

Off the Beaten Track: Predicting Localisation Performance in Visual Teach and Repeat

Julie Dequaire, Chi Hay Tong, Winston Churchill and Ingmar Posner

Abstract—This paper proposes an appearance-based approach to estimating localisation performance in the context of visual teach and repeat. Specifically, it aims to estimate the likely corridor around a taught trajectory within which a vision-based localisation system is still able to localise itself. In contrast to prior art, our system is able to *predict this localisation envelope* for trajectories in similar, yet geographically distant locations where no repeat runs have yet been performed. Thus, by characterising the localisation performance in one region, we are able to predict performance in another. To achieve this, we leverage a Gaussian Process regressor to estimate the likely number of feature matches for any keyframe in the teach run, based on a combination of trajectory properties such as curvature and an appearance model of the keyframe. Using data from real traversals, we demonstrate that our approach performs as well as prior art when it comes to interpolating localisation performance based on a number of repeat runs, while also performing well at generalising performance estimation to freshly taught trajectories.

I. INTRODUCTION

Vision-based localisation is an active area of research due to its reliance on ubiquitous and comparatively low-cost sensors (see, for example, [1], [2], [3], [4], [5]). From amongst several approaches, *teach and repeat* has distinguished itself as being an effective solution in a number of domains such as space exploration [6] and autonomous driving [7]. In this framework, a robot is driven along a desired path in a *teach* phase, during which it builds a topometric map [8], [9] of the world by estimating its relative pose between consecutive keyframes. During a *repeat* run, localisation is performed by matching features between the live feed and those of the map, estimating the position of the robot relative to the route previously traversed. In doing so, it is commonly observed that some environments are less lenient than others to even small lateral or rotational deviations from the known trajectory, so that localisation can be lost if the taught trajectory is not followed within some tolerance dictated by appearance.

Despite this strong coupling between navigation, planning and control of a vehicle, teach and repeat systems are often developed in isolation from planning and control. A notable exception here is [10] which aims to learn controllers which reduce path-tracking errors over time. However, situations commonly arise where environment conditions or vehicle characteristics may not allow for an exact path repetition. In addition, path deviations may be tolerated in order to pursue alternative mission goals such as energy savings (e.g. [11], [12]) or more accurate situational awareness (e.g. [13]).

In these cases, *a-priori* knowledge of the degree to which deviations from a taught trajectory can be tolerated *as soon as a route is acquired* is pivotal for planning and control. Yet this problem has thus far received little attention.

This paper, therefore, advocates an approach for mapping directly from a teach trajectory to likely localisation performance in the surrounding area – its *localisation envelope*. Given the nature of visual teach and repeat, and as demonstrated in [14], trajectory shape can be used to accurately predict the performance of the localisation system once a number of repeat passes have been performed. Here, we extend this work by proposing an appearance-based approach to *predict* localisation performance, in an environment where only a single teach path has been recorded. In particular, we leverage Gaussian Process regression [15] to predict the number of features likely to be matched by the localisation system against a particular map keyframe based on an appearance model of the keyframe as well as on factors such as intended path deviation and trajectory shape. The regressor is trained using data from a prior experience and, due to its appearance-based nature, thus allows the system to *generalise* to predicting the localisation envelope of a novel teach run in a different location. To the best of our knowledge this is the first approach able to do so.

II. RELATED WORK

Furgale and Barfoot’s [6] vision based Teach and Repeat system demonstrated autonomous retro-traverse over 32km in an outdoor setting. As part of their experiments, they calculated how robust the localisation was to lateral deviations. The camera was manually placed at offsets from the original teach path until the failure point was found. By testing at several locations along the route, they computed a single, constant lateral offset for the map.

Churchill *et al.* [14] extended this work by aiming to capture the local variation of the localiser performance as a function of position in the map. This was modelled using a Gaussian Process, with one of the input variables being position. In this context as elsewhere, Gaussian Processes provide a framework amenable to incorporating new measurements as they become available while providing uncertainty estimates in a principled way. However, the limitation of this approach lies in its use of position as an input variable, which means that the model is specific to a particular route, and cannot be generalised to other maps or places.

Krüsi *et al.* [16] compared the accuracy of a laser based Teach and Repeat system to a vision based one. They find the laser based system is more robust to path deviations and

Authors are from the Mobile Robotics Group, Oxford University, Oxford, England. {julie, chi, winston, ingmar@robots.ox.ac.uk}



Fig. 1. A Clearpath Husky A200 with a Bumblebee2 stereo camera.

ambient conditions. However they do not model or predict localiser performance.

Our work is most closely related to Churchill *et al.* By dropping the position term and replacing it with a function of appearance, we are able to generalise to other, similar environments. Consequently, we show our approach to subsume the previous position-based work in that it is commensurate in performance in a like for like comparison, while additionally allowing us to generalise successfully.

III. VISUAL TEACH AND REPEAT

We begin with a brief description of our Visual Teach and Repeat (VT&R) system [6], [7]. At its core, there is a sparse-feature, stereo visual odometry (VO) pipeline. This works as follows: given a stereo image pair, the algorithm first extracts landmarks from each image and then computes their 3D positions. This is followed by landmark matching across sequential frames. The resulting information is then used to compute a 6 Degree-of-Freedom (DoF) pose estimate which best explains the relative movement of landmarks in each frame. The process is repeated for every frame to compute the egomotion of the camera. Our implementation uses FAST [17] corners combined with BRIEF [18] descriptors, RANSAC [19] for outlier rejection and nonlinear least-squares optimisation.

VT&R is built upon saving keyframes containing 3D landmarks at regular intervals along the robot’s route during the teach pass. The same VO pipeline can then be used during a repeat pass by matching landmarks to the keyframes and computing 6 DoF pose estimates relative to the teach path.

We initialise localisation at repeat time with FAB-MAP [20], a large-scale loop-closure framework that uses a visual bag-of-words (BoW) [21] model for place recognition. In this framework, a visual vocabulary is first constructed on a training set by clustering features extracted from SURF detectors and descriptors [20]. Every image can then be represented by a binary vector indicating which *words* from this vocabulary are present, and in what proportion. Finally, loop closures are performed by comparing BoW representations to find the map keyframe that best matches the live image. This BoW representation is also used to model the appearance of the scene as described in Section V-A.

IV. APPEARANCE-BASED PERFORMANCE PREDICTION

A. Modelling

To characterise the localisation envelope, we first define a feature vector, \mathbf{x} , which is a function of both the visual map, \mathcal{M} , and a robot pose, \mathbf{p} . These elements are encoded by the bag-of-words vector, \mathbf{b} , extracted from the map keyframe closest to \mathbf{p} , the signed cross-track error to the path, d , and the local path curvature, c . The feature vector is stacked together and constructed as

$$\mathbf{x} = \begin{bmatrix} \mathbf{b} \\ d \\ c \end{bmatrix} \quad (1)$$

While our VO implementation is able to provide 6 DoF pose estimates, we chose to model only the cross-track error due to the fact that the localisation envelope in rotation is relatively small. As a result, this model assumes that the heading of the robot is always parallel to the path. This is similar to the approach taken by [14], where we have replaced the distance along the path component with BoW appearance features.

These input vectors are then used to learn a mapping, $y = f(\mathbf{x})$, where y is a measure of localisation quality. In this work, we define y to be the number of matched keyframe landmarks, model the function $f(\cdot)$ as a Gaussian Process (GP) and regress over the input-output training pairs.

B. Gaussian Process Regression

GPs are an attractive modelling option as they do not require an explicit model of the localisation envelope and naturally offer measures of uncertainty [15]. This is desirable for operations downstream of localisation such as planning, where a robot may decide to be conservative if the estimate is highly uncertain, or may choose to explore to gather more information.

A GP is described by a mean function, $\mu(\mathbf{x})$ and a covariance function, $k(\mathbf{x}, \mathbf{x}')$. We set the mean function to be zero, to capture the assumption that we cannot localise in the absence of information. As a result, the behaviour of the GP is defined entirely by the covariance function, which encodes how two input feature vectors, \mathbf{x} and \mathbf{x}' , relate to each other.

In this work, we use the Matérn covariance function [15] with $\nu = \frac{3}{2}$, which takes the form

$$k(\mathbf{x}, \mathbf{x}') = \sigma_f^2 \left(1 + \sqrt{3}\tau \right) \exp \left(-\sqrt{3}\tau \right), \quad (2)$$

where $\tau = \sqrt{(\mathbf{x} - \mathbf{x}')^T \mathbf{M} (\mathbf{x} - \mathbf{x}')}$ is the weighted l_2 -norm between the two input vectors, and $\mathbf{M} = \text{diag}\{\ell_b^{-2}, \ell_d^{-2}, \ell_c^{-2}\}$ is a diagonal matrix composed of the length scales associated to each feature type, and we have chosen to use the same length scale for all components of the BoW vector. The collection of noise and length-scale parameters which govern the behaviour of the GP function are known as hyperparameters, typically denoted by the symbol θ .

Given a training set of input vectors, $\mathbf{X} = \{\mathbf{x}_i\}_{i=1}^n$, and corresponding localisation scores, $\mathbf{y} = \{y_i\}_{i=1}^n$, the GP

regression task is to determine \mathbf{y}_* , the localisation scores at a set of query points, \mathbf{X}_* . This is accomplished by considering the joint distribution over all variables, which can be expressed as

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{y}_* \end{bmatrix} \sim \mathcal{N} \left(\begin{bmatrix} \boldsymbol{\mu}(\mathbf{X}) \\ \boldsymbol{\mu}(\mathbf{X}_*) \end{bmatrix}, \begin{bmatrix} \mathbf{K} & \mathbf{K}_* \\ \mathbf{K}_*^T & \mathbf{K}_{**} \end{bmatrix} \right), \quad (3)$$

where $\mathbf{K}_* = \mathbf{K}(\mathbf{X}, \mathbf{X}_*)$ is the $n \times n_*$ matrix consisting of covariance function evaluations between the sets \mathbf{X} and \mathbf{X}_* . Similarly we use the abbreviations $\mathbf{K}_*^T = \mathbf{K}(\mathbf{X}_*, \mathbf{X})$, $\mathbf{K} = \mathbf{K}(\mathbf{X}, \mathbf{X})$, and $\mathbf{K}_{**} = \mathbf{K}(\mathbf{X}_*, \mathbf{X}_*)$ to define the other covariance functions.

To account for measurement noise, we add $\sigma_n^2 I$ to \mathbf{K} and \mathbf{K}_{**} , where σ_n^2 is the variance, and I is the identity matrix.

Conditioning the predictions, \mathbf{y}_* , on the training data results in the distribution

$$p(\mathbf{y}_* | \mathbf{y}, \mathbf{X}, \mathbf{X}_*) \sim \mathcal{N}(\bar{\mathbf{y}}_*, \mathbb{V}[\mathbf{y}_*]). \quad (4)$$

The mean can be expressed as

$$\bar{\mathbf{y}}_* = \mathbf{K}_* \mathbf{K}^{-1} \mathbf{y}, \quad (5)$$

and the covariance as

$$\mathbb{V}[\mathbf{y}_*] = \mathbf{K}_{**} - \mathbf{K}_* \mathbf{K}^{-1} \mathbf{K}_*^T. \quad (6)$$

The final component of GP regression is the training procedure, which determines the value of the hyperparameters suitable for the regression scenario. In our model, the hyperparameters are $\boldsymbol{\theta} = [\ell_b \ \ell_d \ \ell_c \ \sigma_f \ \sigma_n]^T$. These are learned by maximising the log-likelihood of the training data given the hyperparameters, which is given by

$$\log p(\mathbf{y} | \mathbf{X}, \boldsymbol{\theta}) = -\frac{1}{2} \mathbf{y}^T \mathbf{K}^{-1} \mathbf{y} - \frac{1}{2} \log |\mathbf{K}| - \frac{n}{2} \log 2\pi. \quad (7)$$

V. EXPERIMENTS

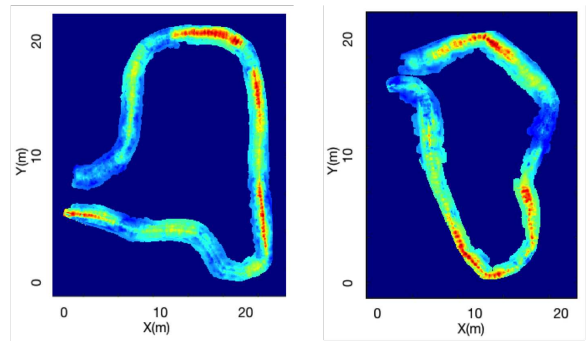
A. Data Collection

To validate our approach, we collected data from University Parks in Oxford, UK, using a Clearpath Husky A200 platform equipped with a Bumblebee2 stereo camera. This platform is depicted in Figure 1.

Two routes in separate locations of the park were identified, and for each route, we conducted a single teach pass followed by seven repeat passes with lateral offsets set to $\{-1, -0.5, -0.25, 0, +0.25, +0.5, +1\}$ m. These offsets were achieved by adding fixed values to the path tracking controller. We refer to the two routes in this section as *Route A* and *Route B*, which were approximately 100m and 70m in length, respectively.

While collected in similar environments, we observe a range of localisation performance, as shown in Figure 2. In both routes, some areas have as little as 5 matching landmarks, whereas other places yield more than 80. To provide intuition for why the performance can vary in such a small workspace, we visualise the number and distribution of keyframe landmarks for locations with high and low localisation scores for each route.

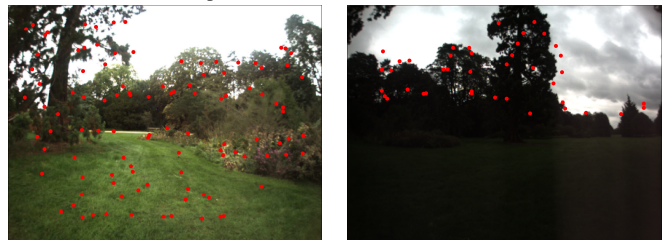
In Figure 4, we see that poor localisation performance can be attributed to the fact that many landmarks are detected



(a) Route A.

(b) Route B.

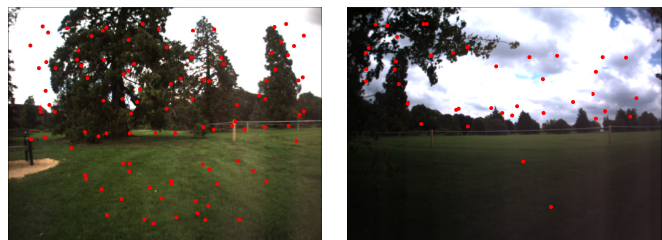
Fig. 2. Localisation quality (number of matched keyframe landmarks) constructed from all collected data for both routes. Red represents a high number of landmark matches (>80), blue represents a low number of matches (<20). It can be seen that the localisation quality varies drastically over the course of the path.



(a) High localisation quality.

(b) Low localisation quality.

Fig. 3. Landmark distribution for keyframes with high and low localisation quality in Route A. As can be seen, poor localisation results when there is an insufficient spread of landmarks throughout the image and poor contrast.



(a) High localisation quality.

(b) Low localisation quality.

Fig. 4. Landmark distribution for keyframes with high and low localisation quality in Route B. As can be seen, low contrast and unstable features detected in the cloudy sky leads to poor localisation scores.

in the cloudy sky, whereas a good location has landmarks spread throughout the image. Similarly, Figure 3 shows that low contrast leads to low feature count, whereas clear texture provides stable features for matching.

These images show us that localisation performance is correlated with appearance, and support our intuition that an appearance-based model may be appropriate for this task.

The appearance of the keyframes is described with the BoW model presented in Section III where the visual *dictionary* is constructed considering a third dataset collected in a similar, yet different area to the two routes. The BoW descriptor is a 1310-bit binary vector representing the relative occurrence of the 1310 words making the visual *dictionary*, where the vocabulary is built using SURF detectors and descriptors. Possible alternative descriptors for keyframe appearance are discussed in Section VI.

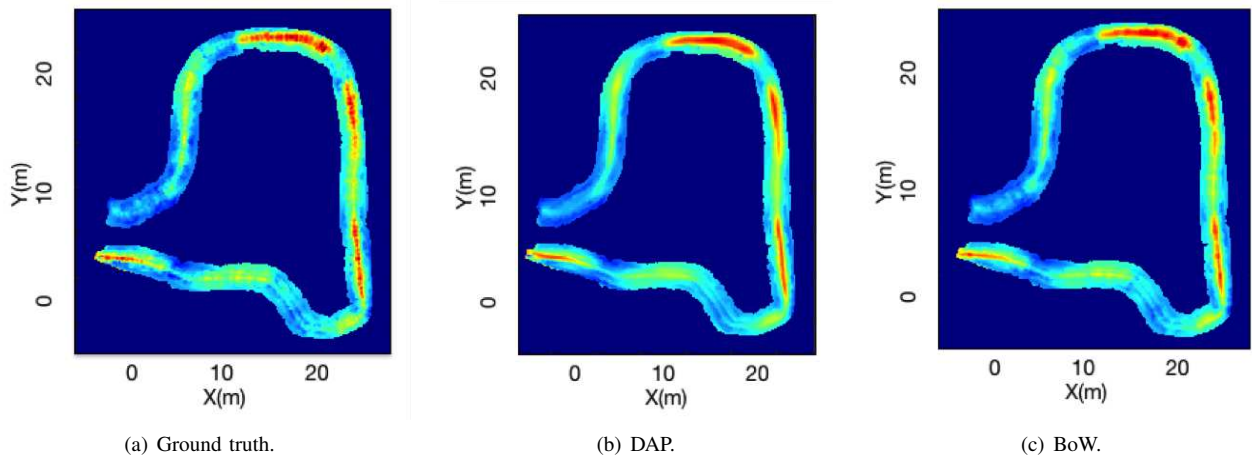


Fig. 5. Scenario I: Localisation quality predictions for a scenario where the Route A data was split into a number of repeat passes for training, and the remainder for testing. Red represents a high number of landmark matches (>80), blue represents a low number of matches (<20). Both models are commensurate in capturing the localisation envelope along the path, suggesting that BoW appearance features are a suitable replacement for the distance-along-the-path parameterisation.

B. Comparison to Baseline Approach

To evaluate our model, we first looked at how the introduction of BoW appearance features compared to the baseline performance of Churchill *et al.* [14], which parameterises the feature vector with distance along the path (DAP). Since the baseline model is unable to generalise to new routes, we considered a scenario where we gathered some training data during a few repeat passes, and wished to predict the localisation performance for future repeat passes. This was accomplished by splitting the data gathered from Route A into training and testing sets. We refer to this test case as *Scenario I*.

Both models’ hyperparameters were optimised by maximising the log-likelihood of the training data. The results from both models along with the ground truth localisation quality values are shown in Figure 5.

The performance of both models were commensurate, supporting the use of BoW features as a surrogate for the notion of a “place”. We quantified this similarity by defining a threshold of 25 matched landmarks as a successful localisation, which resulted in an F_1 measure of $F_{1,BoW} = 0.949$ for our BoW model, and an $F_{1,DAP} = 0.939$ for DAP.

C. Generalising to New Routes

To investigate how well our model does at generalising predictions to new paths, we considered two additional scenarios, *Scenario II* and *Scenario III*, where we trained our model on a selection of repeat passes for one route and tested on a subset of repeat passes for the other. *Scenario II* consisted of training on Route A and testing on Route B, and vice versa for *Scenario III*. The results for both scenarios are shown in Figures 6 and 7.

Both experiments show that we were able to capture the general evolution of the localisation envelope along the path. Quantitatively, using the same localisation success threshold of 25 landmark matches, we obtain F_1 measures of $F_{1,II} = 0.915$ and $F_{1,III} = 0.799$ for *Scenarios II* and *III*, respectively. The lower performance in *Scenario III* may be attributed to two factors. On the one hand, the dataset of

Route B is smaller than that of Route A, making it more challenging for the model to extrapolate than in *Scenario II*. On the other hand, the weather was more changeable during collection of the data in Route B than during that of Route A, which affected localisation negatively and resulted in the robot getting temporarily lost at several locations along the path. We hypothesize that we may have learned a pessimistic model when training on the data of Route B.

VI. DISCUSSION

While the overall results show promise, there were sections of Route B where our model was overconfident. In contrast, the system was generally underconfident on Route A. To form an intuition for this, we first consider a section on Route B where localisation was very poor, yet our model yielded an optimistic prediction. We indicate this section by a white arrow in Figure 6a. The corresponding keyframe obtained from the teach pass of Route B is depicted in Figure 8, along with a selection of repeat frames which matched against it. Interestingly, there were 102 widely distributed landmarks detected in the keyframe. However, the overall appearance of the scene changed dramatically during the repeat runs, introducing shadows and changing the colour of the grass. As a result, the model was optimistic given the feature-rich keyframe, but was unable to account for the reduced performance due to the environmental effects.

Conversely, we consider a section of Route A where our model under-predicted performance. We indicate this section by a white arrow in Figure 7a. As before, we show the corresponding keyframe from the teach pass of Route A along with a selection of successful matches from the repeat passes in Figure 9. The keyframe presents only 42 landmarks with very low contrast. Our model therefore inferred from this landmark-deprived keyframe that performance will be poor at that location. However, we found that the scene appearance remained consistent throughout the repeat runs of the training set, which provided a higher than expected localisation score. Despite their low number, the landmarks lie mostly on the skyline, providing robust matching.

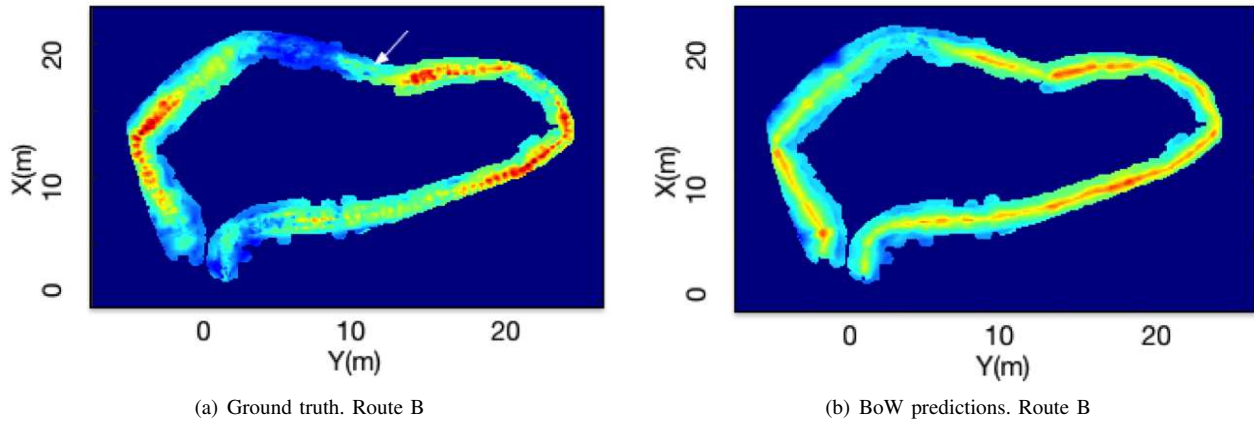


Fig. 6. Scenario II: Localisation quality predictions for training on Route A and testing on Route B. Red represents a high number of landmark matches (>80), blue represents a low number of matches (<20). The BoW model captures the variability of the envelope well along the path, but is overconfident in a few select areas. The white arrow on the ground truth figure, indicates a section of the path where we over-estimate our ability to localise. We refer the reader to the discussion section where we further elaborate on this example

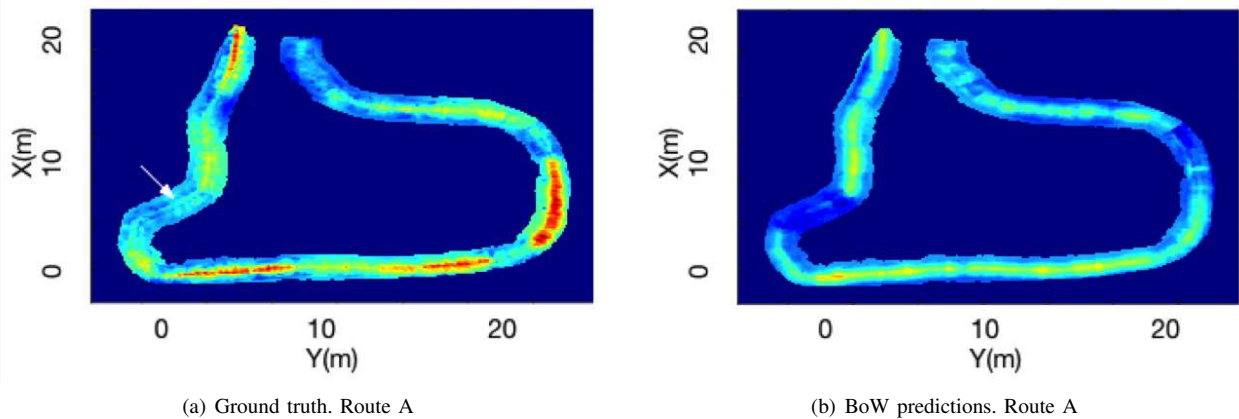


Fig. 7. Scenario III: Localisation quality predictions for training on Route B and testing on Route A. Red represents a high number of landmark matches (>80), blue represents a low number of matches (<20). The model accurately captures the localisation performance for the turns along the trajectory but is overall under confident. The white arrow on the ground truth figure, indicates a section of the path where we particularly under-estimate our ability to localise. We refer the reader to the discussion section for a more detailed study of this example.

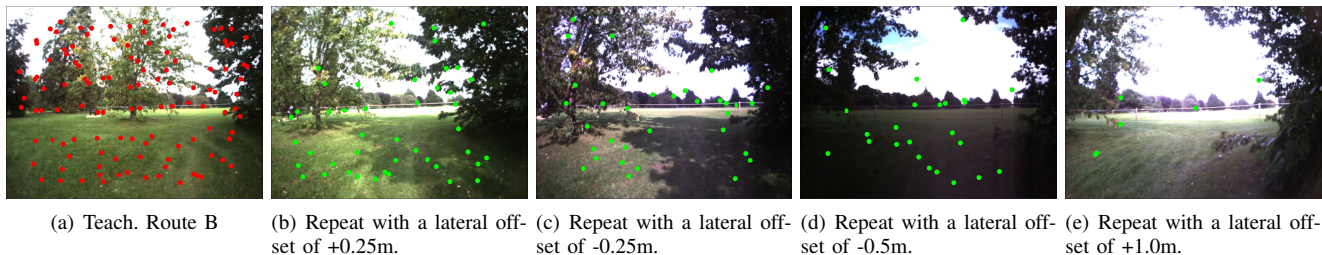


Fig. 8. Teach and repeat images for a location of Route B where the localisation prediction was optimistic in *Scenario II*, but the ground truth performance was poor. We overplot in red the landmarks in the teach pass keyframe, and in green the landmarks that effectively found a match in the repeat images. The keyframe shows good contrast and as many as 102 landmarks to localise against which poses it as a good candidate for localisation. The appearance of the repeat passes is however dramatically altered by environmental conditions resulting in a low number of matches. The two repeats shown to the right of the features effectively did not success in matching to the keyframe, and we here represent the match to the closest keyframe in the map. The model was not able to capture these environmental effects.

There exist a number of limitations to our model. The primary limitation is that we consider the keyframes collected during the teach pass but do not account for a potential change in appearance during repeat. This can result in confusion for the regressor where both high and low localisation scores result for the same input feature vector, which would be accounted for in the model as measurement noise instead of modelling error. Other limitations include our choice of keyframe representation, and localisation performance

metric. Although FAST/BRIEF descriptors are considered for localisation, we model the appearance with BoW features based on SURF detectors and descriptors. This mismatch may cause poorer performance. In addition, localisation performance seems to correlate with the quality rather than the quantity of detected features. An interesting follow-up to this work would be to unify the representation to which features are particular amenable to robust localisation, as indicated in Figure 9.

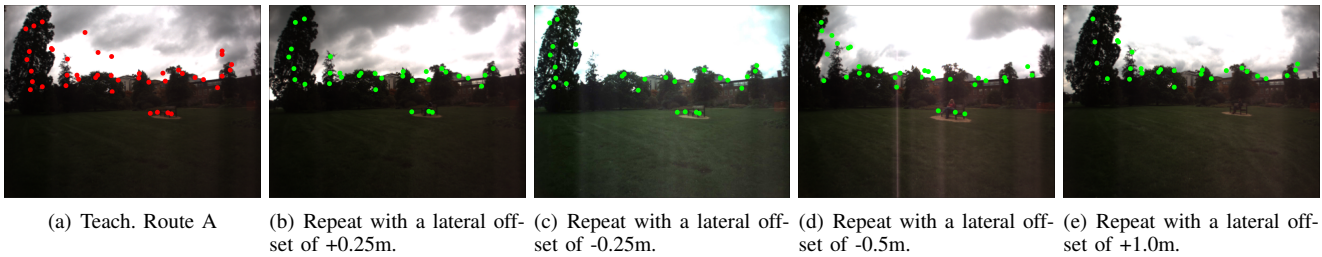


Fig. 9. Teach and repeat images for a location of Route A where the localisation prediction was pessimistic with regards to the ground truth in *Scenario III*. We overplot in red the landmarks in the teach pass keyframe, and in green the landmarks that effectively found a match in the repeat images. The keyframe shows low contrast due to overcast conditions and offers only 43 landmarks for localisation. The appearance of the repeat passes remains however admirably similar to that of the teach phase resulting in nearly perfect matching during the repeats.

VII. CONCLUSIONS

This paper proposes an appearance-based approach to predicting the likely localisation envelope given only a single teach pass at an otherwise novel location. In doing so it aims to aid the planning process in the context of visual teach and repeat systems. Our results indicate that our method successfully generalises to provide if not an exact, then nevertheless a faithful representation of the localisation envelope of novel trajectories – the key objective of this work. While the consideration of appearance features provides the capacity to generalise, it is also the source of our method’s main limitations. Just as vision-based localisation itself, our approach relies on the matching of features which are prone to change due to environmental conditions such as location, lighting or occlusion. This is likely the area where most stands to be gained in improving this method. In the spirit of experience-based navigation as introduced in [7], we posit that implicitly learning a number of representations for dissimilar circumstances such as location, time or weather will alleviate this shortcoming. We leave this to future work.

ACKNOWLEDGEMENT

The authors would like to gratefully acknowledge support of this work by the UK Engineering and Physical Sciences Research Council (EPSRC) under grant numbers EP/M019918/1 and EP/J012017/1 as well as by the UK Space Agency (grant number ST/L002981/1) as part of the CREST-2 Initiative.

REFERENCES

- [1] C. Linegar, W. Churchill, and P. Newman, “Work smart, not hard: Recalling relevant experiences for vast-scale but time-constrained localisation,” in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, 2015.
- [2] G. H. Lee, M. Pollefeys, and F. Fraundorfer, “Relative pose estimation for a multi-camera system with known vertical direction,” in *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014.
- [3] P. Muehlfellner, P. Furgale, W. Derendarz, and R. Philippsen, “Evaluation of Fisheye-Camera Based Visual Multi-Session Localization in a Real-World Scenario,” in *IEEE Intelligent Vehicles Symposium (IV)*, 2013.
- [4] K. Konolige, J. Bowman, J. Chen, P. Mihelich, M. Calonder, V. Lepetit, and P. Fua, “View-based maps,” *The International Journal of Robotics Research (IJRR)*, 2010.
- [5] F. Dayoub and T. Duckett, “An adaptive appearance-based map for long-term topological localization of mobile robots,” in *Intelligent Robots and Systems (IROS). IEEE/RSJ International Conference on*, 2008.
- [6] P. Furgale and T. D. Barfoot, “Visual teach and repeat for long-range rover autonomy,” *Journal of Field Robotics (JFR)*, 2010.
- [7] W. Churchill and P. Newman, “Experience-based navigation for long-term localisation,” *The International Journal of Robotics Research (IJRR)*, 2013.
- [8] G. Sibley, C. Mei, I. Reid, and P. Newman, “Planes, trains and automobiles - autonomy for the modern robot,” in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, 2010.
- [9] H. Strasdat, A. J. Davison, J. Montiel, and K. Konolige, “Double window optimisation for constant time visual slam,” in *Computer Vision (ICCV). IEEE International Conference on*, 2011.
- [10] C. J. Ostafew, A. P. Schoellig, and T. D. Barfoot, “Visual teach and repeat, repeat, repeat: Iterative learning control to improve mobile robot path tracking in challenging outdoor environments,” in *Intelligent Robots and Systems (IROS). IEEE/RSJ International Conference on*, 2013.
- [11] S. Martin and P. Corke, “Long-term exploration & tours for energy constrained robots with online proprioceptive traversability estimation,” in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, 2014.
- [12] P. Ondruska, C. Gurau, L. Marchegiani, C. H. Tong, and I. Posner, “Scheduled perception for energy-efficient path following,” in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, 2015.
- [13] J. Velez, G. Hemann, A. S. Huang, I. Posner, and N. Roy, “Planning to perceive: Exploiting mobility for robust object detection,” in *International Conference on Automated Planning and Scheduling*, 2011.
- [14] W. Churchill, C. H. Tong, C. Gurău, I. Posner, and P. Newman, “Know your limits: Embedding localiser performance models in teach and repeat maps,” in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, 2015.
- [15] C. E. Rasmussen, “Gaussian processes for machine learning,” 2006.
- [16] 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 (JFR)*, 2014.
- [17] E. Rosten, G. Reitmayr, and T. Drummond, “Real-time video annotations for augmented reality,” *International Symposium on Visual Computing (ISVC)*, 2005.
- [18] M. Calonder, V. Lepetit, M. Ozuysal, T. Trzcinski, C. Strecha, and P. Fua, “Brief: Computing a local binary descriptor very fast,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 2012.
- [19] M. A. Fischler and R. C. Bolles, “Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography,” *Communications of the ACM*, 1981.
- [20] M. Cummins and P. Newman, “Appearance-only slam at large scale with fab-map 2.0,” *The International Journal of Robotics Research (IJRR)*, 2011.
- [21] J. Sivic and A. Zisserman, “Video Google: A text retrieval approach to object matching in videos,” in *Proceedings of the International Conference on Computer Vision (ICCV)*, 2003.