Outdoor SLAM using Visual Appearance and Laser Ranging

P. Newman, D. Cole and K. Ho
Oxford University Robotics Research Group. Email: \{pnewman,klh,dmc\}@robots.ox.ac.uk

Abstract—This paper describes a 3D SLAM system using information from an actuated laser scanner and camera installed on a mobile robot. The laser samples the local geometry of the environment and is used to incrementally build a 3D point-cloud map of the workspace. Sequences of images from the camera are used to detect loop closure events (without reference to the internal estimates of vehicle location) using a novel appearance-based retrieval system. The loop closure detection is robust to repetitive visual structure and provides a probabilistic measure of confidence. The images suggesting loop closure are then further processed with their corresponding local laser scans to yield putative Euclidean image-image transformations. We show how naive application of this transformation to effect the loop closure can lead to catastrophic linearization errors and go on to describe a way in which gross, pre-loop closing errors can be successfully annulled. We demonstrate our system working in a challenging, outdoor setting containing substantial loops and beguiling, gently curving traversals. The results are overlaid on an aerial image to provide a ground truth comparison with the estimated map. The paper concludes with an extension into the multi-robot domain in which 3D maps resulting from distinct SLAM sessions (no common reference frame) are combined without recourse to mutual observation.

I. INTRODUCTION

We would like to have a robot perform Simultaneous Localization and Mapping (SLAM) outdoors. We aim to replace the now ubiquitous 2D, planar maps generated from moving indoors on flat floor with rich 3D maps built from vehicles moving on more general outdoor terrain in which estimated vehicle trajectories are now embedded in $\mathbb{R}^6$. Moreover we wish to do this over large areas and accommodate large loops. We do not constrain ourselves to try and do this using only one sensor, for example laser or camera. Instead this work brings together two complementary threads of research - SLAM using a 3D laser scanner and loop closure detection using photometric information. The resulting system which we describe here is, to our knowledge, the first time a 3D laser-vision SLAM system with automated loop closure detection has been showcased working in a typical outdoor urban environment.

We begin in Section II by describing the 3D laser scanner and the scan-match based SLAM framework we choose to employ. In Section III we introduce the sequential, appearance based loop-closure detection algorithm. We then describe how, following the detection of a potential loop closure, a rigid transformation is produced that relates the current location of the vehicle to a previous one. In Section IV we describe how this transformation can be fused with current map and trajectory estimates despite the presence of potentially gross errors in vehicle location and effect the loop closure. Section V presents results generated by this system when applied to a data set of an outdoor site and superimposes the results on an aerial site photograph for comparison. In Section V-A the work is extended to show how the visual loop closing system can be used to guide the fusion of multiple maps (for example built simultaneously by a different vehicle or the same vehicle on a different day) with no common co-ordinate frames to yield a larger combined map. The paper concludes with a discussion of current shortcomings and further work.

II. GEOMETRY FROM LASER

Previous work on 3D SLAM on mobile platforms includes work by Davison and Kita [5], who use purely vision, and more recent work by Weingarten and Siegwart [24] who use 3D laser range data. They subsequently extract planar features from the scans and maintain parameterized estimates in a probabilistic manner using the SP formulation [2]. The work by Surmann, Nuchter, Lingemann and Hertzberg in [22] and [23] is notable. They use sequential registrations to fuse multiple 3D laser range scans together, and when loop-closing, use batch update techniques to redistribute registration error. This work presents strong results, but it is not clear how loop closures are detected. The 3D SLAM system we use is a 3D extension of the delayed state formulation in [1] and [17] and has much in common with work in [8], [14] and more recently [7]. The underlying SLAM representation is a state vector of past vehicle poses. At a suitable interval the state vector is augmented with a new vehicle pose using odometry information $u(k)$ between times $k$ and $k+1$:

$$x(k+1|k) = \left[ \begin{array}{c} x(k|k) \\ x_v(k|k) \oplus u(k+1) \end{array} \right]$$

$$= \left[ \begin{array}{c} x_v \, T \\ \cdots \\ x_v \, T \\ x_{v+1} \, T \end{array} \right] (k+1|k)$$

where $\oplus$ is the $SE_3$ transformation composition operator and $x_v$ is the $i$th vehicle pose in the state vector. Here we are using the standard conditional $(p|q)$ notation to denote an estimate at time $p$ conditioned on observations up until time $q$. Associated with this state vector is a covariance matrix $P$ with the following structure:

$$P(k+1|k) = \left[ \begin{array}{c} P(k|k) \\ P_v(k+1|k) \\ P_v(k+1|k) \end{array} \right]$$

where $P_v$ is the covariance of the newly added vehicle state. As the vehicle moves, a ‘nodding’ laser scanner returns a stream of planar scans at different elevations. Each pose
The underlying joint p.d.f. is loop closure. However, this assumes that the estimate of the Mahalanobis distance between two poses could then indicate p.d.f to detect loop closure as in [16]. A sufficiently small sometimes effective approach would be to look to the SLAM recurrence of a loop closure can be detected. A simple and visually.

...will be used throughout this paper. The figure shows the evolution of the state’s individual poses, from our vehicle’s initial position, until the moment before the loop closer prompted a loop closure between the last (current) and first pose. The corresponding 1σ marginal ‘x, y, z’ uncertainty ellipsoids are plotted around each individual pose state. Note the vehicle has actually returned to the origin — far outside the possibilities admitted by the uncertainty ellipses. The grid is marked in intervals of 20m.

III. DETECTING LOOP CLOSURE WITH VISION

In this section we will describe a new approach to detecting loop closure using images. The method is appearance based and uses the similarity between local scene descriptions to find statistically significant pairings between sequences of images. The images we are considering here are ordered — the vehicle captures them sequentially as it moves through its workspace meaning images captured close in time also have a spatial proximity.

...that they are indeed two views of the same place. If \( I_u \) and \( I_v \) look the same then there is an elevated probability that they are indeed two views of the same place. If \( I_u \) was captured at time step \( k_u \) and \( I_v \) at \( k_v \) then the vehicle’s poses at times \( k_u \) and \( k_v \) should be close\(^1\). However, finding just one image pair correspondence may not always be a strong enough basis on which to suggest a loop-closure event. As a simple illustrative example, the exterior of buildings frequently present repetitive architectural structure, and indoors many doorways look the same. However if we can chain correspondence pairs together to form sequences of paired images we can increase our confidence that the two strands are multiple views of the same scene — they are similar throughout an extended spatial area.

The method described in this section explicitly tackles difficulties in differentiating scenes due to repetitious low-level descriptors (common texture) or broad, background interscene similarity (common large-scale features\(^2\)). The procedure produces sequences of paired, time stamped images. Each sequence represents a putative loop closure, and has an associated probability, conditioned on the totality of collected images, that this is a genuine loop closure.

\(^1\)assuming that not all of the scene content is in the far field

\(^2\)like windows, plants or Victorian architecture — see Figure 4
This similarity function \(^3\) allows the creation of a Similarity Matrix \(M\) which encodes the similarity between all \(N\) images. Each element \(M_{i,j}\) is \(S(i, j)\) the similarity score between \(I_i\) and \(I_j\). A typical Similarity Matrix (and one that will be processed extensively as an example throughout this paper) is shown in Figure 2. Two things are immediately apparent: firstly the strong diagonal which stems from all images being self similar, and secondly the presence of off-diagonal streaks. These are the loop closures we are seeking to detect — sequences of temporally separated, yet similar scenes.

\[ H_{i,j} = \begin{cases} H_{i-1,j-1} + M_{i,j} & \text{if } M_{i,j} \leq \tau, \langle A_i, B_j \rangle \text{ maximal} \\ H_{i-1,j-1} + M_{i,j} - \delta & \text{if } M_{i,j} > \tau, \langle A_i, B_j \rangle \text{ maximal} \\ \alpha \max(H_{i-1,j-1}, H_{i,j-1} - H_{i-1,j-1}) & \text{otherwise} \end{cases} \]

The fourth case, where \(0 < \alpha < 1\), allows for gaps in the sequence of matches typically caused by an obscured field of view (e.g. a pedestrian walking in front of the camera). The constant \(\tau\) is a tolerance threshold which can be set using the statistical analysis in section III-D. When every cell has been visited the maximally scoring sequence \(\langle A, B \rangle\) of paired images can be unwound by back-tracing through \(H\) starting at the maximum element in \(H\) whose value we denote as \(\eta_{A,B}\).

\[ S(u, v) = \frac{\sum_{i=0}^{\mid V \mid} u_i v_i}{\sqrt{\sum_{i=0}^{\mid V \mid} u_i^2 \sum_{i=0}^{\mid V \mid} v_i^2}}. \quad (4) \]

\(^3\)Implementation note : because the majority of the elements in the vectors \(I_u\) and \(I_v\) will be zero (they generally use only a fraction of the entire vocabulary), the calculation in 4 benefits greatly from an efficient implementation that iterates only over non-zero members.
Figure 2). We now ask how the algorithm will perform in an environment that is more visually confusing, such as the one resulting in the visual similarity matrix shown on the left hand side of Figure 3. An off-diagonal dark line starts at around image 400 — this is the start of the genuine loop closure. However there are also numerous dark (mutually similar) off-diagonal regions. These are typically caused by repetitive imaging of architectural features like windows, long brick walls or broadly homogenous foliage. We loosely describe these as “themes”. Themes cause collections of words to appear together across multiple scenes resulting in a (correct) degree of mutual similarity. The resulting blocks and stripes in $M$ make reliable sequence extraction problematic. We wish to remove the effects of these themes in the similarity matrix so that the sequence detection relies more on the unusual similarities between scenes than the common-mode terms. We achieve this by a rank reduction technique which we now describe.

The similarity matrix is decomposed into a sum of rank one matrices formed from outer products

$$M = \sum_{i=1}^{N} v_i \lambda_i v_i^T$$  \hspace{1cm} (5)$$

where $\lambda_i$ is the $i^{th}$ eigenvalue of $M$ and $v_i$ the corresponding eigenvector. The left hand column of Figure 4 shows the first three outer products for the $M$ in the L.H.S of Figure 3. The Matrix $M$ is a real symmetric matrix and so each $v_i \lambda_i v_i^T$ is a rank one approximation to $M$. If a theme is responsible for the dominant structure in $M$ then, because $\sum_{i=1}^{r} v_i \lambda_i v_i^T$ is the best rank-$r$ approximation to $M$ under the Frobenius norm, we should expect its effect in $M$ to be captured in the dominant eigenvalues / vectors. Thus, we can diminish the effect of visual ambiguity / repetitive scene structure by reconstructing $M$ by omitting the first $r$ terms of the summation in 5. We shall now discuss how to choose $r$.

For an $n \times n$ $M$, we define the relative significance, $\rho(r)$ of $\lambda_r$ as

$$\rho(r) = \frac{\lambda_r}{\sum_{k=r}^{n} \lambda_k}$$  \hspace{1cm} (6)$$

Using this we can measure the complexity of decomposition of $M$ as an entropy

$$H(M, r) = -\sum_{k=r}^{n} \rho(r) \log(\rho(r)).$$  \hspace{1cm} (7)$$

The case that $H(M) = 0$ corresponds to an ordered and redundant $M$ which can be represented by a single eigenvector. $H(M) = 1$ corresponds to a similarity matrix where all eigenvectors are equally expressive capturing equally important structure. Hence we are motivated to sequentially remove outer-products from $M$ until $H(M)$ is maximised leaving a similarity matrix in which no one single theme dominates.

We may replace $M$ with a rank reduced version

$$M' = \sum_{i=r^*}^{N} v_i \lambda_i v_i^T$$  \hspace{1cm} (8)$$

The typical effects of this procedure are illustrated in Figure 3. The left hand column of Figure 4 shows the first three rank-1 matrices removed when rebuilding $M$ according to Equation 8 and the typical images responsible for them are shown on the right. Note that this is a soft association used for illustration only and is derived by looking at which cells in the removed $v_i v_i^T$ are maximal and that the off diagonals in Figure 3 have been thickened to make them visible in this p.d.f document.

Fig. 3. The left hand figure shows a similarity matrix constructed from images collected while executing the trajectory shown in Figure 9. The right hand figure shows the structure of the “cleaned” similarity matrix after the rank reduction procedure described in Section III-C.

Fig. 4. The first three (of four) rank one matrices removed from the $M$ shown in Figure 3 and the nine images most responsible for their dominance.

**D. Sequence Significance**

Having performed the rank reduction described above, the sequence extraction can be performed as before with the
effects repetitive visual structure now removed. The last requirement of our loop closure detection scheme is to assign some meaningful significance to the maximal sequence $\langle A, B \rangle$. It could, after all, be only marginally better than a randomly chosen route through $M$. To achieve this we use the extreme value distribution (E.V.D.) to model the distribution of the maximum alignment score $\eta_{A,B}$. The rows and columns of $M$ are randomly shuffled and for each perturbation, a new maximum alignment score calculated. The E.V.D. describes the distribution of $\eta_{A,B}$ as

$$p(\eta_{A,B}) = \frac{1}{\beta} \exp^{-z} \exp^{-\exp^{-z}}, \quad z = \frac{\eta_{A,B} - \mu}{\beta} \tag{9}$$

We proceed by estimating the mode and width parameters, $\mu$ and $\beta$, by fitting (we use a Levenburg-Marquadt scheme) to a histogram of the Montecarlo-generated alignment scores. Equipped with estimates $\hat{\mu}$ and $\hat{\beta}$ and the closed form C.D.F of the E.V.D., we can evaluate the probability of scores greater than or equal to $\eta_{A,B}$ conditioned on all $N$ images:

$$P(\eta \geq \eta_{A,B} \mid M) = 1 - \exp^{-\exp^{-z}} \tag{10}$$

Equation 10 allows the evaluation of the probability that an extracted sequence of image matches $\langle A, B \rangle$ with score $\eta_{A,B}$ could have been generated at random from $M$. The differences between the sequence score $\eta_{A,B}$ obtained from the original, temporarily ordered $M$ and those obtained from the randomly shuffled versions are solely attributable to the topology or connectedness of the spatial locations at which the vehicle captured the images. Thus Equation 10 can be used to evaluate the probability, conditioned on all previous scene appearances, that the detected sequence does indeed indicate a bona-fide loop closure.

E. How Many Loops?

An important distinction between global localisation and loop-closing is that in the former it is often known a-priori that a correspondence between a vehicle’s local scene and a stored representation of the workspace exists. In the case of loop-closing this is not the case — the vehicle may never revisit the same location. It is also possible that within the totality of images, $I$, multiple loop closure events are captured. The probabilistic formulation in Section III-D allows for both these situations. After rank reduction and then distribution fitting, sequences are extracted from $M$ in decreasing order of alignment score, $\eta_{A,B}$, until the probability of false positives associated with $\eta_{A,B}$ becomes excessive. We typically set a threshold of 0.5%.

F. Estimating the Loop Closure Geometry

The sub-system described in the previous section produces a set of images and times (from the time-stamps of the images) which suggest the occurrence of loop closure. In other words we have a strong suspicion that the vehicle is near to a previously known location. One option would be to use the times to index into the state vector (which is after all a sequence of past poses) to find which previous pose $i$ occupied the scene we are now revisiting at time $j$. However this approach has problems when it comes to undertaking a laser scan match to deduce a precise estimate of the interpose transformation $T_{i,j}$. Without a reliable prior or “seed solution” the iterative scan matching method we adopt [4] frequently converges to an incorrect minima. At the same time exhaustive search in 6D is prohibitively slow. In this work we use the putative loop closure image sequences $\langle A, B \rangle$ and 3D laser data to estimate $T_{i,j} –$ the later being used to remove scale ambiguity.

Consider the following common projective model of two identical cameras with projection matrices $P$ and $P’$ [9]. A homogenous 3D image scene point $X = [X, Y, Z, 1]^T$ imaged at $x = PX$ for the first camera and $x’ = P’X$ for the second camera. Without loss of generality the origin can be fixed at the center of the first camera and if the second camera center is parameterized by a rotation matrix $R$ and a translation $t$ with respect to the origin then $P$ and $P’$ can be written $K[R|t]$ and $K[R|t]$ respectively, where $K$ is the matrix of intrinsic camera parameters. In the case of calibrated cameras ($K$ known) the image points, $x$ and $x’$, are related by the “Essential matrix” $E$ such that $x^TEx = 0$. This is a linear constraint on the elements of $E$ which can be obtained by solving the equivalent constraint $\tilde{x}^TE\tilde{x} = 0$ where $\tilde{x}$ is a linear combination of the elements of $x$ and $x’$, and $\tilde{E}$ is the vector of elements of $E$.

The determination of relative pose between camera positions by decomposing an essential matrix has been used to good effect in robot localization [12], [19] and SLAM navigation [6]. Given two image views of the same scene, five point of correspondences are selected for use in an implementation of the “five point algorithm” described in [18] which, notably, is capable of dealing with coplanar correspondence points.

Given two views of the same scene, five or more point of correspondences (this whole procedure should run inside a RANSAC routine) are selected and by stacking the vector $\tilde{x}^T$ for all five points, a $5 \times 9$ matrix is obtained allowing four solutions for the nine elements of $\tilde{E}$ (and thus $E$) to be obtained. The matrix $E$ has a particularly convenient structure. It can be written in terms of $R$ and $t$ as $[t]^\times R$ where $[t]^\times$ denotes the $3 \times 3$ skew symmetric (cross product) matrix constructed from $t$. Given the elements of $E$ this decomposition yields four possible solutions for $R$ and $t$ up to scale. The correct solution is selected by application of suitable constraints on the world points $X$.

Because we know the rigid transformation between the laser scanner and the camera, 3D laser data can be expressed in the camera frame and projected onto the imaging plane (see Figure 5). The range of the correspondence points used to deduce $E$ can now be estimated from neighboring laser range points. This immediately leads to the removal of the scale ambiguity in $t$.

Given estimates of $R$ and $t$, the iterative laser scan matching can proceed with these estimates as an initial solution to $T_{i,j}$. We note that $R$ and $t$ are not perfect owing to inaccuracies in $K$ and imprecise knowledge of the instantaneous laser to camera transformation and the fact that the centers of the
interest points may not coincide exactly with back-projected laser data. Nevertheless, this is a sufficiently accurate relative pose estimation to enable the scan matching to converge and fine tune the transformation estimate.

There remains a finite, albeit small, possibility that the loop closure indicated so far is in fact a false positive. It appears, conditioned on all images, to be statistically significant, while \( R \) and \( t \) describe the views of the scene with tolerable back-projection error. One could resort to the estimates of vehicle location to glean an idea of the credibility although how one could define “credible” in the presence of gross p.d.f errors is an open question. As an alternative we can use the quality of the final scan match to make the final accept/reject decision. This is described in the companion paper [4].

IV. ENFORCING LOOP CLOSURE

Once a loop closure with a corresponding loop closing transformation \( T_{i,j} \) is found, it is used to update the state vector. The system described in Section II might be expected to redistribute (as dictated by the state covariance matrix \( P \)) any gross errors around the circuit. This is exactly what would happen if the update equations involved were linear - however in this case \( T_{i,j} \) expresses a non-linear constraint on the state vector.

Figure 6 shows the effect of naively enforcing the loop closure constraint on the pre-loop closing state vector using the EKF update equations. The result is a discontinuous ‘\( x, y, z \)’ trajectory with highly unrealistic orientations.

This is a well known problem, caused by linearization which is considered in detail in [3]. Motivated by this, we side-step the problem by performing constrained non-linear optimisation around the loop, essentially performing multiple, incremental and iterative changes to the map instead of a single application of a constraint. Prior to loop closing the state vector is a set of \( n \) stacked \( SE_3 \) poses referenced to the origin. However simple application of reference frame transformations yields an equivalent, open chain of \( n \) sequential pose to pose transformations, \( T_{i,i+1}^{\oplus} \in \{0 : n\} \), around the loop. Assuming, without loss of generality, that loop closure is detected between pose \( n \) and pose \( 1 \) a final \( (n + 1)^{th} \) transformation \( T_{n,1} \) can be added to the set which closes the chain of transformations. We can also obtain a measure of the uncertainty in each \( T_{i,i+1} \) in the form of a covariance matrix \( \Sigma_{i,i+1} \) by recalling the results from the original scan-matching procedure and the nominal inter-pose odometry. Given the set of initial set of relative transformations \( T = [T_{1,2}, T_{2,3}, \ldots, T_{n,n+1}] \) and associated uncertainties \( \Sigma = [\Sigma_{1,2}, \Sigma_{2,3}, \ldots, \Sigma_{n,n+1}] \) we set ourselves the task of finding some new set of inter-pose \( SE_3 \) transformations \( T^* \) such that

\[
C(T^*) = \sum_{i=1}^{n+1} (T_i^* - T_i)^T \Sigma_i^{-1} (T_i^* - T_i) \tag{11}
\]

is minimised subject to the loop closure constraint

\[
T_1 \oplus T_2 \oplus \cdots T_n \oplus T_{n+1} = 0 \tag{12}
\]

where, with a slight abuse of notation, \( T_i \) and \( T_i^* \) are the \( i^{th} \) elements of sets \( T \) and \( T^* \) respectively and \( \oplus \) is the \( SE_3 \) composition operator.

The procedure just described allows each inter-pose transformation to be adjusted following loop closure. We have experimented with reducing the degrees of freedom in the optimisation by arbitrarily segmenting the state vector into rigid sub-maps containing some small number, \( m > 1 \), of vehicle poses. The optimisation then occurs over the reduced space of inter-submap transformations. While this is of course to the detriment of overall map quality we have subjectively found its effect to be small and worthwhile compared to the increase in loop-closing speed.

V. RESULTS

The complete system we have described in this paper has been successfully applied to several data sets, one of which is described in detail in this section. Figure 9 shows the final estimated map and trajectory of the vehicle as it traverses the perimeter of a cluster of buildings. The inset to the figure shows a detail of part of the overall picture. The “stripping” on the floor is an artifact of the nodding motion of the laser scanner. The final estimated vehicle trajectory is superimposed upon an aerial photograph of the workspace. Additionally a metric grid has been placed over the area of interest. The
| TABLE I |
|----------------------|------------------|
| Driven Path         | 370.95m          |
| Traversal Time      | 1200s            |
| Computation Time (2Ghz machine) | 3457s |
| Laser Range Points  | 4,030,000        |
| Poses               | 403              |
| Mean Pose Registration | 8.6s         |
| Loop Closing Optimization | 1.01s |
| Poses per sub-map when loop closing ($n_p$) | 9 |

An astute reader will notice a discrepancy between the inset of the figure and the plot (where the trajectory undergoes a semicircular perturbation on the western leg of the circuit). This is because the original buildings present in the photograph have now been replaced with a new building — the steps of which can be seen in the inset. Table I gives some pertinent statistics about this system and its application to this data set. The final map contains just over four million range points and four hundred 6DOF vehicle poses. Over 98% of the total run time is spent performing the 3D interpose registration. Overall our current implementation runs at only one third of real-time however we are confident that with further work we can markedly widen the current registration bottleneck. On a 2GHz PC, the most expensive part of the loop closure detection component is the extraction of Harris Affine regions and converting them to SIFT features (around 0.5 seconds per image). For 450 images, vocabulary generation takes a further three minutes and sequence extraction, rank reduction and significance testing takes around 20 seconds.

A. Extension to Multiple Sessions / Vehicles

On a separate occasion another data set was gathered that partially intersected the data set processed to produce the upper loop in Figure 9. The vehicles started from different locations, initialising different global coordinate frames. This is equivalent to having two independent vehicles $A$ and $B$. In a manner similar to that in [10] we use the vision-based system of Section III to detect shared image sequences. The resulting similarity matrix is shown in Figure 7. Here each element $M_{i,j}$ is the similarity score between image $i$ from robot $A$ and image $j$ from robot $B$. Every image from robot $A$ is compared with all images from robot $B$. Note the off-center dark line indicating overlapping image sequences. In a manner identical to that described in Section III, pairing between image sequences leads (via the estimation of the essential matrix and then scan matching) to an accurate estimate of the transformations between any two vehicle poses lying in areas common to both data sets. Figure 8 shows the matched image sequences from each vehicle. From here it is a simple matter to align both maps/loop in a common coordinate frame. The important point and contribution is that this operation occurs without mutual observation or the existence of a common coordinate frame for all participating robots.

VI. CONCLUSIONS AND FUTURE WORK

This paper has described a SLAM system for outdoor applications and shown it working on a challenging data set in an urban environment. A complementary combination of laser and vision has been used — vision for loop closing and laser data for geometric map building. We showed our system closing a challenging loop — instigated not by geometric considerations but by visual similarity of image sequences. With a potential loop closure detected, constrained non-linear optimisation effected the actual loop closure. To our knowledge this is the first time a 3D ranging sensor has been used in SLAM alongside a vision system for automatic loop closing outdoors. There are however several improvements which we are in the process of researching and implementing. Although convenient, the SLAM formulation we use here is not a efficient one and we would be well served by replacing it with an inverse formulation such as that proposed in [7]. The laser scan matching works well but remains a bottle neck in terms of computation. Finally we are moving towards learning static visual-vocabularies for distinct domains (urban, parkland, indoors etc) and switching between them as the local domain changes. On the grounds that park-like scenes are unlikely to provide evidence for loop closure when working indoors, we propose maintaining a set of domain specific similarity matrices avoiding the cost of maintaining an ever-growing single similarity matrix.
Fig. 9. The resulting estimated map and vehicle trajectory. The inset shows some local detail within the map which contains just over four million range points and the trajectory is decomposed into just over four hundred 6DOF poses. The intersection between the independently executed loops was found in the same way that the loop closures themselves were found - sequential appearance based similarity as described in Section III.

ACKNOWLEDGMENTS

This work is supported in part by the EPRSC through Grant GR/S62215/01 and in part by the Rhodes trust.

REFERENCES

[12] J. Kosecka and X. Yang, “Global localization and relative pose estima-
[18] D. Nister. “An efficient solution to the five-point relative pose prob-