2018 NDIA GROUND VEHICLE SYSTEMS ENGINEERING AND TECHNOLOGY SYMPOSIUM
AUTONOMOUS GROUND SYSTEMS (AGS) TECHNICAL SESSION
AUGUST 7-9, 2018 - NOVI, MICHIGAN

CROSS CORRELATING GROUND-LEVEL PANORAMAS WITH SATELLITE IMAGERY FOR GPS-DENIED LOCALIZATION OF AUTONOMOUS GROUND VEHICLES

Calvin Cheung
U.S. Tank Army
Research Development
Engineering Center
Warren, MI

Stanley Baek
Department of Electrical and
Computer Engineering
University of Michigan
Dearborn, MI

ABSTRACT

In this paper, we explore the usage of normalized cross-correlation to perform localization in GPS-denied environments. Spherical panoramic images from a ground vehicle are transformed to top-down viewpoints and compared with satellite imagery using normalized cross-correlation to find the vehicle's location within the satellite image. The implementation of this system has yielded positive results when tested upon publicly available panoramic images and satellite imagery, with the identified locations being within an average of 46.95 meters from the ground truth.

INTRODUCTION

Localization is one of the fundamental building blocks required for mobile autonomous ground vehicles [1]. For an autonomous system to properly navigate within an environment, it must first determine where it is at. While the frame of reference for localization may differ based on the size and functions of the autonomous system, large scale autonomous ground vehicles (AGV) are primarily concerned with their absolute position with the Earth as the reference frame. In ideal scenarios, Global Positioning System (GPS) sensors can be used to provide this information to AGVs. However, ideal GPS reception cannot be guaranteed, since GPS signal degradation can occur for several reasons, including occlusion from buildings and vegetation, GPS spoofing, and GPS denial attacks [2]. To ensure proper AGV operation during times of GPS signal degradation, it is vital to have methods to perform GPS-denied localization.

GPS-denied localization is an active area of research, with many different techniques being used to attack the problem. Progress has been made through the use of additional sensors, such as ground penetrating RADAR to perform localization with priori ground maps [3]. In addition, visual based methods such as geo-tagged landmark recognition [4] and graph-based street representations [5] have been successfully demonstrated as a viable option for visual based localization. Despite the progress with these methods however, the reliance on specialized sensors suites and labeled map data is not optimal for lower cost systems operating in infrequently traveled environments. The purpose of the work presented in this paper is to allow for GPS-denied localization without the high cost of specialized sensors and priori labeled maps, needing only widely available satellite images, and spherical panoramic images from the vehicle. This method of localization will allow for continued utilization of mobile autonomy in areas of GPS signal degradation.

RELATED WORK

GPS-denied localization has had a wide breadth of research efforts, due to the wide variety of localization needs of the different mobile robots that are in use. One of the most widely utilized methods of localization without GPS is Simultaneous Localization and Mapping (SLAM). In general, SLAM is the process by which an AGV builds a global map of its environment, all the while localizing to the map that is being generated [6]. Several different sensors have been used with this process, including LIDAR [7], SONAR [8], and monocular cameras [9]. All this is possible without the use of GPS in the system. While this type of localization works for many different applications, the efforts described here focus on finding absolute position with Earth as the frame of reference.

In addition to the use of SLAM, AGV localization has been performed using additional sensors or communication modalities that are custom designed to solve this specific problem. Special sensors, such as localizing ground penetrating radar [2], has been used to compare an AGV's location against features that are not easily measurable through the normal AGV sensing modalities of RADAR, LIDAR, and cameras. In other cases, specialized communication methods are leveraged to track relative positions of vehicles while GPS is not available, such as ultrawide band radios [11] and groundwave radio frequency signals [10]. Even though the addition of specialized sensors and communications equipment has been able to help with the localization problem, the additional hardware can be costly or logistically prohibitive, making a more barebones solution more attractive depending on mission needs.

To meet the needs of finding absolute positioning without custom sensors or radios, vision-based localization methods have been shown to be able to perform this task at varying levels of success. Landmark recognition has been used to compare ground level objects to images in geotagged databases to accurately find location [4]. Another similar method used visual odometry to determine road segments, which were then compared to crowdsourced maps contain a priori road segment information to determine a vehicles location [5]. Despite the success of these methods, they constrain the problem space to regions in which detailed a priori data, such as geotagged images and user updated graph-based street representations, are readily available. Another vision-based localization method looked at used scale-invariant feature transform (SIFT) to match feature descriptors of satellite images segmented into tiles against warped panoramic images to find matching locations [12]. The warping and matching concepts described in this method served as the foundation for the work presented here, with a focus on different recognition and localization techniques that are more robust to the natural changes that are present due to the temporal differences between satellite image capture and ground image capture.

LOCALIZATION PROCESS

The localization process developed for the work presented in this paper has four primary steps:

1. Capture a spherical panoramic image on a vehicle and warp it to produce a top-down view.
2. Run Canny edge detection on both the top-down view and the satellite image and perform normalized cross correlation on the resulting images.
3. Filter matches based on distance from the last known good point.
4. Apply Kalman filter step to matched point and report the location.

The following will go into detail regarding the process, assumptions, and limitations for each step.

Spherical Panoramic to Top-down View

The first step in the localization process is to create a top-down view of an area based on a spherical panoramic image captured from the vehicle. This is done by constructing a rough 3D model from the panoramic image, as described by Xiao [13]. At a high level, that algorithm uses panoramic image width W, panoramic image height H, and camera height c^h to transform every image coordinate point (x, y) to a 3D point $P = (X, Y, Z)$. The vanishing line of the ground plane 0 is assumed to be the bottom half of the transformed 3D sphere, as shown in Figure 1.

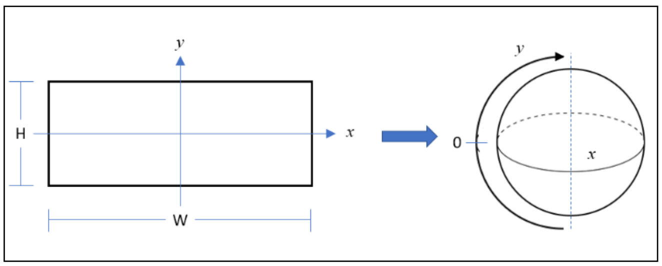

Figure 1. Spherical Panoramic to 3D Model.

Using the 3D model, we can take images from any desired perspective within the model. By setting the capture point to be at the top center of the model facing downward, we are able to create a top-down perspective of the image and save it for comparison with the satellite images, as shown in Figure 2.

Edge Detection and Correlation Matching

Finding the best match between the warped top-down view and the satellite map is a two-part process. First, Canny edge detection is performed on both the top-down view and the satellite map. Canny edge detection is a multi-stage detector algorithm that both finds edges and suppresses noise [14]. By comparing the Canny detected edges instead of the raw images, the results are more resistant against in color, lighting, and season that naturally occur due to satellite imagery being captured at a point in the past and not present time, as shown in Figure 3.

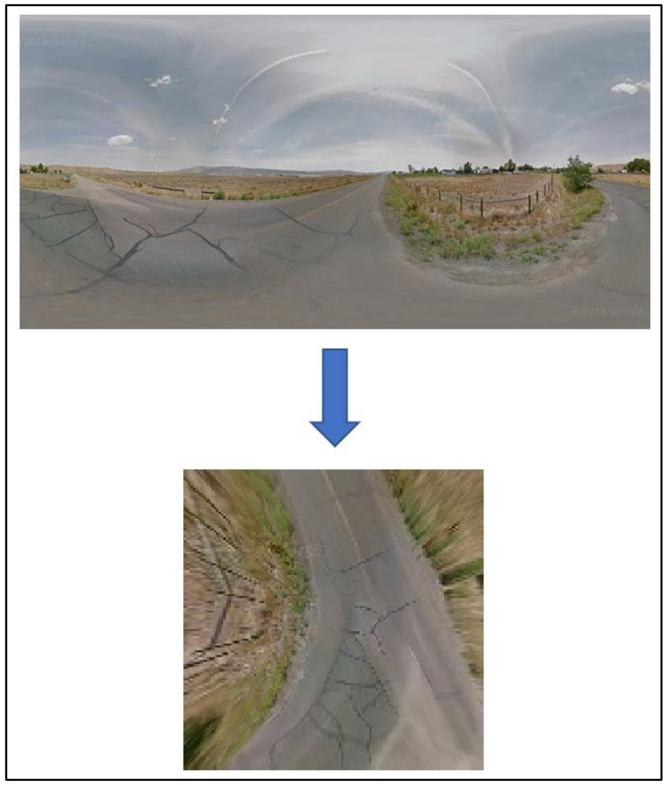

Figure 2. Spherical Panoramic to Top-down View.

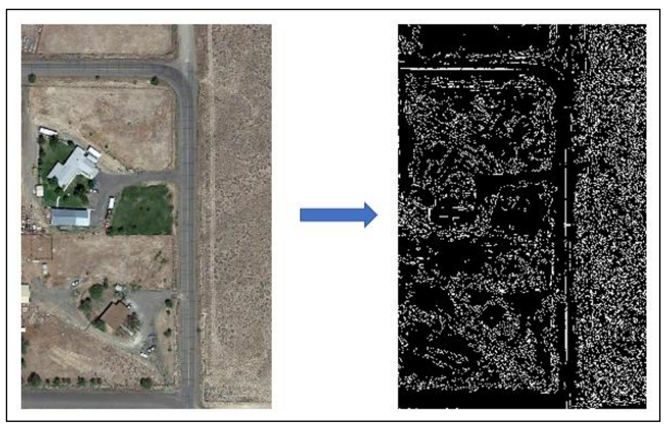

Figure 3. Canny Edge Detection on Satellite Map.

After edges have been detected in both the top-down view and the satellite map, a fast normalized cross-correlation is run between the two images to find the proper match. Fast normalized cross-correlation is a method to perform template matching from transform domain convolutions, which yields a significant performance increase from spatial domain computations of normalized cross-correlation [15]. In this method, the correlation coefficients are calculated by

$$\gamma(u,v) = \frac{\sum_{x,y}[f(x,y)-\bar{f}_{u,v}]\,[t(x-u,y-v)-\bar{t}]}{\{\sum_{x,y}[f(x,y)-\bar{f}_{u,v}]^2\sum_{x,y}[t(x-u,y-v)-\bar{t}]^2\}^{0.5}} \quad (1)$$

where γ is the correlation coefficient, f is the image, $\bar{t}$ is the mean of the template, and $\bar{f}_{u,v}$ is the mean of f(x, y) in the region under the feature template.

Template matching algorithms such as normalized cross-correlation has drawbacks, such as computational intensity and sensitivity to rotation variance, but it was determined to be optimal for this effort due to the uniform colors and lack of rich features that prevalent in potential operational environments for the AGV. To avoid the rotation variance problem, compass bearing information from the AGV is used to determine the proper orientation of the warped top-down view with respect to the upward north oriented satellite images.

The area with the highest correlation coefficient is determined to be the best match and used carried through the process for further filtering.

Distance Filtering

As a means to reduce error, the distance of the current matched point from the last known good point is measured and used as a filter. This presumes that the starting point is known and can be accepted as truth to begin the filtering process. Any matches that are greater than 242 meters away from the last known good point are filtered out.

Kalman Filtering

A Kalman filter is applied to the results that fall within the maximum distance filtering described above. The output of the Kalman filter is the final matched value between the top-down view and the satellite map, giving the localization of the vehicle for one spherical panoramic image.

EXPERIMENT

To test the localization process, we leveraged Google Street View to represent an AGV collecting spherical panoramic images along a driven path. The spherical panoramic images were extracted using "Street View Download 360," a software application that interfaces with the Google Maps API and allows users to download the panoramic images at different resolutions [16]. The satellite images corresponding to the areas traveled were acquired from Google Maps. The area chosen for the test was a suburb near Sonoma Peak and the path traveled is shown in Figure 4. The satellite map shows an area of 559.02 meters by 111.75 meters. The vehicle traveled 965.6 meters and collected 98 spherical panoramic images. Each warped top-down image represents an area of 30.98 meters by 30.98 meters.

Figure 4. Sonoma Peak Experiment Area.

One limitation in using representative data from Google Maps and Google Street View is that the compass bearings of the vehicle are unknown. To account for this lack of information, optimal bearings were manually found by rotating the top-down view images in comparison to the ground truth of the satellite map and finding the bearing value that produced the best match. This list of bearings was coupled with the spherical panoramic images so that all the necessary information needed for the localization process was accounted for.

RESULTS

With the representative data acquired from Google Maps and Google Street View, the localization process was run with different settings to determine if the usage of Canny edge detection and distance filtering improved results. Figures 5-7 show the results of the different settings and the associated results. The green circles represent ground truth, the red squares represent the cross-correlation matches, and the blue diamonds represent the Kalman filtered results.

Figure 5. Distance Filtering Only.

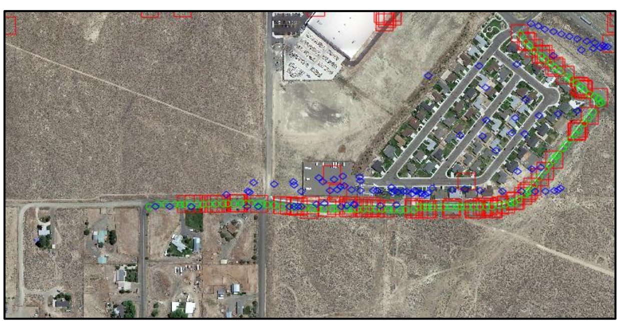

Figure 6. Canny Edge Detection Only.

Figure 7. Canny Edge Detection with Distance Filtering.

Figure 8 overlays a plot of the ground truth position and the localized position on the same figure for both the *X* and *Y* axis. Note that the *X* axis represents East/West, while the *Y* axis represents North/South. The *X* and *Y* position of the graph refers to pixel position on the satellite map. As seen in Figure 8 and Table 1, there was greater error in the North/South axis compared to the East/West axis.

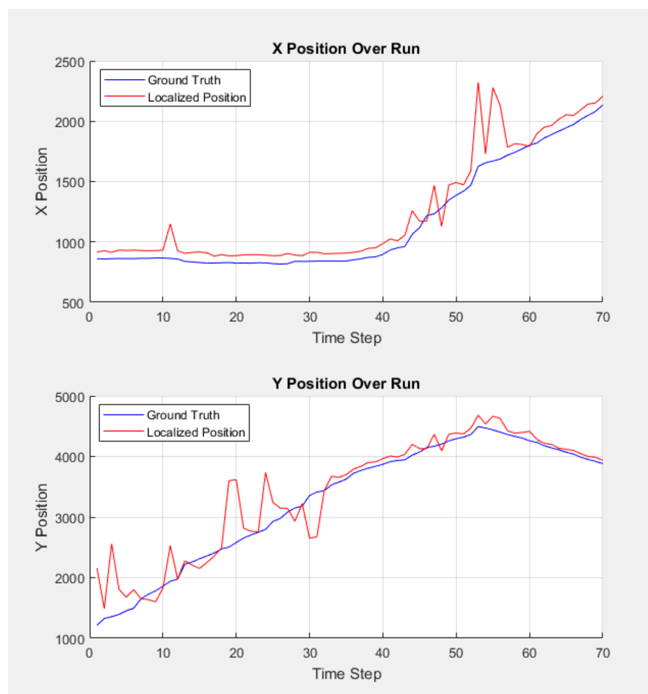

Figure 8. Ground Truth Position Compared to Localized Position.

Table 1 shows the root mean square error (RMSE) of each run from the cross-correlation (CC) matching, the RMSE of the Kalman Filtered (KF) results, and the KF result's mean distance from ground truth.

	RMSE for CC along East/ West (m)	RMSE for CC along North/ South (m)	RMSE for KF along East/ West (m)	RMSE for KF along North/ South (m)	KF Mean Distance from Truth (m)
Distance Filtering	72.32	480.33	71.97	398.49	253.78
Canny Edge Detection	97.72	194.70	36.33	136.21	115.65
Canny Edge Detection, Distance Filtering	36.47	81.53	20.69	52.91	46.95

Table 1. Localization Performance.

As seen in Figure 5 and Table 1, the performance when doing only distance filtering was the worst of all the options shown. Filtering based on distance without performing Canny edge detection removed 91 of the images, meaning that only 7% images could be localized. In addition to the low match rate, that setting resulted in the furthest mean distance from the ground truth, and significantly higher RMSE errors along the North/South axis of the satellite image. Performing the cross-correlation without filtering for distance, but keeping Canny edge detection, displayed improved accuracy for average mean distance to ground truth, despite the slightly higher RMSE along the East/West axis of the image. Combining the two for the full localization process produced the most accurate results in terms of lowest RMSEs and lowest distance from ground truth.

CONCLUSIONS

In this effort, we developed a method of GPS-denied localization by performing normalized cross-correlation on satellite imagery, with warped spherical panoramic images as the template. The ability to localize in this method would allow for greater utilization of AGVs in GPS-denied environments. By leveraging representative image and map data from Google Maps and Google Street View, we demonstrated that on average, the full localization process, including Canny edge detection, distance filtering, and Kalman filtering, can localize a vehicle within 50 m of ground truth. Overall, if there is GPS signal available, the localization process can potentially work with existing GPS solutions to help maintain an AGVs coordinates during intermittent GPS drops. However, the system will need further refinements in the matching and filtering process to rival the government GPS accuracy standard error of less than 7.8 m [17]. For complete GPS-denied environments however, the localization process can provide a great benefit for AGVs looking to approximate their position in the world to perform autonomous mobility.

REFERENCES

[1]Siegwart, R.; Nourbakhsh, Illiah. Introduction to Autonomous Mobile Robots. MIT Press, 2011.

[2]Nighswander, Tyler; Ledvina, Brent; Diamond, Jonathan; Brumley, Robert; Brumley David. "GPS Software Attacks," Proceedings of the 2012 ACM Conference on Computer and Communications Security - CCS 12, 2012, doi:10.1145/2382196.2382245.

[2]Cornick, Matthew; Koechling, J; Stanley, Byron; Zheng, Beijia. "Localizing Ground Penetrating RADAR: A Step Toward Robust Autonomous Ground Vehicle Localization," Journal of Field Robotics, vol. 33, no. 1, 2015, pp. 82–102., doi:10.1002/rob.21605.

[3]Majdik, András L.; Damiano, Verda; Albers-Schoenberg, Yves; Scaramuzza, Davide. "Air-Ground Matching: Appearance-Based GPS-Denied Urban Localization of Micro Aerial Vehicles," Journal of Field Robotics, vol. 32, no. 7, 2015, pp. 1015–1039., doi:10.1002/rob.21585.

[4]Brubaker, Marcus A.; Geiger, Andreas; Urtasun, Raquel. "Lost! Leveraging the Crowd for

Probabilistic Visual Self-Localization," 2013 IEEE Conference on Computer Vision and Pattern Recognition, 2013, doi:10.1109/cvpr.2013.393.

[5]Fuentes-Pacheco, Jorge; Ruiz-Ascencio, José; Rendón-Mancha, Juan Manuel. "Visual Simultaneous Localization and Mapping: a Survey," Artificial Intelligence Review, vol. 43, no. 1, 2012, pp. 55–81., doi:10.1007/s10462-012-9365-8.

[6]Lenac, Kruno; Kitanova,Andrej; Cupecb, Robert; Petrovića, Ivan. "Fast Planar Surface 3D SLAM Using LIDAR," Robotics and Autonomous Systems, vol. 92, 2017, pp. 197–220., doi:10.1016/j.robot.2017.03.013.

[7]Fallon, Maurice F.; Folkesson,John; McClelland, Hunter; Leonard, John. "Relocating Underwater Features Autonomously Using Sonar-Based SLAM," IEEE Journal of Oceanic Engineering, vol. 38, no. 3, 2013, pp. 500–513., doi:10.1109/joe.2012.2235664.

[8] Davison, Andrew J.; Reid, Ian; Molton, Nicholas; Stasse, Olivier. "MonoSLAM: Real-Time Single Camera SLAM," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 29, no. 6, 2007, pp. 1052–1067., doi:10.1109/tpami.2007.1049.

[9]Ma, Xiao; Djouadi , Seddik M.; Crilly, Paul; Smith, Stephen. "Navigation Strategy with the Aid of the Theater Positioning System," Intelligent Control and Automation, vol 4, 2013, pp.206-216. 10.4236/ica.2013.42025.

[10]Guo, Kexin; Qiu, Zhirong; Meng, Wei; Xie, Lihua; Teo, Rodney. "Ultra-Wideband Based Cooperative Relative Localization Algorithm and Experiments for Multiple Unmanned Aerial Vehicles in GPS Denied Environments," International Journal of Micro Air Vehicles, vol. 9, no. 3, 2017, pp. 169–186., doi:10.1177/1756829317695564.

[11] Viswanathan, Anirudh; Pires, Bernardo R.; Huber, Daniel. "Vision Based Robot Localization by Ground to Satellite Matching in GPS-Denied Situations," 2014 IEEE/RSJ International Conference on Intelligent Robots and Systems, 2014, doi:10.1109/iros.2014.6942560.

[12] Ehinger, Krista A.; Oliva, Aude; Torralba, Antonio. "Recognizing Scene Viewpoint using Panoramic Place Representation," Proceedings of 25th IEEE Conference on Computer Vision and Pattern Recognition, 2012.

[13] Canny, J., "A Computational Approach to Edge Detection," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 8, no. 6, 1986, pp. 679–698

[14] J. P. Lewis, "Fast Template Matching," Vision Interface, 1995, pp. 120-123, 1995.

[15]Orlita, Thomas. "Street View Download 360," Street View Panorama Download, 2.1.0, streetviewdownload.eu/.

[16]"GPS Accuracy," GPS.gov, National Coordination Office for Space-Based Positioning, Navigation, and Timing, 5 Dec. 2017, www.gps.gov/systems/gps/performance/accuracy/.