., 2021b] Inc., G. (2021b). *Explore through StreetView.* https://www.google.com/streetview/explore/.

[Inc., 2021c] Inc., G. (2021c). *StreetView Representation.* https://www.google.com/streetview/.

[Inc., 2021d] Inc., G. (2021d). *Tesseract OCR.* https://github.com/tesseract-ocr/r.

[Jansen et al., 2018] Jansen, K., Schäfer, M., Moser, D., Lenders, V., Pöpper, C., and Schmitt, J. (2018). Crowd-gps-sec: Leveraging crowdsourcing to detect and localize gps spoofing attacks. In *2018 IEEE Symposium on Security and Privacy (SP)*, pages 1018–1031. IEEE.

[Jiang et al., 2021] Jiang, H., Li, J., Zhao, P., Zeng, F., Xiao, Z., and Iyengar, A. (2021). Location privacy-preserving mechanisms in location-based services: A comprehensive survey. *ACM Computing Surveys (CSUR)*, 54(1):1–36.

[Kamel and Faloutsos, 1993] Kamel, I. and Faloutsos, C. (1993). On packing R-trees. In *ACM CIKM*, pages 490–499.

[Kuhn, 2004] Kuhn, M. G. (2004). An asymmetric security mechanism for navigation signals. In *International Workshop on Information Hiding*, pages 239–252. Springer.

[Lee et al., 2017] Lee, K. Y., Hong, J., and Chung, S. K. (2017). Smart self-cleaning lens cover for miniature cameras of automobiles. *Sensors and Actuators B: Chemical*, 239:754–758.

[Li et al., 2017] Li, H., Wang, P., and Shen, C. (2017). Towards end-to-end text spotting with convolutional recurrent neural networks. In *Proceedings of the IEEE international conference on computer vision*, pages 5238–5246.

[Maps., 2021] Maps., T. (2021). *Pseudo Random Code.* https://www.trimble.com/gps_tutorial/sub_pseudo.aspx.

[Moser et al., 2016] Moser, D., Leu, P., Lenders, V., Ranganathan, A., Ricciato, F., and Capkun, S. (2016). Investigation of multi-device location spoofing attacks on air traffic control and possible countermeasures. In *Proceedings of the 22nd Annual International Conference on Mobile Computing and Networking*, pages 375–386.

[Nielsen et al., 2011] Nielsen, J., Broumanxdan, A., and Lachapelle, G. (2011). Gnss spoofing detection for single antenna handheld receivers. *NAVIGATION, Journal of the Institute of Navigation*, 58(4):335–344.

[Pat and Kanza, 2017] Pat, B. and Kanza, Y. (2017). Where's waldo? geosocial search over myriad geotagged posts. In *Proceedings of the 25th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pages 1–10.

[Perrine et al., 2020] Perrine, K. A., Kockelman, K. M., and Huang, Y. (2020). Anticipating long-distance travel shifts due to self-driving vehicles. *Journal of Transport Geography*, 82:102547.

[Petropoulos and Srivastava, 2021] Petropoulos, G. P. and Srivastava, P. K. (2021). *GPS and GNSS Technology in Geosciences*. Elsevier.

[Rao et al., 2014] Rao, J., Lin, J., and Samet, H. (2014). Partitioning strategies for spatio-textual similarity join. In *Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Analytics for Big Geospatial Data*, pages 40–49.

[Ribeiro et al., 2021] Ribeiro, M. A., Gursoy, D., and Chi, O. H. (2021). Customer acceptance of autonomous vehicles in travel and tourism. *Journal of Travel Research*, page 0047287521993578.

[Seco-Granados et al., 2021] Seco-Granados, G., Gomez-Casco, D., López-Salcedo, J. A., and Fernandez-Hernandez, I. (2021). Detection of replay attacks to gnss based on partial correlations and authentication data unpredictability. *Gps Solutions*, 25(2):1–15.

[Simoni et al., 2019] Simoni, M. D., Kockelman, K. M., Gurumurthy, K. M., and Bischoff, J. (2019). Congestion pricing in a world of self-driving vehicles: An analysis of different strategies in alternative future scenarios. *Transportation Research Part C: Emerging Technologies*, 98:167–185.

[Sun and Zhang, 2019] Sun, Z. and Zhang, Y. (2019). Accuracy evaluation of videogrammetry using a low-cost spherical camera for narrow architectural heritage: An observational study with variable baselines and blur filters. *Sensors*, 19(3):496.

[Troja and Bakiras, 2015] Troja, E. and Bakiras, S. (2015). Efficient location privacy for moving clients in database-driven dynamic spectrum access.

[University., 2020] University., P. (2020). *The Almanac, Time to First Fix and Satellite Health.* https://www.e-education.psu.edu/geog862/node/1739.

[Wang et al., 2021] Wang, W., Xie, E., Li, X., Liu, X., Liang, D., Zhibo, Y., Lu, T., and Shen, C. (2021). Pan++: Towards efficient and accurate end-to-end spotting of arbitrarily-shaped text. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

[Webster and Ivanov, 2019] Webster, C. and Ivanov, S. (2019). Future tourism in a robot-based economy: a perspective article. *Tourism Review*.

[Wesson et al., 2012] Wesson, K., Rothlisberger, M., and Humphreys, T. (2012). Practical cryptographic civil gps signal authentication. *NAVIGATION, Journal of the Institute of Navigation*, 59(3):177–193.

[Wesson et al., 2011] Wesson, K. D., Shepard, D. P., Bhatti, J. A., and Humphreys, T. E. (2011). An evaluation of the vestigial signal defense for civil gps anti-spoofing. In *Proceedings of the 24th International Technical Meeting of the Satellite Division of The institute of navigation (ION GNSS 2011)*, pages 2646–2656.

[Woodman, 2007] Woodman, O. J. (2007). An introduction to inertial navigation. Technical report, University of Cambridge, Computer Laboratory.

[Yan et al., 2008] Yan, G., Olariu, S., and Weigle, M. C. (2008). Providing vanet security through active position detection. *Computer communications*, 31(12):2883–2897.

[Zeng et al., 2018] Zeng, K. C., Liu, S., Shu, Y., Wang, D., Li, H., Dou, Y., Wang, G., and Yang, Y. (2018). All your {GPS} are belong to us: Towards stealthy manipulation of road navigation systems. In *27th {USENIX} Security Symposium ({USENIX} Security 18)*, pages 1527–1544.