2007 IEEE International Conference on
Robotics and Automation
Roma, Italy, 10-14 April 2007

Recognising and Modelling Landmarks to Close Loops in Outdoor SLAM

Fabio T. Ramos Juan Nieto Hugh F. Durrant-Whyte

Abstract— **In this paper, simultaneous localisation and mapping (SLAM) is combined with landmark recognition to close large loops in unstructured, outdoor environments. Camera and laser information are fused to recognise and create appearance models for landmarks. The representation is obtained through a non-linear probabilistic regression model encoding a neighbourhood preserving dimensionality reduction. A new data association algorithm is proposed where landmarks are associated based on both position and appearance. The resulting system is more robust and able to recover from possible misassociations. Experiments demonstrate the benefits of this approach in challenging problems involving mapping with large loop closings in irregular terrain, and with dynamic objects.**

I. INTRODUCTION

Simultaneous localisation and map building (SLAM) has been a long-standing problem in robotics. Although the computational complexity has been addressed in previous work [1], [2], [3], [4], reliable operation in unstructured environments is still difficult in practice. In SLAM, the robot must identify and associate landmarks with current observations to close loops correctly. For large loops, as the uncertainty in vehicle and landmark position increases quickly, a data association algorithm based entirely on position estimates is likely to fail. A further difficulty when operating in natural environments is the existence of extraneous people and moving objects. If the robot erroneously considers one of these moving objects as a landmark, the map and the trajectory estimates can become inconsistent and the localisation algorithm will fail.

In this paper, landmark recognition is incorporated into SLAM to address these two issues. Recognition of landmarks is performed using both camera and laser, two of the most common sensors used in robotic platforms. The system is trained to recognise common static outdoor landmarks, such as trees, which directly eliminates the problem of operating in dynamic environments. Trees are common objects in both urban and rural areas which makes the system flexible for operation in many different environments. As trees differ in their shape, texture or colour they make excellent landmarks for appearance-based association. Once recognised, landmarks are used to build probabilistic models for further data association. Experimental results in a complex environment show that the new approach significantly improves over conventional data association approaches based entirely on poses.

F. Ramos, J. Nieto and H. Durrant-Whyte are with ARC Centre of Excellence for Autonomous Systems, School of Aerospace, Mechanical and Mechatronic Engineering, The University of Sydney 2006 NSW, Australia {f.ramos,j.nieto,hugh}@acfr.usyd.edu.au

The use of visual clues in identifying features has been previously investigated in [5]. In their work, visually salient features are extracted from images and used in combination with laser scans for loop-closing. The use of feature representations improves data association but does not provide meaningful representations of the objects in the scene. In this sense, it is unlikely that feature-based approaches would cope with dynamic environments. Se et al. [6] have proposed an entirely image-based SLAM with SIFT features and stereo camera. The approach was tested in an office environment over a trajectory of few metres. Feature selection for pose estimation was addressed in [7]. Landmark tracking was performed using principal component analysis over a sequence of frames in a laboratory environment, but no map was constructed.

Recent results from Moreels and Perona [8] have shown that no feature detector-descriptor combination performs well for viewpoint changes of more than $25 - 30^{\mathrm{o}}$. However, in outdoor environments, landmarks can be observed at a much wider viewpoint range. Richer and more robust representations of landmarks are thus investigated in this work. These representations must also be general, not relying on ad-hoc features, to be applicable to different field robotics tasks, in aerial, underwater and all terrain domains. The solution adopted involves dimensionality reduction techniques and, especially, non-linear methods. Several manifold learning approaches have been proposed in recent years [9], [10], [11], [12]. Due to its isometric and convergence properties with the landmark extension [13], which enables it to process large datasets, Isomap [9] is used to create landmark representations. Out-of-sample points from Isomap are obtained by fitting a mixture of linear models from the input and output sets of Isomap. This probabilistic model also provides uncertainty measures to perform compatibility tests for data association.

The novelty and importance of the work described is the introduction of landmark recognition and modelling into SLAM. The approach is demonstrated in a large and complex unstructured environment where current SLAM implementations fail. The vehicle trajectory includes multiple loops with up to 400 metres in length. The illumination changes from direct sunlight in a partly cloudy day to dark areas where the vehicle is underneath foliage and trees. The total trajectory length is approximately 1.5 kilometre with the vehicle travelling up to 25 km/h. The environment also includes moving objects such as cars, buses and people. Around 120 different landmarks are detected and mapped. Some of these are observed from viewpoint changes of 180^{o}. Experimental results show that some loops are not correctly

1-4244-0602-1/07/$20.00 ©2007 IEEE.

closed with conventional SLAM algorithms due to incorrect data association. However, when the proposed algorithm is applied to the same data, misassociations are much reduced and when they do occur the system is still able to recover. Recognition and representation of landmarks significantly improved robustness of the SLAM method.

II. Landmark Recognition

To recognise landmarks far from the robot, a high-resolution camera is necessary to provide a sufficient number of pixels for processing. However, real-time recognition in high-resolution images remains a cumbersome task even for speed-optimised algorithms such as [14]. In the proposed approach, a laser scan is used to select regions of interest (ROI) in the image. Only these regions are processed which allows the algorithm to recognise landmarks in real-time. Laser is thus responsible for allocating visual attention to landmark recognition.

The recognition process has four steps to be described as follows: laser clustering, sub-image processing, dimensionality reduction and classification.

A. Laser Clustering and Feature Extraction

In outdoor environments, a typical laser scan is characterised by the existence of clusters of laser points sparsely distributed. Thus, the first step of the algorithm is to identify and associate points belonging to the same cluster. This clustering operation is unsupervised in the number of clusters and must be performed in real-time. The solution adopted uses a simple recursive algorithm that cluster laser points whenever the distance of a point to its nearest neighbour in a particular cluster is smaller than a defined value - 1 metre is used in the experiments.

To extract shape information from identified clusters, points are first translated to the origin by subtracting the cluster mean. They are then rotated by the angular position of the cluster centre with respect to the robot orientation. This operation tries to make laser points invariant to affine transformations. A second-order polynomial is fitted to the points and the coefficients are the final features used for recognition. This shape descriptor is very simple when compared to spline, Fourier or snakes descriptors used in computer vision. However, it is very efficient and can be computed with only three or more laser points in the same cluster.

B. Defining ROI with Laser

Given a calibrated camera with intrinsic parameters K, the corresponding pixel coordinates $p = [u, v]^T$ of points in the world coordinate system $P = [X, Y, Z]^T$ can be computed as $p \sim K(RP + t)$, where R is a 3×3 rotation matrix and t is a 3D-vector representing translation. Assuming that P_l is a point in the laser coordinate system and P_c in the camera, the equation $P_l = \Phi P_c + \Delta$ represents their transformation, where Φ and Δ are rotation and translation parameters for the laser-camera calibration. To compute Φ and Δ, the method described in [15] is used. The outputs are the optimised parameters Φ and Δ that, in conjunction with the intrinsic parameters K of the camera, allow the projection of laser points into images.

Laser points belonging to the same cluster are used to define a ROI of size $W \times H$, where W is set to twice the distance between the projected laser points at the extremes of the cluster. H is set to twice the size of W. The width of the ROI is larger than the size of the object to account for no-cylindrical shapes and misalignments between camera and laser readings when obtained at slightly different time. The ROI is then resized to 75×150. This process also makes the ROI less sensitive to scale variations since the actual size of the object obtained from laser is used to define the size of the ROI.

C. Dimensionality Reduction and Classification

ROI are then convolved with Gabor wavelets at 2 scales and 2 orientations to introduce texture information. Similar results can be obtained with steerable pyramids [16]. The RGB values are converted to HSV and histogram equalisation is applied to the intensity values to account for illumination changes.

The resulting patch is a point in a 78753-dimensional space $(75 \times 150 \times 7 + 3)$ which includes HSV colour information, texture and laser shape coefficients. This high-dimensional observation is used for both object recognition and data association. Classification of ROI is obtained using an extension to the Fisherfaces [17] which is described below. This approach performed better than [14] in our tests at no significant additional computational cost. One reason for that is the inclusion of colour and texture information which enables the system to distinguish between objects of similar shape such as tree trunks and light posts.

Classification of patches is performed in two steps. First, high-dimensional points are projected to a lower dimensional space where classes are more distinguishable. Then, logistic regression is applied. In Fisherfaces a linear projection of high-dimensional points is obtained by first applying principal component analysis to the data to reduce the dimensionality to $N - c$, where N is the number of points and c the number of classes. This process avoids the problem of computing Fisher discriminant analysis (FDA) with an always singular within-class scatter matrix S_W which has more dimensions than samples. The resulting points are then mapped to a low-dimensional space with $c - 1$ dimensions with normal FDA. Formally, the problem is defined as the computation of W_{opt}^T with

$$W_{opt}^T = W_{fda}^T W_{pca}^T, \tag{1}$$

where

$$W_{pca} = \arg\max_W \left| W^T S_T W \right| \tag{2}$$

and

$$W_{fda} = \arg\max_W \left| \frac{W^T W_{pca}^T S_B W_{pca} W}{W^T W_{pca}^T S_W W_{pca} W} \right|. \tag{3}$$

Fig. 1. Recognised landmarks mapped during SLAM.

The within-class scatter matrix is defined as

$$S_W = \sum_{i=1}^{c} \sum_{\mathbf{x}_k \in X_i} (\mathbf{x}_k - \mu_i)(\mathbf{x}_k - \mu_i)^T \quad (4)$$

and the between-class as

$$S_B = \sum N_i (\mu_i - \mu)(\mu_i - \mu)^T, \quad (5)$$

where μ_i is the mean of class X_i, and N_i is the number of samples in class X_i. The total scatter matrix is thus $S_T = S_B + S_W$. The problem is solved by first computing a singular value decomposition for PCA and a generalised eigenvalue problem for FDA.

Landmark recognition involves the classification of observations into two classes, with or without a landmark. Consequently, the resulting dimensionality of the projected points after Fisherfaces would be just one $(c-1)$. Experiments demonstrate, however, that incorporating more dimensions through the projection of the scatter matrices S_B and S_W into orthonormal sub-spaces of the initial Fisher dimension results in classification improvement. Therefore, the approach of [18] is used to include 4 more dimensions into the projected points, where orthonormalisation is obtained through the Gram-Schmidt procedure. The number of dimensions is obtained with cross-validation.

The final classification is performed with logistic regression [19]. The classifier is trained with 129 labelled observations containing images and laser scans. The average detection performance is approximately 95% (area under the receiver-operator curve) for a 10-fold cross-validation. Figure 1 shows some of the landmarks detected and mapped during SLAM.

III. Landmark Representation

For each new landmark detected, an appearance model is created. Each model is defined using a low-dimensional representation provided by Isomap. Initially, Isomap [9] is applied to a set of landmark images and shape coefficients from a training dataset obtained from previous missions. The choice for Isomap over other non-linear techniques is due to its isometric property, which is important for computing distance comparisons in the feature-space for data association. Also, the landmark extension [13] reduces the computational complexity considerably, making it suitable for large datasets.

Isomap is a deterministic technique that does not directly provide out-of-sample embeddings nor uncertainty measurements. To cope with that the resulting high and low-dimensional points are used to learn a mixture of linear models through EM in a procedure similar to [20]. The mixture is used to compute out-of-sample embeddings of new landmarks and to estimate covariances for compatibility tests. Inferences in a mixture of linear models result in mixtures of Gaussians that can be computed faster than out-of-samples extensions to Isomap [21], with the additional advantage of providing uncertainty estimation. Besides, experiments using the out-of-sample extension for Isomap provided poor approximations for our dataset. This might be due to insufficient number of samples - 330 points - for a high-dimensional space of 78753. The dimensionality of the training dataset was 10, estimated from Isomap by evaluating residuals. Hence, each landmark is represented by a mixture of Gaussians in a low-dimensional space obtained by Isomap and approximated by the mixture of linear models.

When a landmark is re-observed, its appearance model is updated. This is performed by updating the sufficient statistics in a conventional manner,

$$\hat{\pi}_\omega = \frac{M\pi^e_\omega + \bar{\pi}_\omega}{M+1}, \quad (6)$$

$$\hat{\mu}_\omega = \frac{M\mu^e_\omega + \bar{\mu}_\omega}{M+1}, \quad (7)$$

$$\hat{\mathbf{\Sigma}}_\omega = \frac{M\mathbf{\Sigma}^e_\omega + \bar{\mathbf{\Sigma}}_\omega}{M+1} \quad (8)$$

and $M = M + 1$, where $\hat{\pi}_\omega$, $\hat{\mu}_\omega$ and $\hat{\Sigma}_\omega$ are the expected weights, means and covariances of the mixture, and M is the number of past observations of the landmark. The number of components of the mixture of linear models is estimated by learning models with different number of components and comparing them with the Bayesian information criterion [19].

IV. Data Association

Data association is performed using both pose estimation of the landmarks from SLAM and their appearance information. The gated nearest neighbour (NN) [22] is applied to choose the best hypothesis for association. This method is probably the simplest and most widely used data association algorithm. However, if formulated with pose and appearance representations of landmarks, it can be successful even for difficult problems involving large loop closings.

From the Kalman filter equations, on the arrival of a new observation, the state-vector estimation $\hat{\mathbf{x}}(k \mid k-1)$ and covariance $\mathbf{P}(k \mid k-1)$ are updated as:

$$\begin{aligned} \hat{\mathbf{x}}(k \mid k) &= \hat{\mathbf{x}}(k \mid k-1) + \mathbf{W}(k)\,\nu(k) \quad \text{and} \\ \mathbf{P}(k \mid k) &= \mathbf{P}(k \mid k-1) - \mathbf{W}(k)\,\mathbf{S}(k)\,\mathbf{W}(k)^T, \end{aligned} \quad (9)$$

where $\mathbf{W}(k)$ is the Kalman gain, $\nu(k)$ is the innovation vector and $\mathbf{S}(k)$ is the innovation covariance matrix at

Fig. 2. Utilitary car used in the experiments. For the results reported, the top laser and the high-resolution camera were employed.

time k. The gated NN evaluates Mahalanobis distances on the innovation vector ν_{ij} and accepts a hypothesis ij when $D^2_{ij} = \nu^T_{ij}\mathbf{S}^{-1}\nu_{ij} < \chi^2_{d,\alpha}$, where d is the feature dimensionality and α the desired confidence level of the χ^2 distribution.

The combined appearance-position version of NN works in two steps. First, the compatibility test is applied to position only. An association hypothesis is accepted whenever the best D^2_{ij} is within the gate with 95% of confidence. This saves computation as only when there is no trivial association - for example when closing loops - the appearance model is used.

In the second step, the innovation is augmented with appearance. The new innovation is $\nu_I = [\nu_P, \nu_A]^T$ with innovation covariance $\mathbf{S}_I = \begin{bmatrix} \mathbf{S}_P & 0 \\ 0 & \mathbf{S}_A \end{bmatrix}$, where P is used to indicate position and A appearance. ν_A is obtained by computing the difference between the appearance model for the landmark and the low-dimensional embedding of the observation, computed from the learnt mixture of linear models. The mixtures are collapsed using the moment matching technique that provides the best approximation in a KL-divergence sense. Experiments have shown that this procedure is faster than computing more sophisticated associations from mixtures at no major cost on accuracy. This can be explained by the fact that most inferences result in mixtures with one and occasionally two components only. The innovation covariance for appearance $\mathbf{S}_A$ is obtained in the same manner, from the appearance modes. A hypothesis is accepted if the smallest augmented distance $D^2_{I,ij}$ is within the pose-appearance gate with 95% confidence. A new landmark is included in the map if the shortest distance is larger than five times the gate.

V. Experiments

Experiments are conducted with the utilitary car shown in Figure 2. The vehicle is equipped with a laser scan, a high-resolution colour camera (1134×756), GPS for pose ground truth and odometry sensors for steering and wheels. An embedded PC104 running QNX is used to collect time stamps for each of the sensors and to trigger the camera[1]. The car is driven for approximately 15 minutes, travelling a total distance of about 1.5 kilometre. The speed ranges from 0 to 25 km/h. Images and laser scans are grabbed at 5 Hz which causes a maximum possible disynchronisation of 100 ms[2]. Multiple loops were executed during the data acquisition ranging from few metres up to 400 metres. The environment is an urban park in a partly cloudy day. The park has dark areas under trees and open areas with direct sunlight. The car was driven most of the time on an uneven terrain covered by grass.

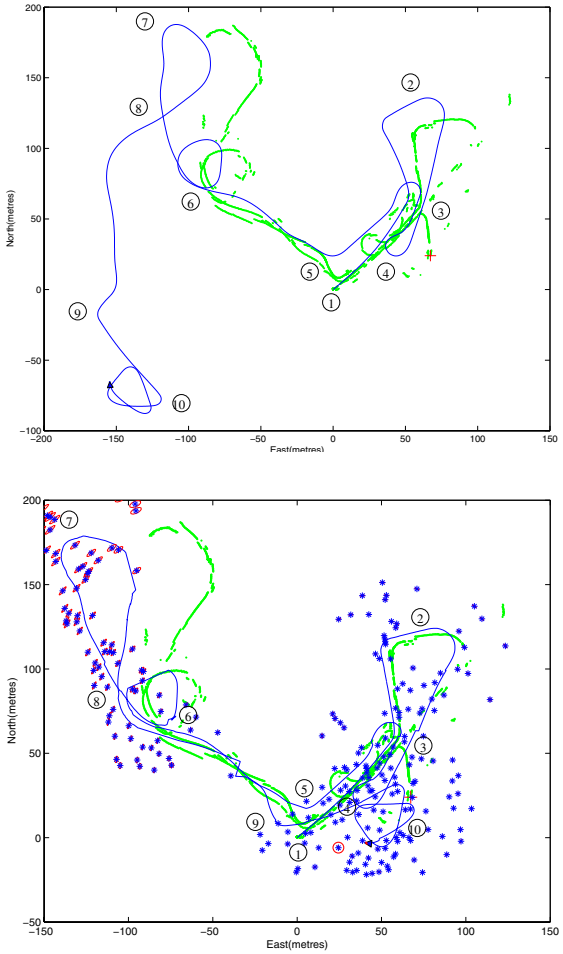

Fig. 3. (Top) Estimated trajectory using odometry only. (Bottom) Estimated trajectory and landmarks position with conventional EKF-SLAM. For comparison, the estimated trajectory with GPS is represented with the green line. The EKF-SLAM becomes inconsistent after the first large loop which is not closed properly. The error in landmarks pose and vehicle trajectory is thus propagated to the rest of the map. The numbers in the graph indicate successive positions of the trajectory, starting with position 1 at (0,0). Note that the GPS signal is not available in some areas of the park.

Estimation of the trajectory using only odometry is depicted in Figure 3 (top). As expected, odometry by itself does not provide an accurate trajectory estimation especially

[1]The real-time software used was developed by J. Guivant and F. Masson and is described at "http://www.acfr.usyd.edu.au/people/academic/jguivant/PAATV/Paatv.html"

[2]The maximum framerate for the camera, at the resolution required, is 5 Hz. The laser is thus set up to scan at the same speed, and the camera is triggered at the reception of a new laser scan. This method tries to minimise the disynchronisation between the two sensors, however, in practice, time differences of up to 80 ms do occur occasionally.

for the irregular terrain where the dataset is collected. The trajectory and the map obtained using standard EKF-SLAM with pose-based data association is depicted in Figure 3 (bottom). The trajectory is closer to the GPS estimation but the system fails to close large loops due to incorrect associations. The result of that propagates throughout the map, causing large pose errors and an inconsistent filter. Another source of problems is the false identification of landmarks caused by people or cars that were erroneously included in the map.

The starting point is indicated as position 1 (0,0) in Figure 3. After starting, the vehicle goes straight and begins the first loop which passes through position 2 and finishes at position 3. Two other small loops are then performed next to position 4. The vehicle goes back to the starting point (position 5) and turns right for the largest loop. Next to position 5 and heading to 6, the vehicle navigates in an area with few trees. Thus, localisation relies mostly on odometry which introduces the largest error in the trajectory estimation. Another loop is performed at position 6 and the largest loop passes through position 7 and finishes at 8. The vehicle goes back to position 9 and at 10 performs the last loop in a region with traffic and buildings (see Figure 4 (bottom right)). To the best of the authors' knowledge, this dataset is one of the most complex and challenging outdoor dataset where SLAM was ever applied in terms of both the trajectory (with multiple loops), velocity of the vehicle, dynamic objects, uneven terrain and illumination conditions.

Extended Kalman filter (EKF) is used for SLAM due to its convergence properties and accuracy in uncertainty estimation [23]. The techniques presented here are, however, independent of a particular filter and are beneficial to all of them. Map and trajectory obtained using the proposed approach are depicted in Figure 4. The system is able to close all loops correctly resulting in a much more accurate map. The number of landmarks detected is smaller than conventional SLAM demonstrating that the algorithm is able to eliminate dynamic objects from the map. This can be particularly seen from the area next to the road (position 10), where conventional SLAM included cars and buses in the map. The left part of the map is slightly rotated from the GPS ground truth due to the reduced number of trees between poses 5 and 6, causing the vehicle to rely only on odometry when turning. Some straight lines represent points where the system is adjusting its position when re-observing landmarks (loop next to pose 6, closing the large loop next to 8). The images next to the graph show the landmark recognition algorithm performing in the environment. In the bottom-left image, three trees are recognised as landmarks while the person and the post are correctly assigned as non-landmarks. In the bottom-right image, buses are correctly classified as non-landmarks but the algorithm misclassified the pile. The reason for this is the partial occlusion of the pile with a branch of a tree. Occlusions by tree parts are indeed responsible for most of the misclassifications encountered. Quantitatively, the RMS error for the estimated trajectory using conventional EKF-SLAM was 20.5 metres. With our approach the RMS error was 8.6 metres. The error was calculated using GPS measurements in areas where it was available. This represents a significant improvement over the conventional approach. The result is also close to the GPS error itself which is around 5 metres.

Fig. 5. Multiple views of one of the landmarks acquired during SLAM. As with this particular landmark, during outdoor exploration, the vehicle observes landmarks at completely different viewpoints. The yellow crosses represent laser points projected on the object. As laser scans and images are not perfectly synchronised, misalignment exists especially when the vehicle is turning (last image on the right).

The capability to associate landmarks observed at different viewpoints is one of the main properties of the algorithm. Figure 5 illustrates this for a landmark observed 411 times at multiple viewpoints during exploration. When appearance is not enough for association, the estimated position from SLAM can compensate, improving the robustness of the algorithm. With an observation from a different viewpoint correctly associated, the appearance model for the landmark is updated to account for the new information.

VI. Discussion

The proposed algorithm improves significantly the performance of outdoor SLAM. This improvement results in the additional computational cost of detecting and creating appearance models of landmarks. Although the computational complexity has been taken into account in the design of the algorithms, the system is still not real-time. The current implementation can process about 1 frame per second where most of the computation is spent in performing inferences on the mixture of linear models for dimensionality reduction. Current research tries to reduce this time by using specific processor instructions for vector operations.

The approach was tested in a very challenging urban environment with dynamic objects, irregular terrain and different illumination conditions. The system performs better than pose-based data association and is able to correctly close loops of more than 400 metres while recognising and associating 120 different landmarks. An interesting property of the algorithm is the capacity to recover from occasional misassociations. This is possible with the help of the appearance models that can compensate from incorrect position estimates.

In addition to the map and trajectory estimation, our approach also provides images of landmarks with the relative positions where they are obtained. Given a calibrated camera, this information can be combined to create 3D models of landmarks. We plan to investigate techniques for landmark 3D modelling in future work.

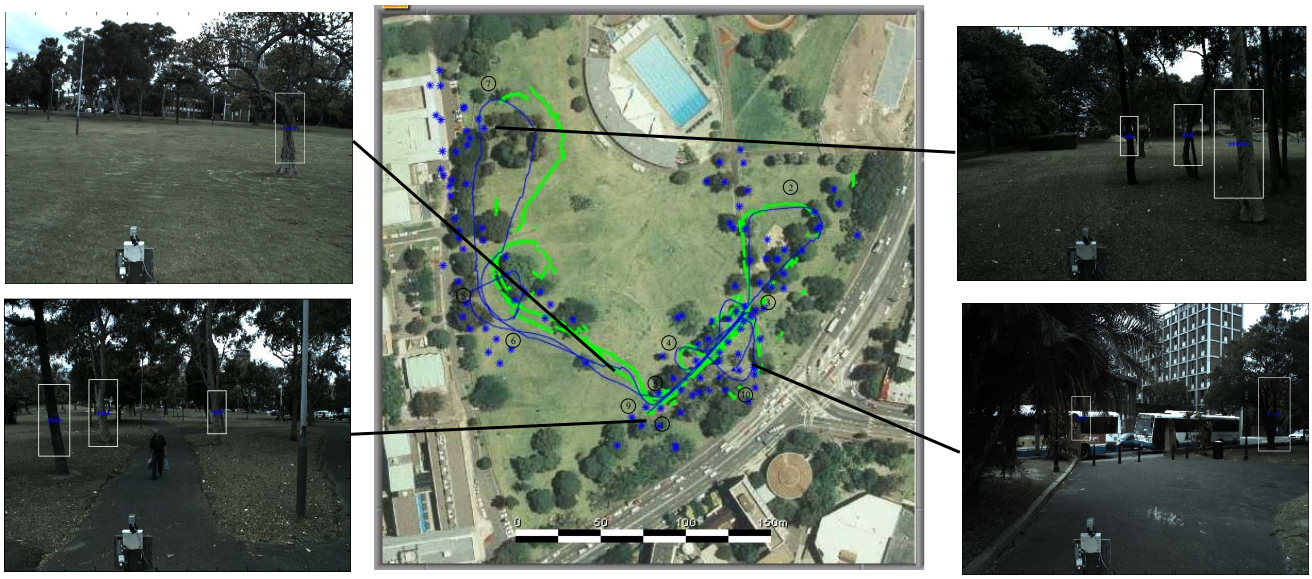

Fig. 4. Map and vehicle trajectory obtained using the proposed approach. Results overlay an aerial picture of the area. The images next to the map are screenshots taken from the indicated positions. The white boxes define regions where landmarks were detected.

ACKNOWLEDGEMENTS

This work is supported by the ARC Centres of Excellence programme funded by the Australian Research Council (ARC) and the New South Wales State Government. The authors would like to thank Roman Katz, Oliver Frank, Jose Guivant and Christel-Loic Tisse for helping with the dataset acquisition.

REFERENCES

[1] J. A. Castellanos, M. Devy, and J. D. Tardos, "Simultaneous localisation and map building for mobile robots: A landmark-based approach," in *Proceedings of IEEE international Conference on Robotics and Automation: Workshop on Mobile robot Navigation and Mapping*, San Francisco, USA, 2000.

[2] M. Montemerlo and S. Thrun, "Simultaneous localization and mapping with unknown data association using FastSLAM," in *Proceedings of the International Conference on Robotics and Automation (ICRA)*, Taipei, Taiwan, 2003.

[3] S. B. Williams, G. Dissanayake, and H. F. Durant-Whyte, "An efficient approach to the simultaneous localization and mapping problem," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, Washington, USA, 2002.

[4] M. Bosse, P. Newman, J. Leonard, and S. Teller, "Simultaneous localization and map building in large-scale cyclic environments using the Atlas framework," *International Journal of Robotics Research*, vol. 23(12), pp. 1113–1139, 2004.

[5] P. Newman and K. Ho, "SLAM - loop closing with visually salient features," in *Proceedings of the International Conference on Robotics and Automation (ICRA)*, Barcelona, Spain, 2005.

[6] S. Se, D. Lowe, and J. Little, "Mobile robot localization and mapping with uncertainty using scale-invariant visual landmarks," *International Journal of Robotics Research*, vol. 21(8), pp. 735–758, 2002.

[7] R. Sim and G. Dudek, "Learning and evaluating visual features for pose estimation," in *Proceedings of the Seventh International Conference on Computer Vision (ICCV)*, Kerkyra, Greece, 1999.

[8] P. Moreels and P. Perona, "Evaluation of features detectors and decriptors based on 3D objects," in *Proceedings of the International Conference on Computer Vision (ICCV)*, 2005.

[9] J. Tenenbaum, V. DeSilva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality redution," *Science*, vol. 290, pp. 2319–2323, 2000.

[10] S. Roweis and L. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290(5500), pp. 2323–2326, 2000.

[11] M. Belkin and P. Niyogi, "Laplacian eigenmaps for dimensionality reduction and data representation," *Neural Computation*, vol. 15(6), pp. 1373–1396, 2003.

[12] K. Q. Weinberger and L. K. Saul, "Unsupervised learning of image manifolds by semidefinite programming," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR-04)*, Washington, USA, 2004, pp. 988–995.

[13] V. DeSilva and J. Tenenbaum, "Global versus local methods in nonlinear dimensionality reduction," *Advances in Neural Information Processing Systems M.SBecker, S., Thrun, S., and Obermayer, K. (eds).*, vol. 15, pp. 705–712, 2002.

[14] P. Viola and M. J. Jones, "Robust real-time face detection," *International Journal of Computer Vision*, vol. 57(2), pp. 137–154, 2004.

[15] Q. Zhang and R. Pless, "Extrinsic calibration of a camera and laser range finder (improves camera calibration)," in *Proceedings of the International Conference on Intelligent Robots and Systems (IROS)*, Sendai, Japan, 2004.

[16] W. Freeman and E. Adelson, "The design and use of steerable filters," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13(9), pp. 891–906, 1991.

[17] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19(7), pp. 711–720, 1997.

[18] T. Okada and S. Tomita, "An optimal orthonormal system for discriminant analysis," *Pattern Recognition*, vol. 18(2), pp. 139–144, 1985.

[19] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference and Prediction*. New York, USA: Springer-Verlag, 2001.

[20] F. Ramos, S. Kumar, B. Upcroft, and H. F. Durrant-Whyte, "Representing natural objects in unstructured environments," in *NIPS 05 - Workshop on Machine Learning Based Robotics in Unstructured Environments*, Whistler, Canada, 2005.

[21] Y. Bengio, J.Paiement, P. Vincent, O. Delalleau, N. L. Roux, and M. Ouimet, "Out-of-sample extensions for LLE, isomap, MDS, eigenmaps, and spectral clustering," in *Advances in Neural Information Processing Systems*, Vancouver, Canada, 2004.

[22] S. Blackman and R. Popoli, *Design and Analysis of Modern Tracking Systems*. Norwood, USA: Artech House Publishers, 1999.

[23] G. Dissanayake, P. Newman, S. Clark, and H. F. Durrant-Whyte, "A solution to the simultaneous localisation and map building (SLAM) problem," *IEEE Transactions on Robotics and Automation*, vol. 17(3), pp. 229–241, 2001.