Towards Reprojection Flow for Image Registration Across Seasons

Shane Griffith and Cédric Pradalier

I. INTRODUCTION

The geometry of a natural environment may provide a basis for visual data association across seasons. Time brings changes in appearance due to illumination, weather, and foliage, which can make visual data association challenging even for humans [1]. Yet, extremely precise visual data association across seasons is possible, which some seed-caching corvids demonstrate from the autumn into the winter, spring, and summer [2]. Indeed, the use spatial information is a primary mechanism with which these birds identify places [2]. A similar feat may be possible using mobile robots if scene structure is made a primary mechanism of visual data association.

This abstract presents initial work towards Reprojection Flow: exploiting scene structure for visual data association. In this new approach, image registration provides the initial data association between consecutive surveys, SLAM among all surveys captures scene structure in a globally consistent map, and scene structure within image registration provides data association across longer spans of time. This last step is Reprojection Flow, in which 1) co-visibility is defined according to the visibility of reprojected points; and 2) reprojected map points provide anchors to guide image registration. We applied these methods to four surveys of a lakeshore captured by a mobile robot and acquired promising results.

II. RELATED WORK

Environment monitoring is a topic of several recent papers, but many solutions still lack robustness to variation in appearance of natural environments. Consumer–priced digital cameras have resulted in an explosion of freely available, repetitive scans of many places over long periods of time, which makes it possible to see trends (e.g., global warming) [3]. Yet, in relying on point–based features to construct time-lapses, which lack robustness to variation in appearance of the environment, only a subset of images may be used [3].

Cheap video cameras have also increased the spatial scale of observations for monitoring applications. Large swaths of an environment can be captured and analyzed if the same path is walked multiple times. Once images of the same places are identified, a homography that is estimated using point–based feature matches may be used to more closely align images [4]. The alignment precision could, however, be lower than image registration techniques [3].

Both authors are with GeorgiaTech Lorraine - UMI 2958 GT-CNRS, Metz, France {sgriffith7, cedric.pradalier}@gatech.edu

Mobile robots may be able to capture large swaths of an environment autonomously, which could lead to more consistent observations of a place. The use of robots also means a larger sensor suite may be utilized to further simplify data association. Dense laser scans have led to detecting changes on a willow tree across seasons [5], and identifying crop growth for precision agriculture [6]. In case a laser is unavailable, precise visual data association is possible between two images of the same scene using image registration (e.g., SIFT Flow [7]).

Our prior work demonstrated precise visual data association for *consecutive* surveys of a natural environment [1]. There, visual SLAM and SIFT Flow (applied in that order, and SIFT Flow was applied without extra constraints) outperformed other state-of-the-art techniques for registering surveys. In registering non-consecutive surveys, however, the number of precise alignments decreased with time. We aim to overcome this limitation of appearance–based techniques using Reprojection Flow.

III. METHODOLOGY

A. *Image Registration*

The best techniques for data association in a natural environment utilize scene structure. SIFT Flow is among the best, which registers whole images of point–based features using an MRF. In minimizing the alignment energy, scene structures are matched and used as anchors to align the rest of the image. Because the method is designed for non-rigid registration, it can align images of different scenes, but it lacks basic feature matching constraints. Thus, in using it for monitoring applications, in which images of the same scenes are aligned, we add epipolar constraints and match consistency constraints to improve the alignment quality.

Enforcing epipolar geometry with data association is different for an MRF than for point–based features. First, SIFT Flow is run without constraints at the top layer of its image pyramid. Between layers, feature matches from the previous layer are used with RANSAC to estimate a fundamental matrix. This is used to weight the data terms of the MRF in order to constrain image alignment to epipolar lines.

Match consistency constraints ensure that the flow image that aligns one image to another is consistent with the flow image for the reverse alignment. It is, e.g., useful for eliminating alignment errors due to reflective surfaces. This constraint is enforced using images from the top layer of the image pyramid. Given a flow image for one direction, the hypothesis space in the reverse direction is weighted to be more consistent with it. After several iterations, the most consistent flow is used for full-resolution registration.

B. Visual SLAM

A map of the environment also encodes scene structure, which is obtained using visual SLAM. Visual SLAM jointly optimizes the estimates of the camera pose and the visual features in the environment (i.e., the map). In this case, the map consists of features from visual odometry; up to 300 KLT feature tracks per image. To acquire one consistent map for multiple surveys, the output from image registration is used to map each KLT feature onto images from other surveys. Thus, several visual odometry features are observed several times both within and between surveys.

Observations of visual odometry features across surveys are obtained using the highest quality image alignments. First, the image pairs that align best between two surveys are found. Given an image from one survey, the search for the image that aligns best in the other survey is run on the set of images from the same general position (similar GPS and compass values). During the search, image registration is only run at the top layer. Given a full–resolution alignment, the KLT features are mapped to the registered image, and only the inliers according to epipolar geometry are utilized as measurements for SLAM.

C. Reprojection Flow

Given scene structure, the first question is what images to align between surveys. This is non-trivial if appearance–based matching becomes unreliable. Because the robot's poses are accurate (due to SLAM), the image with the closest pose may match best. However, this heuristic does not maximize co-visibility. Instead, we define co-visibility using the reprojected 3D map points and the robot poses.

That is, for an image in one survey, the one from a different survey that it is aligned with has the highest co-visibility of map features. Co-visible points are those that each pose sees or that neither sees. Other points are those that one sees but not the other. We set up a contingency table in this way, and then apply the G-statistic to compute co-visibility. The G-statistic comes from statistical analysis and is used to measure independence between two variables.

Given two images of the same scene, the second question is how to obtain robust image registration. In aligning consecutive surveys, image registration works because appearance–based scene structure acts as anchors, which pull the rest of the image into alignment. The more surveys are separated in time, however, the fewer the anchors. Fortunately, because the map and the poses are consistent (after SLAM), points in the map from one survey can be projected onto an image from a different survey to provide the anchors for image registration. Reprojected points define matching constraints at particular pixels (as weights to the data terms of the MRF), but they also provide epipolar constraints for the entire image before appearance-based matching, and can anchor the match consistency constraint.

IV. EXPERIMENTS

Both Reprojection Flow techniques were evaluated using four surveys of a 1km lakeshore that were captured using the Kingfisher Autonomous Surface Vehicle. The surveys were captured between Sep. 11 and Oct. 3, 2014. The data we used consisted of the robot's pose from GPS and the compass, as well as 704x480 images, captured twice per second. The evaluation was limited to a survey section about 1/5 the size of the full lakeshore to reduce the size of the optimization problem. Image alignment energy is used for the analyses, which is lower for better alignments.

TABLE I

ALIGNMENT ENERGY (LOW-RES) FOR TWO CO-VISIBILITY METHODS BETWEEN IMAGES FROM A SEP. 11 SURVEY TO OTHER SURVEYS.

Survey Date	Sep. 19	Sep. 26	Oct. 3
Closest Pose	$1.14 \pm .2$	$.99 \pm .2$	$1.16 \pm .1$
Co-Visibility	$1.11 \pm .1$	$.97 \pm .1$	$1.14 \pm .1$

Results are multiplied by a factor of 10^6.

TABLE II

ALIGNMENT ENERGY FOR TWO IMAGE REGISTRATION METHODS BETWEEN IMAGES FROM A SEP. 11 SURVEY TO OTHER SURVEYS.

Survey Date	Sep. 19	Sep. 26	Oct. 3
SIFT Flow	$6.28 \pm .4$	$5.94 \pm .4$	$6.47 \pm .3$
Reprojection Flow	$6.26 \pm .4$	$5.94 \pm .4$	$6.34 \pm .3$

Results are multiplied by a factor of 10^8.

Image pairs selected using Reprojection Flow reach lower alignment energies (on average) than those selected using the closest pose, as shown in Table I. This is because they are closer to the same scenes. They are found because co-visibility is a function of both the camera pose and the observed scene points, rather than only a function of pose.

Using the co-visible image pairs, an additional improvement in alignment energy is achieved using Reprojection Flow and other image registration constraints, as shown in Table II. A slight improvement for all three surveys is due to the epipolar constraints. Reprojection Flow adds a little noise to the Sep. 19 and the Sep. 26 survey, which already align very well. Match consistency significantly improves the alignments for the Oct. 3 survey, which has a lot of reflections. Reprojection Flow also significantly improves the alignment energy of the Oct. 3 survey.

These results support Reprojection Flow, and warrant a larger study with more surveys across a longer span of time.

REFERENCES

[1] S. Griffith and C. Pradalier, "Survey registration for long–term lakeshore monitoring," *JFR special issue on FSR 2015 (submitted)*, 2016.

[2] N. S. Clayton and J. R. Krebs, "Memory for spatial and object-specific cues in food-storing and non-storing birds," *J. of Comparative Physiology A*, vol. 174, no. 3, pp. 371–379, 1994.

[3] R. Martin-Brualla, D. Gallup, and S. M. Seitz, "Time-lapse mining from internet photos," *ACM TOG*, vol. 34, no. 4, p. 62, 2015.

[4] M. Milford, J. Firn, J. Beattie, A. Jacobson, E. Pepperell, E. Mason, M. Kimlin, and M. Dunbabin, "Automated sensory data alignment for environmental and epidermal change monitoring," in *ACRA*. Australian Robotic and Automation Association, 2014, pp. 1–10.

[5] G. Hitz, F. Pomerleau, F. Colas, and R. Siegwart, "State estimation for shore monitoring using an autonomous surface vessel," in *ISER*, 2014.

[6] L. Carlone, J. Dong, S. Fenu, G. C. Rains, and F. Dellaert, "Towards 4d crop analysis in precision agriculture: Estimating plant height and crown radius over time via expectation-maximization," in *ICRA Workshop on Robotics in Agriculture*, 2015.

[7] C. Liu, J. Yuen, and A. Torralba, "SIFT Flow: Dense correspondence across scenes and its applications," *PAMI*, vol. 33, no. 5, pp. 978–994, 2011.