Vision Based Robot Localization by Ground to Satellite Matching in GPS-denied Situations

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

***Abstract*— This paper studies the problem of matching images captured from an unmanned ground vehicle (UGV) to those from a satellite or high-flying vehicle. We focus on situations where the UGV navigates in remote areas with few man-made structures. This is a difficult problem due to the drastic change in perspective between the ground and aerial imagery and the lack of environmental features for image comparison. We do not rely on GPS, which may be jammed or uncertain. We propose a two-step approach: (1) the UGV images are warped to obtain a bird's eye view of the ground, and (2) this view is compared to a grid of satellite locations using whole-image descriptors. We analyze the performance of a variety of descriptors for different satellite map sizes and various terrain and environment types. We incorporate the air-ground matching into a particle-filter framework for localization using the best-performing descriptor. The results show that vision-based UGV localization from satellite maps is not only possible, but often provides better position estimates than GPS estimates, enabling us to improve the location estimates of Google Street View.**

I. Introduction

Robotic systems deployed in the field rely on effective mapping and localization methods to achieve their goals. Aerial maps, taken from satellites or high-altitude air vehicles, can be supplemented by up-to-date, ground-based maps constructed by single or multiple robots operating cooperatively to create an integrated world model [10]. In order to create such an integrated model, the ground-based maps must be precisely localized with respect to the aerial maps. This paper addresses this problem of visual matching and localization of an unmanned ground vehicle (UGV) with respect to a satellite map.

The obvious solution – use GPS to align the maps – is not feasible in many practical situations. The GPS signal may be jammed, it may be blocked by vegetation or buildings, or it may simply not be accurate enough for a given application. We elect, instead, to focus on visual matching, which can be used as an alternative to GPS-based localization, or as a supplement to GPS estimates when they are available. Our approach uses wide-angle or panoramic camera images from the UGV, which are warped to obtain a bird's-eye view of the nearby ground (Figure 1). The warped ground images are then matched to the satellite map using descriptor-based feature-matching techniques common to object-recognition algorithms.

Matching UGV images to satellite maps is challenging for several reasons. First, the ground-based images often appear significantly different from the aerial maps (Figure 1). The aerial map typically taken at a different time of day, in a different season, with different lighting conditions, and at a vastly different resolution from the ground-based imagery. Furthermore, any objects in the environment that violate the implicit flat ground assumption in the warping process will not match well between the maps. For example, a UGV will observe the ground under a tree or the side of a building, whereas a satellite view will observe the tree canopy or the building's roof. While our current approach does not explicitly address these non-ground artifacts, we find that the method works well even in situations that violate the flat ground assumption. A third challenge is the limited viewing horizon of a ground vehicle. Depending on the sensor placement and terrain shape, a relatively small area around the vehicle may be visible from a single position (typically about 10 m radius). This limited-size ground map can be addressed by integrating individual images into larger maps. A final challenge is that we focus on wilderness or off-road situations that contain few or no man-made structures and minimal unique distinguishing visual features with which to localize.

Various image descriptors have been proposed for wide baseline image matching [14], recognizing objects [23], and matching low-flying unmanned aerial vehicle (UAV) imagery to satellite imagery [16], [25]. Temporal integration of image-based feature matching has been effective for visual localization [2] and loop-closure detection in FAB-MAP 2.0 [7] using only ground-based imagery. It is an open question whether the concept can be adapted to air/ground image matching. Furthermore, it is not known which of the myriad of image descriptors would perform best at such a task. The primary purpose of this paper is to address these questions. We begin by proposing a method to warp UGV imagery to form ground-based maps. We then evaluate the performance of a variety of descriptors at matching these ground-based maps to a satellite map for various map sizes and terrain types. Based on these results, we integrate these components into a Bayesian localization framework using a particle filter. Finally, we show that our localization algorithm can actually outperform Google Street View's localization by comparing the position estimates from our algorithm to Street View's GPS-based estimates.

The main contributions of this paper are: 1) a characterization of the performance of different image-based descriptors for the problem of matching ground imagery to satellite imagery, 2) a framework for Bayesian localization of a ground vehicle with respect to satellite imagery, and 3) a

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

Fig. 1. Top: Panoramic imagery captured at Balboa Park, CA (at the location of the red marker.) Bottom: Satellite map of the same region. This paper studies matching between both images. Note that the drastic change in perspective between ground and air views makes image matching challenging. The location of the vehicle as indicated by GPS is shown by the red marker. The green marker is an estimate of vehicle location using visual matching. The blue dots represent individual particles. The inset images show location details using a particle filter to track the vehicle position.

demonstration of improvement of localization over GPS in unstructured environments.

II. Related Work

A number of researchers have tackled the problem of matching ground-based data to satellite imagery in various contexts, including 3D matching of air/ground data, image geo-location, and feature-based air/ground image matching. The survey paper [24] provides insight into recent developments and open problems in the field of UGV/UAV cooperation.

One approach to matching aerial data to ground data is to use 3D geometry. For example, Vandapel et al. [20] register lidar data from a low-flying UAV with lidar data from a UGV using shape-based matching using spin-image shape descriptors.

Image geo-location (i.e., the problem of determining where in the world an image was obtained) is another application that can benefit from matching between ground and aerial data. Early work on this problem relied on geo-tagged images [8] or the combination of geo-tags and text information [6] to estimate image locations. Recently, Lin et al. [11] developed a data-driven geo-location algorithm that learns the relationship between ground-level images, their corresponding aerial view, and land use attributes. Their method provides a ranked list of matches, with the correct match being in the top 1% of candidate matches 40% of the time.

Various geometric approaches have addressed matching ground imagery to aerial views. For example, Bansal et al. [3] match ground-level imagery to obliquely viewed facades of buildings seen from an aerial viewpoint. The building outline is used in extracting edge information, and a self-similarity descriptor is used in evaluating potential matches. Vidal et al. [21], [22] use monocular vision as the primary sensing modality. As part of their SLAM framework, edges from building boundaries were extracted and matched between ground and aerial imagery. These approaches are limited to areas with buildings for matching.

The DARPA PerceptOR program addressed ground/air localization challenges for a low-flying aerial vehicle that served as a "flying-eye" to aid a UGV in navigation [9], [17]. The UAV captures advance 3D information of the environment, allowing the UGV to traverse the region faster. A-priori information allowed faster hazard detection compared to the on-board UGV sensors. In [9], the authors address the issue of localization jumps caused by intermittent GPS availability by a lazy registration technique. The positions of obstacles on the global map were made immune to GPS jumps by updating the local map onto the global map at a high refresh rate. More recently, Majdik et al. [13] explored image-based matching between a UAV and panoramic ground imagery using a modified version of affine SIFT to match the UAV imagery with Google Street View images, though the viewpoints were largely from the same perspective.

An analysis of feature spaces like SIFT and SURF in the context of vision-based localization has been explored in [19] for the case of a single ground vehicle. The benefits of using a single image descriptor for the task of matching was studied in [1]. An application using wSURF for vehicle localization was investigated by our group in [2] and motivates the use of whole image descriptors in this paper.

The method proposed in this paper differs from other approaches in two fundamental ways: 1) we formulate the problem as a visually matching images from drastically different perspectives, and do not use 3D sensory information, and 2) the test environment is outdoors, with little man-made structure as a visual aid.

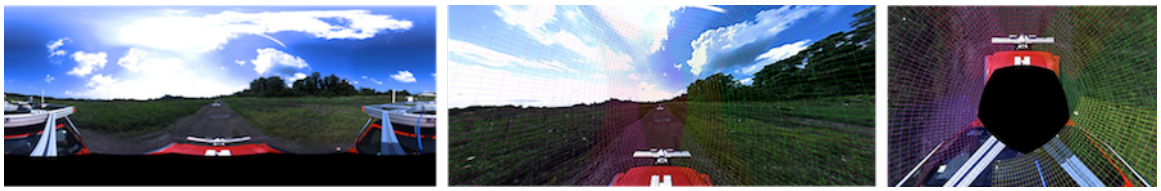

Fig. 2. Images of the panorama, mapping to sphere, and top-down view. The panorama is captured using the Ladybug5 camera. Images captured by different cameras are stitched to form the panorama. The overlaid grid shows the section corresponding to individual cameras.

III. Ground to Satellite Image Matching

The ground based panoramas are warped to form the top-down view of the scene. A single image descriptor is computed for the warped image (query). Similarly, a database of descriptors are computed for a grid of locations on the satellite image. The query is compared to the entire database. Descriptors in the database are considered to be a correct match to the query, if the location of the match is within 5 meters of the UGV.

A. Terminology

All the images in the paper are assumed to be full color images, with an 8-bit representation per channel. The following notation is adopted, as outlined below.
The relevant images are denoted by:

- S – high flying UAV or satellite image (dimensions m x n)
- P – wide field-of-view or panoramic image from the UGV (dimensions p x q)
- T – bird's eye view of the ground (dimensions r x r)

The computed descriptors and associated locations are denoted by:

- $l = \{(u, v)|u \leq m, v \leq n\}$ – pixel location in the satellite image
- $L = \{l_1, l_2, \ldots l_n \in S\}$ – grid of pixel locations in the satellite image
- $D_S^m(l) = \{d_{l_1}^m, d_{l_2}^m, \ldots, d_{l_n}^m | l_1, l_2, \ldots l_n \in S\}$ – the database of satellite image descriptors computed for the grid of locations L corresponding to class $m = \{$SIFT[12], SURF[4], FREAK[15], PHOW[5]$\}$
- q_T^m – image descriptor (query) computed for the bird's eye view, belonging to class $m = \{$SIFT, SURF, FREAK, PHOW$\}$

The matching quality is evaluated using:

- δ – threshold to define proximity of a correct match to the vehicle position
- h – histogram of frequency of proximal matches
- F – cumulative distribution of the likelihood of observing a correct match, conditioned on δ

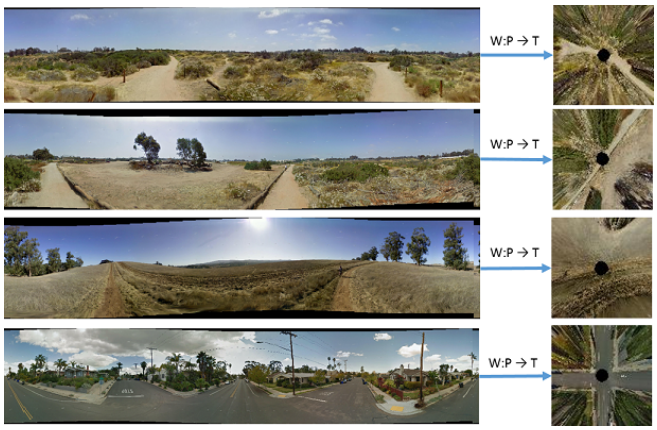

Fig. 3. The UGV panorama is warped to form a bird's eye view of the scene. Warping allows an approximate perspective alignment with the satellite view, and facilitates the matching process.

B. Generation of the Bird's Eye View

The panoramic image from the UGV is warped to obtain a bird's eye view of the scene. Specifically, the warping function is defined by $W : P \rightarrow T$. For each pixel in the top-down view, we lookup the position of the corresponding pixel in the panorama using the inverse warping. The color for the ground location is then obtained using bi-linear interpolation from the panorama pixels. The panorama also captures segments of the vehicle (Figure 2). These pixels are removed by applying a mask on the top-down view (Figure 3).

C. Forming Image Descriptors

Once the template is obtained for the UGV view, a descriptor is computed over the entire image. Whole image descriptors offer the advantage of reducing the computation overhead incurred at the feature detection stage. The descriptor q_T is computed for the template, and is considered the query descriptor. A uniformly sampled rectangular grid of locations, $L = \{l_1, l_2, \ldots, l_n \in S\}$, is overlaid on the satellite image. A square sub-region is extracted centered at each location l_i on S, spanning $\frac{w}{2}$ pixels from a w x w window. Subsequently, a whole image descriptor d_{l_i} is computed for the window. An overview is provided by Figure 4. The descriptors are collected in a database D_S for every location $l_i \in L$. The process is repeated for each of the m descriptor types.

D. Descriptor Matching

The query descriptor q_T is matched against the entire database of descriptors D_S to find the most similar match, as shown in Figure 5. A successful match minimizes the L_2 norm between q_T and $d_{l_i} \in D$. The count of the number of top matches, indexed by K, is defined by

$$h_k = \{c_k(i)|k = 1, 2, 3, \ldots, K, i = 1, 2, \ldots, n\} \quad (1)$$

$$c_k(i) = c_k(i) + 1, \text{ if } \arg\min_{\forall l,i} ||d_{li} - q_T||_2^2 \leq \delta \quad (2)$$

As an implementation detail, $K = 100$ and $\delta = 50$, which correspond to the top-100 matches within a distance of 10 meters from the vehicle. To compare m different feature descriptors, the cumulative distribution of h_k is computed as:

$$F_m = \eta \Sigma_k h_k^m \quad (3)$$

where η is the normalizer.

IV. Experimental Results

Our experiments use data from Google Maps for the aerial maps and Google Street View for the ground-based imagery. Google Street View data is normally acquired using a car with a panoramic camera rig mounted on top at a height of 2.5 m. The platform is also equipped with a GPS unit for localization and a laser scanner for capturing 3D data.

Google recently introduced the Google Trike, which a Street View platform based on a large tricycle. The trike is capable of exploring trails, sidewalks, and other areas that would be difficult or impossible for the Street View Car to traverse.

Our experiments focus on two sites that were captured primarily using the Google Trike. The first site is Balboa Park, in San Diego, CA, and the second site is Arastradero Preserve, in Palo Alto, CA. The sites were chosen because they consist primarily of trails and unstructured terrain. The data in our experiments consists only of panoramic imagery and aerial maps. The GPS information associated with the ground-based imagery is only used for establishing ground truth.

Our experiment demonstrates the basic capabilities of our ground-to-air matching algorithm and compares several different image descriptors. Then, we analyze the performance as the map size varies (equivalently, the region of uncertainty of the sensor location varies). Finally, we demonstrate localization over image sequences for several test routes and compare the performance with GPS-based position estimates.

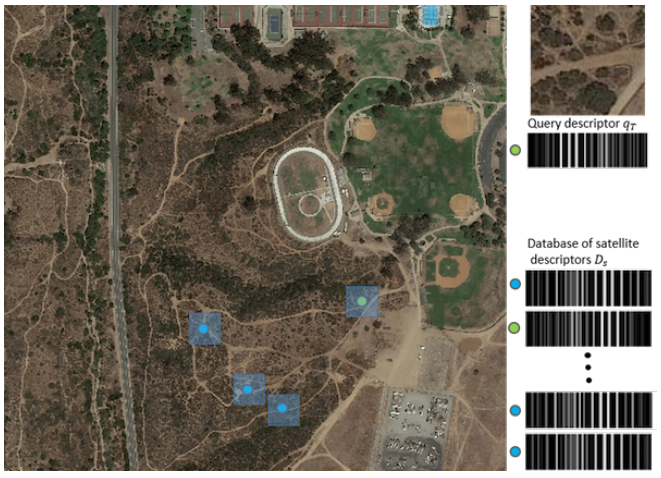

Fig. 4. The query descriptor q_T is compared against a database of descriptors D_S at a grid of locations on the satellite image. The transparent squares on the satellite image are sample locations at which satellite descriptors are computed. The closest match is indicated by the green position on the satellite.

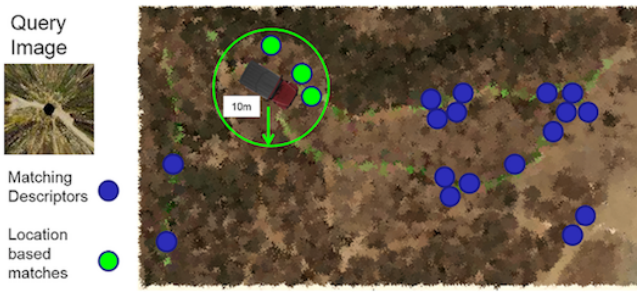

Fig. 5. An example showing the descriptor catching framework. Satellite descriptors with match the template are shown in blue. Locations in green correspond to matching descriptors within 10m of the UGV. The specific instance contributes to three correct matches at the location of the UGV.

A. Comparison of Feature Descriptors

The localization problem is formulated as matching image descriptors of the warped UGV image, with a database of descriptors computed over the satellite map. A descriptor is considered to be a correct match from the satellite, if it is similar to the ground-view descriptor, and also lies within 5 meters of the UGV location. Figure 6 shows the performance of SIFT, SURF, FREAK, and PHOW in matching ground images onto the satellite map for Balboa Park.

For the instance of the large map, 77 locations were found along a continuous trail in Balboa Park. These were instances where Street View imagery was available with GPS information. Each of the 77 bird's eye view templates were queried across a grid of 380,000 locations on the satellite.

The curves are generated as follows: For a given data set, and a set of x,y positions on the map, corresponding ground-based panoramas, and descriptors for those panoramas are computed (labeled ground samples). A database of descriptors for locations at regular intervals on an overlapping grid (spacing delta-$\delta = 8$ pixels) is computed for the satellite map. For each labeled ground sample, the similarity between the ground sample descriptor q_T and every descriptor from the satellite map $D_S^m(l)$ is computed. Each match is labeled as correct or incorrect based on the Euclidean distance of the x,y coordinates of q_T and $D_S^m(l)$. Labeled ground samples are sorted based on the similarity. K_i is the index of the first correct match in the sorted list. A histogram h_k of fraction of ground samples for each value of K is computed, from which the cumulative distribution is plotted in Figure 6.

1) Effect of Map Size: The curves in Figure 6 are the top-K or, F_m plots referred to earlier. One descriptor outperforms another when the area under the curve is larger, or a one-shot comparison may be done at the k^{th} location. SIFT has the best performance in this specific instance, as can be seen by the fact that it has higher fraction of correct matches for all indices k. The two lower rows in Figure 6 show additional experiments conducted to determine if the

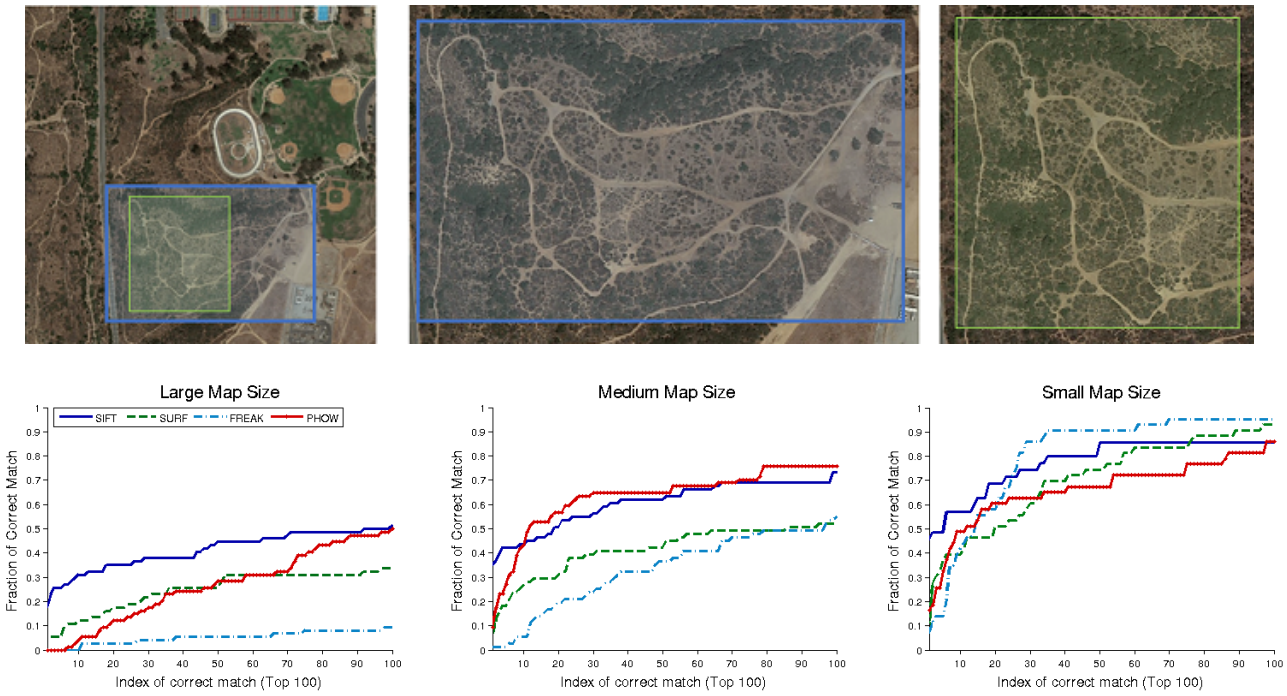

Fig. 6. Results of comparing different feature descriptors in matching ground to satellite imagery. The size of the satellite is highlighted in each case. For smaller satellites, the feature descriptors have similar performance. SIFT is found to be the most consistent at the top-100 matches, across varying satellite image size, and is used in the measurement step of the particle filter.

descriptor generalizes to different map sizes. To validate the performance of SIFT, the map was divided into three sizes: small, medium, and large (original size.) Smaller maps have greater likelihood to match the query descriptor at the correct location. Further, different descriptors perform equally well when the map size is restricted. This motivates the use descriptors that use fewer computations, like FREAK, for the small map. However, the graphs in 6 show that SIFT obtains the overall best performance, even with increasing complexity of the satellite map. Consequently, we use SIFT for all subsequent experiments in the paper.

2) Location Details: An investigation into the location details of an incorrect match provides additional insights into the unique nature of the matching problem. Figure 7 shows such a case. Non-ground areas like the a group of bushes, appear distorted when warped to a bird's eye view and cause a star-like artifacts in the template. The dominant gradients associated with the star-like pattern overpower the pattern of the ground region, causing an increased chance of being mismatched. The results show the necessity to distinguish between regions of ground against non-ground. Explicitly dealing with these non-ground regions is a subject of ongoing research.

V. Applications

The particle filter is an over the shelf tool (Algorithm 1), used to demonstrate the efficacy of using the SIFT descriptor for the application of UGV to satellite localization. The details of the particle filter are adapted from [18], with all probability distributions modeled by Gaussians. The vehicle state at time t is indicated by $\mathbf{x}^t$, which includes the position

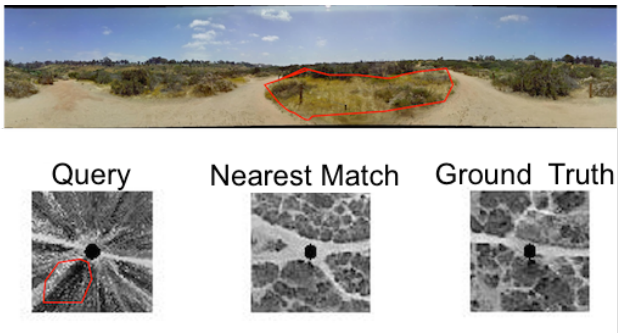

Fig. 7. Non-ground regions, like bushes, cause star-shaped artifacts on being warped to form the template, marked out in red. The dominant gradient due to the artifacts cause an incorrect match to an section of the trail that has a star-shaped intersection. The ground truth, however, shows a straight path, and provides an insight into the fact that non-ground regions need to be treated differently from the rest of the image.

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

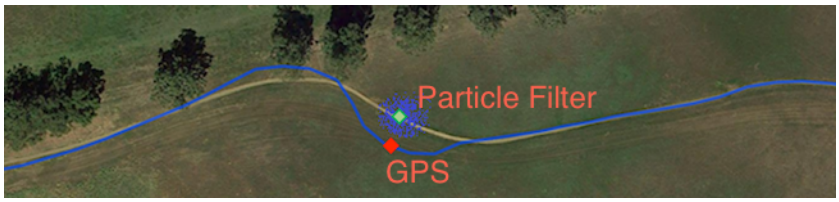

Fig. 8. The Arastradero Preserve dataset location detail shows that GPS is not aligned with the trail in the meandering region between the successive crests. The particle filter estimates the true location of the vehicle to be closer to the trail, demonstrating the efficacy of air to ground image matching.

Fig. 9. Balboa Park location (1). The image above is magnified below, to shown an instance where Google Street View indicates an incorrect GPS location, in this case there are duplicate paths on a section of the trail. Visual matching places the location of the vehicle near the true location of the trail.

on the satellite map. The set of weights, indexed by j is denoted by w_j. Vehicle motion is denoted by u, observation (panorama captured) by z, and set of particles by P. For the initial step $t = 0$ the particles locations $\mathbf{x}$ are distributed randomly over the whole map and the weights are set to a uniform distribution $w = 1/N$.

The particle filter is applied to the Balboa and Arastradero datasets, and is shown to have improved localization compared to using only GPS data.

A. Improving Google Street View

Comparing our image-based localization algorithm to position estimates from Google Street View, we find that, in many cases, our algorithm provides a subjectively more accurate estimate than the GPS estimate used by the Street View sensor. The Google Street View locations on Google Maps are indicated by light blue lines overlaid on the map.

Fig. 10. Balboa Park location (2). GPS locates the vehicle at the center of the road divider line associated with a section of trees, highlighted by the transparent orange box. The position estimated using vision based matching provides a more feasible location for the vehicle on the road.

At certain locations, the tracks are incorrectly placed on the satellite map, presumably due to GPS errors or drift. Sample locations where the position of the track is incorrect are shown in Figures 9, 10, and 8. Our vision-based localization algorithm provides a more accurate estimate of the camera location on the map as shown in the following examples.

1) Balboa Park Location Detail: The data in Balboa Park has incorrectly marked tracks in two cases. 1) The lower section of the off-road trail has two potential paths associated with a single route. 2) The section of road near the parking lot, places the position of the vehicle on a region of trees dividing the road. The particle filter correctly determines the location in both cases. The method successfully converges to the correct location of the trail in the off-road region. In the case of the parking lot, the position of the vehicle is estimated to be on the road, while GPS indicates it is on a clump of trees dividing the road.

2) Arastradero Preserve Location Detail: The Arastradero dataset is more challenging, since the terrain type is uniformly grassland. Street View data indicates that the vehicle moves off the trail between successive meandering locations as shown in Figure 8. Using the particle filter partially improves vehicle localization by placing more weight on region on the trail.

The online multimedia content shows the particle filter for the Balboa Park dataset. The video file shows the ground vehicle traversing through the environment, the warped panorama, and predicted location on the satellite map.

VI. Conclusion

A comparison of feature descriptors for the task of ground to aerial image matching is discussed in this paper. In particular, the task of matching panoramas from a UGV to a satellite map is investigated. Different feature descriptors

are compared in a top-K matching framework, for varying map complexity. The results indicate that SIFT performs consistently well in matching warped bird's eve view from the panorama to the satellite image. An application to vehicle localization is demonstrated using Google Street View data and it is shown that vision based localization allows for better UGV position estimates compared to stand-alone GPS. Future work includes extension of the method to associate semantic information into the matching procedure. Additionally, non-ground regions which violate the dominantly flat-ground assumption cause warping artifacts, that lead to incorrect image matching. This is currently being explored in conjunction with non-ground segmentation schemes.

ACKNOWLEDGMENT

This work was supported by the Agency for Defense Development, Jochiwongil 462, Yuseong, Daijeon, Korea.

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. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (SURF)," *Comput. Vis. Image Underst.*, vol. 110, no. 3, pp. 346–359, Jun. 2008.

[5] A. Bosch, A. Zisserman, and X. Munoz, "Image classification using random forests and ferns," 2007.

[6] 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.

[7] M. Cummins and P. Newman, "Appearance-only SLAM at large scale with FAB-MAP 2.0," *The International Journal of Robotics Research*, vol. 30, no. 9, pp. 1100–1123, 2011.

[8] J. H. Hays and A. A. Efros, "IM2GPS: estimating geographic information from a single image," in *Proc. Computer Vision and Pattern Recognition (CVPR)*, June 2008.

[9] A. Kelly, A. Stentz, O. Amidi, M. Bode, D. Bradley, A. Diaz-Calderon, M. Happold, H. Herman, R. Mandelbaum, T. Pilarski, P. Rander, S. Thayer, N. Vallidis, and R. Warner, "Toward reliable off road autonomous vehicles operating in challenging environments," *The International Journal of Robotics Research*, vol. 25, no. 5-6, pp. 449–483, 2006.

[10] 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.

[11] 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.

[12] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vision*, vol. 60, no. 2, pp. 91–110, Nov. 2004.

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

[14] J. Matas, O. Chum, M. Urban, and T. Pajdla, "Robust wide-baseline stereo from maximally stable extremal regions," *Image and Vision Computing*, vol. 22, no. 10, pp. 761 – 767, 2004, british Machine Vision Computing 2002.

[15] R. Ortiz, "FREAK: Fast retina keypoint," in *Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, ser. CVPR '12. Washington, DC, USA: IEEE Computer Society, 2012, pp. 510–517.

[16] S. Scherer, "Low-altitude operation of unmanned rotorcraft," Ph.D. dissertation, Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, May 2011.

[17] A. T. Stentz, A. Kelly, H. Herman, P. Rander, O. Amidi, and R. Mandelbaum, "Integrated air/ground vehicle system for semi-autonomous off-road navigation," in *Proc. of AUVSI Unmanned Systems Symposium 2002*, July 2002.

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

[19] 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.

[20] N. Vandapel, R. R. Donamukkala, and M. Hebert, "Unmanned ground vehicle navigation using aerial ladar data," *The International Journal of Robotics Research*, vol. 25, no. 1, pp. 31–51, January 2006.

[21] 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.

[22] 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.

[23] P. Viola and M. Jones, "Robust real-time object detection," *International Journal of Computer Vision*, vol. 4, 2001.

[24] S. Waslander, "Unmanned aerial and ground vehicle teams: Recent work and open problems," in *Autonomous Control Systems and Vehicles*, ser. Intelligent Systems, Control and Automation: Science and Engineering, K. Nonami, M. Kartidjo, K.-J. Yoon, and A. Budiyono, Eds. Springer Japan, 2013, vol. 65, pp. 21–36.

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