2016 IEEE International Conference on Robotics and Automation (ICRA)
Stockholm, Sweden, May 16-21, 2016

Vision-based Robot Localization Across Seasons and in Remote Locations

Anirudh Viswanathan, Bernardo R. Pires, and Daniel Huber[1]

***Abstract*— This paper studies the problem of GPS-denied unmanned ground vehicle (UGV) localization by matching ground images to a satellite map. We examine the realistic, but particularly challenging problem of navigation in remote areas using maps that may correspond to a different season of the year. The problem is difficult due to the limited UGV sensor horizon, the drastic shift in perspective between ground and aerial views, the absence of discriminative features in the environment due to the remote location, and the high variation in appearance of the satellite map caused by the change in seasons. We present an approach to image matching using semantic information that is invariant to seasonal change. This semantics-based matching is incorporated into a particle filter framework and successful localization of the ground vehicle is demonstrated for satellite maps captured in summer, spring, and winter.**

I. Introduction

Lower cost and complexity of autonomous systems has made it practical to field single or multiple robots, which operate cooperatively, to build an integrated world model or map [9] that is a layer-based world representation of the environment. Aerial maps, captured by satellite or high-flying unmanned aerial vehicles (UAV), are augmented by ground-based maps, which are build incrementally by UGV robots. The correspondence between the aerial map and ground-based map is obtained by precisely localizing the ground vehicle on the aerial map. This paper addresses the problem of localizing a UGV with respect to a satellite map by leveraging visual semantic matching.

The simplest solution to localize the ground-based vehicle to the satellite image is to use GPS. However, using GPS is not always feasible, since it may be jammed. GPS dropouts may occur in regions with dense vegetation, or in regions where GPS is blocked by buildings. This paper assumes that the robot operates in a GPS-denied environment and presents a vision-based localization technique that can operate in-place of, or as a complement to, GPS-based localization.

Vision-based localization between ground-based and aerial views is challenging for several reasons. The first challenge is associated with the UGV's limited sensor horizon. Depending on the sensor configuration and terrain, the UGV is able to only view up to a limited distance (typically a 10 meter radius). The limited visibility can be somewhat addressed by stitching together local UGV maps to form a larger integrated map. The second challenge is the drastic shift in perspective, e.g., while a UGV observes the branches of a tree at a specific location, a UAV observes the tree canopy at the same location. The viewpoint shift can be mitigated by warping the ground-based image to a bird's-eye view (top-down view) of the scene. The third challenge relates to the environment itself. The UGV operates in a remote outdoor environment, with little discernable man-made structure and few visually distinguishable features.

Our previous work [20] addressed these challenges and proposed a solution to matching aerial and ground-based views. This work used wide-angle or panoramic cameras to obtain a bird's-eye view of the scene, which is viewpoint-aligned with the satellite map. The localization problem was cast into an image-matching problem and an image descriptor-based method was used for feature-matching, similar to object-recognition algorithms. The paper successfully demonstrated localization of a ground vehicle in different outdoor environments, and the vision-based localization was able to provide a more accurate GPS estimate than the sensor used by Google Street View.

However, our approach in [20] failed to localize the UGV across seasons. The feature-matching framework in [20] implicitly assumed that ground-based images and the aerial maps were captured during the same time period, since the visual appearances between the UGV images and the satellite were compared directly as an input to the localization task. This paper eliminates that assumption and allows for drastic seasonal changes in the satellite map.

Designing a vision-based localization system robust to seasonal change in overhead data, encompasses identifying an image representation invariant to appearance change, in addition to overcoming all earlier challenges. The main contributions of this paper are: 1) an approach to image matching using semantic information, which is invariant to

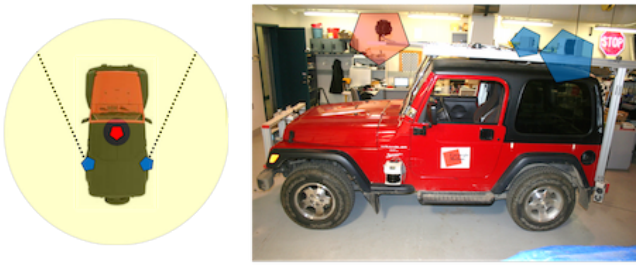

Fig. 1. The NavLab11 platform is used for data collection and localization experiments in this paper. The vehicle is equipped with two outward facing line-scanning lidars (indicated by blue polygon) and a panoramic camera (red polygon).

[1]All authors with the Department of Computer Science, Carnegie Mellon University, anirudh@cmu.edu, {bpires,dhuber}@cs.cmu.edu

Fig. 2. Satellite images of the Gascola region near Pittsburgh. The aerial images were taken during different seasons of the year – in spring, summer, and winter. This paper addresses the localization problem between ground and aerial views, when the aerial view are captured in different seasons. The teardrop shaped loop is path traversed by the ground robot. The images shown above are a subsection of a much larger (2 km x 2 km) map.

seasonal change, 2) a framework for Bayesian localization of a ground vehicle on a satellite map, and 3) a demonstration of the localization process using satellite maps captured in three different seasons.

II. Related Work

The problem of estimating the location of ground-based imagery on an aerial map has been investigated previously. Air-ground localization approaches predominantly fall into three categories: geo-location approaches, geometric approaches, and appearance-based approaches.

The geo-location approaches originated with the ICCV'05 competition titled "Where am I?" A collection of color images, with camera calibration parameters, was provided to the contestants. The images contained overlapping fields of view, and some even contained common objects. A subset of images was augmented with GPS location labels. The goal of the contest was to estimate the GPS locations of the unlabeled images. The winning entry [23] used an image-descriptor based approach to find potential matches, followed by a filtering stage using the fundamental matrix and homography. The work was followed-up by geo-location approaches on significantly larger scales, such as [8], [10], and [7]. However, these types of approaches are not applicable to a GPS-denied environment, since they require GPS labels for a portion of the data.

Geometric approaches were shown to be successful in cities. Typically, the edges associated with building façades are extracted, and, together with camera calibration information, are matched to an aerial view [3]. Alternative approaches use the self-similarity descriptor to evaluate potential matches between edges extracted from the boundary of the building matched to the aerial view [18], [19]. These approaches break down for remote outdoor locations, such as the ones considered in this paper, since there is not enough man-made structure to allow for reliable correspondence.

The approach proposed in this paper is most similar to [15], in which sidewalk edges are matched with the road map to localize the robot. However, several significant differences exist between the proposed approach and [15]. First, this paper considers remote, outdoor environment, with little man-made structure, such as sidewalks. Second, the proposed method does not require a prior "road map" to obtain the sidewalk locations, since no such map exists for remote locations. Other related research includes, a direct approach to learn a ground-satellite feature dictionary and associated projection matrices for the two views [6], and where-CNN, which explores the use of several features for cross-view geolocalization [11].

Appearance-based approaches have been demonstrated in [13] to localize images taken from a UAV with a front-facing camera to Google Street View images using a variant of affine-SIFT [21] to perform matching. Localization of ground-based views across seasons is investigated in [2] and [17] using whole-image descriptors [1]. In both approaches, however, the viewpoint remained relatively the same, and is not directly applicable to the air-ground localization problem.

III. Proposed Method

Figure 3 illustrates the proposed method. First, lidar data is classified into ground/non-ground categories. Next, ground-based panoramas are segmented into semantic regions by

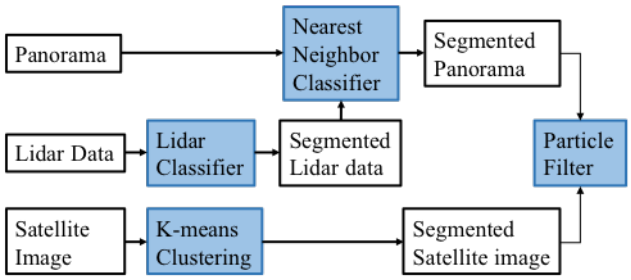

Fig. 3. Overview of the proposed semantics-based approach for localization across seasons. The minimal set of ground/non-ground semantics is used to localize the UGV on the satellite map.

using lidar data, and then warped to obtain a bird's-eye view. The satellite map is also segmented into ground/non-ground regions using k-means clustering. Regions on the satellite map with a similar spatial distribution of semantic categories as in the warped panorama, correspond to likely positions of the UGV on the map. Finally, an iterative estimation technique using a particle-filter framework is used to localize the UGV on the satellite map.

The localization experiments and data collection used in this paper were carried out at Gascola, a remote location, near Pittsburgh. The site was chosen because of its proximity to the CMU campus.

A. Image Matching Invariant to Seasonal Change

The satellite images of the Gascola region were captured during summer, spring, and the end of winter (Figure 2). The three satellite images exhibit drastically different visual appearance, owing to the presence of perennial vegetation. The satellite images in summer and spring show the presence of green cover. The satellite image captured at the end of winter is visually the most dissimilar and has completely lost all green cover. The only vegetation on the image comprises regions attached to houses, which are presumably well-maintained lawns. The central challenge addressed in this paper is the localization of the UGV across maps captured in different seasons, and the identification of a minimal semantic representation robust to seasonal changes.

In order to solve the localization problem, one needs to identify a representation of images that is robust to seasonal changes. To solve this problem, the images are represented using semantic categories rather than appearance (as encoded by, e.g., the SIFT vector). The key idea is that even though the appearance of the map might change, the underlying semantics remain the same.

The next question to be addressed relates to the number of semantic categories used to represent the image. The semantic categories can be fine-grained (e.g., individual categories for grass, shrubs, trees, dirt-road, etc.) or coarse-grained (e.g., ground/non-ground). Associating fine-grained semantic information with image data involves higher computation cost as the labeling problem becomes exponential. The more practical limitation of using find grained techniques involves the choice of assigning labels. For example, there is no clear distinction between a large shrub and a small tree. Instead, this paper elects to use coarse-grained semantics. The semantic classes consist of ground regions (or dominantly flat terrain) and non-ground regions (comprising grass, shrubs, trees, etc.). The deliberate choice of using a minimal set of semantics is also motivated by the fact that the traversable regions on the map remain unchanged during different seasons. The localization task is simplified to matching semantic categories between ground and aerial views.

The localization framework can be broadly divided into the following modules:

- Assignment of semantic categories to pixels in both UGV and satellite images
- Matching ground to satellite images using semantics
- Using a particle filter to localize the UGV

B. Data Collection

The ground-based imagery for the Gascola dataset was captured during the end of summer. The perennial vegetation at Gascola was in bloom during the period of data collection. Figures 5 (top) and 6 (top) show sample panoramic imagery at Gascola.

We used the NavLab11 platform shown in Figure 1 for the ground-based data collection. The NavLab11 is a Jeep Wrangler equipped with a wide variety of sensors for short and mid-range obstacle detection. The on-board sensor suite includes:

- Differential GPS with L-band satellite differential
- Odometer and compass (orientation estimation)
- Gyroscopes – Solid-state vertical gyro, roll and pitch angle measurement in dynamic environments
- Two outward-facing proximity lidar scanners (SICK LMS 221)
- Panoramic camera (LadyBug5)

The data used by the localization algorithm consists of only panoramic imagery, odometer, compass, and lidar information. GPS data is recorded for establishing ground truth, and is not otherwise used by the localization system.

C. Terminology

The images used in the paper are represented in the L-a-b color space. The convention followed in this paper to represent a transformation from a base frame to a target frame is represented by ${}^{Target}_{Base}T$. The terminology used in this paper is as follows:

- S – high flying UAV or satellite image (dimensions m x n), corresponding to 2 km x 2 km region
- I – wide field-of-view or panoramic image from the UGV (dimensions p x q), with leftmost column oriented North using compass information
- B – bird's-eye view of the ground (dimensions r x r)
- L – lidar coordinate frame
- C – LadyBug5 camera coordinate frame
- ${}^{C}_{L}T$ – transformation of lidar to camera coordinate frame
- ${}_{F}l^{t}$ – lidar return from SICK LMS 221 at timestamp t in coordinate frame $F = \{C, L\}$

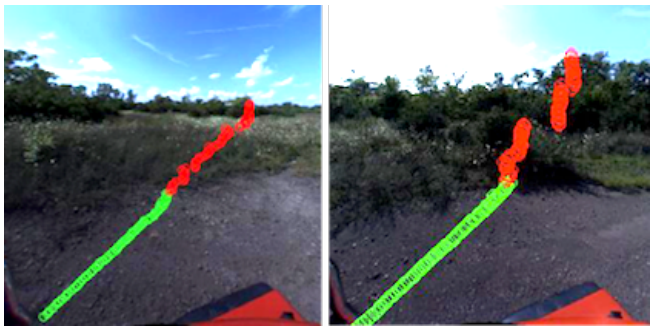

Fig. 4. Classification of lidar points as ground (green) and non-ground (red) using a linear regression model. The lidar data is projected on the undistorted image to help visualize the semantic labeling results.

- $h({}_F l^t)$ – classifier for labeling lidar data into semantic categories
- ${}_F l^t_{semantic}$ – labeled lidar into semantic categories $semantic = \{ground/non-ground\}$
- (u, v) – index of pixel in image
- $\chi^t = \{S(u_1, v_1), \ldots, S(u_n, v_n)\}$ – training pixel a-b color components at locations of labeled lidar points
- K – scalar representing past K observations (seconds)
- μ_{ab} – mean a-b values incorporating history $\chi^{(t-K):t}$
- $I_{semantic}$ – panorama with semantic labels
- B_e – Binary image obtained from edge-detection on $T_{semantic}$
- S_e – Binary image obtained from edge-detection on S
- D_T – Distance transform applied to S_e

D. Classification of lidar data

The lidar data expressed in the camera coordinate frame is classified into semantic categories of ground/non-ground. The assumption used when classifying the lidar data, is that the immediate region (within 1 m) surrounding the UGV is planar. The assumption allows us to learn the parameters of a linear regression function on a subset of n lidar data points $\{(x_1, y_1), \ldots, (x_n, y_n)\} \subset {}_C l^t$ where $\sqrt{x_i^2 + y_i^2} < 1$ m of the vehicle. The best-fit regression function is a straight line $ax + by + c = 0$ and the parameters are obtained using the least-squares solution.

Subsequently, the distance from the best-fit line $d_i = \frac{|ax_0 + by_0 + c|}{\sqrt{a^2 + b^2}}$ is computed for all lidar points $(x_i, y_i) \in {}_C l^t$. The classifier $h : {}_C l^t \rightarrow {}_C l^t_{semantic}$ assigns lidar points that fit the learned model with $d_i < 20$ cm (a measure of the planarity of the ground) to the "ground" semantic category, or otherwise to the "non-ground" semantic category. Figure 4 shows the classification of lidar points.

E. Assignment of semantic categories to pixels in panorama

The panorama obtained from the LadyBug5 camera consists of six individual images stitched together [14]. The individual (undistorted) images from each of the six cameras can also be directly accessed. Extrinsic calibration of the SICK lidar to an individual camera of the LadyBug5 was carried out prior to the localization experiment [5], [22]. The transformation ${}^C_L T$ allows us to obtain the correspondence between lidar points and pixels in the undistorted image.

At each labeled lidar point ${}_C l_{semantic}$, the a-b color components of the associated pixel are added as training data to χ^t. In order to incorporate a history of lidar classification into training data, we consider the set of past K observations of $\chi^{(t-K):t}$. In practice, we incorporate a history of 10 seconds into training data.

The mean a-b color components μ_{ab} of $\chi^{(t-K):t}$ are computed to form representative models of the ground/non-ground regions in the undistorted image. Subsequently, we label all the pixels in the panorama with a semantic label by using the nearest neighbor to μ_{ab} for each category. The segmented panorama $I_{semantic}$ is shown in Figure 5.

F. Matching ground to satellite using semantics

The segmented panorama is warped to obtain the bird's-eye view of the scene B. The orientation of the vehicle is determined using compass information, which allows us to correctly match the orientation of the warped image to the satellite map. Subsequently, edge-detection is performed to obtain B_e, which encodes ground-non/ground semantics.

The satellite image S is segmented into clusters using k-means clustering. The cluster centers are initially set to μ_{ab}, which are the learned models of ground/non-ground regions at the start of the localization run. The resulting binary satellite map, after the cluster association step, encodes the ground/non-ground semantics. Edge detection is performed to obtain S_e.

Edges, such as those between a dirt-road and grass, implicitly encode a change in semantic category. Image semantics can be compared using a measure of edge-overlap between

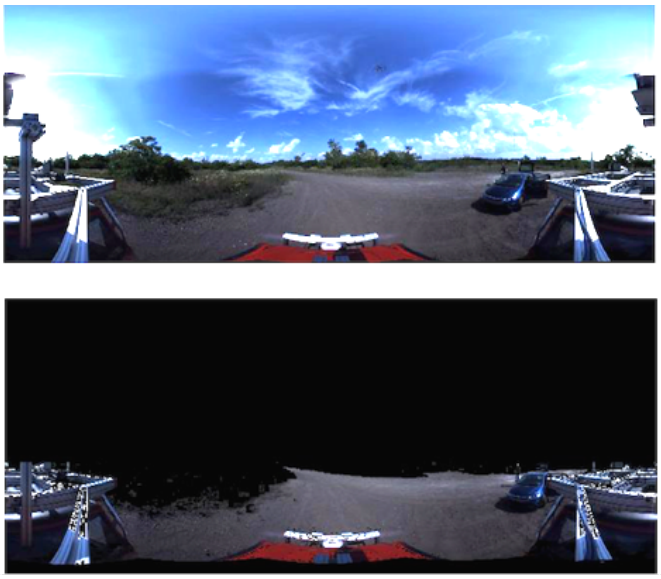

Fig. 5. The top image shows a panorama taken using the LadyBug5 camera at the Gascola region. The panorama is segmented by learning semantic models for the ground/non-ground regions in corresponding lidar data (after masking out the vehicle). The bottom image shows the segment corresponding to the ground region.

images. An alternate formulation to measure semantic similarity would be to compute the overlap area between the same semantic categories. We present both edge-based and area-based semantic localization in this section.

The localization problem consists of determining locations on the satellite map S_e with a similar distribution of edges as in the ground image B_e. Edges are compared using the Chamfer distance [4]. The Chamfer distance between sets of points in template $U = \{u_i\}$ and query $V = \{v_j\}$ is the average distance between n points $u_i \in U$ and the nearest edge V:

$$d_{CM} = \frac{1}{n} \sum_{u_i \in U} \min_{v_j \in V} ||u_i - v_j||$$

d_{CM} can be efficiently be computed using the distance transform D_T. During implementation, a faster variant of traditional Chamfer Matching described in [12] was used. A matching score is associated with every location on the satellite map, and represents a distribution of likely locations of the ground vehicle. The distribution is used in a standard particle filter to localize the UGV.

Semantic matching can also be computed using the overlap area within the same category. A matching score is obtained by computing the correlation between the binary satellite image and binary warped ground image and is used in the particle filter stage. The lidar-based segmentation of the warped panorama, and binary map associated with the ground region on the satellite after k-means clustering are shown in Figure 6.

G. Using standard particle filter to localize the UGV

A particle filter [16] is used to demonstrate localization between ground-based and aerial views. The details of the particle filter are found in our earlier paper [20], and the algorithm is summarized below for convenience. The particle filter is able to successfully localize the ground vehicle across maps in different seasons. The notation used is as follows:

- t – time index
- $\mathbf{x}^t$ – vehicle state (x,y) position on satellite map
- w_j – set of weights indexed by j
- u – vehicle odometry information
- z – sensor observation
- P – set of N particles, randomly distributed over the map at start of localization run

IV. Results

We compared the semantics-based method proposed in this paper to our previous feature-based approach in [20]. A smaller portion of the Gascola map, as outlined by the yellow box in Figure 8, was used for the comparison experiment. The small map covers an area of 0.4 km x 0.6 km and is used to initially test the efficacy of the proposed approach. The results in Table I indicate that the feature-based baseline approach is able to successfully localize the UGV during the spring and summer season maps. However, the particle filter does not converge for the winter season (indicated by N/A in Table I).

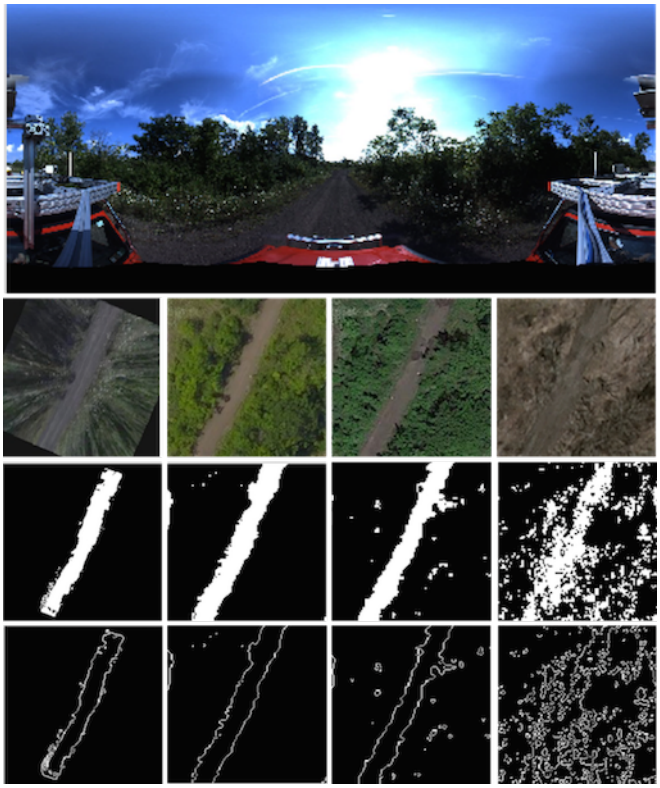

Fig. 6. The first column corresponds to the warped panorama (top). Remaining columns (left to right) correspond to ground truth for the spring, summer, and winter maps. The binary images for the satellite are obtained via kMeans clustering. The area-based or edge-based representation encodes the ground/non-ground semantics.

Algorithm 1 Particle Filter Step for $t > 0$

Require: N particles at $t-1$: $P^{t-1} = [\mathbf{x}_j^{t-1}, w_j]_{j=1...N}$
  for $j = 1$ to N **do**
    sample $\mathbf{x}_j^t \sim p(\mathbf{x}^t | u_t, \mathbf{x}_j^{t-1})$
    compute $w_j = p(z_t | \mathbf{x}_j^t)$
  end for
  normalize weights w
  for $j = 1$ to N **do**
    draw i with probability distribution w
    add $[\mathbf{x}_i^t, w_i]$ to P^t
  end for
  return P^t

TABLE I
Comparing baseline approach [20] to the semantics-based approach on 0.4 km x 0.6 km maps.

Average localization error (m)			
Season	Baseline	Edge-based semantics	Area-based semantics
Spring	4.5	3.4	3.0
Summer	4.3	3.6	3.3
Winter	N/A	5.1	4.4

TABLE II

AVERAGE ERROR USING EDGE AND AREA-BASED SEMANTIC LOCALIZATION ON 2 KM X 2 KM MAPS

Season	Edge-based semantics Avg. localization error (m)	Area-based semantics Avg. localization error (m)
Spring	4.7	3.2
Summer	4.8	3.5
Winter	5.4	4.8

Localization using ground/non-ground semantics is successful for the smaller map. The semantics-based approach converges even for the case of the winter map, and demonstrates localization across seasons. Subsequently, we experimented with an increased map size of 2 km x 2 km to test the effectiveness of the semantics-based approach. In practical situations, it is reasonable to expect that the UGV location is known to within 2 km even if GPS is unavailable. Since we are able to localize on a map of this scale, it is plausible that the system will be able to localize correctly on maps of larger scales. We are currently exploring the scalability of our approach to substantially larger regions.

Figure 8 shows the particle filter converging to the ground truth. The accompanying video file shows the complete localization run in all three seasons. Both edge-based and area-based approaches to localization using semantic information are successful in all three seasons. The results show that the minimal choice of ground/non-ground semantic features is robust to seasonal variation of the satellite map.

The area-based approach achieves lower error compared to the edge-based approach (see Table II) once the particle filter has converged. A likely explanation is that edge detection performed in the satellite image is a noisy representation of ground/non-ground semantics. Using alternate segmentation methods on the satellite, other than simple edge detection, will possibly provide better semantic edges, and implicitly better localization.

The localization framework was also evaluated using robot kidnapping. The UGV in operation was carried to an arbitrary location on the map to test for recovery from localization failure. In one case, for the winter map, the particle filter failed to converge during the lower section of the map in Figure 7 (see video for specific experiment). The section of the trail consists of the path splitting into two parts of similar shape; consequently, the particle filter is equally likely to choose one over the other. The path closer to the split, is initially associated with a higher weight, and is incorrectly estimated to be the UGV location. In all other cases, our algorithm is successfully able to localize ground-based imagery on the aerial map.

V. CONCLUSION

This paper describes a method for localizing ground-based images to aerial images captured in different seasons using semantic information. A minimal set of ground/non-ground semantics is identified that is able to successfully localize the ground vehicle. Even though maps taken in different seasons exhibit significant visual differences, the underlying semantics remain unaltered. The results indicate that the area-based approach to comparing the distribution of semantic categories provides lower average error compared to the edge-based approach. A particle filter is used to successfully localize the UGV on maps captured in summer, spring, and winter. Future work includes an extension to combine both semantic and appearance information to match ground and aerial images.

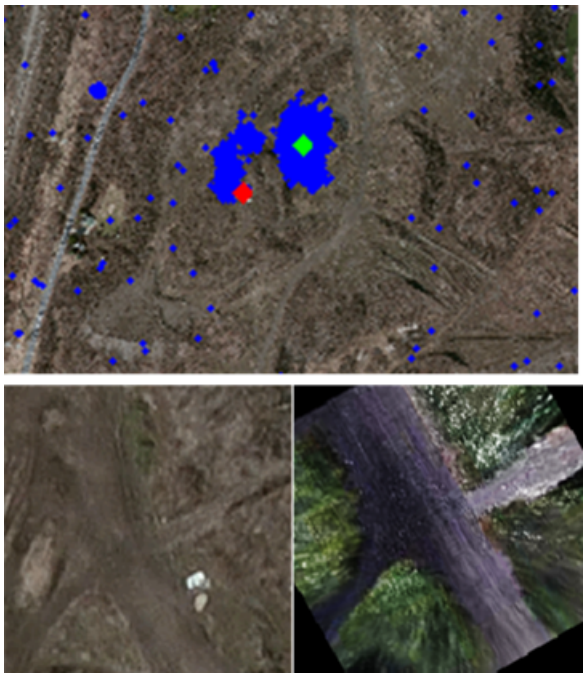

Fig. 7. Incorrect localization of the UGV on the winter map. The map at ground truth location (red dot) is shown on the lower-left. The warped panorama at the estimated location (green dot) is shown on the lower-right. The incorrect localization is due to a similar distribution of semantic edges at both locations.

ACKNOWLEDGMENT

This work was supported by the Agency for Defense Development, Jochiwongil 462, Yuseong, Daijeon, Korea. Work of Bernardo Pires also supported by Office of Naval Research Grant N00014-12-1-0903.

REFERENCES

[1] M. Agrawal, K. Konolige, and M. Blas, "CenSurE: Center surround extremas for realtime feature detection and matching," in *Computer Vision ECCV 2008*, ser. Lecture Notes in Computer Science, D. Forsyth, P. Torr, and A. Zisserman, Eds. Springer Berlin Heidelberg, 2008, vol. 5305, pp. 102–115.

[2] H. Badino, D. Huber, and T. Kanade, "Real-time topometric localization," in *International Conference on Robotics and Automation*, May 2012.

[3] M. Bansal, H. S. Sawhney, H. Cheng, and K. Daniilidis, "Geo-localization of street views with aerial image databases," in *ACM Multimedia*, 2011, pp. 1125–1128.

[4] H. G. Barrow, J. M. Tenenbaum, R. C. Bolles, and H. C. Wolf, "Parametric correspondence and chamfer matching: Two new techniques for image matching," DTIC Document, Tech. Rep., 1977.

[5] J. Y. Bouguet, "Camera calibration toolbox for Matlab," 2008. [Online]. Available: http://www.vision.caltech.edu/bouguetj/calib_doc/

[6] H. Chu, H. Mei, M. Bansal, and M. R. Walter, "Accurate vision-based vehicle localization using satellite imagery," *arXiv preprint arXiv:1510.09171*, 2015.

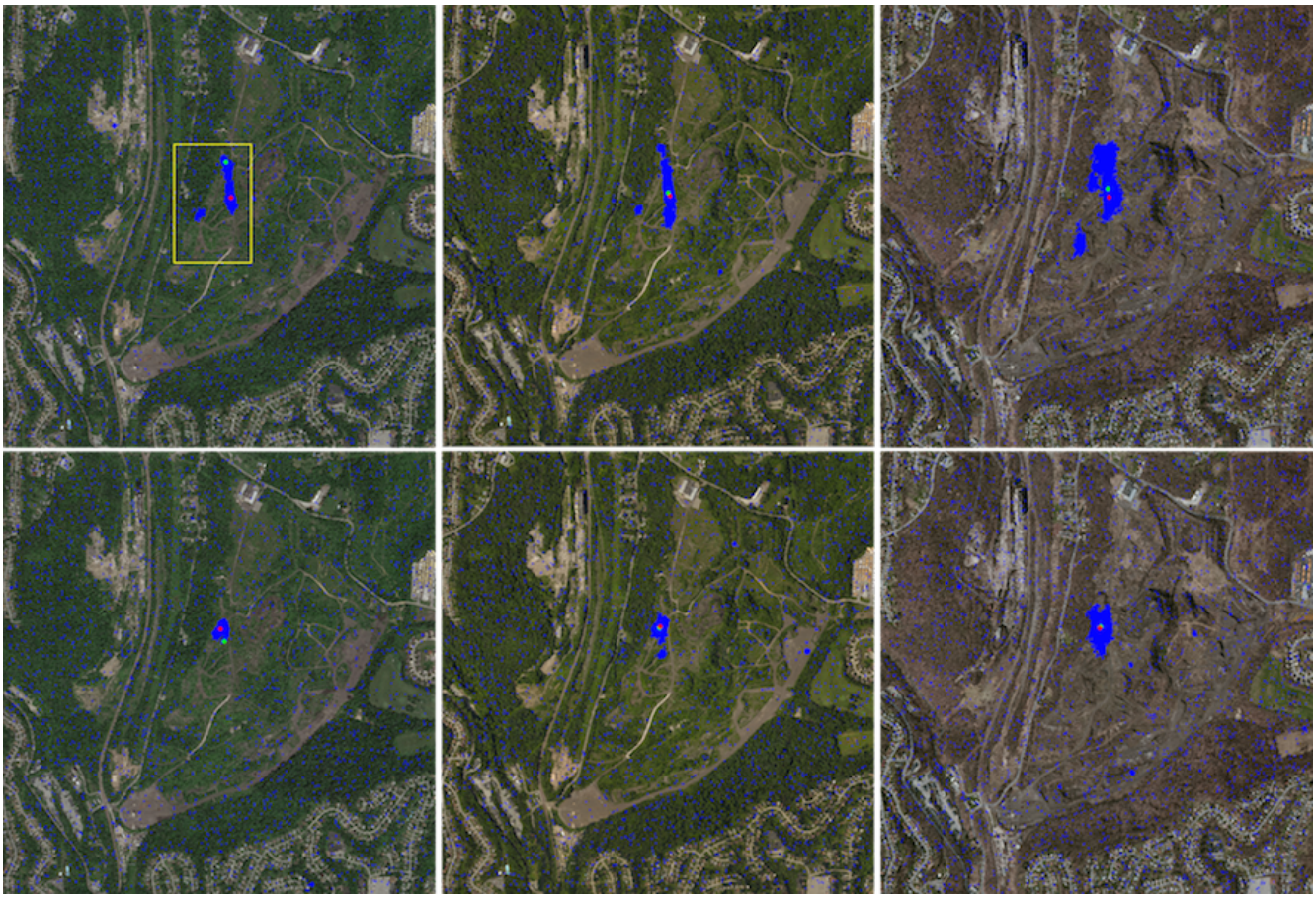

Fig. 8. Localization using a particle filter is successful for 2 km x 2 km maps captured in different seasons. The top row shows the particle filter before convergence for the same vehicle location. The bottom row shows the filter estimate (green dot) converging to the ground truth location (red dot). The yellow box encloses the area shown in Figure 2. Results from both area-based, and edge-based localization are shown in the accompanying video file.

[7] D. J. Crandall, L. Backstrom, D. Huttenlocher, and J. Kleinberg, "Mapping the world's photos," in *Proceedings of the 18th International Conference on World Wide Web*, ser. WWW '09. New York, NY, USA: ACM, 2009, pp. 761–770.

[8] J. Hays and A. A. Efros, "im2gps: estimating geographic information from a single image," in *Proceedings of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2008.

[9] S. Lacroix and G. Besnerais, "Issues in cooperative air/ground robotic systems," in *Robotics Research*, ser. Springer Tracts in Advanced Robotics, M. Kaneko and Y. Nakamura, Eds. Springer Berlin Heidelberg, 2011, vol. 66, pp. 421–432.

[10] T.-Y. Lin, S. Belongie, and J. Hays, "Cross-view image geolocalization," in *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, June 2013, pp. 891–898.

[11] T.-Y. Lin, Y. Cui, S. Belongie, and J. Hays, "Learning deep representations for ground-to-aerial geolocalization," in *Computer Vision and Pattern Recognition (CVPR), 2015 IEEE Conference on*, June 2015, pp. 5007–5015.

[12] M.-Y. Liu, O. Tuzel, A. Veeraraghavan, and R. Chellappa, "Fast directional chamfer matching," in *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, June 2010, pp. 1696–1703.

[13] A. Majdik, Y. Albers-Schoenberg, and D. Scaramuzza, "MAV urban localization from google street view data," in *IROS*, 2013, pp. 3979–3986.

[14] Point Grey Research, "Ladybug5 USB 3.0 Spherical Camera Technical Reference," http://www.ptgrey.com/support/downloads/10128/, 2014.

[15] T. Senlet and A. Elgammal, "Satellite image based precise robot localization on sidewalks," in *Robotics and Automation (ICRA), 2012 IEEE International Conference on*, May 2012, pp. 2647–2653.

[16] S. Thrun, "Probabilistic robotics," *Communications of the ACM*, vol. 45, no. 3, pp. 52–57, 2002.

[17] C. Valgren and A. J. Lilienthal, "SIFT, SURF & seasons: Appearance-based long-term localization in outdoor environments," *Robotics and Autonomous Systems*, vol. 58, no. 2, pp. 149–156, 2010.

[18] T. Vidal, C. Berger, J. Sola, and S. Lacroix, "Environment modeling for cooperative aerial/ground robotic systems," in *Robotics Research*, ser. Springer Tracts in Advanced Robotics, C. Pradalier, R. Siegwart, and G. Hirzinger, Eds. Springer Berlin Heidelberg, 2011, vol. 70, pp. 681–696.

[19] T. A. Vidal-Calleja, C. Berger, J. Sol, and S. Lacroix, "Large scale multiple robot visual mapping with heterogeneous landmarks in semi-structured terrain," *Robotics and Autonomous Systems*, vol. 59, no. 9, pp. 654 – 674, 2011.

[20] A. Viswanathan, B. R. Pires, and D. Huber, "Vision based robot localization by ground to satellite matching in GPS-denied situations," in *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems, Chicago, IL, USA, September 14-18, 2014*, 2014, pp. 192–198. [Online]. Available: http://dx.doi.org/10.1109/IROS.2014.6942560

[21] G. Yu and J.-M. Morel, "ASIFT: An Algorithm for Fully Affine Invariant Comparison," *Image Processing On Line*, vol. 2011, 2011.

[22] Q. Zhang and R. Pless, "Extrinsic calibration of a camera and laser range finder (improves camera calibration)," in *Intelligent Robots and Systems, 2004. (IROS 2004). Proceedings. 2004 IEEE/RSJ International Conference on*, vol. 3, Sept 2004, pp. 2301–2306 vol.3.

[23] W. Zhang and J. Kosecka, "Image based localization in urban environments," in *Proceedings of the Third International Symposium on 3D Data Processing, Visualization, and Transmission (3DPVT'06)*, ser. 3DPVT '06. Washington, DC, USA: IEEE Computer Society, 2006, pp. 33–40. [Online]. Available: http://dx.doi.org/10.1109/3DPVT.2006.80