

PRESENTING A TECHNIQUE FOR REGISTERING IMAGES AND RANGE DATA USING A TOPOLOGICAL REPRESENTATION OF A PATH WITHIN AN ENVIRONMENT

Received 15th June; accepted 20th July.

Filipe Ferreira, Luis Davim, Rui Rocha, Jorge Dias, Vítor Santos

Abstract:

This article presents a novel method to utilize topological representation of a path that is created from sequences of images from digital cameras and sensor data from range sensors. Leading the robot around the environment during a familiarisation phase creates a topological representation of the environment. While moving down the same path, the robot is able to localise itself within the topological representation that has been previously created. The principal contribution to the state of the art is that, by using a topological representation of the environment, individual 3D data sets acquired from a set of range sensors need not be registered in a single, [Global] Coordinate Reference System. Instead, 3D point clouds for small sections of the environment are indexed to a sequence of multi-sensor views, of images and range data. Such a registration procedure can be useful in the construction of 3D representations of large environments and in the detection of changes that might occur within these environments.

Keywords: sensor feature integration, binary data, Bernoulli mixture model, dimensionality reduction, robot localisation, change detection

1. Introduction

In previous work we have been primarily interested in developing a method for place recognition by attempting to recognise previously identified places in the environment using features from multiple sensors. This document applies the same technique for place recognition to solve the problem of registering pairs of 3D point-cloud-sets collected by a robot as it moves in large indoor surroundings. This method of registration of point clouds is currently being applied to the creation of 3D environment-visualization tools and to detect changes that might have taken place in the environment.

1.1 Geometrical and Topological Maps

To be useful for robot navigation, maps must contain information about the topology of the environment. The maps represent the layout of the environment and information of the connectivity of the environment [1], [2], [3].

Some maps can include metric information. Grid-based (geometrical) maps are a popular means of representing environments, not least, because, they easily allow the incorporation of uncertainty in the position of objects in the map (and naturally of the robot) [4], [5]. In other cases other representations of the environment, such as vector-based (geometrical) maps using lines or planes have been used, one of the first reported examples

being the work of Drumheller [6].

Topological maps and geometric maps have been compared in a number of works. These comparisons range from the very definition of when does a topological map begin to be a geometric one, to the relative advantages of each type, vis-à-vis scalability, computing costs and other relevant criteria. Lately, there has been debate on the type of map that is [generally] more effective for achieving localization and imparting autonomy to mobile robots placed within large environments.

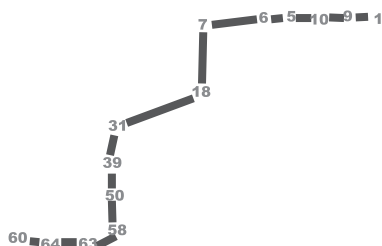
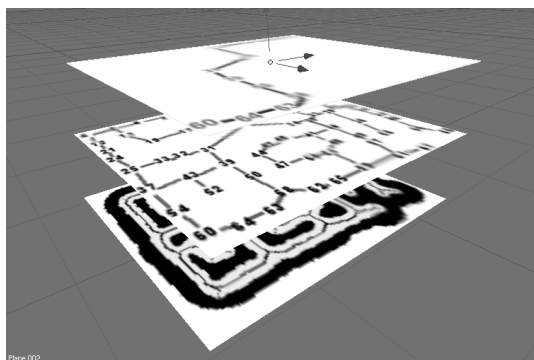
There has been some support for an approach that incorporates (memory consuming) metric information at certain levels and retains only, [topological] layout and connectivity information at others. The Spatial Semantic Hierarchy or the SSH by Ben Kuipers [7] is one such approach that attempts to provide a complete environment representation approach. The framework, described as a "model of knowledge of large-scale space consisting of multiple interacting representations", is presented as hierarchies, each of which perform some abstraction of the perception and interaction of the robot with the environment.

More recent methods maintain fewer hierarchical levels, with [unconnected] grid-based geometrical maps at the lower level and with a high-level topological map containing the connectivity information of the whole environment [8]. In such approaches, the two maps are closely linked since the topological map is actually constructed from the grid-based map.

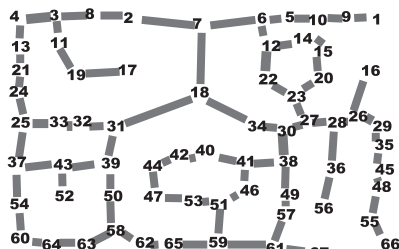
In previous work performed by the authors of this paper was found that the knowledge of the expected sequence of observations to be found along a particular path in the environment is usually sufficient for localization (within this path). Thus, in order to get from a point A to another point B, in an environment, the robot only needs to be told of the sequence of observations it will expect to sense as it proceeds from A to B, Fig. 1 top-right. This is in contrast to the usual situation wherein the robot is provided with a complete map of the environment and localization is defined as the procedure that maximises the consistency of the observations with the [entire] map at any instant.

As stated earlier, the aim of the current work is to register 3D point clouds by first registering the robot within the topological representation created during the familiarisation run Fig. 2, bottom-left. The environment is represented in the form of a line graph where each node of the graph represents a place, at which the environment was sampled, images and scans collected. The sequence of sensor views, thus collected and known as the Reference Sequence, represents a topological path

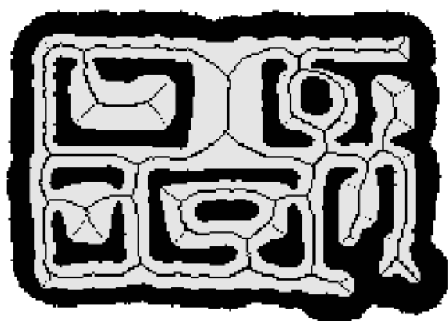
through the environment. During subsequent runs through the environment, the position of the robot is recovered within this Reference Sequence. Once this position within the Reference Sequence is known, the position in the environment that corresponds to the 3D point cloud can be obtained Fig. 2, bottom-right, and the point cloud data registered.



(a) A Topological "path" represented as a Reference Sequence.



(b) The Topological map of the environment.



(c) The Geometrical map of the environment.

Fig. 1. A hierarchy of Geometric, Topological and Reference Sequence Maps that can be used for mobile robot navigation. The grid map [at the bottom] includes the information that can be captured by the sensors and the Reference Sequence contains only a path within the topological map. This comparison of hierarchies has been based on Sebastian Thrun's Hybrid maps [8] and parts b) and c) of the figure are taken from [8], Fig 14.

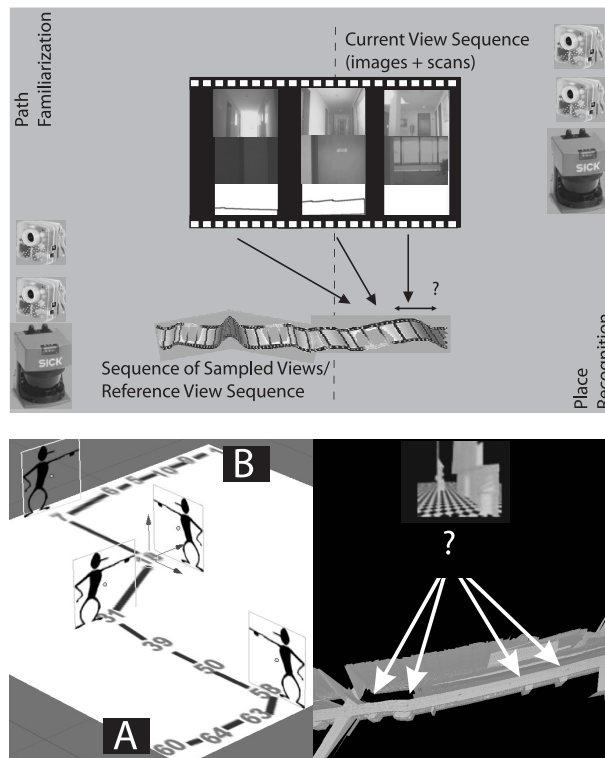


Fig. 2. A depiction of the topological path representation problem, at top. The robot is led through the environment on the Familiarisation run, bottom-left. The 3D point clouds must be registered in the environment, bottom-right.

The experiments in this article were performed on a Segway Robotic Mobility Platform, the RMP 200. Two SICK Laser Range Finder (LRF) and two Firewire cameras have been added to the platform to gather data from the environment. One LRF points upwards, taking vertical sections of the environment as the robot moves forward and a forward-facing LRF that provides a horizontal section through the part of environment in front of the robot. The forward-looking camera, Camera 1, looks in the direction of robot motion while the lateral camera, Camera 2 is mounted at a sufficient height to view posters and other texture appearing on the walls of the building. The cameras are capable of taking VGA-sized images and the SICK laser range finder provides a set of 361 range measurements taken through a 180-degree interval. Features from three sensors, i.e. the forward-facing laser range finder and from the camera images were used for place recognition.

1.2 Features from Ranged Data

Range based methods have been frequently used to index places in indoor environments, first with ultrasound sensors and later with laser range scanners. In an attempt to obtain more reliable features in the environment, many range-sensor based methods extract lines and other primitive features from the laser scan. Cox [9], attempts to match points extracted in the laser range scan with the lines in the map. Drumheller [6] describes a multistage algorithm in which line segments are constructed from data from ultra-sound sensors. Using interpretation trees, the extracted lines are then matched with a given map that is composed of line segments. The extraction of lines from the laser scan

continues to be a popular approach in the robust segmentation of laser scan data, see [10] and [11] for recent reviews of popular line-extraction algorithms. Representing places only in terms of lines (and corners) provides a limited amount of information. Many places in the environment are found to have similar representations and these methods do not scale up easily to larger environments.

Other publications have described the expansion of the set of features extracted from range scans to include, for example, trees, kerbs and more. Manandhar and Shibasaki in [12] extract roads, buildings, tunnels and other outdoor features by modelling 3D range data. In indoor environments too, composite landmarks including lines and other simpler features have been used, [13].

A different approach, which eschews segmentation into simple primitive features, favours the description of a section or sections of the 2D Laser scan in some reduced variable space. This is the approach used, for example, in [14] where each feature extracted from the laser range scan is given a symbol and each scan is described in the form of a string for example mMmMmMmMmDCm. The string alphabet, in this case (M)axima, (D)iscontinuity, (m)inima, (c)onnection, depends on the features extracted from the laser scan. Other methods use 'sections' of the laser range scan so as to minimise the effect that changes in one part of the scan will have representation of the place.

In our work, we have used multiple types of features from the laser range scan, namely 1) wall-like (line) features, 2) scan region properties and 3) scan contour properties in the form of a vector that characterises 2D discontinuities in the plane of the scan using Hu moments [15].

1.3 Local Features from Digital Images

The use of cameras on mobile robots has become widespread over the last few years. Cameras can be viewed as high bandwidth sensors and images can have large redundancy. It is "expensive", in terms of memory and computational costs, to store every raw image that is associated with a place or with an object. Also, certain regions in an image are known to be [more] stable with viewpoint and lighting changes. For this reason, local image features have been used to perform scene and object recognition. The use of local image descriptors based on these stable regions is characterized by two steps: 1) the selection of points of interest and 2) their characterisation. The selection must be repeatable (even with changes in the conditions in which the images are taken) and the characterisation must employ properties that must, again, be tolerant to changes in the viewpoints, lighting and other conditions.

Local-image features based on local image gradients are an important class of vision features. Baker, in [16], attempts to create a generalised descriptor for local image features and the introduction to his thesis provides a perspective on the development of gradient based methods. The stability and repeatability of points extracted at local Maxima (or Minima) in gradient images that have been repeatedly smoothed using operators, has been known for some time [17] [18], and research in the field

finally culminated in the Scale-Space theory proposed by Lindeberg [19].

Building vector descriptors to represent each such feature ensures the uniqueness of the extracted features. In work that combined the lessons of Scale-Space with the reliable characterisation of features, Lowe [20] describes the use of gradient histograms taken at various points close to some point of interest. These features were called Scale Invariant Feature Transforms, SIFT. The work described in this thesis is based on local image features that are based on the SIFT features and to which some modifications that allow us to create the descriptors faster have been made. The procedure for creation of the feature database has been modified in order to simplify the creation of features for image sequences. Since their introduction, SIFT features have been widely applied, among others, to object recognition [21], [22], in the panoramic assembly of images [23] and in image retrieval [24]. Various researchers have used this descriptor in new applications and modifications on the original procedure have appeared (see Weighted Gradient Orientation Histograms [25], Modified SIFT [1], PCA SIFT and Global SIFT).

In this work we have utilised SIFT features to characterise images obtained from the two Firewire cameras. Fast matching of many hundred features is achieved through the use of KDtrees, which can be constructed quickly (see [26] for an fast open-source implementation).

2. Representing a topological path using a Reference Sequence

The problem of localization within a reference sequence is akin, in some ways, to the problem of supervised sequential learning [27]. In the same work, Dietrich identifies three important issues that must be addressed in the case of sequential machine learning, namely specification of a loss function, feature selection and computational efficiency.

The Reference Sequence contains the sensor information that is gathered as the robot moves on its familiarisation run through the environment. In order to improve the results of the localisation algorithm, distinctive places could be maintained in the original Reference Sequence such that they improve the results of localisation and represent the environment in the form of a compact Reference Sequence. The remainder of this section will be devoted to attempting to classify places and then define what makes places "distinct" and how a compact Reference Sequence with distinct places might be created. Further details of the method can be found in [28].

2.1 View Classification using a Bernoulli Mixture Model

Having to reduce the dimension of data from a sensor or multiple sensors in order to recover the original class is a common problem in mobile robot localisation. To solve this problem, methods seek to apply a model to explain the data, allowing the identification of data that is significant (and which must be used during classification).

On the other hand, a data-driven approach will attempt to extract these correlations. Methods that reduce the dimension of features with continuous values are common in many perception fields including face recognition, speech recognition etc. Among these approaches, Mixture Models are a common solution to modelling data that is thought to follow a non-parametric distribution. Mixture models assume that there exists a finite number of distributions which, when mixed together in a particular proportion, result in a distribution that best describes the data. Sajama and Orlitsky in [29] demonstrate the use of Mixture models composed of Gaussian, Bernoulli and exponential distributions as a solution to the classification problem. To a greater or lesser extent these clustering or classification methods seek to identify features that are more correlated with members of their own group than with members from another group. McLachlan and Peel [30] provide a good reference to the general topic of Finite Mixture Models.

Articles by Kaban [31] and Wang [32] provide a healthily different viewpoint and go some way to demonstrated the usefulness of binary features. In [32] the context in which a word is used in a sentence is converted into multiple binary features. Similarly [33] and [34] seek to model some training data as a sample of sets of binary features taken from a population of binary features, each distributed according to a mixture of Bernoulli distributions. The application of binary features to the classification of images and text has motivated us to apply the approach to classifying other types of features, such as barcode/data-matrix features.

While classification of a few binary properties might be easy, using many, correlated features is very difficult. In applications such as mobile robot localisation, image retrieval and robot localisation methods typically make use of a large number of features. In the application to mobile robot localization, we have employed up to 16000 binary features to allow the robot to recover its position within the environment.

The binary features from each of the objects in the pilot set are represented within a Feature Incidence Matrix (FIM), \mathcal{V} . Each row i , of the FIM corresponds to a feature Y_i and each column j , to an object, V_j , from the reference sequence (each entry in the FIM might be represented as $Y_{i,j}$ where the first subscript indicates the feature and the second subscript, the object). $Y_{i,j}$ takes value 1 if feature Y_i appears (is present in the code) in object Y_j , 0 otherwise.

$$\mathcal{V} = \begin{bmatrix} Y_{1,1} & Y_{1,2} & \dots & Y_{1,K} \\ Y_{2,1} & Y_{2,2} & \dots & Y_{2,K} \\ \vdots & \vdots & \ddots & \vdots \\ Y_{N,1} & Y_{N,2} & \dots & Y_{N,K} \end{bmatrix} \quad (1)$$

A simple metric for verifying the similarity between a new object to be classified, and another in the pilot set, could be the number of binary features in each object that are unchanged. This metric would assume that the individual features on each object are independent. Unfortunately, given that features arise in groups and persist/disappear as a result of the behaviour of the object and its exposure to the environment, the assumption of independence between the features does not reflect reality.

Inferences made under this assumption would be biased toward certain objects in the FIM and in practice, some of the features are highly correlated while others are less. In such circumstances we need to employ methods that deal with the correlation between the features. As stated earlier, this thesis describes our use of Mixtures of Bernoulli Distributions to model the binary FIM.

Mixture models assume that there exist a finite number of parametric distributions, which, when mixed together in a particular proportion, result in a distribution that best describes the data we wish to characterize. In this case we model any code that is observed V_{obs} as a vector of binary features $\{0,1\}^N$, which is obtained from a particular mixture of Bernoulli distributions, as in (2).

If \mathcal{V} represents the complete set of objects as collected during the classification of the pilot set, V_k is a single object with an index k within the pilot set, \mathcal{V} , described by multiple features, V_{obs} is a single object that must be classified and F is a single [named] Feature. Z is a Hidden or incomplete data in a Mixture Model, α_i represents the mixture component coefficient or component Prior probability and Θ_i is a single component of the mixture model with the named features \prod represents the product operator while $\sum_{k=1}^K$, represents the sum operator with the index k varying from 1 to K .

$$P(V_{obs} | \Theta) = \sum_{i=1}^M \alpha_i P(V_{obs} | \Theta_i) \quad (2)$$

where Θ denotes the parameters of the distribution of the objects that compose our Mixture Model. These parameters include the M component vectors, the Θ_i s, and the proportions in which these are mixed, the α_i s. Each α_i represents the prior probabilities of the component i in the mixture model, subject to the constraint $\sum_i \alpha_i = 1$. The term $P(V_{obs} | \Theta_i)$, can be determined using (3) where each Θ_i is a multivariate vector of Bernoulli probabilities each of whose N components indicate the probability of success for a particular feature.

$$P(V_{obs} | \Theta_i) = \prod_{j=1}^N \Theta_i^{V_{obs,j}} (1 - \Theta_i)^{(1 - V_{obs,j})} \quad (3)$$

To obtain the mixture parameters that explain a particular FIM \mathcal{V} , consisting of K observations it is assumed that the objects are independent and the likelihood of the mixture satisfying the FIM is expressed thus (4).

$$P(\mathcal{V} | \Theta) = \prod_{i=1}^K P(V_i | \Theta) = \mathcal{L}(\Theta | \mathcal{V}) \quad (4)$$

The optimisation task to find the mixture that best explains this \mathcal{V} can be expressed as in (5), i.e. find the value of Θ that best satisfies the distribution of features in \mathcal{V} .

$$\Theta^* = \operatorname{argmax}_{\Theta} \mathcal{L}(\Theta | \mathcal{V}) \quad (5)$$

The preferred method of solving the Mixture Model problem is the Expectation Maximisation algorithm. McLachlan ([30], page 19) states "...it will be seen that conceptualisation of the mixture model ... (hidden data + component distributions)... is most useful in that it allows the Maximum likelihood estimation of the mixture distribution to be computed via a straightforward appli-

cation of the EM algorithm". The EM algorithm proceeds in two stages: the *Expectation* stage attempts to reach the best value for the missing data Z , by keeping the parameters of the Mixture model constant (6), while the subsequent *Maximization* stage attempts to optimise the components and mixing parameters themselves by using the values of the "missing data" obtained in the expectation step just performed (7), (8). The method then alternates between the two steps until some termination criteria are satisfied.

$$z_{ki} = \frac{\alpha_i P(V_k | \Theta_i)}{\sum_{j=1}^M \alpha_j P_j(V_k | \Theta_j)} \quad (6)$$

$$\alpha_i = \frac{\sum_{k=1}^K z_{ki}}{K} \quad (7)$$

$$\Theta_i = \frac{\sum_{k=1}^K z_{ki} V_k}{\sum_{k=1}^K z_{ki}} \quad (8)$$

This termination criterion is usually a lack of change in the mean error when the Mixture Parameters are applied to the original data. In the case of such applications, where the parameters of the Mixture models are required for the purpose of classification, the process is usually stopped quite early, when the reduction in the Mean Error is not significant.

Mixture models used for classification make use of both, the Mixture parameters and the posterior probabilities over the components, the Z are used to evaluate the likelihood in the space of the objects in the reference sequence as in (9) where $P(V_k)$ represent the prior probabilities on each index k .

$$P(k | V_{obs}) = \frac{\sum_{j=1}^M P(V_k) z_{ki} \alpha_j P(V_{obs} | \Theta_j)}{\sum_{k=1}^K \sum_{j=1}^M P(V_k) z_{kj} \alpha_j P(V_{obs} | \Theta_j)} \quad (9)$$

The *Maximum Likelihood Estimation* approach is used to obtain the matching object, the index k^* , in \mathcal{V} that best describes the object to be matched, V_{obs} .

$$P(k=k^* | V_{obs}) = \max_k P(k | V_{obs}) \quad (10)$$

2.2 Identifying distinct views

In order to match multiple, consecutive, current views by using the context in which these are observed, the movement of the robot along the sampled views can be modelled as a Finite State Machine. The order in which the views were collected in the reference sequence, should, if all goes well in the subsequent runs, be repeated as the robot retraces its path on a latter occasion.

It is conceivable that in certain cases, by identifying distinctive places, better results that were obtained, using the original Reference Sequence. These improvements could lead to better localisation from more suitable prior and place-transition probabilities. Alternatively, by identifying these distinct places a more compact topological representation of the path that does not necessarily provide inferior localization capabilities could be created. Shorter, more compact Reference Sequences are desirable in applications involving communication between robots/persons having different capabilities

and limited computing power or communication bandwidth. Compact representations of the Reference Sequence are also desired when long paths through the environment must be traversed, resulting in faster localisation.

As shown in Fig. 3, the original, sampled Reference Sequence might be broken down into a number of smaller sequences resulting in a compact representation of the topological path. A compacting procedure for the Reference Sequence should keep two types of Views:

- 1 Essential Views defined as those Views in the Reference Sequence at which the robot motion behaviour is altered.
- 2 Non-Essential Views, which, as the term might suggest are those views, that, while not essential, might improve the topological path representation. Non-Essential views should increase the probability of identifying Essential Views and should also increase the probability of correctly detecting that the robot is lost.

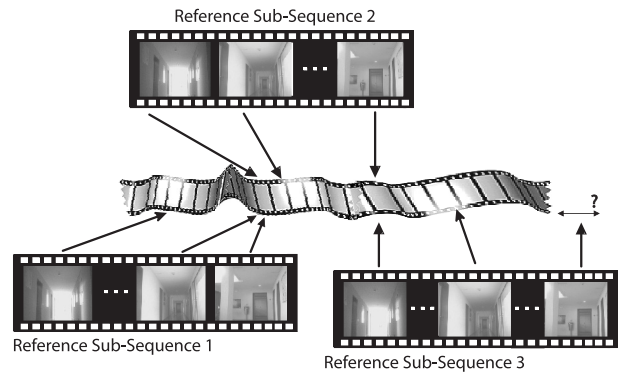


Fig. 3. The original Reference Sequence is now decomposed into a series of sub sequences, which together make up the Compact Reference Sequence.

While the Essential views depend on the behaviours that are adopted along the path, and on any markers laid down by the user, the Non-Essential views reflect the information content of that part of the environment. The problem of selecting the Non-Essential Views entails selecting the most successful sub-set of Views from a *single* sequence. Such a set of parameters will depend on the length chosen between the Views of the HMM and (a related quantity) the number of Non-Essential Views chosen between a pair of Essential Views. Since exact methods could not be found to solve the problems of selecting the distinct views Non-Essential Views were researched within this work.

These approximate methods might use cost functions that seek to maximise one or more of the above-mentioned factors. For example, a strategy for addressing the first and second factors mentioned above would aim to select the earlier *single view* that has the greatest [total] chance of getting selected. The selection of the "Next" non-essential View is not completely straight forward since the Prior probability $P(k)$ in (9) implicitly influences the Place recognition using the Bernoulli Mixture Model.

Since the information in the compact Reference

Sequence is not uniformly distributed in time or distance, a conventional Markov chain (with regular observations) cannot be used to model the hidden states of the Hidden Markov Model. Instead, the distance between the places must be modelled by modelling the duration (or time) between consecutive places.

The principal feature of the HMMs that takes into account the duration of the transition between states is that there some mechanism is found to model the "self-transition" probability of a state (place) to itself. Such variants of HMMs are becoming quite popular in the modelling of phoneme information in speech recognition and synthesis [35]. The conventional way of modelling duration is by building higher order Bayesian networks or by breaking the Markov assumption of HMMs to give Hidden Semi-Markov models.

A simpler method that has been suggested to account for the duration for which the system remains in the same state is the Post-Processing model by [36]. In this approach, the original Viterbi algorithm is maintained but at every step the probability is updated to reflect the probability that the transition actually respects or is in concordance with the expected transition duration, according

to (11), where $p_j(\frac{l_j}{T})$ represents the estimate of the probability of staying in state j for a time l_j given an expected duration of T and a is a constant.

$$\log \hat{f} = \log f + a \sum_{j=1}^N \log p_j(\frac{l_j}{T}) \tag{11}$$

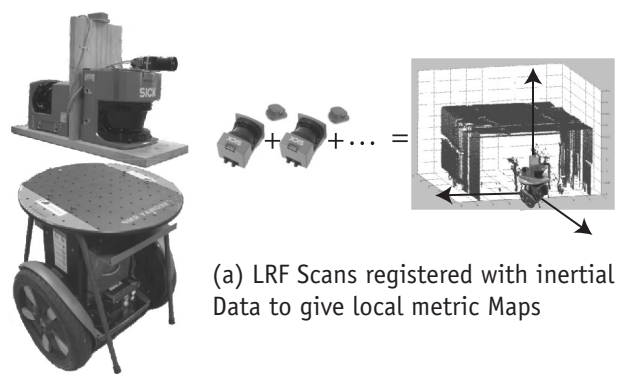
This approach has been employed in the current work, since it allows large variations in the distances between nodes without adding to the computational cost of the Viterbi Algorithm.

3. Feature Registration and Localisation of the Robot

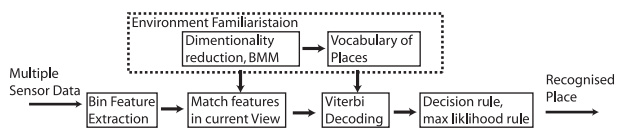
Building on and extending previously existing modules within the Carnegie Mellon Robotic Toolkit, CARMEN, have developed support for the Segway RMP 200 robot. Sensor data is recorded in standardised file formats within CARMEN. The sensor data from the log files is indexed to the XML file representing the Reference Sequence through the "distance-covered" variable. By recovering the correct View within the Reference Sequence, the sensor data corresponding to the position of the robot in the environment is obtained. The range data from the vertically facing LRF is extracted corresponding to the place in the Reference Sequence and consecutive scans registered using the accompanying inertial information.

Reconstruction of the 3D environment from data collected from the vertically facing LRF is performed in a two-step process. In the first step, consecutive individual planar scans are registered with each other by applying the transformation that the LRF suffered between consecutive scans. The transformations required to perform this registration are obtained from an inertial sensor attached to the sensor platform. The registered scans create small sections of reconstructed 3D data each within its own local coordinate system, and known as local metric maps.

In the second step these local metric maps are temporally ordered using the information contained in the Topological map. The robot has been previously led through the environment during the environment familiarisation phase during which the Reference Sequence and the data from the sensors was collected. During subsequent runs, the sensor Views are compared to the Views of the Reference Sequence to estimate the location of the robot within the Reference Sequence according to the method summarised in Fig. 4, bottom-right.



(a) LRF Scans registered with inertial Data to give local metric Maps



(b) Registration of Local Maps with the Reference Sequence

Fig. 4. A loose reconstruction of the environment is created from the 2D scans extracted by the vertically facing, rear LRF using a 2-step process. In the first step, small 3D local metric maps are reconstructed from the 2D LRF scans using data from the inertial sensor that is attached to the sensor set-up. In the second step, local 3D metric maps from the first step are registered within the Reference Sequence, the topological representation of the path from the Environment Familiarisation run. A schematic of the method used to perform place recognition within the Reference Sequence is shown.

The results of localisation along a run through a long corridor of around 120 meters length are shown in Fig. 5. The place in the Reference Sequence are shown in the Y coordinate and the places at which localisation was attempted in the second run is plotted along X coordinate. The scatter plot shows the place in the Reference Sequence against which the observed views were registered. Typical 3D reconstructions for Point cloud data along the Reference Sequence and the Observation Sequence are indexed to this scatter plot.

The Fig. 6 shows two examples of registered sections of the environment with the screen shots at right indicating the reconstructions for the Reference Sequence and those at left representing the reconstructions for the subsequent run.

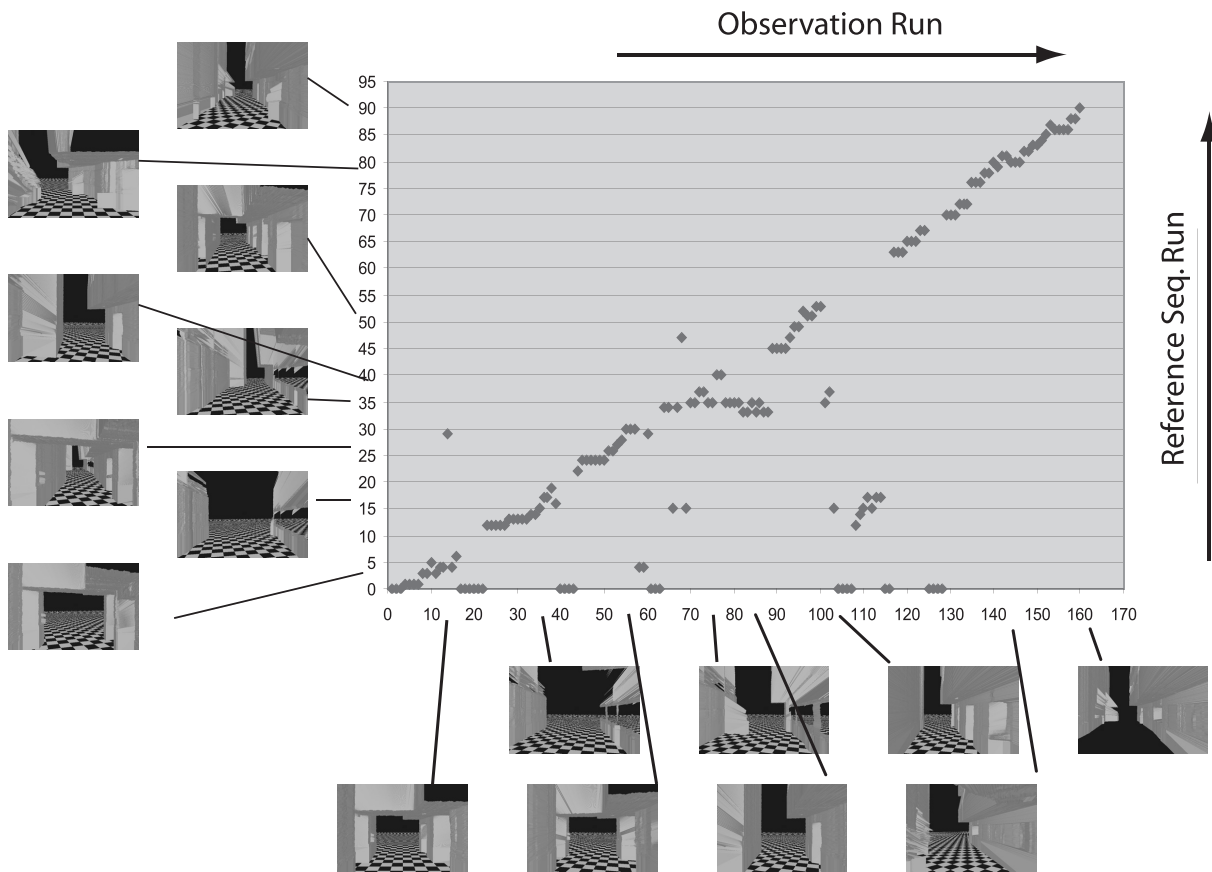


Fig. 5. The scatter graph plots the views in the observation sequence that was matched against views in the Reference Sequence. Sections of Rendered Point Cloud data for the same matched places in the topological path are also shown.

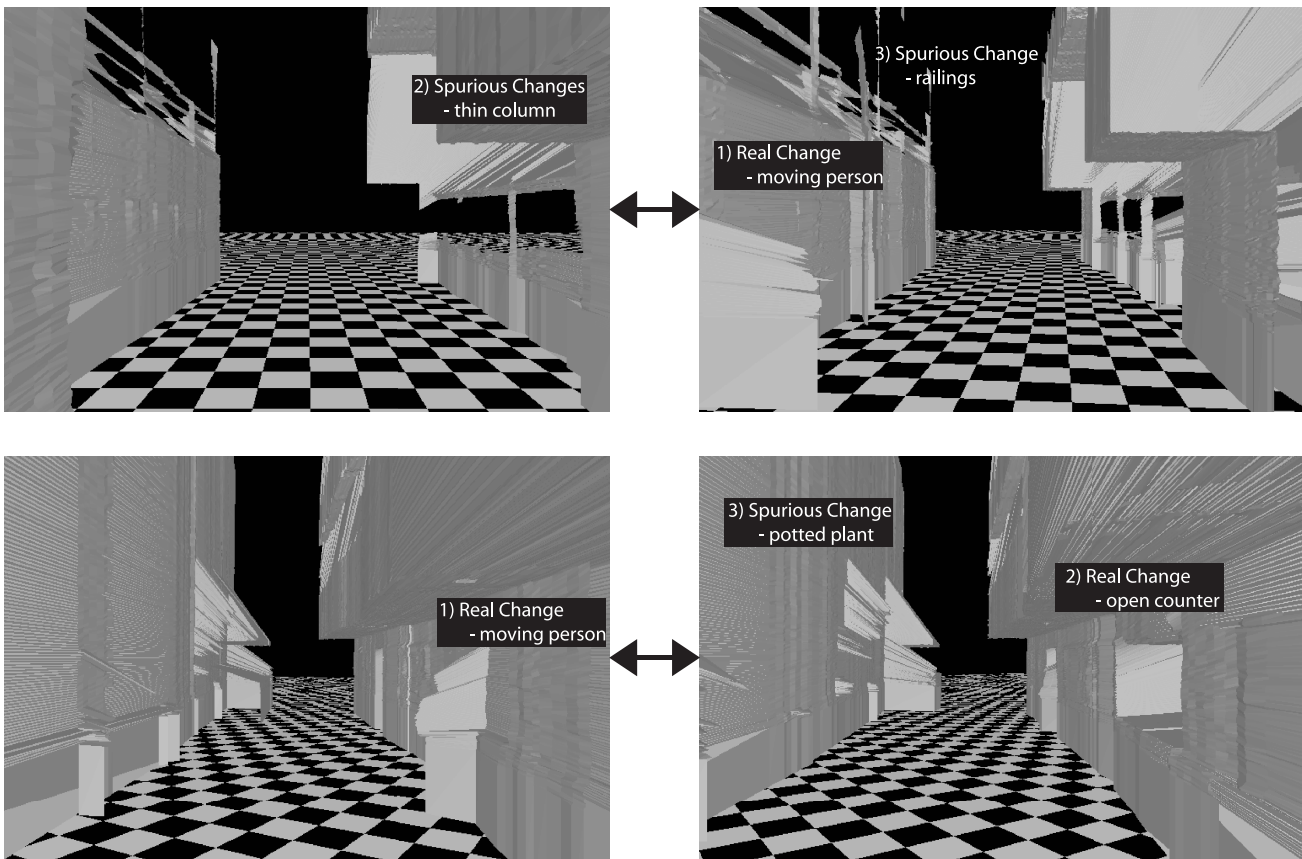


Fig. 6. The screen shots at left depict local metric maps rendered at places that were matched against equivalent places in the Reference Sequence shown on the right. A visual comparison of the screen shots reveals the occurrence of real and spurious changes in the environment.

4. Conclusions

The localization of the robot in the Reference Sequence by reducing the dimensionality of the sensor data has been demonstrated in an earlier publication. Further improvements in place recognition have been made in order to create compact sequences.

Localisation within the topological path represented in the Reference Sequence is found to be a promising way to initiate procedures to detect changes in the environment and to represent sections of the 3D point clouds in this representation of the topological path.

We are currently developing local image features that can be extracted from the 3D point cloud data that can also be added to the set of features to aid the localisation of the robot along the topological path.

Work has also begun to develop algorithms that might autonomously flag the presence of changes in the environment through a comparison of the a pair of point cloud data sets taken of the same place within the environment. The algorithms will attempt to fuse local metric maps containing 3D data and flag the differences that impede correct fusion. Such algorithms will be applied to detect the presence of people and the occurrence of environment-change events that may pose risks to users of the infrastructure.

ACKNOWLEDGEMENTS

The first author is a recipient of a research scholarship from the Faculdade de Ciências e Tecnologia, University of Coimbra (FCTUC) under the BACS (FP6-IST-27140) project. Part of this work was performed within the PhD program of the first author, funded by the University of Aveiro, Portugal. The second author is a recipient of a research scholarship from the FCTUC under the IRPS (FP6-IST-45048) project.

The authors have utilised the CMU-promoted CARMEN toolkit and wish to thank the development community of this project. Special thanks go out to the developers of the 3D visualisation tool [37].

AUTHOR

Filipe Ferreira*, **Jorge Dias**, **Luis Davim**, **Rui Rocha** - Institute of Systems and Robotics, University of Coimbra, 3030-290 Portugal; e-mail: cfferreira@isr.uc.pt.

Vitor Santos - Department of Mechanical Engineering, University of Aveiro 3810-193 Portugal.

* corresponding author

References

- [1] H. Andreasson and T. Duckett, "Topological Localization For Mobile Robots Using Omni-Directional Vision And Local Features". In: *5th IFAC/EURON Symposium on Intelligent Autonomous Vehicles*, Instituto Superior Tecnico, Lisboa, Portugal, July 2004. IFAC/EURON.
- [2] N. Vlassis, Y. Motomura, and B. Kröse, "Supervised linear feature extraction for mobile robot localization". In: *Proc. IEEE Int. Conf. on Robotics and Automation*, 2000, pages 2979-2984.
- [3] A. Tapus, N. Tomatis, and R. Siegwart, "Topological Global Localization and Mapping with Fingerprints and Uncertainty". In: *9th International Symposium on Experimental Robotics*, ISER, 2004.
- [4] A. Elfes, "Sonar-based real-world mapping and navigation", *IEEE Journal of Robotics and Automation*, vol. 3, issue. 3, June 1987, pp. 249-265.
- [5] R. Rocha, *Building Volumetric Maps with Cooperative Mobile Robots and Useful Information Sharing*. PhD thesis, University of Porto, Department of Engineering, 2005.
- [6] M. Drumheller, "Mobile robot localisation using sonar", *IEEE Transactions On Pattern Analysis And Machine Intelligence*, vol. 9, issue 2, March 1987, 325332.
- [7] B. J. Kuipers, "The Spatial Semantic Hierarchy", *Artificial Intelligence*, no. 119, 2000, pp. 191233. Available at: <http://www.cs.utexas.edu/users/qr/papers/Kuipers-aij-00.html>.
- [8] S. Thrun, "Learning metric-topological maps for indoor mobile robot localization", *Artificial Intelligence*, issue 99(1), 1998, pp. 21-71.
- [9] I.J. Cox. Blanche-an experiment in guidance and navigation of an autonomous robot vehicle. *IEEE Transactions on Robotics and Automation*, vol. 7, issue 2, April 1991, 193-204.
- [10] V. Nguyen, A. Martinelli, N. Tomatis, and R. Siegwart, "A comparison of line extraction algorithms using 2d laser rangefinder for indoor mobile robotics". In: *International Conference on Intelligent Robots and Systems*, 2005.
- [11] D. Sack and W. Burgard, "A comparison of methods for line extraction from range data". In: *IAV 2004*, 2004.
- [12] D. Manandhar and R. Shibasaki, "Feature extraction from range data". In: *ACRS 2001 - 22nd Asian Conference on Remote Sensing*, Singapore, November 2001, volume 2, pages 1113-1118.
- [13] Z. Xiang, V.M. F. Santos, and J. Liu, "Locating And Crossing Doors And Narrow Passages For A Mobile Robot". In: *First Int. Conf. on Informatics in Control, Automation and Robotics (ICINCO)*, 2004.
- [14] J. Fellers, S. Lee, N. M. Amato, "Localization based on Visibility Sectors using Range Sensors". In: *Proceedings of the IEEE Int. Conference Robot. Autom. (ICRA)*, January 2000, pages 3505-3511.
- [15] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, Addison-Wesley Pub. Co, 2002.
- [16] S. Baker, *Design and Evaluation of Feature Detectors*. PhD thesis, Columbia University, 1998.
- [17] J. J. Koenderink, "The Structure Of Images", *Biological Cybernetics*, vol. 50, no. 5, 1984, pp. 363-370.
- [18] A.P. Witkin, "Scale-space filtering". In: *Proc. 8th Joint conference on Artificial Intelligence*, Karlsruhe, W. Germany, 1983, pages 1019-1023.
- [19] T. Lindeberg, *Scale-Space Theory in Computer Vision*, Kluwer Academic Press, 1994.
- [20] D. G. Lowe, "Object Recognition from Local Scale-Invariant Features". In: *Proc. of the International Conference on Computer Vision*, Corfu, 1999, pages 1150-1157.
- [21] A. Pope and D. G. Lowe, "Probabilistic Models of Appearance for 3-D Object Recognition", *International Journal of Computer Vision*, vol. 40, no. 2, 2000, pp.149-167.
- [22] D. G. Lowe, "Local Feature View Clustering for 3D Object Recognition". In: *Proceeding of the IEEE Conference on*

- Computer Vision and Pattern Recognition*, Kauai, Hawaii, 2001, pages 682-688.
- [23] M. Brown and D. G. Lowe, "Recognising Panoramas". In: *10th International Conference on Computer Vision (ICCV 2003)*, October 2003.
- [24] Y. Ke, R. Sukthankar, and L. Huston, "An efficient parts-based near-duplicate and sub-image retrieval system". In: *MULTIMEDIA'04: Proceedings of the 12th annual ACM international conference on Multimedia*, New York, USA, 2004. ACM Press, pages 869-876.
- [25] D. M. Bradley, R. Patel, N. Vandapel, and S. M. Thayer, "Real-Time Image-Based Topological Localization in Large Outdoor Environments". In: *Proceedings of the 2005 IEEE Conf. on Intelligent Robotics and Systems*, August 2005.
- [26] M. B. Kennel, KDTREE 2: Fortran 95 and C++ software to efficiently search for near neighbors in a multi-dimensional Euclidean space, 2004.
- [27] T. G. Dietterich, Machine learning for sequential data: A review. Technical report, Oregon State University, 2002.
- [28] F. Ferreira, V. Santos, and J. Dias, "A topological path layout for autonomous navigation of multi-sensor robots", *International Journal of Factory Automation, Robotics and Soft Computing*, no. 1, 2007, pp. 203-215.
- [29] Sajama and Alon Orlitsky, "Supervised dimensionality reduction using mixture models". In: *Proceedings of the 22nd International Conference on Machine Learning*, Bonn, Germany, 2005.
- [30] G. McLachlan and D. Peel, *Finite Mixture Models*, John Wiley and Sons, 2000.
- [31] A. Kaban and M. Girolami, "Initialized and Guided EM-Clustering of Sparse Binary Data with Application to Text Based Documents". In: *Proceedings of the 15th International Conference on Pattern Recognition (ICPR'00)*, volume 2, 2000.
- [32] X. Wang and A. Kaban, "Finding Uninformative Features in Binary Data". In: *Sixth International Conference on Intelligent Data Engineering and Automated Learning (IDEAL '05)*, 2005.
- [33] A. Juan and E. Vidal, "Bernoulli Mixture Models for Binary Images". In: *International Conference on Pattern Recognition (ICPR'04)*, volume 3, 2004, pages 367-370.
- [34] J. García-Hernández, V. Alabau, A. Juan, and E. Vidal, "Bernoulli mixture-based classification". In: A. R. Figueiras-Vidal *et al.*, ed., *Proc. of the LEARNING04*, ISBN 84-688-8453-7, Elche (Spain), October 2004.
- [35] J. Pyllkkönen and M. Kurimo, "Duration modeling techniques for continuous speech recognition". In: *8th International Conference on Spoken Language Processing (Interspeech 2004)*, Jeju Island, Korea, 2004, pages 385-388.
- [36] B. Juang, L. Rabiner, and S. Levinson M. Sondhi, "Recent developments in the application of hidden markov models to speaker-independent isolated word recognition". In: *IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP '85*, vol. 10, 1985, pages 9-12.
- [37] M. Kaess, R.C. Arkin, and J. Rossignac, "Compact encoding of robot-generated 3D maps for efficient wireless transmission". In: *IEEE Intl. Conf. on Advanced Robotics, ICAR*, Coimbra, Portugal, June 2003, pages 324-331.