

Matching of Ground-Based LiDAR and Aerial Image Data for Mobile Robot Localization in Densely Forested Environments

Marwan Hussein, Matthew Renner, Masaaki Watanabe and Karl Iagnemma

Abstract— We present a vision based method for the autonomous geolocation of ground vehicles and unmanned mobile robots in forested environments. The method provides an estimate of the global horizontal position of a vehicle strictly based on finding a geometric match between a map of observed tree stems, scanned in 3D by sensors onboard the vehicle, to another stem map generated from the structure of tree crowns observed in overhead imagery of the forest canopy. This method can be used in real-time as a complement to the Global Positioning System (GPS) in areas where signal coverage is inadequate due to attenuation by the forest canopy, or due to intentional denied access. The method presented in this paper has two key properties that are significant: i) It does not require *a priori* knowledge of the area surrounding the robot. ii) It uses the geometry of detected tree stems as the only input to determine horizontal geolocation.

I. INTRODUCTION

Traditional geolocation of terrestrial vehicles has primarily utilized GPS in different modes to achieve accuracies in the decimeter range (e.g. Differential and Real-time Kinematic techniques) [1]. Despite advances in GPS accuracies and measurement methods, attenuation of GPS signals in dense forest environments renders the service unreliable for continuous and real-time localization purposes. Geolocation of rovers without the use of GPS focuses on dead reckoning techniques that use Inertial Measurement Units (IMU) [2]. Such approaches are prone to continuous drifts in measured position, which can be problematic for rovers operating in close proximity to points of interest.

Techniques such as Simultaneous Localization and Mapping (SLAM) have been successfully utilized to localize rovers in a variety of settings and scenarios [3,4]. SLAM focuses on building a local map of landmarks as observed by a rover and using it to estimate the rover's local position in an iterative way [3,4]. Position errors could be initially high but tend to converge as more landmarks are observed and as errors are filtered. Therefore, SLAM does not require *a priori* knowledge of the locations of landmarks or that of the rover. The method presented in this paper is different from SLAM in that it provides a non-iterative estimate of the global position of a rover by comparing landmarks, in this case tree stems observed on the ground, to tree stems detected in overhead georeferenced imagery of the forest canopy. The presented method is envisaged to be complementary in nature to traditional SLAM frameworks

- Marwan Hussein and Karl Iagnemma are with the Robotic Mobility Group, Department of Mechanical Engineering, Massachusetts Institute of Technology, USA. {marwanh/kdi}@mit.edu
- Matthew Renner is with the U.S. Army Engineer Research and Development Center, USA. matthew.l.renner@usace.army.mil
- Masaaki Watanabe is with the Robotics Group of IHI Corporation, Japan. masaaki_watanabe_2@ihi.co.jp

where it could be invoked on-demand by the rover as the need arises.

The envisioned operational scenario of the presented method is summarized as follows. A rover is initially assumed to be driving autonomously or via teleoperation in a forested area. The rover is anticipated to initially use GPS for localization. However, in situations where the GPS service becomes inaccessible or unreliable, the rover is expected to utilize the method discussed in this paper for the duration of the GPS blackout. In essence, the proposed method is envisaged to act as a backup whenever the primary localization service is down. Alternatively, and to enable real-time redundancy in geoposition estimation, the presented method could be used in tandem with the primary localization service.

The localization method presented in this paper requires two types of input data. The first are 3D scans of the environment surrounding the rover by onboard Light Detection and Ranging (LiDAR) sensors. At a minimum, the rover is expected to employ a LiDAR sensor with 360° horizontal Field-Of-View (FOV) to completely image the geometry of tree stems surrounding the rover. The second data input is an overhead high-resolution image of the exploration site. The overhead image can be acquired either by satellite or aerial means and needs to be orthorectified and georeferenced. The following is a summary of the sequence of operations required to geoposition a vehicle using the vision based localization method:

1. At a particular pose, a ground LiDAR scan is taken of the area surrounding the rover.
2. LiDAR data is processed to detect and to label tree stems. Tree stem center locations are subsequently estimated to create a map of the horizontal locations of tree stem centers relative to the rover.
3. The overhead georeferenced image of the forest canopy above the rover is processed in order to delineate individual tree crowns and estimate tree centers. A map is subsequently generated that contains the absolute horizontal locations of tree stem centers.
4. Maps of tree stem centers estimated from LiDAR and from overhead imagery are matched in this step. The closest match is selected and used to calculate the rover's horizontal geoposition.

The vision based localization algorithm is composed of three main components:

- **Tree Crown Identification, Delineation and Center Estimation from Overhead Imagery:**

Delineates tree crowns and estimates tree centers from the geometric profile of the delineated tree crowns.

- **Tree Stem Center Estimation from Rover LiDAR Data:**

Identifies tree stems and estimates their centers from their 3D profiles.

- **Matching of Tree Centers from Overhead Imagery and LiDAR Data:**

Estimates the rover's horizontal geoposition by matching tree center maps generated from LiDAR to those obtained from overhead imagery.

An important aspect in the above formulation of the algorithm is an assumption about the growth profile of trees. In particular, the tree center estimation algorithms for the LiDAR and overhead image assume that the center of the tree crown coincides or is in close proximity to the center of the stem. This implies that for the purposes of this paper, tree crowns are assumed to have an elliptical profile along the horizontal, while tree stems are assumed to have an upright vertical profile.

This paper is comprised of four main sections. Section II, III and IV discuss the different components of the presented localization algorithm, its properties and constraints. In Section V, test results based on real-world data obtained from an outdoor test campaign are presented and discussed. Concluding remarks are presented in Section VI.

II. TREE CROWN IDENTIFICATION, DELINEATION AND CENTER ESTIMATION FROM OVERHEAD IMAGERY

This component involves creating a map of the horizontal locations of tree stem centers using a high-resolution orthophoto of the forest canopy. The purpose of this map is to provide a basis for comparison to the map generated from the ground LiDAR data. Several automatic tree crown delineation algorithms have been previously developed. Wang et al utilize a multi-step approach where Near Infra-Red (NIR) images are first used to mark tree centers, followed by intensity based segmentation to extract individual tree crowns [5]. Although effective, this approach was not followed because we were interested in developing an algorithm that works with visible imagery only.

The main hypothesis that guided the development of the algorithm is the fact that the horizontal location of a tree center in an overhead image can be estimated from the geometric centroid of the delineated crown. Therefore, if a tree crown is detected and delineated, the location of a tree center can in principle be deduced with a horizontal accuracy limited by the pixel resolution of the image. Wang et al follow the same hypothesis but augment their algorithm with NIR intensity data of the tree canopy [5].

Figure 1 outlines the process of tree delineation and center estimation using a sample image. First, non-green pixels (i.e. non-vegetation) are filtered out using a simple auto threshold approach. Second, the resultant color image is converted to a grayscale image with 8-bit resolution. With treetops as the only visible feature in the resulting image, a copy binary

image is created in order to calculate the Euclidean distance transform for all objects in the image. The Euclidean distance transform d between two points (pixels), p and q , is defined as follows (in 2D):

$$d(p, q) = \sqrt{\sum_{i=1}^2 (q_i - p_i)^2} \quad (1)$$

The next step involves applying watershed segmentation to the 8-bit grayscale image with the distance transform as input. The purpose of the distance transform and watershed segmentation is to segment any cluster of treetops into their constituent trees. This step is necessary in the case of dense forests that have touching tree crowns.

Following the segmentation step, the resultant image is composed of delineated tree crowns. The image is subsequently analyzed to estimate centers of trees by calculating the Euclidean centroid of each delineated object. The final product is a map composed of the pixel coordinates (x, y) of the centroid of each detected crown.

A walkthrough of the algorithm using a sample overhead image is shown in Figure 1. The overhead image is of a pine forest located northeast of Lake Mize in Florida, USA. The image is an orthophoto with 0.3 m resolution (Source: USGS).

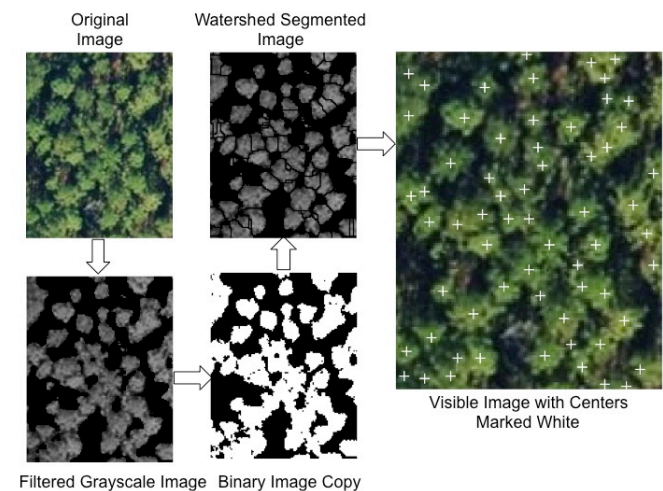


Figure 1: Example Showing Tree Center Estimation Using a Sample Overhead Image of a Pine Forest

Figure 1 shows that for a relatively dense cluster of trees, a visually correct result in terms of the locations of tree centers is obtained. Based on the above, the following are some general constraints:

- Accuracy of estimated tree centers depends on the view angle (perspective) of the image. Therefore for best results and to reduce the effect of parallax, the area of the forest of interest needs to be as close as possible to the nadir of the image.
- The average diameter of a tree crown is recommended to be at least an order of magnitude larger than the pixel resolution of the input image [5]. For example, the pixel resolution of an input image needs to be around 0.3 m in order to adequately detect 3 m crowns.

III. TREE STEM CENTER ESTIMATION FROM ROVER LiDAR DATA

As discussed in Section I, the ground vehicle is expected to acquire LiDAR data of the forest grounds for use by the tree stem center estimation algorithm. It is noted that the algorithm discussed in this section is based on prior work by McDaniel et al [6].

Based on input LiDAR data such as that shown in Figure 2, the ground plane is first estimated in order to constrain the search space. This is accomplished by tessellating the LiDAR point cloud into 0.5×0.5 m cells across the horizontal plane. Each cell is identified by a unique (row, column) index. Based on the cluster of points within each cell, the point with the lowest height is initially labeled as ground.

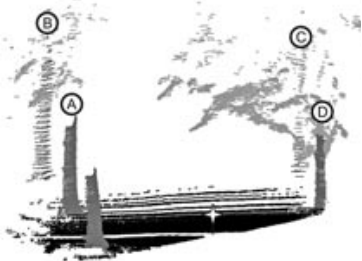


Figure 2: Sample Data Showing a LiDAR Scan of Trees [6]

Due to underbrush, not every point selected in the previous step represents true ground. Therefore, a classifier is utilized to filter data using metrics such as: the number of neighboring points, their geometry and ray tracing scores. Filtered data is then provided to another classifier that utilizes Support Vector Machine (SVM). If a cell contains a point labeled by the classifier, it is treated as true ground. Interpolation is performed for cells that do not have points labeled by the classifier.

The next step involves identifying the vertical extents of stems using thresholding and clustering methods. To identify which LiDAR point belongs to the main stem, a simple height-above-ground filter is used. Any points below the selected height threshold are discarded and classified as underbrush. Based on the dataset obtained, the team empirically determined that a height threshold of 2 m is adequate. Filtered data is then clustered following a single-linkage clustering method, which classifies two points to belong to a single cluster if they are in close proximity to each other (within a predefined distance threshold) [7].

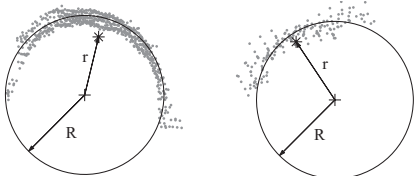


Figure 3: Example Showing Cylinder Fitment [6]

Finally, cylinders primitives are fit to the LiDAR data using a least squares scheme (Figure 3). R denotes the radius of the cylinder primitive while r denotes the estimated radius of curvature of the stem transect containing LiDAR data. The fit between the cylinder primitive and data is found by solving the following minimization problem where n is the

number of points in a cluster, x is the transect containing LiDAR points pertaining to the identified stem and f_i is the distance from the i th point to the surface of the cylinder primitive.

$$\arg \min_{x \in \mathbb{R}^3} = \frac{1}{2} \sum_{i=1}^n f_i^2(x) \quad (2)$$

Following the minimization step, tree stem centers are estimated based on the curvature of each fitted cylinder. Lastly, the location of each estimated tree center in the horizontal plane is incorporated into a tree center map that is used by the matching algorithm discussed in section IV.

IV. MATCHING OF TREE CENTERS FROM OVERHEAD IMAGERY AND LiDAR DATA

The last component of the algorithm involves matching tree center maps obtained from the LiDAR and overhead image in order to estimate the horizontal geoposition of the rover. Prior to invoking the matching process, both maps are converted to the WGS84 reference system. The datasets are then provided to an Iterative Closest Point (ICP) algorithm that estimates the horizontal translation and rotation vectors required to fit them together. More precisely, the tree center map derived from the overhead image is treated as the baseline upon which the LiDAR based tree center map is matched to. This is because the map obtained from the overhead image is usually larger than the local map generated from the LiDAR dataset.

The ICP implementation follows the Besl-McKay (point-to-point) framework given as follows [8,9]:

$$\arg \min E_{p,q \in \mathbb{R}^2} = \frac{1}{N} \sum_{i=1}^N \|\vec{R} \times \vec{p}_i + \vec{t} - \vec{q}_i\|^2 \quad (3)$$

Where E is the mean squared distance error between both maps. R and t are the rotation and translation vectors respectively. p and q are points in the tree center maps from the LiDAR and overhead image datasets respectively. N is the total number of points in the LiDAR tree center map. A match is found by determining the minimum of the error expression E . Figure 4 shows a sample run of ICP at a single rover pose using data acquired from the Lake Mize site.

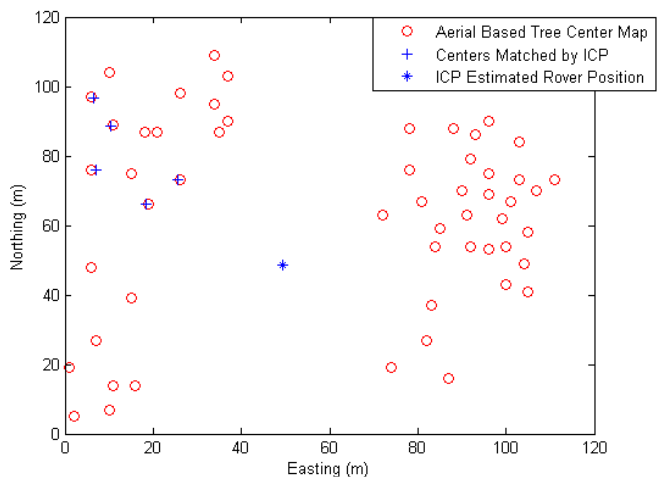


Figure 4: Sample ICP Run for a Single Rover Pose

In Figure 4, the circles denote stem centers generated from the overhead image. The plus signs denote stem centers

generated from a single LiDAR scan at a particular rover pose. The asterisk denotes the estimated rover position as a result from ICP. As seen in Figure 4, the closeness of the points in both datasets shows that a good match was found. In fact, the match was found with a reported average point-pair distance error of ~ 0.9 m (3 pixels). It is noted that in Figure 4, the axes represent easting and northing with respect to a local predefined reference point for that dataset.

The ICP algorithm implemented as part of this paper has the following properties:

- The search space is constrained to a 35x35 m box centered on the last known position of the rover and projected onto the baseline dataset (stem centers from overhead imagery). The optimum search space size was determined following an empirical investigation of the accuracy and processing-time performance of ICP using different search space sizes and shapes. In particular, a 35x35 m search space enabled us to maximize matching accuracy while minimizing processing time.
- The ICP algorithm always provides a result and produces the best-found match along with a mean error metric (parameter E in eq. 3). Good candidate matches produce low E while inadequate matches result in a large E . The error metric E depends on the final geometric configuration of both maps and the number of points in the LiDAR stem center map.

V. RESULTS

The vision based localization algorithm was completely developed in Matlab and was tested with real-world data. Survey-grade LiDAR data was collected at a test site located northeast of Lake Mize in Florida, USA (Coordinates: N29.738°, W82.216°). The data was acquired by a Leica ScanStation-2 LiDAR system that was provided by the University of Florida. The LiDAR was placed at multiple survey stations within the area bounded by the rectangular box in Figure 5. The area is approximately 110x110 m and the LiDAR dataset was collected with an average spatial resolution of approximately 5 cm. The area is exclusively comprised of pine trees with moderate to dense underbrush. It is noted that 561 tree stems were manually identified and labeled within the bounded area.

In total, 4 high-resolution orthorectified images of the test site were acquired from USGS. The images were provided in GeoTiff format and were all captured in the visible spectrum. All images were provided in the UTM coordinate system.

Without access to a real rover, and to simulate the envisaged operational scenario, a ground vehicle was simulated in Matlab traversing a 4-sided polygonal path and using the acquired LiDAR data as input. The simulated rover path is illustrated in red as shown in Figure 5. More specifically, the acquired LiDAR data from the Lake Mize site was gridded and incrementally fed at each pose to the vision based localization algorithm. Figure 6 shows the simulated rover path and all identified stem centers in the LiDAR dataset. It

is noted that the rover path is composed of 234 poses that are spaced at 0.5 m intervals.



Figure 5: Test Site Northeast of Lake Mize (Source: USGS)

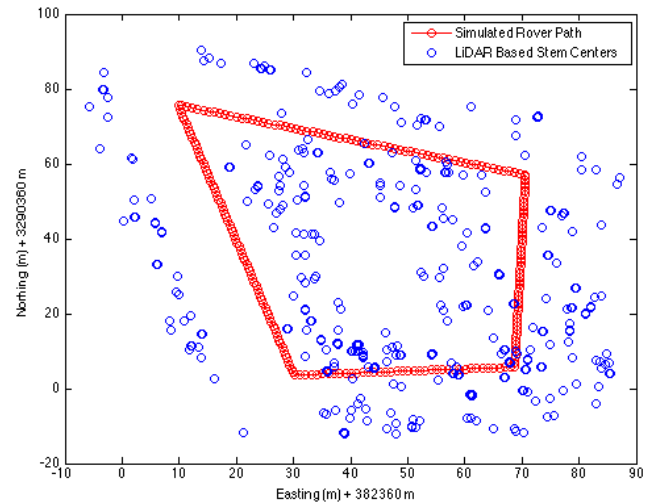


Figure 6: Simulated Rover Path and LiDAR Based Stem Centers

Table 1 summarizes the key properties of the acquired imagery and LiDAR data such as size, resolution and accuracy. It is noted that Aerial Image 1 and Aerial Image 2 were taken of the same area but with different sensors and at different dates.

Table 1: Properties of Input Data

Dataset Type	Properties	Horizontal Accuracy RMS
Aerial Image 1	Visible, 5000x5000 pixels, 0.3 m resolution	2.1 m
Aerial Image 2	Visible, 5000x5000 pixels, 0.3 m resolution	2.1 m
Aerial Image 3	Visible, 8000x8000 pixels, 0.5 m resolution	2 m
LiDAR	5 cm resolution (Average)	0.01 m

Based on the simulation setup, several tests were conducted using the same ground based LiDAR dataset but with 3

different overhead images as listed in Table 1. Since the LiDAR dataset is of high accuracy (1 cm RMS), it was treated as the ground truth to which geolocation estimates from the localization algorithm are compared against. More specifically, for each test run, the algorithm's estimated rover position is graphed with the corresponding GPS location, which is derived from the overlay of the simulated rover path onto the LiDAR data. To assess the accuracy of position estimates, the horizontal position error between the result provided by the localization algorithm and the GPS estimate is calculated. It is noted that due to the convergence properties of ICP, the matching algorithm requires an initial estimate of the location of the rover in order to avoid converging to an incorrect local minimum. Therefore, the GPS location of the first pose of the simulated rover is given to ICP as a seed upon which matching is initiated. For subsequent poses, the latest position estimates by ICP are used as seeds. In cases where large jumps are observed in position estimates provided by ICP, the latest estimate with the least deviation from the average is selected.

4 test runs in total were performed. 3 tests were fully automated (i.e. processing and labeling of input data were all conducted by the vision based algorithm in an end-to-end fashion). One test bypassed the tree crown delineation and center estimation algorithm and instead used a manually handpicked tree center map using Aerial Image 2. This test allowed us to decouple the vision based tree crown delineation and center estimation algorithm from the overall accuracy performance of the matching algorithm.

Figure 7 and Figure 8 show results from an automated test of the localization algorithm using Aerial Image 2. Figure 9 and Figure 10 show results from the manual test that also used Aerial Image 2. The mean position error and the standard deviation for all test runs are summarized in Table 2. Considering the automated test using Aerial Image 2, the mean rover position error is 4.76 m, which is about 14% of the linear dimension of the 35x35 search space box. In particular and as seen in Figure 8, the majority of rover poses (107 to be specific) resulted in less than 2 m positional error between GPS and the estimation provided by the matching algorithm. This is a good result considering that the spatial accuracy of Aerial Image 2 is 2.1 m RMS. It is noted that due to occlusion by underbrush, some tree stems were not detected at several poses. These poses were therefore discarded and shown as gaps in the rover path as seen in Figure 7 and Figure 9.

Three factors play a major role in determining the positional accuracy of the vision based algorithm: i) Accuracy of the input datasets (aerial and LiDAR). ii) Validity of tree stem centers estimated from aerial imagery, and iii) Validity of tree stem centers estimated from LiDAR. The second and third factors are very important in that they can easily affect the positional accuracy reported by the algorithm. If several tree centers were mislabeled, the matching algorithm would have difficulties finding the true matches. In cases where erroneous tree centers are present in a dataset, ICP would certainly provide a biased result. Online filtering techniques could be used to mitigate this issue. This is part of the scope of upcoming work on the algorithm.

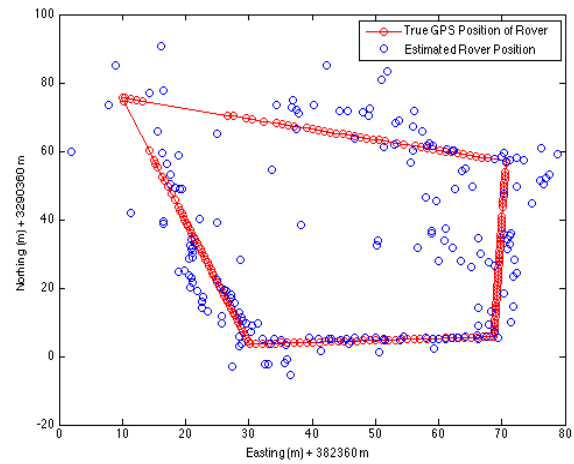


Figure 7: Result from Automated Test using Aerial Image 2

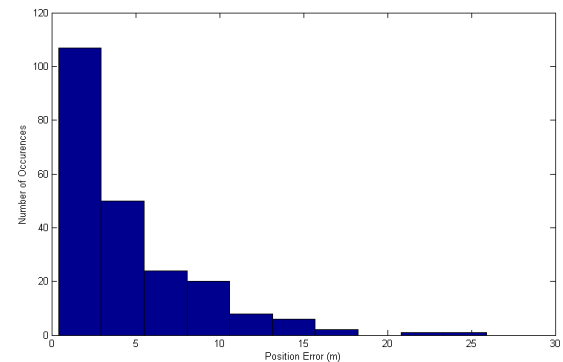


Figure 8: Histogram of Automated Test Using Aerial Image 2

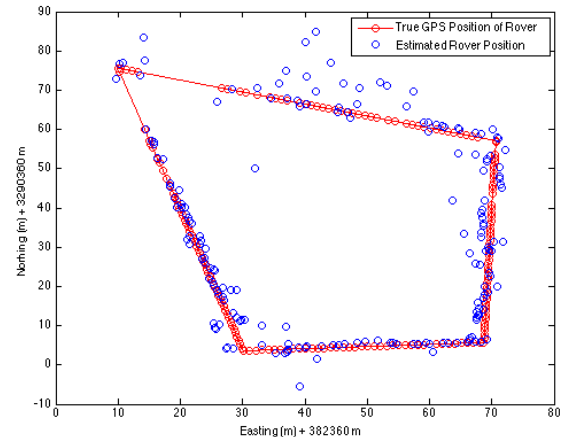


Figure 9: Result from Manual Test Using Aerial Image 2

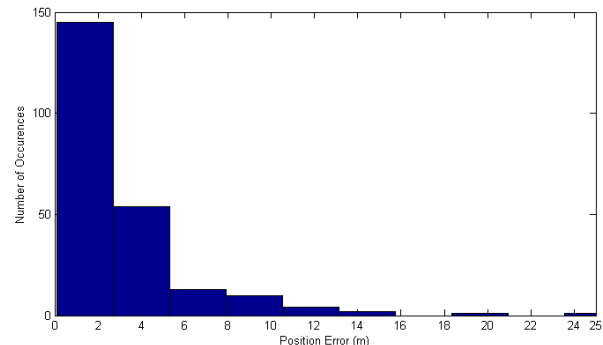


Figure 10: Histogram of Manual Test Using Aerial Image 2

To further understand the effect of mislabeled tree centers on the accuracy of geoposition estimates, a manual test was conducted as discussed previously for benchmarking purposes. Figure 9 and Figure 10 show the accuracy performance of the matching algorithm using a manually handpicked tree center map that was generated from Aerial Image 2. As seen in Figure 9 and Figure 10, geoposition error figures from the manual test show an improved performance compared to results obtained from the automated test. In fact, the mean position error reported by the manual test was reduced by almost 50% compared to the results from the automated test. Results from all tests are summarized in Table 2.

Table 2: Summary of Test Results

Aerial Image	Type of Test	Average Position Error (m)	Standard Deviation (m)
1	Automated	5.41	4.42
2	Automated	4.76	3.97
2	Manual	2.79	3.38
3	Automated	6.83	5.63

From Table 2, it is clear that as the aerial image resolution worsens, the mean position error increases. This is the case with Aerial Image 3. The increase in reported geoposition error could also be attributed to color contrast. Qualitatively speaking, a lower Red-Green-Blue (RGB) contrast in an aerial image could reduce the accuracy of tree crown delineation and center estimation. This has the potential to degrade the map matching performance of the algorithm and hence reduce the accuracy of position estimates.

The difference in performance observed between the automated and manual tests points out that the tree crown delineation and center estimation algorithm produces some mislabeled tree centers. Qualitatively speaking, this behavior can be attributed to multiple factors, mainly: i) Tree crowns tend to merge in dense forests rendering the task of delineation somewhat tricky and not without uncertainty. ii) Some trees that are visible in the LiDAR dataset may not have visible crowns in aerial imagery due to occlusions by taller trees. iii) Some trees that are visible in the aerial image may not be visible in the LiDAR dataset due to occlusions by other trees, underbrush or other obstructions. Therefore, operations in dense forests will usually result in some uncertainty in the estimation of geoposition. Nevertheless, considering situations when GPS is unavailable or unreliable, the benefit of being able to localize using the method of this paper is expected to outweigh its reduced accuracy performance.

VI. CONCLUSIONS AND FUTURE WORK

In conclusion, the localization algorithm presented in this paper has the following properties that make it significant: i) It enables rover position estimation by matching vision data from ground LiDAR and overhead visible imagery of the area of interest. ii) Does not require external georeferenced landmarks to tie data together or perform corrections. iii) Furnishes a positioning capability that is completely

decoupled from GPS to allow on demand localization in situations when GPS becomes unreliable or inaccessible.

The algorithm presented in this paper is considered a prototype that has constraints and limitations. In this phase of the project, the utility of the algorithm has been verified to provide reasonable horizontal position estimates using real-world data. Future improvements to the positioning accuracy of the algorithm are planned. These will involve improvements to the accuracies of the tree crown delineation and center estimation algorithm as well as to the LiDAR-based tree stem estimation algorithm. Aerial imagery with higher resolution and accuracy will be acquired to increase the accuracy of position estimates. In addition, online filtering techniques will be used to smooth position estimates and to discard anomalies. The end goal of the project is to develop the vision based localization algorithm to a level where it could reach accuracies that are comparable to current GPS standards.

ACKNOWLEDGMENT

The research presented in this paper is based upon work supported by the U.S. Army Engineer Research and Development Center (ERDC) under contract number W912HZ-12-C-0027.

REFERENCES

- [1] C. Kee, B.W. Parkinson., P. Axelrad, "Wide Area Differential GPS", *Navigation*, Vol. 38, No. 2, pp. 123-146, 1991
- [2] B. Liu, M. Adams, J. Ibanez-Guzman, "Multi-aided Inertial Navigation for Ground Vehicles in Outdoor Uneven Environments," IEE International Conference on Robotics and Automation ICRA, pp.4703-4708, 2005
- [3] J.J. Leonard, H.F. Durrant-Whyte, "Simultaneous map building and localization for an autonomous mobile robot," IEEE/RSJ International Workshop on Intelligent Robots and Systems - Intelligence for Mechanical Systems, vol. 3, pp.1442-1447, 1991
- [4] J. Guivant, F. Mason, E. Nebot, "Simultaneous Localization and Map Building Using Natural Features and Absolute Information", *Robotics and Automation Systems* 40, pp. 79-90, 2002
- [5] L.Wang, P. Gong, GS. Biging, "Individual Tree-crown Delineation and Treetop Detection in-Spatial Resolution Aerial Imagery", *Photogrammetric Engineering and Remote Sensing*, Vol. 70, No. 3, pp. 351-357, 2004
- [6] M. McDaniel, T. Nishihata, C. Brooks, P. Salesses, K. Iagnemma, "Terrain Classification and Identification of Tree Stems Using Ground-Based LIDAR", *Journal of Field Robotics*, Vol. 29, No. 6, pp. 891-910, 2012
- [7] M. McDaniel, T. Nishihata, C. Brooks, K. Iagnemma, "Ground Plane Identification Using LIDAR in Forested Environments", *IEEE International Conference on Robotics and Automation*, pp. 3831-3836, 2010
- [8] P. Besl, N. McKay, "A Method for Registration of 3D Shapes", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 14, No. 2, pp. 239-256, 1992
- [9] Y. Chen, G. Medioni, "Object Modeling by Registration of Multiple Range Images", *IEEE International Conference on Robotics and Automation*, Vol.3, pp.2724-2729, 1991