Reprojection Flow for Image Registration Across Seasons

Shane Griffith[1,2,3]
sgriffith7@gatech.edu
Cédric Pradalier[2,3]
cedric.pradalier@georgiatech-metz.fr

[1] College of Computing, Georgia Institute of Technology, Atlanta, USA
[2] GeorgiaTech Lorraine, Metz, France
[3] CNRS UMI 2958 GT-CNRS, Metz, France

Abstract

We address the problem of robust visual data association across seasons and viewpoints. The predominant methods in this area are typically appearance–based, which lose representational power in outdoor and natural environments that have significant variation in appearance. After a natural environment is surveyed multiple times, we recover its 3D structure in a map, which provides the basis for robust data association. Our approach is called Reprojection Flow, which consists of using reprojected map points for appearance–invariant viewpoint selection and robust image registration. We evaluated this approach using a dataset of 24 surveys of a natural environment that span over a year. Experiments showed robustness to variation in appearance and viewpoint across seasons, a significant improvement over a state-of-the-art appearance–based technique for pairwise dense correspondence.

1 Introduction

The capacity to register observations of a natural environment across seasons has been achieved in Nature, and may be possible to automate using machines. Evidence suggests it can be solved using information primarily from vision. Clark's nutcrackers cache over 30,000 whitebark pine seeds in autumn, which sustain them through winter, spring, and summer [8, 22]. An average of 1-15 seeds are stored per cache [22], which are spread out over a large geographical area (rather than clustered) [2]. Nutcrackers have an uncanny spatial memory to achieve this feat [11]. They integrate spatial information (before shape and appearance) of environmental cues like trees, rocks, and logs to find caches, including those that are out in the open or buried under snow [1, 19] (also see [5] for a nice video). A similar feat may be possible with mobile robots (and other portable cameras) if they utilize scene structure in a similar way.

Images captured in outdoor and natural environments accumulate many kinds of variation in appearance, which limit the effectiveness of appearance for registering scene contents. The strength of illumination, the locations of shadows, and the type of weather vary between

Pages 67.1-67.12
DOI: https://dx.doi.org/10.5244/C.30.67

Reference Image | Unaligned Image | Target Alignment | Appearance-based Alignment

Figure 1: Odd behavior can be observed when image alignment is appearance–based. Here tree branches are warped into foliage. The target alignment, in contrast, retains scene structure; the foreground trees are aligned despite their significant variation in appearance.

each survey. Over longer time scales, seasonal changes in foliage may manifest as differences in size, shape, and color. Noise accumulates with water droplets, sun glare, and debris. Other changes in natural environments contribute. Furthermore, dynamic processes diminish the possibility of capturing scenes from the same viewpoint. Cumulatively, the variation in appearance of natural environments represents a formidable challenge to data association.

The difficulty in using appearance for correspondence may be due to a misattribution of the key source of information to that. Across seasons, for example, several trees could appear leafy with a bright skyline in the summer, yet bare with the sky shining through them in winter (as in, e.g., Fig. 1). The winter image may best match the *appearance* of the summer image if the tree branches are warped into foliage. As humans, however, we know this is incorrect. We have experienced many trees (in the ways our sensorimotor repertoire allowed). Consequently, we know what a tree is, that it has a structure, that it may abscise its leaves in winter, etc. The true correspondence for these images would align the branches from one to parts of the dense foliage of the other, leaving the tree structures intact.

Identifying accurate correspondences may not, however, require categorical (or semantic) knowledge of things; prior experience may be enough. Trees keep the same position between observations, as do rocks, logs, the landscape, and many other objects that lack agency. The positions of things can, independently of appearance, indicate correspondences between images. A localized observer may be able to exploit this information to acquire fine–grained correspondences of things that appear substantially different.

This paper presents Reprojection Flow, a method that exploits the spatial information in a map to achieve image registration across seasons. To enable its use, first image registration is performed between surveys to acquire inter-survey observations of Kanade-Lucas-Tomasi feature tracker [16] (KLT) points, and then visual simultaneous localization and mapping (SLAM) is applied to acquire one map for all the surveys. Reprojection flow consists of two steps: 1) identifying images of the same scene by the co-visibility of reprojected map points; and 2) using reprojected map points to initialize and constrain image registration. Both techniques are independent of appearance (given the map and the camera poses). In tests with surveys of a natural environment, this approach significantly improves dense correspondence across seasons. It also provides robustness to changes in viewpoint.

2 Related Work

There is a broad literature on robust data association for natural environments. Methods use the environment appearance and structure to varying degrees. A recent survey provides a comprehensive review on visual place recognition [15], which shows that the central components are a map and a belief in whether the incoming visual information is in the map.

Among the state-of-the-art, CNNs for appearance–based place recognition have been shown to outperform other techniques (e.g., SeqSLAM [18]) that are designed for robustness to changes in appearance [21]. The descriptor provided by the third layer of a CNN provides the right balance of specificity and generality to recognize many places across seasons. This supports the idea that methods designed for scene categorization can also work well for scene recognition across major changes in appearance.

Image registration using SIFT Flow is driven by a similar idea, which is designed for scene correspondence [12]. That is, it can register images of different scenes that have a similar appearance, but it also works well for aligning images of the same scene that have significant variation in appearance. For example, [6] showed that the method is robust to some changes in appearance of a natural environment. Scene structure provides the visual anchors through which spatial constraints pull the rest of the image into alignment. More work is, however, needed to precisely align images that are more than a few months apart.

Several recent techniques in image registration have built on the idea of matching whole images worth of point–based features [10, 23]. Instead of using image pyramids, Deformable Spatial Pyramids can be used to achieve scene correspondence much faster than SIFT Flow [10]. The method enforces spatial coherence by registering blocks of an image, rather than using spatial constraints between individual pixels. Because the finest layer of registration also lacks spatial constraints, the method is fast, but it loses precision.

Other techniques have sought to improve SIFT Flow by replacing the descriptor. In case images have large degrees of rotation, image registration based on the Daisy Filter can be applied [23]. Yet, this and many other predominant image descriptors are hand-engineered, which has motivated [13] to investigate the use of CNN features for image registration. For registration tasks, they found that SIFT Flow performs comparably with either SIFT features or CNN features. In some cases, correspondence may be made more precise by matching generically spaced patches between images, rather than grid–sampled keypoints [7].

Large image collections of the same place may enable us to see the general trend of how a scene changes over time (i.e., time-lapse mining [17]). Using that approach, they estimate a depth map to combine images. An image will be part of the time–lapse if it has SIFT features that match the majority set. One challenge is how to utilize the rest of the images. Ideally, every image would be part of the final result.

FlowWeb has been proposed to achieve consistency among a large set of images of the same category of thing [24], which takes as input the flow labels from individual image registrations and outputs the corrected, consensus flow labels. The key idea is that alignment quality between an image pair can be measured using consistency with a third (or more) image. This same idea can be used to exploit 3D semantic appearance for aligning objects [25]. Their aim is to learn the 3D semantic appearance of things (using 3D CAD models) in order to overcome variation in appearance and viewpoint, where other techniques (like SIFT Flow) may fail. A CNN is trained to learn dense correspondences using two images from a category of thing with a 3D model from the same category and viewpoint.

Reprojection Flow is complementary to techniques for pairwise dense correspondence. It builds on high quality alignments between near–time surveys (within a couple months) in order to align surveys further apart in time. Independently of appearance, it identifies which scenes to match between two surveys. It also bootstraps the alignment process by exploiting the sparse correspondences of reprojected map points.

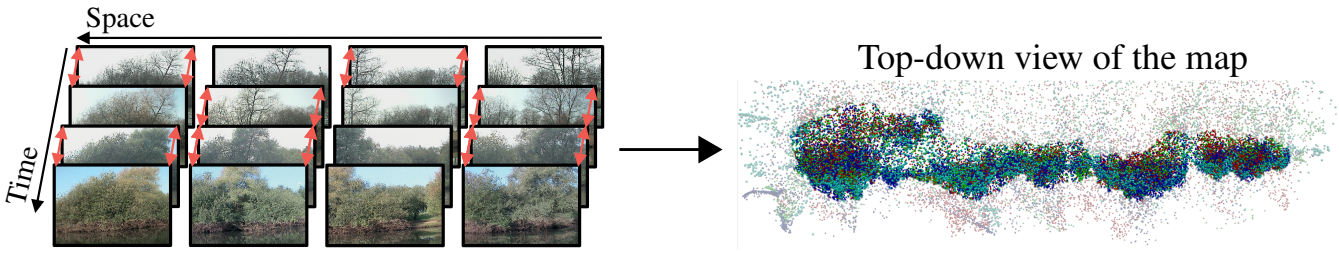

Figure 2: Sparse image registrations (red arrows) between near-time surveys are used to map KLT points between surveys. These measurements are used with camera pose estimates from each survey to create a multi-survey map. The map's highlighted area is the foreground.

3 Methodology

A map can make precise, robust data association possible, but acquiring one that is composed of landmarks from different seasons is a feat in and of itself. First, images are registered (low-res) between near–time surveys to identify images of the same scenes (aided by GPS). Full resolution image registration is performed on the set that aligns well in order to acquire inter-survey observations of KLT–tracked landmarks. A map is recovered from the set of intra- and inter-survey landmark observations using visual SLAM. Reprojection Flow uses the map to perform appearance–invariant viewpoint selection and data association. See Fig. 2.

3.1 Initial Image Registration

This paper applies the SIFT Flow dense correspondence algorithm [12] 1) to perform an initial image retrieval; 2) to acquire inter–survey observations of KLT points; and 3) with Reprojection Flow for dense correspondence across seasons. SIFT Flow aligns whole images using appearance 'anchors' to bring the rest of the image into alignment. Because the method is designed for nonrigid dense correspondence, it works well for matching images of different scenes. In this work, however, we are interested in acquiring dense correspondence for images of the same scenes. Therefore, we extend SIFT Flow by adding basic feature matching constraints to it, including epipolar constraints and match consistency constraints.

SIFT Flow finds a dense correspondence between two images, I^p, I^q, using a Markov Random Field (MRF). In an MRF, the pixels of I^p correspond to a grid of variables with constraints among neighbors, in which the goal is to assign each variable to the most likely hypothesis of a set. For the dense correspondence problem, a hypothesis is a flow vector that matches a pixel in I^p to a pixel in I^q. A flow vector for a pixel, $p = (x^p, y^p) \in I^p$, is $w_p = (u_p, v_p)$, where $u_p, v_p \in [-h \ldots h]$ and the corresponding pixel is $q = (x^p + u_p, y^q + v_p) \in I^q$.

A flow vector, w_p, becomes part of the output dense correspondence if it minimizes the alignment energy, $E(w)$. The alignment energy is the total cost of all the constraints in the MRF for a particular dense correspondence, w. The constraints consist of a data term, a regularization term, and a spatial term, as:

$$E(w) = \sum_p min(|S^p(p) - S^q(p + w_p)|_1, t) * \text{epi} + \text{cyc} \qquad (1)$$
$$+ \sum_p v|u_p + v_p| + \sum_{r \text{ adj. } to \ p} min(\alpha|u_p - u_r|, d) + min(\alpha|v_p - v_r|, d)$$

The data term is defined using the L_1 distance between SIFT descriptors [14], which is extracted at each pixel to form SIFT Images S^p and S^q. The parameters $d = 10200$, $\alpha = 255$,

$\nu = 0.255$ were experimentally determined. The truncation term t is the median of all the descriptor distances computed from I^p to I^q. The hypothesis space, h, was decreased from 11, 5, 3, to 1 in an image pyramid with four layers. SIFT Flow minimizes the alignment energy for the whole image to produce the dense correspondence.

Epipolar constraints are added to this process to improve dense correspondence of real scenes. For real scenes, if matches for a subset of pixels are known, the rest are likely to lie on epipolar lines. This constraint is used to improve dense correspondence as the flow result is propagated up the image pyramid. Given a flow result, epipolar lines are estimated and used to weight the data terms of the MRF. The weight $\text{epi} = 1 - \mathcal{N}(\mu, \delta)$, where μ is the distance to the epipolar line from q and $\delta = 2.5$.

Match consistency is implemented to bring the $I^p \rightarrow I^q$ correspondence in agreement with the correspondence in the reverse direction. It helps to improve correspondence accuracy and reduce correspondence failures (due to perceptual aliasing). The forward and reverse alignments are consistent if the flows of their corresponding pixels sum to zero. The cycle weight is defined as $\text{cyc} = (w_r - w_p(r + w_r)) * c$, where $c = 16$ (a small fraction of α to gradually make the alignments consistent after several iterations). Match consistency is only used at the top layer of the image pyramid in order to reduce computation time.

3.2 Visual SLAM

Graph–based visual SLAM is used to capture the structure of the environment in the form of hundreds of thousands of 3D visual landmarks, which we use in Reprojection Flow to guide data association across seasons (see [6] for a detailed formulation of the SLAM factor graph for one survey; here we add constraints between surveys). The challenge for surveys of outdoor and natural environments is that traditional feature matching between them lacks robustness to variation in appearance. Therefore, inter-survey observations are acquired by mapping KLT points between surveys using image registration, which is more robust to variation in appearance (but only over short time intervals).

Visual features are extracted from each image and then tracked for the duration that they are visible using KLT. To acquire a set of points that are distributed in the scene, images are divided into a 12×20 grid and at most five Harris corners are extracted per cell. New features are only extracted from cells with fewer than five features being tracked. Up to 300 visual features are tracked per image.

Inter-survey observations are acquired by mapping the KLT points between images using image registration. Only the best alignments are used because many images have significant alignment noise. Image alignment quality is measured using alignment energy and match consistency at the top layer of the image pyramid (i.e., low resolution). For an image I^p in survey 1, a local search for the image I^q of survey 2 that best matches I^p is found. If the alignment energy is < 1120000 (which mostly corresponds to high-quality alignments) and the consistency is $\geq 95\%$ (that is, 95% of pixels have error ≤ 1 pixel), full resolution image registration is performed, and the resulting flow field is used to acquire inter-survey observations. Only the inliers according to epipolar geometry and match consistency are used for the optimization. Optimization (as in, e.g., [3]) is applied to the trajectories and the maps from all the surveys in order to bring them into alignment.

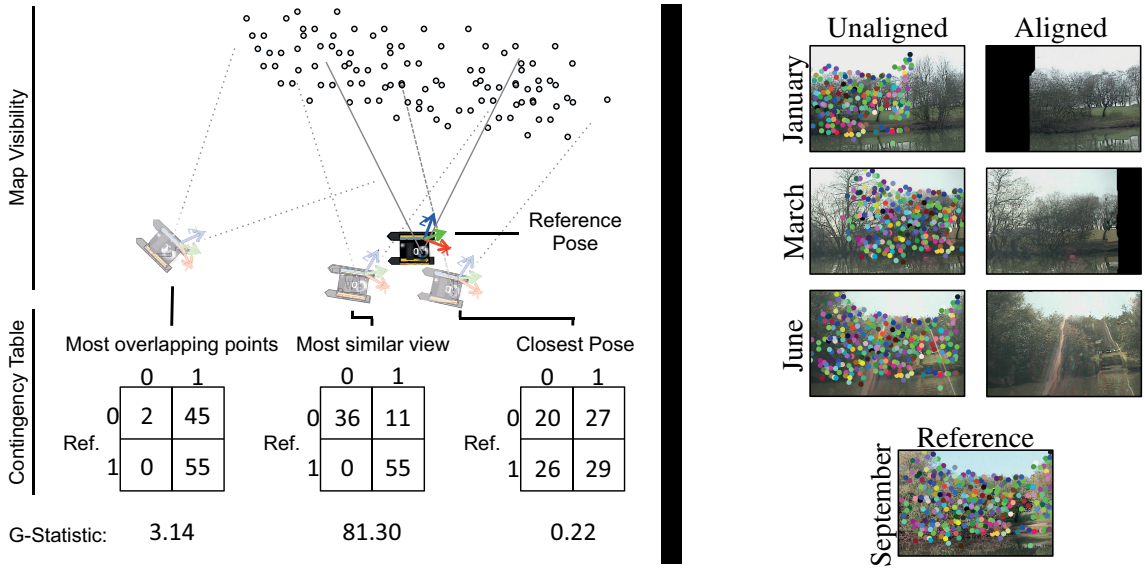

Figure 3: Reprojection Flow: **left)** Finding the most similar view by maximizing co-visibility using the G-statistic, compared to a closest pose heuristic, and a heuristic to maximize the number of overlapping points. The contingency tables are shown for each case. **right)** Using reprojected map points to guide image registration. The KLT points of the reference image are shown projected onto the unaligned images.

3.3 Reprojection Flow

Given the optimized map and camera poses, reprojected map points are used for both viewpoint selection and the registration of images (see Fig. 3). Both provide information that is independent of scene appearance. We call this technique Reprojection Flow.

3.3.1 Viewpoint Selection

A map's first use is in identifying images of the same scene from multiple surveys. An image from one survey is the prototype scene, which guides the selection for the rest. This task is typically nontrivial using appearance–based techniques due to the significant variation in appearance between surveys. A few heuristics are available if a consistent map and poses are available. The pose closest to the reference pose may be an obvious choice, but the weights for the position and the orientation depend on the locations of scene contents. A different heuristic that accounts for the contents of the scene is to use the pose with the most overlapping landmarks. However, the best image in that case may have a large difference in pose and may capture many other landmarks. The best pose maximizes co-visibility.

Co-visibility is the property of two images that the set of landmarks seen and the set of landmarks not seen are similar. That is, high co-visibility indicates that a similar set of map points from the two surveys projects onto both images. For each map point, there are four possibilities: 1) the point projects outside of both images; 2) and 3) the point projects onto one and not the other; or 4) the point projects onto both images. This information is captured in a contingency table of entries N_{ij}, each of which indicates the number of landmarks that match one of the four possible cases of co-visibility:

$$\begin{array}{|c|c|}\hline N_{00} & N_{01} \\ \hline N_{10} & N_{11} \\ \hline \end{array},$$

The degree with which two poses maximize co-visibility is measured using the G-statistic. The method is used to calculate the degree with which two variables are dependent, i.e., the

p-value. In robotics, this technique was originally proposed by Sukhoy *et al.* [20] for co-movement detection. Here, the image with the highest G-statistic maximizes co-visibility:

$$G = 2\sum_{i=0}^{1}\sum_{j=0}^{1} N_{ij} \ln\left(\frac{N_{ij}(N_{00}+N_{01}+N_{10}+N_{11})}{(N_{0j}+N_{1j})(N_{i0}+N_{i1})}\right), \tag{2}$$

3.3.2 Image Alignment

Map points define the positions of things, which can indicate correspondences between images when reprojected onto them. This *reprojection flow* directly provides sparse data association among images of the same scene. Indirectly, it provides epipolar constraints for the rest of the image, it anchors the match consistency constraint, and it centers the hypothesis space on the correct alignment. Collectively, these techniques improve image registration across seasons because they maximize the use of appearance–invariant information.

This approach relies on the correctness of the map points and the camera poses. To minimize the effect of reprojection error, only the original KLT points from the two to-be-aligned images are used (as opposed to all the map points at the same scene). Thus, the reprojection error for each point is only incurred in one direction. Also, the outliers according to epipolar geometry are eliminated from the set. Beyond these constraints, some error is allowable because image registration proceeds through a coarse-to-fine image pyramid.

The direct constraint is applied to image registration through the data term of the MRF. The data term is replaced instead of weighting it to eliminate the effect of appearance on the correspondence. For a reprojected pixel location, r, the constraint is defined as $(1 - \mathcal{N}(r,\sigma)) \times t$, where σ is the ratio of the reprojection error to the image scale.

4 Experiments

4.1 Dataset

This paper analyzes data from 24 100m-long surveys of a natural environment captured over the span of a year. An autonomous surface vehicle (ASV) was deployed roughly bi-weekly on a lake to capture image sequences of its lakeshore. The platform was the Kingfisher ASV from Clearpath Robotics. As it is moving, the boat stores 704x480 color images from the camera, distances to everything within 20m from the laser, and information about its position from the GPS, the inertial measurement unit, and the compass.

The boat autonomously surveyed the lakeshore. As it moved along the perimeter, its pan-tilt camera captured images of the shore while a state–lattice motion planner maintained the boat's distance to it. A 10m distance from the shore was sufficient to avoid most debris in the shallower water, while it maximized the coverage of the shore in each image. Sometimes it was necessary to manually correct the boat's trajectory, e.g., around fishing lines.

4.2 Viewpoint Selection using Reprojection Flow

A large–scale evaluation was performed to determine the benefit of viewpoint selection using Reprojection Flow. Standard SIFT Flow was used as the image alignment method in this analysis. Given an image, I^p, from one survey, it is aligned (low-res) with an image I^q from a different survey, where I^q is determined by the viewpoint selection method. If I^p and I^q

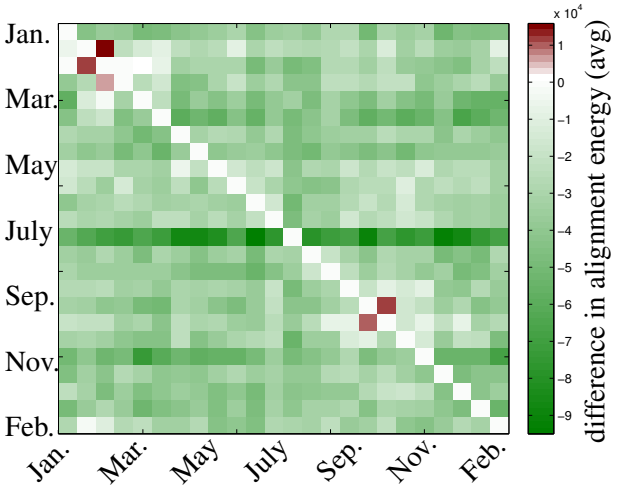

Figure 4: The average improvement in alignment energy of viewpoint selection using Reprojection Flow over viewpoint selection using the closest pose heuristic. Each square represents the result for aligning images from two surveys. A total of 24×23 survey comparisons make up this analysis. Lower is better. (*Best viewed in color.*)

capture the same scene, the alignment energy will be lower than if they are slightly shifted due to the regularization energies. In aligning all the images from one survey to another, both Reprojection Flow and the closest pose heuristic are used to separately determine which I^q is aligned with I^p. The average difference in their resulting alignment energies is computed for each such comparison. A total of 24*23=552 survey comparisons make up this analysis.

Figure 4 shows the result. Out of 552 survey comparisons, 547 reach lower alignment energies on average using Reprojection Flow. Our method chooses images that are, on average, closer to the same scenes. This is also true across seasons. Thus, the task of image registration is simplified because the best images are used for the alignment.

4.3 Consistency of Image Registration Across Seasons

We next tested how well Reprojection Flow can drive image registration across seasons. Although hand–labeled point correspondences are typically used to evaluate alignment quality (as in [24]), finding any correspondences is difficult if the images capture a natural environment and span seasons. Instead, the set of alignments that were used to create the map, specifically their KLT points, were used here. For one such image registration $I^p \rightarrow I^q$, the image I^j from survey j with the highest co-visibility to I^p is found, and then image registration is performed for $I^p \rightarrow I^j$ and $I^q \rightarrow I^j$. This makes a three-cycle between the images, which indicates alignment quality based on how consistent the flows are [24]. The more points closer to zero, the more consistent the image registrations. Using surveys that span $j = 4$ to $6, 9$ to $11, \ldots$, and 34 to 36 week differences, we computed the average alignment consistency of standard SIFT Flow and our method.

Table 1 shows the results. Reprojection Flow with constraints makes image registration significantly more consistent, on average, than SIFT Flow. Whereas SIFT Flow loses consistency across seasons, Reprojection Flow retains about the same level of performance. This is because the structures are better preserved across seasons.

The difference in alignment quality for the two approaches is shown in Fig. 5. The alignment quality was manually labeled for three different surveys across space and time. Large errors in scene structures are apparent in the images marked red. For example, the

Table 1: The average ratio of points within 15 pixels after three-cycle consistency. The values are high given that the analysis involves a three-cycle and 704x480 resolution images.

Number of Weeks Between Surveys	4-6	9-11	14-16	19-21	24-26	29-31	34-36
SIFT Flow	0.27	0.23	0.24	0.23	0.21	0.21	0.21
Reprojection Flow	0.31	0.30	0.31	0.31	0.29	0.30	0.32

Table 2: The average ratio of points within 15 pixels after three-cycle consistency, with added viewpoint variation.

Number of Weeks Between Surveys	4-6	9-11	14-16	19-21	24-26	29-31	34-36
SIFT Flow	0.19	0.15	0.16	0.17	0.14	0.15	0.16
Reprojection Flow	0.26	0.20	0.25	0.25	0.21	0.26	0.25

June images at section 375 show perceptual aliasing with reflections in the water, albeit more so with SIFT Flow. Reprojection Flow lost the scene structure near the shoreline of that image due to the error of the map points there. The more accurate map points at section 325, however, keep the scene structures in place during the alignment. Similarly, the other images marked green better retain the foreground structure.

4.4 Consistency of Image Registration Across Seasons and Viewpoint

Because Reprojection Flow centers the hypothesis space around the correct alignment, it can make image registration more robust to variation in viewpoint. To evaluate this, the same test as in Section 4.3 was performed, except using images that are offset from I^j. An offset of -10 is used, which corresponds to a displacement of the scene in the image by roughly 25-50% of the image width. The result is shown in Table 2 (Figures 1 and 3 show examples of manually added viewpoint variation). The improvement over SIFT Flow is retained.

5 Discussion

The positions of things in an environment provide a basis for image registration across seasons. With Reprojection Flow, a localized observer knows what scene it is viewing. It also knows how the scene corresponds to views of the same place it acquired in the past. Map points reproject onto images to provide this appearance-invariant information. Constraints help to maximize it.

Knowing the rough location of the correct alignment in the image is powerful, for robustness to variation in both appearance and viewpoint. The hypothesis space can be made much tighter if it is centered around the correct alignment. This advantage means that in the cases when few features match well, the best match is still often the correct one. SIFT Flow uses a large hypothesis space to compensate for the fact that it is not centered at the correct one. With the parameters we used, its hypothesis space is large enough to compensate for correspondences up to 50% of the image width. It may have been made larger, but that would also add many more candidate matches. This can lead to significant artifacts in the image due to perceptual aliasing.

The two steps required to create a consistent map are the primary limitations of this work. Map consistency highly depends on the accuracy of the initial image registrations. Also, the optimization does not scale well; here the optimization was performed on a machine with 192 GB of RAM. Both may be improved, however, by utilizing Reprojection Flow in

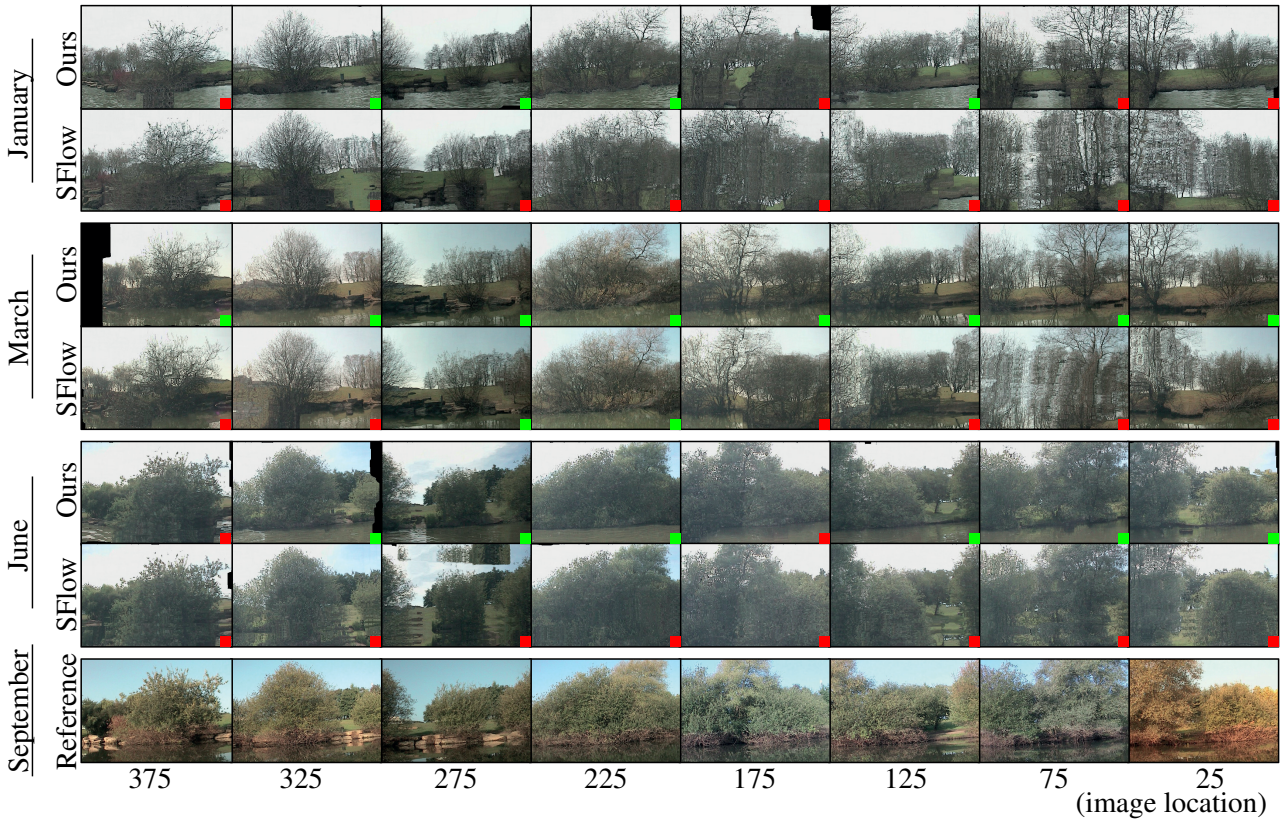

Figure 5: Images from a September survey shown here aligned with images from January, March, and June surveys using our method and SIFT Flow. Green and red flags indicate alignment quality as manually labeled by a human. The foreground mostly aligns well in images marked green whereas significant artifacts appear in images marked red. (*Best viewed in color.*)

an incremental data association and optimization approach. Rather than perform one large optimization, incremental optimization, using e.g. iSAM2 [9], may allow for optimizing one survey at a time. As the map and the poses are incrementally made consistent, Reprojection Flow may be applied to improve the set of initial data associations (analogous to [4]).

References

[1] Nicky S Clayton and John R Krebs. Memory for spatial and object-specific cues in food-storing and non-storing birds. *Journal of Comparative Physiology A*, 174(3): 371–379, 1994.

[2] Selvino R De Kort and Nicola S Clayton. An evolutionary perspective on caching by corvids. *Proceedings of the Royal Society of London B: Biological Sciences*, 273 (1585):417–423, 2006.

[3] Frank Dellaert. Factor Graphs and GTSAM: A Hands-on Introduction. Technical Report GT-RIM-CP&R-2012-002, GT RIM, Sept 2012. URL https://research.cc.gatech.edu/borg/sites/edu.borg/files/downloads/gtsam.pdf.

[4] Jakob Engel, Thomas Schöps, and Daniel Cremers. LSD-SLAM: Large-scale direct monocular SLAM. In *European Conference on Computer Vision (ECCV)*, pages 834–849. Springer, 2014.

[5] David Gonzales and Sunni Brown. Symbiosis: A surprising tale of species cooperation. http://ed.ted.com/lessons/symbiosis-a-surprising-tale-of-species-cooperation, 2012. Accessed: 2016-04-27.

[6] Shane Griffith, Frank Dellaert, and Cédric Pradalier. Robot-Enabled Lakeshore Monitoring Using Visual SLAM and SIFT Flow. In *RSS Workshop on Multi-View Geometry in Robotics*, 2015.

[7] Bumsub Ham, Minsu Cho, Cordelia Schmid, and Jean Ponce. Proposal flow. *arXiv*, 2016. arXiv:1511.05065.

[8] Harry E Hutchins and Ronald M Lanner. The central role of clark's nutcracker in the dispersal and establishment of whitebark pine. *Oecologia*, 55(2):192–201, 1982.

[9] Michael Kaess, Hordur Johannsson, Richard Roberts, Viorela Ila, John J Leonard, and Frank Dellaert. iSAM2: Incremental smoothing and mapping using the Bayes tree. *IJRR*, 31(2):216–235, 2012.

[10] Jaechul Kim, Ce Liu, Fei Sha, and Kristen Grauman. Deformable spatial pyramid matching for fast dense correspondences. In *Computer Vision and Pattern Recognition (CVPR)*, pages 2307–2314. IEEE, 2013.

[11] John R Krebs, Nicky S Clayton, Susan D Healy, Daniel A Cristol, Sanjay N Patel, and Anna R Jolliffe. The ecology of the avian brain: Food-storing memory and the hippocampus. *International Journal of Avian Science*, 138(1):34–46, 1996.

[12] Ce Liu, Jenny Yuen, and Antonio Torralba. SIFT Flow: Dense correspondence across scenes and its applications. *PAMI*, 33(5):978–994, 2011.

[13] Jonathan L Long, Ning Zhang, and Trevor Darrell. Do convnets learn correspondence? In *Advances in Neural Information Processing Systems*, pages 1601–1609, 2014.

[14] David Lowe. Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2):91–110, 2004.

[15] Stephanie Lowry, Niko Sunderhauf, Paul Newman, John J Leonard, David Cox, Peter Corke, and Michael J Milford. Visual place recognition: A survey. *IEEE Transactions on Robotics*, 30(1):1–19, 2016.

[16] Bruce D Lucas and Takeo Kanade. An iterative image registration technique with an application to stereo vision. In *IJCAI*, volume 81, pages 674–679, 1981.

[17] Ricardo Martin-Brualla, David Gallup, and Steven M Seitz. Time-lapse mining from internet photos. *ACM Transactions on Graphics (TOG)*, 34(4):62, 2015.

[18] Michael J Milford and Gordon F Wyeth. SeqSLAM: Visual route-based navigation for sunny summer days and stormy winter nights. In *International Conference on Robotics and Automation (ICRA)*, pages 1643–1649. IEEE, 2012.

[19] Jennifer D Scott. *Clark's nutcracker occurrence, whitebark pine stand health, and cone production in the waterton-glacier international peace park*. PhD thesis, University of Colorado, 2013.

[20] Vladimir Sukhoy, Shane Griffith, and Alexander Stoytchev. Toward imitating object manipulation tasks using sequences of movement dependency graphs. In *RSS Workshop on the State of Imitation Learning*, 2011.

[21] Niko Sunderhauf, Sareh Shirazi, Feras Dayoub, Ben Upcroft, and Michael Milford. On the performance of convnet features for place recognition. In *Intelligent Robots and Systems (IROS), 2015 IEEE/RSJ International Conference on*, pages 4297–4304. IEEE, 2015.

[22] Diana F Tomback. Foraging strategies of clark's nutcracker. *Living Bird*, 16(1977): 123–160, 1978.

[23] Hongsheng Yang, Wen-Yan Lin, and Jiangbo Lu. Daisy Filter Flow: A generalized discrete approach to dense correspondences. In *Computer Vision and Pattern Recognition (CVPR)*, pages 3406–3413. IEEE, 2014.

[24] Tinghui Zhou, Yong Jae Lee, Stella X Yu, and Alexei A Efros. FlowWeb: Joint image set alignment by weaving consistent, pixel-wise correspondences. In *Computer Vision and Pattern Recognition (CVPR)*, pages 1191–1200. IEEE, 2015.

[25] Tinghui Zhou, Philipp Krähenbühl, Mathieu Aubry, Qixing Huang, and Alexei A Efros. Learning dense correspondence via 3d-guided cycle consistency. *arXiv preprint arXiv:1604.05383*, 2016.