

# Automated Sensory Data Alignment for Environmental and Epidermal Change Monitoring

Michael Milford<sup>12</sup>, Jennifer Firn<sup>3</sup>, James Beattie<sup>3</sup>, Adam Jacobson<sup>12</sup>, Edward Pepperell<sup>12</sup>, Eugene Mason<sup>3</sup>, Michael Kimlin<sup>4</sup> and Matthew Dunbabin<sup>2</sup>

<sup>1</sup>ARC Centre of Excellence for Robotic Vision, Queensland University of Technology, <sup>2</sup>School of Electrical Engineering and Computer Science, <sup>3</sup>Earth, Environmental and Biological Sciences School, <sup>4</sup>School of Public Health and Social Work and Institute of Health Biomedical Innovation

Queensland University of Technology

<sup>1</sup>michael.milford@qut.edu.au

## Abstract

In this paper we present research adapting a state of the art condition-invariant robotic place recognition algorithm to the role of automated inter- and intra-image alignment of sensor observations of environmental and skin change over time. The approach involves inverting the typical criteria placed upon navigation algorithms in robotics; we exploit rather than attempt to fix the limited camera viewpoint invariance of such algorithms, showing that approximate viewpoint repetition is realistic in a wide range of environments and medical applications. We demonstrate the algorithms automatically aligning challenging visual data from a range of real-world applications: ecological monitoring of environmental change, aerial observation of natural disasters including flooding, tsunamis and bushfires and tracking wound recovery and sun damage over time and present a prototype active guidance system for enforcing viewpoint repetition. We hope to provide an interesting case study for how traditional research criteria in robotics can be inverted to provide useful outcomes in applied situations.

## 1 Introduction

In recent years a number of innovative new robotic technologies have been developed to deal with a problem of increasing relevance; how to enable robots to persist autonomously over long periods of time in

challenging and always changing environments. One of the key challenges in robot persistence is robotic navigation – enabling a robot to map and navigate its environment even though it may change or appear completely different over a long enough period of time. While conventional SLAM and navigation algorithms have proven highly capable in unchanging environments [Cummins and Newman, 2009], MonoSLAM [Davison, et al., 2007], FrameSLAM [Konolige and Agrawal, 2008], V-GPS [Burschka and Hager, 2004], Mini-SLAM [Andreasson, et al., 2007], [Andreasson, et al., 2008, Paz, et al., 2008, Royer, et al., 2005, Zhang and Kleeman, 2009, Konolige, et al., 2008, Milford and Wyeth, 2008], they have not readily been adapted to operate in changing environments.

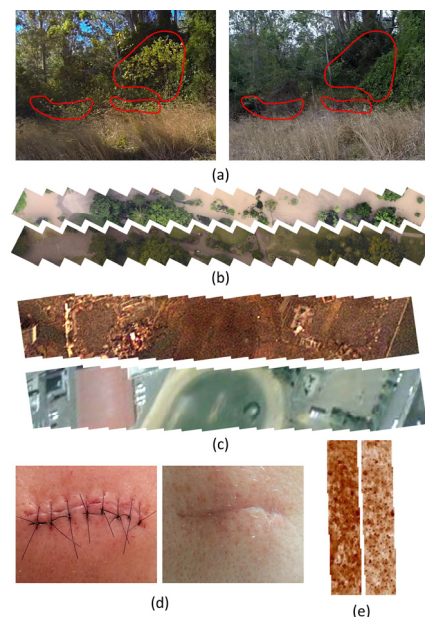


Figure 1: Examples of automatically aligned imagery output by our system both at an inter-image and intra-image scale, for (a) an ecological context with foliage removal, (b) simulated low altitude flood disaster monitoring, (c) pre- and post-tsunami damage, (d) wound healing and (e) sun damage to skin under a UV camera.

The authors are with the School of Electrical Engineering and Computer Science, Queensland University of Technology, Brisbane, Australia. <http://www.roboticvision.org/>. This work was supported by an Australian Research Council DECRA grant DE120100995 and Microsoft Faculty Fellowship both awarded to MM, and a QUT School of Electrical Engineering and Computer Science Pilot Project Grant awarded to MM, JF and MD.

Instead, a range of new algorithms have been developed that address the problem of navigation in changing environments in a number of ways including using learning techniques [Johns and Yang, 2013, Biederman, 1988, Neubert, et al., 2013, Lowry, et al., 2014], condition-invariant ‘features’ [Milford, et al., 2014] or integrating information over temporal sequences of images SeqSLAM [Milford, 2013, Milford and Wyeth, 2012, Pepperell, et al., 2014]. These new algorithms generally solve the problem of changing environment appearance but typically have poor viewpoint invariance, which limits their applicability in robotic applications where similar viewpoint repetition is not guaranteed, such as in aerial or open field applications.

In this paper we explore the utility of adapting these condition-invariant robotic algorithms to provide automated data alignment in a range of non-robotics scenarios (Figure 1). Our key insight is that in many usage scenarios, viewpoint invariance becomes a minor problem, because of the use of a “motivated” human ecologist or patient in the loop; in the case of performing transects of ecological environments, the human attempts or is guided to follow approximately the same path; for skin monitoring the patient takes regular photos of approximately the same region of skin. In all these scenarios, dealing with extreme appearance change rather than viewpoint change becomes the dominant concern.

We present a number of novel contributions:

- A modified sequence-based algorithm (inspired by SeqSLAM) more suited to producing correct individual frame correspondences across spatio-temporal sequences of imagery obtained from a camera moving through or over an environment.
- A two-dimensional intra-image sequence-based algorithm for robustly finding feature correspondences between two images despite significant appearance change.
- A prototype active guidance system leveraging the condition-invariant properties of sequence-based localization techniques to repeatedly guide ecologists along the same transect through an always-changing environment.

The primary contribution of this work is the development of inter- and intra-image alignment algorithms that function despite significant appearance change; we do not yet include automatic image analysis algorithms (for example extracting foliage density, wound healing status, or percentage of urban destruction); we hope to provide a widely applicable set of algorithms that enable future work in these areas.

The paper proceeds as follows. In Section 2 we provide a brief background of the current state of affairs in the application scenarios we target in this work including ecological monitoring, disaster observation and skin damage monitoring. In Section 3 we describe the implementation of the three primary contributions; the inter- and intra-image sequence-based alignment techniques and the prototype active guidance algorithm. Section 4 describes the various

testing scenarios including details of the datasets and sensory acquisition rigs, while Section 5 presents the results. Finally we conclude in Section 6 with a discussion of the most profitable areas of future work building on the results and capabilities presented here.

## 2 Background

In this section we briefly review the motivation behind and current best practice techniques for change monitoring in ecological, environmental and skin monitoring applications, in order to provide context for the possible improvements that automated data alignment can offer.

### 2.1 Ecological Monitoring

Because of the impacts of climate change and increased anthropogenic pressure on biodiversity and ecosystem services, never before in history has there been such a need for cost-effective, rapid but accurate monitoring techniques for assessing changes to ecosystems [Parmesan, 2006]. Ecologists use a range of monitoring techniques [Spellerberg, 2005] to study how organisms are impacted by interactions with other organisms, and, changing disturbance regimes, resource availability and climatic conditions.



Figure 2: Photo of traditional Daubenmire technique for assessing plant species richness and cover in a grassland community. By marking the corners of a 1m<sup>2</sup> quadrat the same area of grassland can be sampled each year. (Photo taken by Joslin Moore at the Bogon High Plains Nutrient Network site, <http://www.nutnet.umn.edu/>)

Current ecological monitoring techniques for quantifying change over time or space are laborious and many have remained largely unchanged over many decades. For example, in grassland ecology, the most common method for quantifying plant species richness and cover is the Daubenmire canopy coverage method of vegetation analysis that was published in 1959 [Daubenmire, 1959]. This method involves repeatedly visiting a quadrat of generally 1 m<sup>2</sup> at the beginning and/or end of a growing season to track changes in species composition overtime (Figure 2). Other generic methods for estimating plant cover in grasslands that are often adapted for forests include point intercept methods, line intercepts and cover-class estimates (of which the Daubenmire method is an example) [Braude and Low, 2010]. All of these methods are time consuming, require expert skill gained only through experience and present different types and degrees of bias and error [Floyd and Anderson, 1987].

Recent advances in technology including data loggers, automatic radiotelemetry, acoustic monitoring and autonomous and remote photography offer much

promise with respect to developing faster and more accurate change monitoring techniques. In particular there is much potential to move beyond a 1 square metre quadrat process to repeatedly logging data over long transects through an environment, resulting in orders of magnitude more data. The challenge of automatically aligning this data over time is similar to the challenges faced in localizing navigating robots; GPS degrades under thick forest canopies and foliage structure and appearance can change significantly over time, limiting the practicality of 3D structure alignment-based techniques or vision-based alignment techniques.

## 2.2 Robotic Environmental Monitoring

Although our initial focus here is on automated techniques for ecologists working in the field, a brief review of robotics-based techniques is also useful. The primary challenges for robotic monitoring of change are (1) perception, (2) persistence, and (3) repeatability. While manual survey can provide the first and third in favourable weather conditions, robotics and associated sensor technology provide the potential for a complete solution to all three. A survey of the state-of-the-art in robotic environmental monitoring systems [Dunbabin and Marques, 2012] identified the same challenge for robotic monitoring as for larger scale manual surveys; an automated method for alignment and subsequent analysis or classification of very large quantities of data.

Vision has been the primary sensor for detecting change in natural environments. Processing these large data sets has been accomplished using satellite remote sensing techniques with in-situ validation [Roelfsema, et al., 2013] and machine learning [Friedman, et al., 2012]. However, challenges still exist in matching imagery under significant change and lighting/visibility conditions, and when GPS or other localisation information is not readily available.

## 2.3 Low Altitude Disaster Monitoring

In Australia and around the world natural disasters including bushfires, flooding and tsunamis have a significant impact on the lives of people living in affected areas. Over the past decade the capability of low cost small quad-rotor and fixed wing aircraft has significantly improved, prompting investigations into their utility for monitoring and assessing disaster zones from the air [Eugster and Nebiker, 2008]. One of the standard contributions is to automate the mapping of live aerial imagery onto reference aerial maps in order to provide up to date information to disaster response personnel. Typical approaches involve a combination of on-board positioning information and image-based alignment, with image-based alignment playing a larger role on small, low cost craft where very accurate GPS is not feasible for cost and bulk considerations, or because local infrastructure providing GPS corrections is not available or destroyed by the disaster.

## 2.4 Skin Damage Monitoring

Skin cancer accounts for approximately 40% of all cancer cases globally [Cakir, et al., 2012], and resulted in approximately 55,000 deaths globally in 2012

[Stewart and Wild, 2014]. Skin cancer is a particular concern in Australia where incidence rates are one of the highest in the world, with an annual fatality rate of approximately 2000. The single most significant factor in treating skin cancer is early detection and identification. While regular detection in a controlled clinical environment is available, such processes take time and require a special visit to a centre (Figure 3). The success of in-home monitoring depends on motivation and ability of the inspector to detect changes in skin, a process which is not trivial to perform especially when the interval between inspections is significant. Current technological solutions require expensive (\$500) optic accessories that attach to a smartphone and require the user to have already identified the region of interest [Stewart and Wild, 2014]. The development of a robust, in-home skin cancer detection system using existing consumer hardware remains elusive, for many reasons including the inability to easily align skin photos over time taken in uncontrolled conditions with changing skin appearance and inconsistent camera viewpoint.

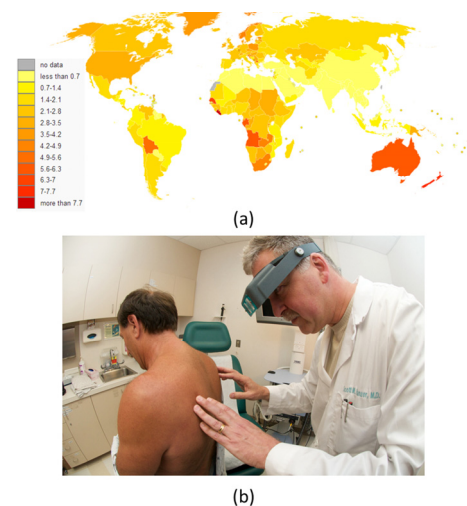


Figure 3: (a) Worldwide age-standardized death from melanoma and other skin cancers per 100,000 inhabitants in 2004 (image source: [http://en.wikipedia.org/wiki/Skin\\_cancer](http://en.wikipedia.org/wiki/Skin_cancer)) and (b) a typical skin cancer clinic examination process (picture copyright Christiana Care on Flickr).

## 2.5 Active Guidance

Traditional robotic teach and repeat systems utilise a human guide to learn a path through an environment. Once this path has been taught, the robot can repeat the journey without human intervention. Many teach and repeat systems utilise GPS [Hogg, et al., 2002], lasers [Krüsi, et al., 2014, McManus, et al., 2013, Sprunk, et al., 2013], panoramic cameras [Zhang and Kleeman, 2009] and stereo cameras [Furgale and Barfoot, 2010] to enable robots to repeat a previously taught path. These robot control systems all perform localization using onboard sensors to determine the error between the current location of the robot and the desired path and attempt to minimise the error to keep the robot on the same path.

The usage of a human operator, as opposed to a robot platform, allows a different approach to be taken. A motivated human can leverage situational awareness and human dynamics to traverse diverse and difficult



terrain that a robot may not be capable of traversing, while attempting to follow a path they have never previously traversed. As will be shown here, humans are capable of repeating a path using a topological map, removing the constraints imposed by attempting to create a metrically accurate map.

### 3 Approach

This section describes the new methods for an environmental monitoring-specific implementation of SeqSLAM for aligning sensor data from multiple transects of an environment, a new intra-frame method based on SeqSLAM for aligning two images to each other despite significant appearance change, and the active guidance methods building on SeqSLAM that guide a user along a previously traversed route.

#### 3.1 Inter-Frame Best Match SeqSLAM

The SeqSLAM algorithm performs place recognition by integrating single frame place match hypotheses over many frames in order to produce an overall sequence matching hypothesis [Milford and Wyeth, 2012]. Such an approach must be modified when attempting to align data for tasks such as environmental monitoring, because the correctness of individual frame matches becomes more important. For example, an ecologist may need precisely aligned camera snapshots in order to identify and analyse foliage change in a specific sub-region of the image.

To address this requirement, we developed a new ‘best match’ implementation of SeqSLAM. The approach identifies the top ranked frame matches and then searches for coherent sequences of these top ranked matches.

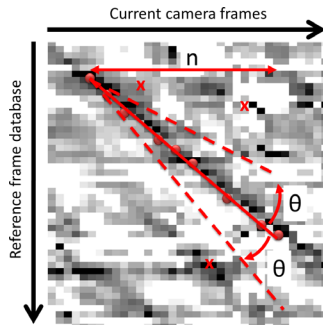


Figure 4: Best match SeqSLAM, which identifies semi-coherent sequences of top ranked frame match hypotheses (red dots).

Figure 4 illustrates the process. After constructing a difference matrix comparing recent camera frames to all frames stored in the reference database, the strongest match candidates for each recent frame are identified (red dots). Starting at each best match candidate, a search forward in time is performed, identifying which best match candidates fall within an allowable match zone defined by angle  $\theta$ . If the fraction of best matches over a search length of  $n$  frames exceeds a threshold  $\delta$ , the matches falling within the zone are validated as correct matches. For all experiments in this paper, parameters were set as in Table I.

#### 3.2 Intra-Frame SeqSLAM and Alignment

The inter-frame alignment matches camera frames that were obtained in approximately the same spatial location, but the matched frames are often not aligned precisely. We perform an intra-frame SeqSLAM search in two-dimensions to find ‘feature’ correspondences between images matched by the inter-frame step, in order to align the image match pairs (Figure 5). Images are then aligned by performing an affine transformation based on the feature correspondences. To perform the fine alignment, we make the following assumptions:

- The matched frames are taken from approximately the same (x, y, z) location in space
- Variations in pitch, roll and yaw are relatively small

As the results will show, these assumptions are valid in a wide range of applications.

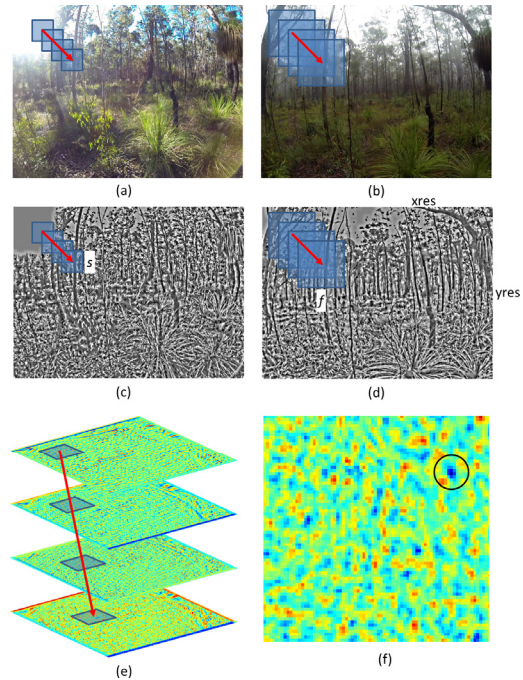


Figure 5: Intra-frame SeqSLAM. After the (a-b) raw images are patch-normalized,  $q$  patches evenly spaced along a matching trajectory of length  $z$  in the (c) new image are correlated across the (d) entire reference image to produce a (e) sequence of difference matrices. These difference matrices are aligned based on the search trajectory and then cropped to a (f) sub-region difference matrix. The minimum difference score within this matrix is used as the matching location.

#### 3.3 SeqSLAM Active Guidance

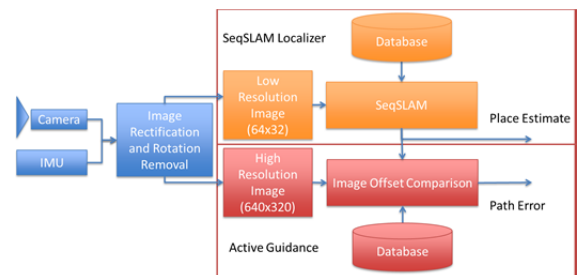


Figure 6: SeqSLAM Active Guidance System Schematic. The SeqSLAM localiser determines a user’s location along a path and the Active Guidance system determines the user’s deviation from that path in image space, allowing the user to follow the path with minimal metric error.

The SeqSLAM Active Guidance system has three major components; image acquisition and pre-processing, SeqSLAM Localisation and the Active Guidance module (Figure 6).

### Image Acquisition and Pre-Processing

The image input for this system is obtained from a camera and IMU rig that un-distorts the camera image and removes the roll of the camera (Figure 7). Localization along a previously traversed route is performed as in the original SeqSLAM paper [Milford and Wyeth, 2012]. Using a constant velocity assumption, an approximate path is calculated using the reference transect data and a simple patch-based visual odometry technique [Milford, et al., 2011].

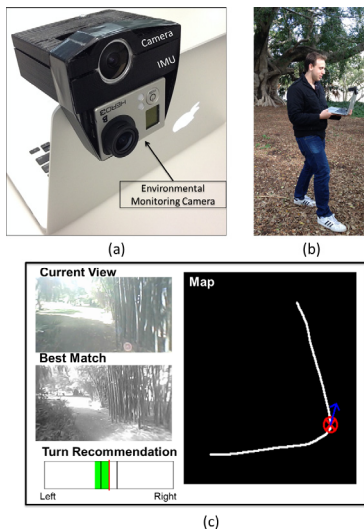


Figure 7: SeqSLAM Active Guidance Setup. (a) Camera and IMU for active guidance and high resolution Go-Pro for capturing environmental change data. (b) User in the field following the guidance instructions in the (c) user interface.

The Active Guidance module guides the user using two mechanisms. The first mechanism takes the current place match hypothesis, looks ahead approximately 2 seconds in time and compares the spatial shift of the predicted reference image to the current camera image. Any spatial discrepancy in the horizontal image direction is communicated to the user via the interface (Figure 7, bottom left). The second guidance mechanism involves localizing the user on the reference path to provide them with the ability to see what upcoming turns, much like a navigator provides future turn directions to a rally car driver.

## 4 Experimental Setup

This section describes the experimental environment, dataset acquisition and pre-processing, ground truth creation and key parameter values.

### 4.1 Datasets

#### Ecological Datasets

Datasets were obtained by ecologists performing repeated transects across two active monitoring environments; Kroombit tops, a national park in Central Queensland, Australia and the Samford Ecological Reserve, located in Brisbane.

### Wound Healing

After facial surgery to remove a lipoma, the author captured images of the wound zone daily over a period of two weeks using a Samsung Galaxy S3 phone camera (an approximately 2 year old, 2012 era smartphone). Images were captured under a range of different lighting conditions and with only approximate viewpoint repeatability (viewing distance, camera pitch, roll and yaw all varying somewhat from image to image) (Figure 8). The equipment and capture conditions were consistent with what might be expected of a normal user in a domestic environment using standard consumer devices rather than specialist equipment.

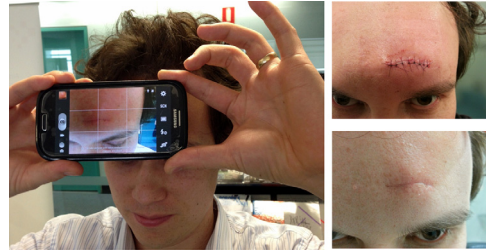


Figure 8: Wound images were taken using a Samsung Galaxy S3 phone camera daily for a period of two weeks. Camera distance, roll, pitch and yaw were only approximately repeated, mimicking an in-home monitoring scenario.

### Sun Damage

Sample before and after images were sourced from the internet and anonymized, with references given to the source.

### Intra-frame SURF Comparison

To provide a qualitative comparison to a traditional feature-based technique applied to the intra-frame alignment problem, we tuned a SURF-based feature matcher to produce a comparable number of feature correspondences (where possible) across matching image pairs. Feature correspondences found in this way were then run through the same geometric transformation calculation to produce the aligned images.

### Aerial-based Datasets

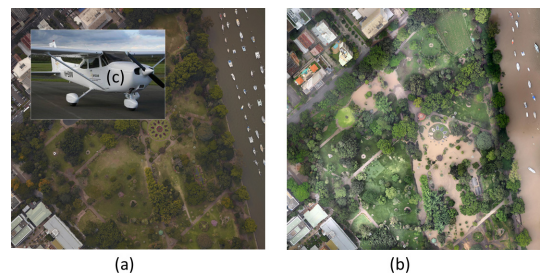


Figure 9: One of the aerial flooding imagery datasets was developed by combining aerial imagery from a downward facing camera mounted on a Cessna aircraft overflying the area with Nearthmaps data captured during the Brisbane floods (copyright Nearthmaps 2011). Data courtesy of the Australian Research Centre for Aerospace Automation (ARCAA).

Aerial imagery was sourced from a variety of sources including Nearthmaps, Google Maps and from a downward facing high resolution camera mounted on a Cessna 172R aircraft (Figure 9). Images were approximately standardized to a northward facing



orientation and standard altitude by rotation and scaling, which could be achieved in an online manner by using a compass and altimeter. In the experiments presented here, the dataset obtained from the Cessna was compared against Nearmaps imagery of the same area during the 2011 Brisbane floods. Qualitative results are provided for before and after aerial imagery of flooding, tsunami damage, bushfires and day-night cycles.

### Active Guidance Experimental Procedure

Each of the two transects was first learnt by one person performing a mapping run along a path, logging data onto a carried laptop to form the reference transect. Testing was performed by recruiting a second person who had *no knowledge of the path* except for the starting location and orientation. Testing was performed across a range of forested and open areas, with the two learnt transects shown alongside the results.

TABLE I  
PARAMETER LIST

| Parameter    | Value | Description  |
|--------------|-------|--|
| $\theta$     | 26°   | Search tolerance   |
| $n$          | 10    | Search range in frames   |
| $\delta$     | 0.55  | Fraction of best matches required within search range for overall match validation |
| $s$          | 40×40 | Width of patch comparison area in pixels   |
| $xres, yres$ | 400   | Length of maximum dimension of patch normalized image in pixels                    |
| $q$          | 20    | Number of patches used in the intra-frame sequence search                          |
| $z$          | 50    | Length of sequence search within a frame in pixels                                 |
| $f$          | 0.1   | Fractional range of image over which patch correspondences are sought              |

## 5 Results

We present frame correspondence graphs and sample aligned images for the majority of datasets, and include comparison to images aligned using conventional SURF-based feature correspondences and the unaligned images.

### 5.1 Automated Environmental Change Alignment

#### Kroombit Tops

Figure 10 shows the frame correspondences for the Kroombit Tops transects, with red crosses showing the raw individual frame matches and blue dots showing the validated frame matches. More than half of the frames are correctly matched to the reference transect with no errors. An examination of the missing matches reveals that the ecologist's path diverged from the reference path, motivating the development of the prototype guidance system.

Figure 11 shows the feature correspondences across a pair of images matched by the inter-frame alignment process. Figure 11a shows the feature matches output using a conventional SURF-based approach, while Figure 11b shows those output by the intra-frame sequence-based method. Finally, Figure 12 shows corresponding image regions from the (a) unaligned images and images aligned using the (b) SURF-aligned and (c) intra-frame sequence-based

approach. The image alignment is better across all 16 image regions using the sequence-based method than the SURF-based method.

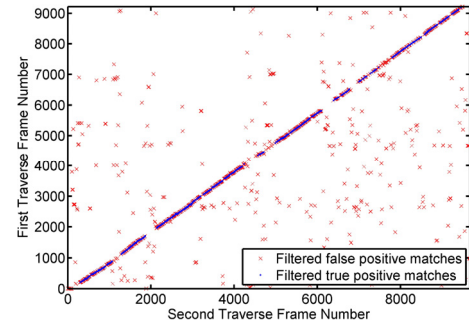


Figure 10: Frame correspondences for Kroombit Tops. Maximum recall at 100% precision was 57.23%, with a maximum coverage gap of 300 frames, or approximately 7 metres.

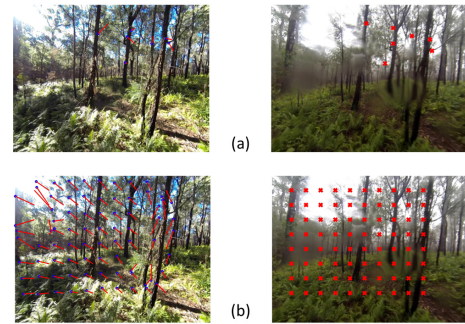


Figure 11: Feature correspondences found using SURF and using (b) intra-frame SeqSLAM. Right frames shows source feature locations and left frames show relative movement of features from their location in the right frame to their new location in the left frame (shown by a blue circle).

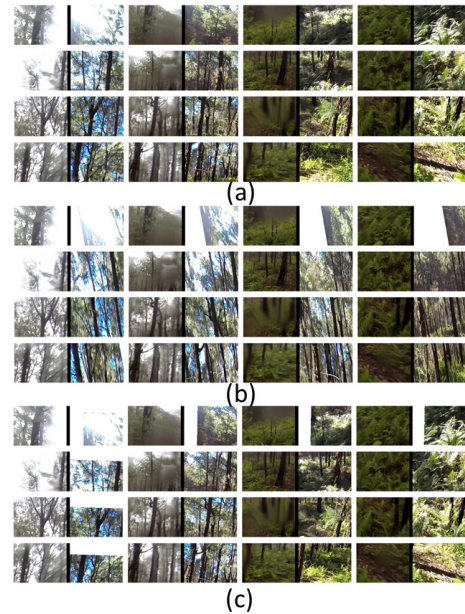


Figure 12: (a) Sample inter-frame matched frames and (b) frames from intra-frame alignment with SURF and (c) frames resulting from alignment with intra-frame SeqSLAM.

#### Samford Ecological Reserve

Figures 13-15 show the frame correspondence graph, feature correspondences and aligned image regions for the Samford Ecological Reserve dataset. More than 40% of the frames are correctly aligned using the inter-frame alignment technique, despite

significant appearance change due to lighting and weather conditions. Close examination of the aligned image pairs show that the sequence-based alignment is superior to the SURF-based technique. Finally, Figure 15c shows an example of manually identified foliage change using the aligned images, where a significant amount of Lantana plant has been removed from the environment.

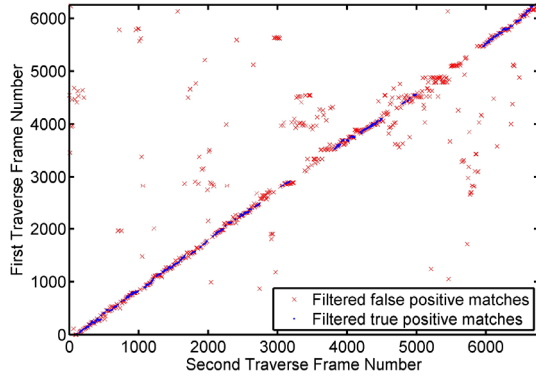


Figure 13: Frame correspondences for the Samford Ecological Reserve transects. Maximum recall at 100% precision was 42.5%, with a maximum coverage gap of 970 frames, or approximately 22 metres.

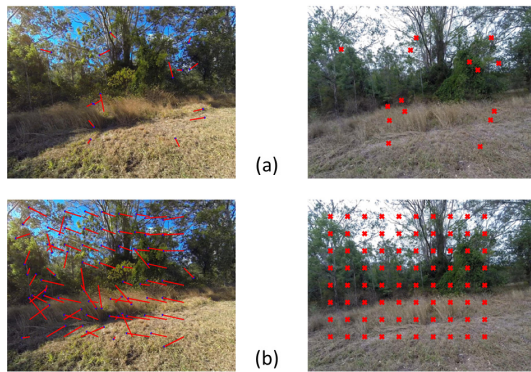


Figure 14: Feature correspondences found using (a) SURF and using (b) intra-frame SeqSLAM.

## 5.2 Skin Damage Monitoring

We present alignment results for imagery of wound healing and sun damage over time using the intra-frame sequence-based method with comparison to a SURF-based method.

### Wound Healing

Figure 16 shows the feature correspondences using (a) SURF features and (b) using the sequence-based method. While challenging, the sequence-based method is able to pick up enough coherent feature matches to produce the alignment shown in Figure 17c, with an average alignment error approximately five times smaller than that achieved using SURF features.

### Sun Damage

Figure 18 shows facial scans under a UV camera revealing pigmentation due to sun damage over a period of time. Two skin patches aligned using the intra-frame sequence alignment algorithm are shown. Note that no local matching assumption is made in this case; matches can be found between any image locations. A user specifies a desired patch of skin using a mouse-driven interface, and the sequence-based

method finds the corresponding region in the other image. The resulting finely aligned images can be examined for change either automatically or by a human expert, reducing the chance of a human making an incorrect correspondence. Figure 19 shows the same process but for skin under a normal camera, revealing changes such as a pimple and slight skin damage.

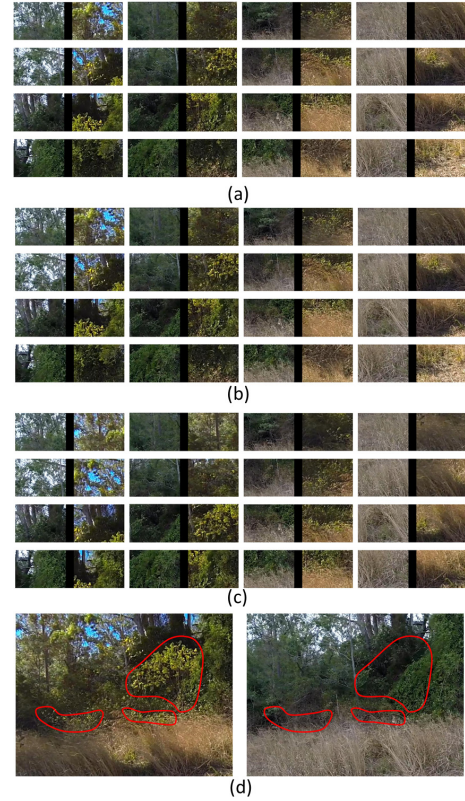


Figure 15: (a) Sample inter-frame matched frames and (b) frames from intra-frame alignment with SURF and (c) frames resulting from alignment with intra-frame SeqSLAM. (d) Ecologists are easily able to identify foliage change manually using the automatically aligned images; in future work we aim to automate this change detection.

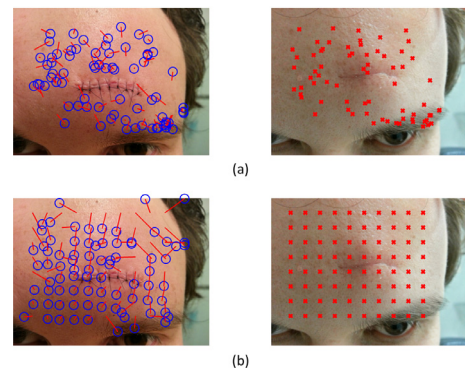


Figure 16: Feature correspondences found using (a) SURF and using (b) intra-frame SeqSLAM.

## 5.3 Aerial Image Alignment across Change

In this section of the results, we present examples (Figures 20-23) of automatically aligned before and after aerial imagery of bushfires, flooding, tsunami destruction and day-night cycles. These results are intended simply to provide a qualitative indication of the capabilities of a sequence-based alignment algorithm, especially when approximate planarity requirements are met. We re-iterate here that such a



technique would complement coarse GPS localization available on small, low altitude quad rotor and fixed wing craft, by providing an image-driven method for finer alignment of disaster imagery to reference imagery. All images represent crudely simulated low altitude traverses of the environment. Matching was performed using only the imagery presented.

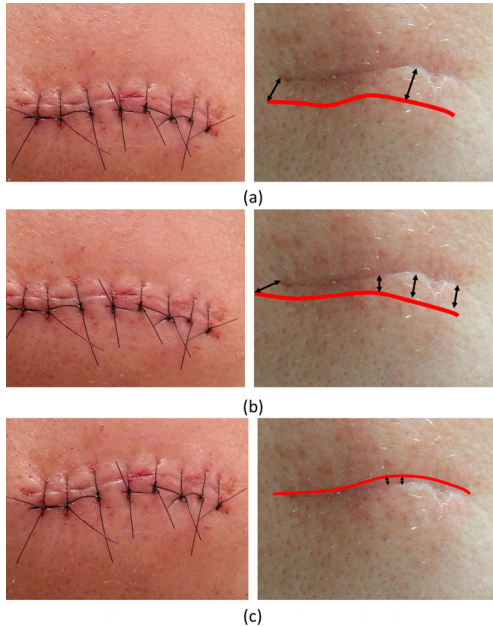


Figure 17: (a) Sample inter-frame unaligned frames and (b) frames from intra-frame alignment with SURF and (c) frames resulting from alignment with intra-frame SeqSLAM, with zoom insets showing alignment. The approximate average alignment error reduced from 4.0 mm for the unaligned images to 2.1 mm for the SURF-aligned images and 0.4 mm for the SeqSLAM aligned images.

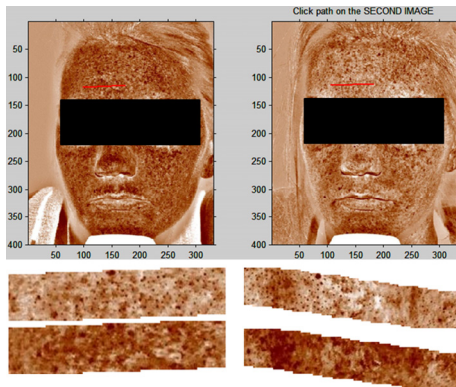


Figure 18: Aligned skin imagery. Image from <http://www.dbreath.com/medical-skin-care-knoxville-tennessee>.



Figure 19: Aligned skin imagery. Image from <http://mamalotion.com/before-and-after.html>.

#### 5.4 Active Guidance for Route Repetition

Finally, we present preliminary results showing the active guidance system operating over two transects in an open and forested environment (Figure 24). The

blue solid line shows GPS tracking during the test run, the orange long dashed line shows the path of the first transect, and the red dotted line shows the approximate path walked by the naïve tester using the active guidance system. Image insets show frames from the initial and guided transects, showing that the approximate same route was traversed, despite the tester having no knowledge of the path taken through the environment. Finally, we present a confusion matrix for the longer transect which shows a matching diagonal, indicating the route was repeated successfully.

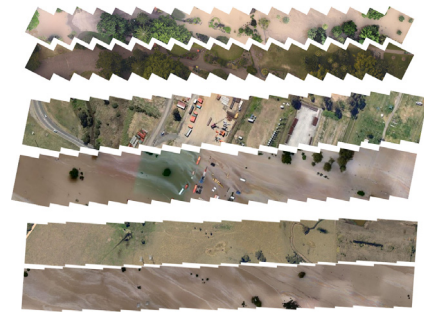


Figure 20: Sample aligned before and after aerial imagery of flooded environments. First image courtesy of ARCAA, other images copyright Nearmaps.

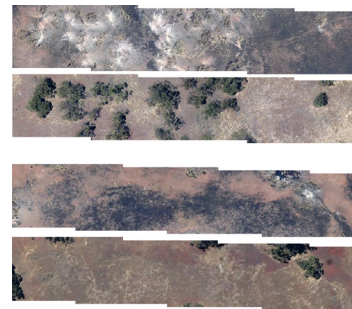


Figure 21: Sample aligned before and after aerial imagery of bushfire affected environments. Imagery copyright Nearmaps.

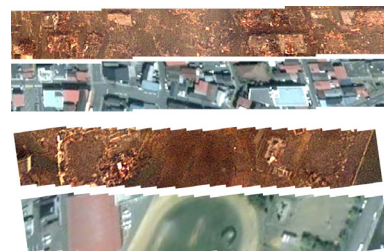


Figure 22: Sample aligned before and after aerial imagery of a tsunami damaged area. Imagery copyright Google, DigitalGlobe.

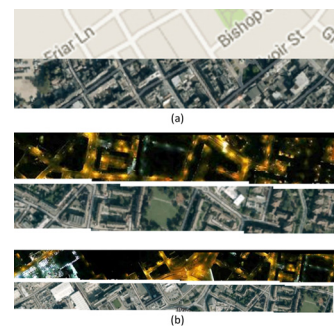


Figure 23: Sample aligned (a) schematic-day and (b) night-day-night aerial imagery. Imagery copyright Google, DigitalGlobe.



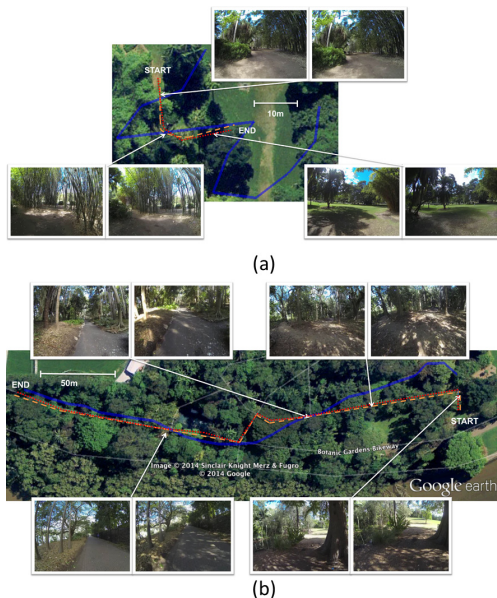


Figure 24: Reference (orange dashed line) and active guidance (red dotted line) routes through the environment, with GPS track shown in blue.

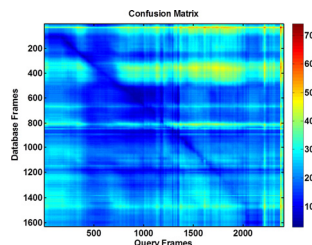


Figure 25: Confusion matrix matching imagery from the actively guided longer transect back to the reference transect.

## 5.5 Compute

The algorithms are currently primarily implemented in Matlab but have similar computational limitations to the algorithms presented in [Milford, et al., 2014, Milford, 2013] – real-time operation on a standard PC being feasible on datasets of up to many tens of thousands of images (equivalent to several kilometre-long transects on foot), with potential for further improvements through the use of GPU-based computing.

## 6 Discussion and Future Work

This paper has explored the potential for adapting a condition-invariant sequence-based place recognition algorithm originally developed in a robotics context for automatic alignment of sensory data obtained repeatedly over time. The premise of the paper has been that there are many potential applications in environmental monitoring and skin change monitoring where the primary weakness of the algorithm – viewpoint variance – is not an issue. We presented two new sequence-based algorithms for finding frame correspondences and subsequently finding feature correspondences between matched frames in order to align them more closely. Results presented on a wide range of datasets suggest that the approach is quite feasible. We also presented a prototype active guidance algorithm that can facilitate ecologists repeating the same transect through an environment over time.

There are several areas of future work. First and foremost, we are working with computer vision researchers, ecologists and skin cancer researchers to develop automated analysis algorithms; given the presented approach produces correctly aligned observation data over time, what automated analysis can be performed? Current ideas undergoing exploration include using aligned frames to learn a model of environmental appearance change, so as to be able to produce “normalized” images that facilitate analysis. We are also integrating other sensing modalities including lasers – once the data streams from multiple traverses have been aligned, the laser can provide 3D structural change data especially under conditions where 3D visual reconstruction may fail such as in light fog or poor lighting. In the ecological domain, we are examining ways to automatically quantify and localize foliage change, (such as that visible in Figure 15), to be validated using traditional methods such as drying and weighing the removed vegetation. For disaster monitoring, we are currently working on acquiring very low altitude datasets from aerial craft of the environment before, during and after an incident such as a bushfire. Finally, we are working towards the publication of an open source codebase providing researchers with easy-to-use access to the algorithms, an especially important requirement given the wide range of research disciplines involved.

## References

- [M. Cummins and P. Newman, 2009] M. Cummins and P. Newman, "Highly scalable appearance-only SLAM - FAB-MAP 2.0," in *Robotics: Science and Systems*, Seattle, United States, 2009.
- [A. J. Davison, I. D. Reid, N. D. Molton, and O. Stasse, 2007] A. J. Davison, I. D. Reid, N. D. Molton, and O. Stasse, "MonoSLAM: Real-Time Single Camera SLAM," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, pp. 1052-1067, 2007.
- [K. Konolige and M. Agrawal, 2008] K. Konolige and M. Agrawal, "FrameSLAM: From Bundle Adjustment to Real-Time Visual Mapping," *IEEE Transactions on Robotics*, vol. 24, pp. 1066-1077, 2008.
- [D. Burschka and G. D. Hager, 2004] D. Burschka and G. D. Hager, "V-GPS (SLAM): Vision-based inertial system for mobile robots," 2004, pp. 409-415 Vol. 1.
- [H. Andreasson, T. Duckett, and A. Lilienthal, 2007] H. Andreasson, T. Duckett, and A. Lilienthal, "Mini-SLAM: Minimalistic Visual SLAM in Large-Scale Environments Based on a New Interpretation of Image Similarity," in *International Conference on Robotics and Automation*, Rome, Italy, 2007, pp. 4096-4101.
- [H. Andreasson, T. Duckett, and A. Lilienthal, 2008] H. Andreasson, T. Duckett, and A. Lilienthal, "A Minimalistic Approach to Appearance-Based Visual SLAM," *IEEE Transactions on Robotics*, vol. 24, pp. 1-11, 2008.
- [L. M. Paz, P. Pinies, J. D. Tardos, and J. Neira, 2008] L. M. Paz, P. Pinies, J. D. Tardos, and J. Neira, "Large-Scale 6-DOF SLAM With Stereo-in-Hand," *IEEE Transactions on Robotics*, vol. 24, pp. 946-957, 2008.
- [E. Royer, J. Bom, M. Dhome, B. Thuilot, M. Lhuillier, and F. Marmoiton, 2005] E. Royer, J. Bom, M. Dhome, B. Thuilot, M. Lhuillier, and F. Marmoiton, "Outdoor autonomous navigation using monocular vision," in *IEEE International*

- Conference on Intelligent Robots and Systems*, 2005, pp. 1253-1258.
- [A. M. Zhang and L. Kleeman, 2009] A. M. Zhang and L. Kleeman, "Robust Appearance Based Visual Route Following for Navigation in Large-scale Outdoor Environments," *The International Journal of Robotics Research*, vol. 28, pp. 331-356, 2009.
- [K. Konolige, M. Agrawal, R. Bolles, C. Cowan, M. Fischler, and B. Gerkey, 2008] K. Konolige, M. Agrawal, R. Bolles, C. Cowan, M. Fischler, and B. Gerkey, "Outdoor mapping and navigation using stereo vision," 2008, pp. 179-190.
- [M. Milford and G. Wyeth, 2008] M. Milford and G. Wyeth, "Mapping a Suburb with a Single Camera using a Biologically Inspired SLAM System," *IEEE Transactions on Robotics*, vol. 24, pp. 1038-1053, 2008.
- [E. Johns and G. Z. Yang, 2013] E. Johns and G. Z. Yang, "Feature Co-occurrence Maps: Appearance-based Localisation Throughout the Day," in *International Conference on Robotics and Automation*, Karlsruhe, Germany, 2013.
- [I. Biederman, 1988] I. Biederman, "Aspects and extension of a theory of human image understanding," *Computational processes in human vision: An interdisciplinary perspective*, 1988.
- [P. Neubert, N. Sünderhauf, and P. Protzel, 2013] P. Neubert, N. Sünderhauf, and P. Protzel, "Appearance change prediction for long-term navigation across seasons," presented at the European Conference on Mobile Robots, Barcelona, Spain, 2013.
- [S. Lowry, M. Milford, and G. Wyeth, 2014] S. Lowry, M. Milford, and G. Wyeth, "Transforming Morning to Afternoon using Linear Regression Techniques," in *IEEE International Conference on Robotics and Automation*, Hong Kong, China, 2014.
- [M. Milford, W. Scheirer, E. Vig, A. Glover, O. Baumann, J. Mattingley, and D. Cox, 2014] M. Milford, W. Scheirer, E. Vig, A. Glover, O. Baumann, J. Mattingley, and D. Cox, "Condition-Invariant, Top-Down Visual Place Recognition," in *IEEE International Conference on Robotics and Automation*, Hong Kong, China, 2014.
- [M. Milford, 2013] M. Milford, "Vision-based place recognition: how low can you go?," *International Journal of Robotics Research*, vol. 32, pp. 766-789, 2013.
- [M. Milford and G. Wyeth, 2012] M. Milford and G. Wyeth, "SeqSLAM: Visual Route-Based Navigation for Sunny Summer Days and Stormy Winter Nights," in *IEEE International Conference on Robotics and Automation*, St Paul, United States, 2012.
- [E. Pepperell, P. Corke, and M. Milford, 2014] E. Pepperell, P. Corke, and M. Milford, "All-environment visual place recognition with SMART," in *IEEE International Conference on Robotics and Automation*, Hong Kong, China, 2014, pp. 1612-1618.
- [C. Parmesan, 2006] C. Parmesan, "Ecological and evolutionary responses to recent climate change," *Annual Review of Ecology, Evolution, and Systematics*, pp. 637-669, 2006.
- [I. F. Spellerberg, 2005] I. F. Spellerberg, *Monitoring ecological change*: Cambridge University Press, 2005.
- [R. Daubenmire, 1959] R. Daubenmire, "A Canopy-Coverage Method of Vegetational Analysis," *Northwest Science*, vol. 50, p. 431, 1959.
- [S. Braude and B. S. Low, 2010] S. Braude and B. S. Low, *An introduction to methods and models in ecology, evolution, and conservation biology*: Princeton University Press, 2010.
- [D. A. Floyd and J. E. Anderson, 1987] D. A. Floyd and J. E. Anderson, "A comparison of three methods for estimating plant cover," *The Journal of Ecology*, pp. 221-228, 1987.
- [M. Dunbabin and L. Marques, 2012] M. Dunbabin and L. Marques, "Robots for environmental monitoring: Significant advancements and applications," *Robotics & Automation Magazine, IEEE*, vol. 19, pp. 24-39, 2012.
- [C. Roelfsema, E. M. Kovacs, M. I. Saunders, S. Phinn, M. Lyons, and P. Maxwell, 2013] C. Roelfsema, E. M. Kovacs, M. I. Saunders, S. Phinn, M. Lyons, and P. Maxwell, "Challenges of remote sensing for quantifying changes in large complex seagrass environments," *Estuarine, Coastal and Shelf Science*, vol. 133, pp. 161-171, 2013.
- [A. Friedman, O. Pizarro, S. B. Williams, and M. Johnson-Roberson, 2012] A. Friedman, O. Pizarro, S. B. Williams, and M. Johnson-Roberson, "Multi-scale measures of rugosity, slope and aspect from benthic stereo image reconstructions," *PloS one*, vol. 7, p. e50440, 2012.
- [H. Eugster and S. Nebiker, 2008] H. Eugster and S. Nebiker, "UAV-Based Augmented Monitoring-Real-Time Georeferencing and Integration of Video Imagery with Virtual Globes," *IAPRSSIS*, vol. 37, pp. 1229-1235, 2008.
- [B. Ö. Cakir, P. Adamson, and C. Cingi, 2012] B. Ö. Cakir, P. Adamson, and C. Cingi, "Epidemiology and economic burden of nonmelanoma skin cancer," *Facial plastic surgery clinics of North America*, vol. 20, pp. 419-422, 2012.
- [B. W. Stewart and C. P. Wild, 2014] B. W. Stewart and C. P. Wild, "World Cancer Report 2014," International Agency for Research on Cancer 2014.
- [R. W. Hogg, A. L. Rankin, S. I. Roumeliotis, M. C. McHenry, D. M. Helmick, C. F. Bergh, and L. Matthies, 2002] R. W. Hogg, A. L. Rankin, S. I. Roumeliotis, M. C. McHenry, D. M. Helmick, C. F. Bergh, and L. Matthies, "Algorithms and sensors for small robot path following," in *Robotics and Automation, 2002. Proceedings. ICRA '02. IEEE International Conference on*, 2002, pp. 3850-3857 vol. 4.
- [P. Krüsi, B. Bücheler, F. Pomerleau, U. Schwesinger, R. Siegwart, and P. Furgale, 2014] P. Krüsi, B. Bücheler, F. Pomerleau, U. Schwesinger, R. Siegwart, and P. Furgale, "Lighting-invariant Adaptive Route Following Using Iterative Closest Point Matching," *Journal of Field Robotics*, pp. n/a-n/a, 2014.
- [C. McManus, P. Furgale, B. Stenning, and T. D. Barfoot, 2013] C. McManus, P. Furgale, B. Stenning, and T. D. Barfoot, "Lighting-invariant Visual Teach and Repeat Using Appearance-based Lidar," *Journal of Field Robotics*, vol. 30, pp. 254-287, 2013.
- [C. Sprunk, G. D. Tipaldi, A. Cherubini, and W. Burgard, 2013] C. Sprunk, G. D. Tipaldi, A. Cherubini, and W. Burgard, "Lidar-based teach-and-repeat of mobile robot trajectories," in *Intelligent Robots and Systems (IROS), 2013 IEEE/RSJ International Conference on*, 2013, pp. 3144-3149.
- [P. Furgale and T. D. Barfoot, 2010] P. Furgale and T. D. Barfoot, "Visual teach and repeat for long-range rover autonomy," *Journal of Field Robotics*, vol. 27, pp. 534-560, 2010.
- [M. Milford, F. Schill, P. Corke, R. Mahony, and G. Wyeth, 2011] M. Milford, F. Schill, P. Corke, R. Mahony, and G. Wyeth, "Aerial SLAM with a Single Camera Using Visual Expectation," in *International Conference on Robotics and Automation*, Shanghai, China, 2011.