

All-Environment Visual Place Recognition with SMART

Edward Pepperell, Peter I. Corke, *Fellow, IEEE*, Michael J. Milford, *Member, IEEE*

Abstract— This paper presents Sequence Matching Across Route Traversals (SMART); a generally applicable sequence-based place recognition algorithm. SMART provides invariance to changes in illumination and vehicle speed while also providing moderate pose invariance and robustness to environmental aliasing. We evaluate SMART on vehicles travelling at highly variable speeds in two challenging environments; firstly, on an all-terrain vehicle in an off-road, forest track and secondly, using a passenger car traversing an urban environment across day and night. We provide comparative results to the current state-of-the-art SeqSLAM algorithm and investigate the effects of altering SMART's image matching parameters. Additionally, we conduct an extensive study of the relationship between image sequence length and SMART's matching performance. Our results show viable place recognition performance in both environments with short 10-metre sequences, and up to 96% recall at 100% precision across extreme day-night cycles when longer image sequences are used.

I. INTRODUCTION

Visual place recognition systems have been increasingly popular within robotics, due in part to the wide availability of visual sensors, as well as their lack of reliance on global cues (c.f. GPS). These properties make them applicable in areas where satellite coverage is not available or the scale of the environment is small. Many of such systems reside under the banner of Simultaneous Localisation And Mapping (SLAM) – an area that is the subject of substantial research [1-5].

The increasing deployment of visual sensors as the primary sensor modality in state-of-the-art place recognition systems is motivated by their inherently high information potential, small size, modest power requirements, wide availability and low cost [6-8]. There has also been substantial focus on long-term autonomy for service robots with persistent navigation being a desirable outcome. However, persistent visual navigation remains a challenge, due to vision being inherently sensitive to changes in illumination, such as between different times of the day (such as the extreme case of day to night), and across weather and seasonal variations. Attempts to solve this problem have shown impressive results [9-12], but these existing systems either require prior training in similar environments [10, 11],

require impractically long image sequences [9, 12], or make assumptions about the repeatability of vehicle pose and velocity between traversals [9, 12]. Overcoming these limitations is essential for a persistent system with applicability to the real world.

Sequence Matching Across Route Traversals (SMART) is a route-based place recognition algorithm, which improves the general applicability of Sequence SLAM (SeqSLAM) by integrating self-motion information to form spatially consistent sequences (Fig. 1), and new image matching techniques to handle greater perceptual change and variations in translational pose. Together, these developments allow the sequence-based algorithm to successfully recognise places across difficult environmental conditions, and for the first time, enable image sequences to be reduced to lengths practical for real-world use. SMART achieves these aims without professional-grade sensors, calibrated imagery or a training phase. In this paper, we build on a preliminary study in [13]; we extend the pose invariance capabilities of SMART and compare its performance against SeqSLAM on a new, varied-pose and visually bland off-road forest dataset, and conduct an extensive study of the relationship between sequence matching length and performance.

The paper proceeds as follows: Section II provides background into prior work on visual place recognition and odometry. Section III summarises the SMART algorithm, reviewing its improvements over SeqSLAM. The experimental setup and results are presented in Section IV and Section V, respectively. Finally, Section VI discusses the outcomes of this paper and presents suggestions for future work.

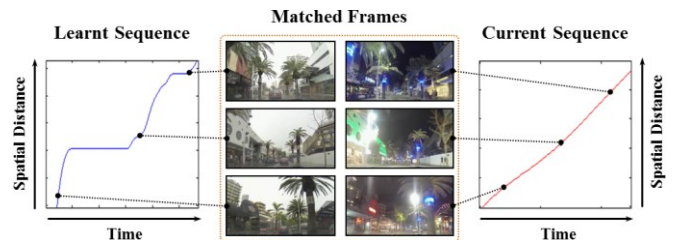


Fig. 1: Using condition-invariant, whole-image matching techniques and odometry, SMART compares image sequences with constant spatial separation between images.

II. BACKGROUND

With vision as the primary sensor, SLAM systems have previously shown impressive place recognition, as judged by accuracy and environmental size. One state-of-the-art example is FAB-MAP, which has been successfully demonstrated on long routes up to 1000 km [14], but requires

E. Pepperell, P. I. Corke and M. J. Milford are with the School of Electrical Engineering and Computer Science at the Queensland University of Technology, Brisbane, Australia, e.pepperell@qut.edu.au. This work was supported by an Australian Research Council DECRA grant (DE120100995) awarded to M. J. Milford.

a training phase before deployment. The use of features in such systems, such as SIFT [15] and SURF [16] allows invariance to scale and rotation, but leads to poor performance across extreme perceptual change [9].

Sensor fusion is a common technique to improve the perceptual invariance of vision systems, with notable examples using thermal camera [17] and laser augmentations [18, 19]. The disadvantage of these approaches is that systems become more expensive and cumbersome, which has motivated further investigation into more invariant vision-only place recognition.

SeqSLAM is a place recognition algorithm that uses sequences of whole images and local-best matching techniques to cope with extreme perceptual change [9], even over very long (3000 km) journeys [12]. Its lack of reliance on features enables SeqSLAM to work with high levels of motion blur [20], allowing slower shutter speeds for greater exposure in low light. The linear sequence search process in SeqSLAM assumes similar speeds in repeated route traversals and negligible accelerations, limiting its performance when these criteria are not met. These shortcomings have been somewhat addressed by Cooc-MAP, which uses feature codebooks for various times of the day and allows non-linear sequence searching [10]. However, the disadvantage of this approach is that multiple training traversals are required, including those in similar conditions and similar time of day to the query traversal. Non-linear sequence searching also reduces search constraints, potentially allowing more spurious sequence matches and increasing computational complexity.

Combining visual place recognition and odometry is a demonstrably effective technique for motion estimation and pose filtering [4, 21]. Odometry effectively allows constraints in searching for loop-closures, improving accuracy and reducing computational complexity. An alternative system, OpenStreetSLAM, is an odometry-only approach to place recognition, which fits trajectory shapes of its path to streets on roadmaps [22]. However, locations exhibiting self-similarity, such as highways and gridded streets pose a serious challenge to this approach.

III. THE SMART ALGORITHM

This section summarises the SMART algorithm and describes the addition of several new components over its progenitor, SeqSLAM.

A. Sky Blackening

The sky regions of an image offer no localisation information and degrade matching performance if conditions change between traversals. In such cases, we effectively remove the sky in daytime traversals by a process we call *Sky Blackening*. The process begins by applying a transform [23] to the original RGB image to enhance the contrast of the brighter, bluer sky regions, giving C :

$$C = -1.16R + 0.363G + 1.43B - 82.3 \quad (1)$$

This image is then converted into an image mask by thresholding at the sky-ground boundary using valley-emphasis [24] – a technique which does not assume a bi-modal distribution of sky-ground pixel occurrences. The

mask is then used to effectively zero all pixels of the sky regions in a standard greyscale version of the original image, which is then used for image comparison (Fig. 2). Despite the significantly different types of datasets used in this paper, the method leads to consistent performance improvements.

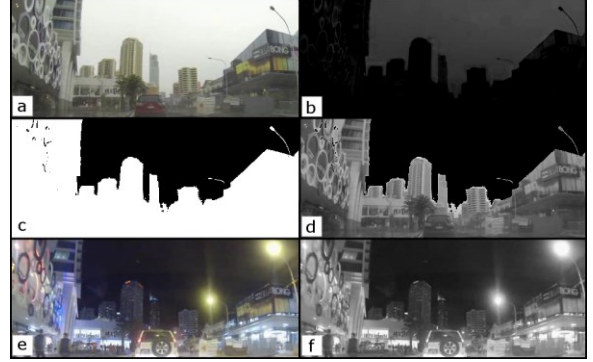


Fig. 2: The sky blackening algorithm takes a daytime frame (a), automatically enhances the contrast between sky and ground (b), forms an image mask (c) and zeroes sky pixels of the image in its greyscale form (d). For comparison, the corresponding night time frame (e) and its greyscale form (f) are shown. The similarity of (d) and (f) is vastly greater than the original frames (a) and (e).

B. Image Comparison

To cope with changes in camera pose, SMART implements variable offset image matching – a technique absent from SeqSLAM, but used in other SLAM systems [25]. Each query frame is compared to each database frame (template) over a range of offsets up to a horizontal and vertical maximum (notated here as x_{max} and y_{max}) such that the Sum of Absolute Differences (SAD) score of the overlapping region is minimised.

As with SeqSLAM, comparisons are made with the resolution reduced, patch-normalised frames – a process performed by dividing the image into patches, then taking each pixel and subtracting its patch mean and dividing by its patch standard deviation [9]. All difference scores are assembled into a difference vector, \mathbf{D} for each frame.

C. Distance-Based Template Learning and Querying

SMART learns and queries resolution-reduced, patch-normalised images at constant distance intervals, rather than constant time intervals. To function effectively, a consistent, but not necessarily metric source of translational velocity (e.g. encoder-based or visual) is required.

D. Localised Sequence Searching

Like SeqSLAM, SMART employs local neighbourhood normalisation to remove biases from lighting variations between route traversals. Each element in the difference vector \mathbf{D} is normalised within a fixed range (l) by subtracting the local mean and dividing by the local standard deviation to enhance the local matching contrast [9].

To search for image sequences, difference vectors are joined to form an image-matching matrix, \mathbf{M} :

$$\mathbf{M} = [\hat{\mathbf{D}}^{T-d_s}, \hat{\mathbf{D}}^{T-d_s+1}, \dots, \hat{\mathbf{D}}^T] \quad (2)$$

where d_s is the sequence length and T is the current frame number. Straight-line trajectories are then projected from

each element in $\hat{\mathbf{D}}^{T-d_s}$ to find the lowest-cost sequence, which has a normalised difference score, S :

$$S = \frac{\sum_{t=T-d_s}^T \mathbf{D}_k^t}{d_s} \quad (3)$$

where k is the index of the column vector \mathbf{D}^t that the search trajectory passes through:

$$k = s + (d_s - T + t) \tan \phi \quad (4)$$

where s is the number of the originating template and ϕ is the search trajectory angle. As a final step, a global cost threshold is applied to determine which sequences are accepted or rejected.

Using straight lines for sequence searching assumes that both route traversals were performed at similar speeds (depending on the search angle range) with negligible relative acceleration. These assumptions are not valid in many environments, such as busy city streets. SMART's use of odometry for learning and querying effectively linearises sequences in \mathbf{M} and constrains the search space to approximately 45° straight lines (Fig. 3). The result is a faster (fewer angles to search) and more accurate search process as spurious sequences are less likely to be selected.

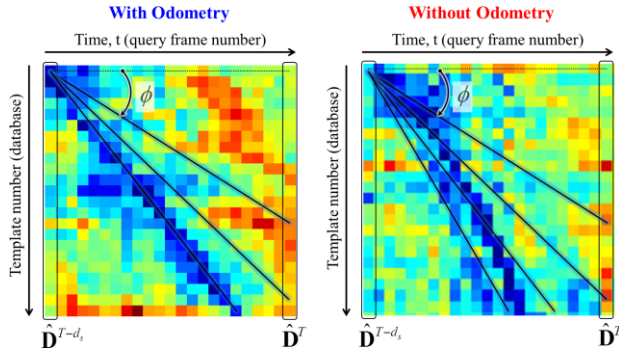


Fig. 3: The search algorithm finds the lowest-cost (dark blue) straight-line segment across recent image difference vectors as the search angle, ϕ , is varied. For clarity, only one sequence starting point is shown in each case. SMART's use of odometry (left) linearises the sequence trajectory, allowing a faster, more constrained search, with less chance of spurious matching. Without odometry (right), the linear trajectory lines do not correctly span over all sequence frames.

IV. EXPERIMENTAL SETUP

In this section, we describe the experimental setups, vehicular datasets, ground truth measures and the studies performed.

A. Equipment

A GoPro Hero3 Black Edition camera was used to record all video footage, chosen for its short focal length of 2.77 mm (15 mm full-frame equivalent). For the off-road dataset, a 4WD Summit Traxxas 1:10 scale Radio-Controlled (RC) car was used to traverse the environment. The car was equipped with a Hall Effect sensor to log odometry to an onboard laptop computer.

For the road dataset, the camera was mounted to the dashboard of an unmodified car, facing through its front

windscreen. Odometry information was collected via the car's OBDII port (standard on all modern cars) using an OBDPro USB Scantool and a laptop computer. Fig. 4 shows the experimental setup for both datasets.



Fig. 4: Experimental setup, showing the RC car with GoPro camera and on-board laptop computer (top) and the road vehicle with GoPro camera (bottom-left), OBDPro and laptop computer (bottom-right).

B. Datasets

We tested SMART on two datasets¹, as shown in Fig. 5.



Fig. 5: Maps of the Alderley off-road dataset (top) and Surfers Paradise road dataset (bottom). The paths taken are shown in red. Note the U-turn in the north-most part of the road dataset, which was necessary for street access due to construction and road closures. Imagery © 2014 DigitalGlobe, Sinclair Knight Merz & Fugro, Map data © 2014 Google.

¹ Video datasets are available from tiny.cc/milforddatasets

The first dataset was collected with the RC car in an off-road, forest environment in Brisbane, Australia, and spanned approximately 850 metres. This dataset was structured to deliberately test both the pose variance and velocity invariance capabilities of the new algorithm. The first traversal followed the left side of a dirt path with deliberate, regular and large speed variations (Fig. 6). The second traversal followed the right side of the same path at a relatively constant speed. The two traversals took approximately 20 and 10 minutes respectively. The ratio between vehicle speeds at corresponding locations during the first and second traverses varied from infinite (stopped vs. moving) to approximately 50%, while lateral pose variance ranged from 1 to 3 metres (average of 2 metres).

The second dataset consisted of a 3.5 km drive on-road through the busy streets of Surfers Paradise on the Gold Coast in Australia [13]. The first traversal of this dataset was performed during the day in overcast weather with light rain, and the second was performed in clear weather at night. As both traversals were done in peak traffic, the speed variation within and between traversals was significantly greater than the original SeqSLAM dataset [9]. Travel times across this dataset were approximately 20 and 15 minutes for the daytime and night time traversals, respectively.

C. Ground Truth

For both datasets, ground truth was constructed by manually finding pairs of frame correspondences at regular frame intervals. Approximately 200 pairs were found for each dataset, with linear interpolation used for in-between frames. Ground truth can be considered correct to within approximately 1 metre of forward translation for the road dataset and 3 metres for the off-road dataset (due to difficulty in identifying similar scenes across the changed pose and camera perspective).

D. Data Pre-Processing

All original videos were recorded at a resolution of 1920 by 1440 pixels at frequencies of 30 and 24 frames per second for day and night, respectively. The road night time video was recorded at a lower frame rate to enable a longer exposure time (approximately 1/48th of a second) and later converted to 30 frames per second. For the road dataset, a rectangular crop was then performed to remove visible internal areas of the car. For image matching, the lower half of each frame was cropped out of the off-road videos to prevent spurious localisation from the highly-aliased dirt track.

In cases where it was used, sky blackening was performed prior to downsampling and patch-normalisation, which was the final step before SAD image comparison.

Odometry information was used to generate lists of frame numbers with fixed-distance separations as inputs into SMART. Separations of 0.25 metres and 1 metre (f_{dist}) were used between frames for the off-road and road datasets, respectively. In the tests without odometry, similar lists were generated with fixed frame separations (f_{jump}) to give approximately the same number of frame comparisons as in the odometry cases.

E. Studies and Parameters

For the off-road dataset, we conducted studies investigating how odometry, sky blackening and variable offset image matching (in increasing amounts) affect performance. The following scenarios were tested:

1. Classic SeqSLAM
2. Sky blackening and offset matching up to 2 pixels horizontally and 1 pixel vertically (no odometry)
3. Odometry only
4. Odometry and sky blackening
5. Odometry, sky blackening and offset matching up to 1 pixel horizontally and vertically
6. Odometry, sky blackening and offset matching up to 2 pixels horizontally and 1 pixel vertically
7. Odometry, sky blackening and offset matching up to 2 pixels horizontally and vertically
8. Odometry, sky blackening and offset matching up to 3 pixels horizontally and 1 pixel vertically

An additional set of studies investigated the effect of changing the sequence length from 5 to 200 metres on the road dataset. Table I shows the parameter settings for all studies.

TABLE I
PARAMETER LIST

Parameter	Value	Description
R_x, R_y	64,32	Reduced image size (road dataset)
R_x, R_y	24,8	Reduced image size (off-road dataset)
f_{dist}	1 metre	Frame learning rate (road dataset)
f_{jump}	10 frames	Frame learning rate (off-road dataset, scenarios 1-2)
f_{dist}	0.25 metres	Frame learning rate (off-road dataset, scenarios 3-8)
l	80 templates	Neighbourhood length
d_s	Varied	Sequence length (road dataset)
d_s	40 frames (10 metres)	Sequence length (off-road dataset)
P	8x8 pixels	Patch size
x_{max}, y_{max}	1,1	Maximum shift offsets (off-road scenario 5, road dataset)
x_{max}, y_{max}	2,1	Maximum shift offsets (off-road scenarios 2 and 6)
x_{max}, y_{max}	2,2	Maximum shift offsets (off-road scenario 7)
x_{max}, y_{max}	3,1	Maximum shift offsets (off-road scenario 8)
ϕ	[40°, 45°, 50°]	Trajectory search angles

V. RESULTS

In this section, we present the results showing speed variation and place recognition performance in the off-road dataset, and the results of the sequence length study on the road dataset. A video accompaniment to this paper demonstrates the matching techniques and results.

A. Speed Variation

There was significant speed variation between the two traversals of the off-road dataset (Fig. 6), with the first (left) traversal having a significant amount of arbitrary acceleration, stops and starts.

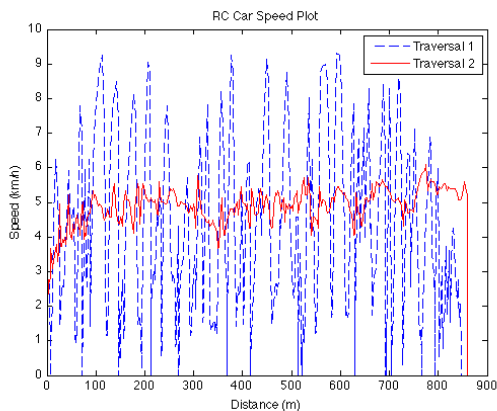


Fig. 6: Speed plots of each traversal of the RC car dataset. The significant speed variation between each traversal makes sequence matching difficult without odometry.

B. Place Recognition Performance

Fig. 7 shows example frame matches on the off-road dataset. Despite the large variations in vehicle velocity, SMART attained good matching performance, especially when compared with SeqSLAM, as well as coping with moderate changes in shadowing, sky appearance and pose.

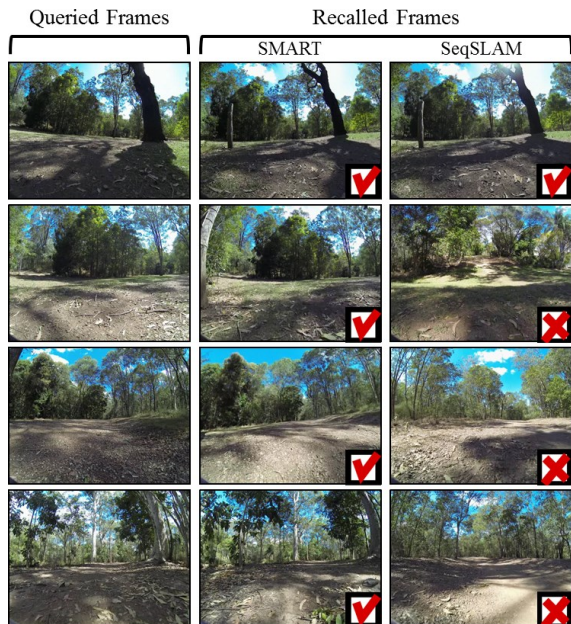


Fig. 7: Example frame matches from the off-road dataset. The left column shows frames queried from the second traversal, and the middle and right columns show the frames recalled by SMART and SeqSLAM, respectively. Both algorithms performed well in distinct locations (top row), but SeqSLAM struggled to correctly match bland scenes (bottom 3 rows). Each frame is the midpoint of a sequence.

Precision-recall curves (Fig. 8) were generated by varying the sequence cost threshold and comparing the chosen sequences to ground truth. Sequences were deemed correct if

their midpoints were within 10 metres of ground truth – a stricter threshold than that used in previous studies (40 metres in [9, 14, 20]). Here, precision is the proportion of returned frame pairs that were correct, and recall is the proportion of total correct frame pairs that were returned. Recall performance at 100% precision is the criterion of interest (despite missing some correct matches), as false positives are undesirable in a place recognition system. As the first and last $d_s/2$ frames do not form part of complete sequences, it is not possible to achieve 100% recall. As shown in Table II, scenario 6 had the best performance, achieving 36% recall at 100% precision.

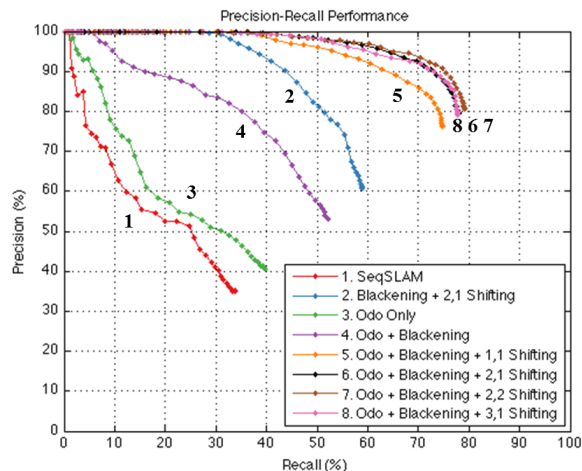


Fig. 8: Precision-Recall curves for the 8 testing scenarios in the off-road dataset, showing the performance improvement as the new features of SMART are added to the SeqSLAM baseline.

TABLE II
OFF-ROAD SCENARIO TESTING RESULTS

Scenario	Best Recall at 100% Precision
1	1%
2	25%
3	1%
4	5%
5	30%
6	36%
7	21%
8	28%

Odometry alone did not improve performance at the 100% precision level (scenario 1 vs. 3), but did make a significant improvement when used in conjunction with optimal settings (scenario 2 vs. 6). Even though both traversals were performed at similar times of the day, sky blackening made a noticeable performance improvement (albeit not to the same enormous recall benefit of 50% as on the road dataset between day and night [13]), as it prevented cloud pattern variations from diluting matching performance. Image offsetting up to 2 pixels horizontally and 1 pixel vertically was found to be optimal, with further increases in either direction worsening performance. We attribute this to larger shifts resulting in a smaller image overlap (comparison region), making it more prone to aliasing and returning

spurious matching scores. Ground truth performance plots for the best cases with and without odometry are shown in Fig. 9.

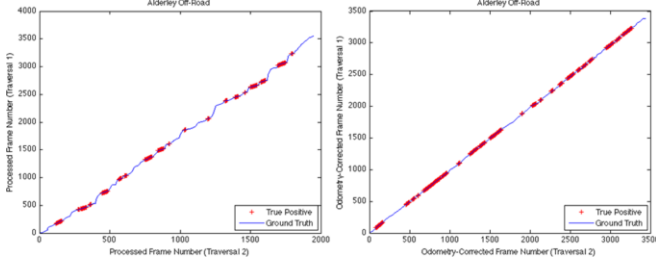


Fig. 9: Ground truth performance plots for the off-road dataset for optimised no-odometry scenario 2 (left – note the scale) and optimised odometry scenario 6 (right) at 100% precision. The horizontal axis represents frame numbers on the query traversal and the vertical axis represents those on the database traversal. Blue lines connect ground truth frame pairs and red crosses indicate the correctly reported frame matches.

C. Sequence Length Study

Table III shows the performance achieved by SMART as the sequence length is varied between 5 and 200 metres, as well as the theoretical maximum (to the nearest integer percentage) for each case. All tests were performed with sky blackening and optimal parameters (see Table I). Note that increasing the sequence length also reduces the maximum possible recall, as the first and last $d_s/2$ frames are not midpoints (recalled frame pairs) of complete sequences. Thus, over the 2350 query frames, the maximum possible recall drops to 96% for 100-metre sequences, and 91% for 200-metre sequences. This recall limitation applies for each new section of contiguous matching frames, with this dataset being a special case as the query traversal is a direct repeat of the database traversal.

TABLE III
SEQUENCE LENGTH PERFORMANCE

Sequence Length (1 frame/metre)	Best Recall Achieved at 100% Precision	Maximum Possible Recall at 100% Precision
5	12%	100%
10	37%	100%
20	47%	99%
30	81%	99%
40	86%	98%
50	88%	98%
75	91%	97%
100	96%	96%
200	68%	91%

Significant performance improvement was achieved by increasing sequence lengths up to 30 metres, after which the improvements became more modest. Performance peaked at a sequence length of 100 metres, where SMART correctly identified all possible matches within complete sequences. The drop in performance at a sequence length of 200 metres was due to accumulated odometry error – fitting straight lines across slightly non-linear sequences resulted in several incorrect matches.

All sequence lengths had strong enough recall performance to be viable in a navigation system, although the actual required recall at 100% precision rate would depend on the specific application. Using 10-metre sequences (37% recall), the median and mean match-to-match distances were

1 metre and 2.6 metres, respectively, with a maximum gap of 91 metres. Reducing the sequence length to 10 m thus allows faster searching and less initialisation lag (5 metres). Fig. 10 shows ground truth performance plots of the 10, 30, 100 and 200 metre sequence lengths, and Fig. 11 compares a 10-metre and 100-metre sequence from the same location.

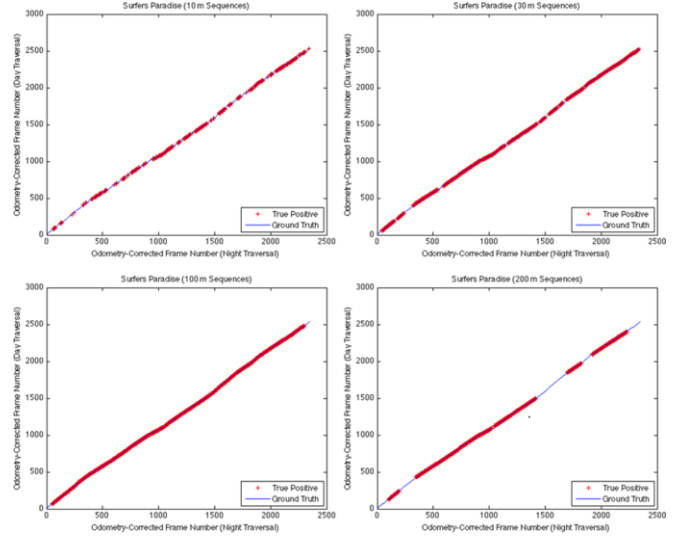


Fig. 10: Ground truth performance plots at 100% precision for the road dataset with optimal parameters and sequence lengths of 10, 30, 100 and 200 metres from top-left to bottom-right, respectively. Blue lines connect ground truth frame pairs and red crosses indicate the correctly reported frame matches.

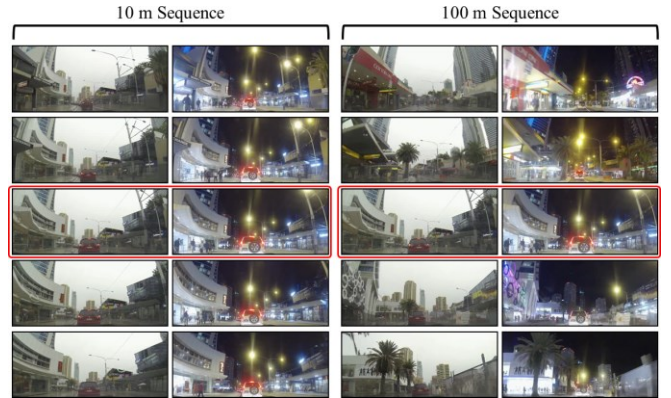


Fig. 11: Example frames equally-spaced along a 10-metre (left) and 100-metre (right) matched sequence. Both sequences report the same location match (outlined in red), but the 10-metre sequence spans a significantly smaller portion of the dataset and requires an order of magnitude fewer image comparisons.

VI. DISCUSSION AND FUTURE WORK

Our results showed that the new image matching techniques and use of odometry in SMART enabled successful place recognition in challenging environments where SeqSLAM struggled. The sky blackening technique proved beneficial for both day-night and day-day cycles, even when appearance change was relatively minor. Combining a wide field of view camera with image offsetting was effective at overcoming moderate and varied changes in lateral pose.

Odometry provided velocity invariance and reduced the potential for matching aliased sequences by tightly

constraining the search for image sequences, unlike alternative methods that search a wide range of possible velocities and accelerations [10]. Obviously, a source of odometry information is required, but many platforms have either wheel-based or visual odometry-based odometry sources available.

The final study showed the implications of varying the sequence length on the road dataset, and demonstrated that sequences as short as 10 metres could provide viable localisation. Shorter sequences enable faster searching and less localisation lag, which is equal to half a sequence length. This criterion becomes particularly important in instances where the second traversal is not a direct overlap of the first, such as in a navigation system where frames are unlikely to be queried in the same order as database frames.

In both datasets, SMART achieved regular localisation, but had difficulty over a number of regions. The system had lower performance in areas with significant self-similarity, such as long stretches of road or track with common distal features. In a navigation system, this could be mitigated with dead reckoning using odometry, or by automatically increasing the sequence length – a technique we are currently investigating.

One of the most promising applications for SMART is as a real-time navigation system that potentially augments GPS. Storage and computational concerns in such an application have been previously discussed [26], with databases of low-resolution imagery shown feasible in environments up to the size of a small city. By constraining the search space (after an initial global localisation phase) to search database frames within only the local area (such as a city), real-time computation is feasible on standard embedded hardware. The contributions described in this paper represent a significant step towards achieving generic applicability across all environments and platforms.

REFERENCES

- [1] G. Dissanayake, P. M. Newman, S. Clark, H. Durrant-Whyte, and M. Csorba, "A solution to the simultaneous localisation and map building (SLAM) problem," *IEEE Transactions on Robotics and Automation*, vol. 17, pp. 229-241, June, 2001 2001.
- [2] M. Cummins and P. Newman, "FAB-MAP: Probabilistic Localization and Mapping in the Space of Appearance," *International Journal of Robotics Research*, vol. 27, pp. 647-665, 2008.
- [3] 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.
- [4] K. Konolige and M. Agrawal, "FrameSLAM: From Bundle Adjustment to Real-Time Visual Mapping," *IEEE Transactions on Robotics*, vol. 24, pp. 1066-1077, 2008.
- [5] M. J. Milford, G. Wyeth, and D. Prasser, "RatSLAM: A Hippocampal Model for Simultaneous Localization and Mapping," in *IEEE International Conference on Robotics and Automation*, New Orleans, USA, 2004, pp. 403-408.
- [6] F. Chaumette, P. Corke, and P. Newman, "Editorial Special Issue on Robotic Vision," *The International Journal of Robotics Research*, vol. 29, pp. 131-132, 2010.
- [7] G. Hager, M. Hebert, and S. Hutchinson, "Editorial: Special Issue on Vision and Robotics, Parts I and II," *International Journal of Computer Vision*, vol. 74, pp. 217-218, 2007.
- [8] J. Neira, A. J. Davison, and J. J. Leonard, "Guest Editorial Special Issue on Visual SLAM," *IEEE Transactions on Robotics*, vol. 24, pp. 929-931, 2008.
- [9] 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.
- [10] E. Johns and G.-Z. Yang, "Feature Co-occurrence Maps: Appearance-based Localisation Throughout the Day," in *IEEE International Conference on Robotics and Automation*, Karlsruhe, Germany, 2013.
- [11] N. Sünderhauf, P. Neubert, and P. Protzel, "Predicting the Change – A Step Towards Life-Long Operation in Everyday Environments," in *Robotics: Science and Systems (RSS) Robotics Challenges and Vision Workshop*, 2013.
- [12] N. Sünderhauf, P. Neubert, and P. Protzel, "Are We There Yet? Challenging SeqSLAM on a 3000 km Journey Across All Four Seasons," in *Proc. of Workshop on Long-Term Autonomy, IEEE International Conference on Robotics and Automation (ICRA)*, 2013.
- [13] E. Pepperell, P. I. Corke, and M. J. Milford, "Towards Persistent Visual Navigation using SMART," in *Australasian Conference on Robotics and Automation (ACRA)*, Sydney, Australia, 2013.
- [14] M. Cummins and P. Newman, "Highly scalable appearance-only SLAM - FAB-MAP 2.0," in *Robotics: Science and Systems*, Seattle, United States, 2009.
- [15] D. G. Lowe, "Object Recognition from Local Scale-Invariant Features," in *Proceedings of the International Conference on Computer Vision-Volume 2 - Volume 2*, 1999.
- [16] H. Bay, T. Tuytelaars, and L. Van Gool, "SURF: Speeded Up Robust Features," in *Computer Vision – ECCV 2006*, 2006, pp. 404-417.
- [17] S. Vidas and W. Maddern, "Towards Robust Night and Day Place Recognition using Visible and Thermal Imaging," in *Beyond laser and vision: Alternative sensing techniques for robotic perception workshop at RSS2012*, Sydney, Australia, 2012.
- [18] P. Newman, D. Cole, and K. Ho, "Outdoor SLAM using Visual Appearance and Laser Ranging," in *International Conference on Robotics and Automation*, Florida, United States, 2006.
- [19] G. Gallegos, M. Meilland, P. Rives, and A. I. Comport, "Appearance-Based SLAM Relying on a Hybrid Laser/Omnidirectional Sensor," in *Intelligent Robots and Systems (IROS)*, 2010, pp. 3005-3010.
- [20] M. J. Milford, I. Turner, and P. I. Corke, "Long Exposure Localization in Darkness Using Consumer Cameras," in *IEEE International Conference on Robotics and Automation*, Karlsruhe, Germany, 2013.
- [21] W. Maddern, M. Milford, and G. Wyeth, "CAT-SLAM: probabilistic localisation and mapping using a continuous appearance-based trajectory," *The International Journal of Robotics Research*, vol. 31, pp. 429-451, 2012.
- [22] G. Floros, B. v. d. Zander, and B. Leibe, "OpenStreetSLAM: Global Vehicle Localization Using OpenStreetMaps," in *IEEE International Conference on Robotics and Automation*, Karlsruhe, Germany, 2013.
- [23] S. Thrunrowgood, D. Soccol, R. J. D. Moore, D. Bland, and M. V. Srinivasan, "A Vision Based System for Attitude Estimation of UAVs," in *Intelligent Robots and Systems*, St. Louis, USA, 2009.
- [24] H.-F. Ng, "Automatic thresholding for defect detection," *Pattern Recognition Letters*, vol. 27, pp. 1644-1649, 2006.
- [25] 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.
- [26] M. Milford, "Vision-based place recognition: how low can you go?," *The International Journal of Robotics Research*, vol. 32, pp. 766-789, 2013.