Werner, Felix and Maire, Frederic and Sitte, Joaquin (2009) *Topological slam using fast vision techniques.* In: Advances in Robotics : FIRA RoboWorld Congress 2009, Incheon, Korea, August 16-20, 2009, Proceedings. Lecture Notes in Computer Science, 5744. Springer Berlin Heidelberg, pp. 187-198. (In Press)

# Topological SLAM Using Fast Vision Techniques

Felix Werner, Frederic Maire, and Joaquin Sitte

[1] School of Information Technology,
Faculty of Science and Information Technology,
Queensland University of Technology,
1 George Street, Brisbane, QLD 4001, Australia
[2] NICTA, Queensland Lab,
Staff House Road, St Lucia, QLD 4072, Australia
{f.werner,f.maire,j.sitte}@qut.edu.au
http://www.scitech.qut.edu.au
http://www.nicta.com.au

**Abstract.** In this paper we propose a method for vision only topological simultaneous localisation and mapping (SLAM). Our approach does not use motion or odometric information but a sequence of noisy visual measurements observed by traversing an environment. In particular, we address the perceptual aliasing problem which occurs using external observations only in topological navigation.

We propose a Bayesian inference method to incrementally build a topological map by inferring spatial relations from the sequence of observations while simultaneously estimating the robot's location. The algorithm aims to build a small map which is consistent with local adjacency information extracted from the sequence measurements. Local adjacency information is incorporated to disambiguate places which otherwise would appear to be the same.

Experiments in an indoor environment show that the proposed technique is capable of dealing with perceptual aliasing using visual observations only and successfully performs topological SLAM.

**Keywords:** Autonomous mobile robots, SLAM, correspondence problem, topological navigation, panoramic vision, colour histograms.

## 1 Introduction

Simultaneous localisation and mapping (SLAM) is one of the most researched areas in robotics. Two different approaches exist to the SLAM problem: *Metric* and *topological* [1]. The fromer aims to model the environment using a metric map so geometrically accurate position estimation is achieved. Topological maps are graphical models of the environment that capture *key places* and their *connectivity* in an abstract and compact manner for localisation and path planning.

In both cases, probabilistic approaches have been successfully applied to deal with the inherent uncertainties associated with robot perception, that would otherwise trouble the map-building process. Beside measurement noise, topological

J.-H. Kim et al. (Eds.): FIRA 2009, LNCS 5744, pp. 187–196, 2009.

mapping is complicated through *perceptual aliasing* which occurs when physically different parts in in the environment appear to be the same to the robot. This phenomenon occurs as sensors may supply insufficient data to identify the current state of the world because of sensory noise, limited field of view (aperture problem) and repeated structures in the environment. Perceptual aliasing makes it difficult for a robot to decide when it is visiting a new place or revisiting a memorised place (*loop closing*) [1,2].

In this work, we address the problem of topological SLAM from a sequence of visual measurements obtained from visited places only. In particular, we are interested in the problem of loop closing in environments which contain physically different places which appear to be the same. We approach this problem using Bayesian inference to estimate the posterior distribution on topological maps while simultaneously determining the place the robot currently occupies. The inference method embeds a strategy to reliably deal with percptual aliasing by distinguishing similar places on the basis of neighbouring information [3]. Using neighbouring information to disambiguate physically different places which appear identical was proposed by Werner et al. [3]. However, it was only examined on artificial topological graphs with deterministic observations.

Our approach is purely vision based so we suppose that actions (e.g. turn left) and odometry cannot be sensed directly. In contrast to other vision based methods [4,5] which aim to represent places highly distinctive using sophisticated visual features such as SIFT [7], or SURF [8,9], we use colour histograms only [10,11,12,13,14]. Colour histograms in conjunction with panoramic images exhibit several attractive properties such as invariance to rotation around the vertical axis. Moreover, colour histograms represent salient colour information of images in a very compact manner and are very fast to extract and process. Clearly, colour histograms are not very distinctive image features so we have the scenario that similar visual appearance is shared by different places.

Research in topological SLAM has mainly been concerned with a particular aspect to avoid the perceptual aliasing problem by improving the distinctiveness of the appearance of places [4,5]. These approaches do not properly address situations where places are indistinguishable even with perfect sensing. Other approaches support the robot's perceptual abilities using metric information gained from odometry measurements or the robot's actions [15,16,17].

## 2 Neighbourhood Information for Topological Mapping

Our algorithm represents a topological map by a *labelled graph* where *vertices* represent places and *edges* reflect the connectivity between places [1]. The *labels* of vertices refer to *fingerprints* which characterise the place in terms of sensor data. The graph which corresponds to the surrounding is denoted *environment graph* $G_{env}$ and the *map graph*, which we want to infer, $G_{map}$. The environment graph is unknown and the only available information about it is a finite history $h_{env} = l^0_{env}, l^1_{env}, ... \in L^*_{env}$ of labels of visited vertices obtained from the traversal of the environment graph (Here, * is the Kleene star).

Our method exploits the neighbourhoods of places to disambiguate places with similar appearance. The neighbourhoods of the environment graph are not accessible directly as it is unknown. Local neighbouring information which is contained in the history is accessible through sequences of length $n$, called $n$-grams [3]. Consecutively visited vertices are represented by consecutive labels in the history and, in turn, consecutive labels in the history must originate from adjacent vertices in the environment graph. Hence, the set of $n$-grams $Grams(h, n)$, which can be obtained from a history, corresponds to a *feature space* on the history.

In order to achieve reliable navigation, a robot requires an internal representation that exhibits the properties of the environment with respect to the selected representation (e.g. topological). In our case, a map graph is required to be isomorphic with the environment graph. The map graph is isomorphic with the environment graph if there is a bijective mapping such that each neighbourhood of the map graph corresponds to a neighbourhood of the environment graph and vice versa. However, it is not possible to compare neighbourhoods of the map graph directly with neighbourhoods in environment graph as the latter one is unknown. Consequently, we propose to measure the consistency of graphs in the feature space; that means, the sets of $n$-grams of the graphs. Hence, two graphs are $n$-consistent if they share the same set of $n$-grams [3]. For noisy data the Hausdorff distance is used to measure the $n$-consistency of two graphs $G_0$ and $G_1$ using $\Gamma_0 = Grams(h_0, n)$ and $\Gamma_1 = Grams(h_1, n)$ generated from $h_0 \in L_0^*$ and $h_1 \in L_1^*$, so

$$d_H(\Gamma_0, \Gamma_1) = \max(\max_{\gamma_i \in \Gamma_0} \min_{\gamma_j \in \Gamma_1} d(\gamma_i, \gamma_j), \max_{\gamma_j \in \Gamma_1} \min_{\gamma_i \in \Gamma_0} d(\gamma_j, \gamma_i)). \tag{1}$$

The smaller the Hausdorff distance the more $n$-consistent are graphs $G_0$ and $G_1$. The distance of two $n$-grams $\gamma_i$ and $\gamma_j$ is computed using the maximum norm

$$d(\gamma_i, \gamma_j) = ||\gamma_i - \gamma_j||_\infty = \max_{k=0...n-1}(|\gamma_{i,k} - \gamma_{j,k}|) \tag{2}$$

so the distance between two sets of $n$-grams is determined by the most significant distance of two labels which are mapped to the same vertex.

## 3 Topological SLAM

In this section we describe our method for topological SLAM from a sequence of visited places. The history of observations of visited places is the only information about the environment. In particular, the robot has no access to metric information such as odometry and no information about its actions but is aware of performed U-turns. We suppose that the robot has explored the environment and has recorded a history $h_{env} = l_{env}^0, ..., l_{env}^{M-1}$ of $M$ fingerprints of visited places. After the exploration run the set $\Gamma_{env} = Grams(h_{env}, n)$ is derived from the history.

### 3.1 Bayesian Map Inference

The space of topologies grows hyper-exponentially with the number of measurements [15]. Thus, we use a Bayesian inference method to only infer topological maps which are consistent with the observations. In Bayesian map inference observations are used to update or to newly infer the probability that a hypothesis may be true using Bayes' theorem. For the purpose of inferring a topological map graph $G_{map}$ from a history $h_{env}$ we have

$$P(G_{map}|h_{env}) \propto P(h_{env}|G_{map})P(G_{map}). \tag{3}$$

We can assume the process of incrementally building a topological map is Markovian – that is the current topological map contains all relevant information and is conditionally independent of all earlier states. Consequently, we write Equation 3 in an incremental way

$$P(G^{t+1}_{map}|l^{0:t+1}_{env}) \propto P(l^{t+1}_{env}|G^{t}_{map})P(G^{t}_{map}) \tag{4}$$

to estimate the posterior distribution on topological maps $P(G^{t+1}_{map}|l^{0:t+1}_{env})$ from the prior distribution $P(G^{t}_{map})$ using the measurement likelihood $P(l^{t+1}_{env}|G^{t}_{map})$. It is difficult to represent uncertainty directly in a topological map so we use a sequential Monte-Carlo technique to represent uncertainty by maintaining a collection of $N$ map candidate samples which are randomly drawn from the probability density function in the space of topological environment maps. A collection of map samples $\left\{G^{t+1}_{map,i}, w^{t+1}_{i}\right\}^{N-1}_{i=0}$ is used to model the posterior distribution $P(G^{t+1}_{map}|h^{0:t+1}_{env})$ on topological maps. The weights $w^{t+1}_{i}$ are importance factors which are normalised such that

$$\sum_{i=0}^{N-1} w^{t+1}_{i} = 1. \tag{5}$$

Bayesian filters recursively estimate the posterior distribution $P(G^{t+1}_{map}|l^{0:t+1}_{env})$ from the proposal distribution $P(G^{t}_{map})$ and the perception $l^{t+1}_{env}$ using two distinct phases: Prediction and Update.

The prediction phase uses the map estimate from the previous time step to estimate the map at the next time step. In our case, we can not predict the next observation as the map is not known in advance so the probability which vertex corresponds to the place the robot visits next is uniformly distributed over the vertices contained in the map and an additional vertex which refers to a new place. Therefore, for each sample $G^{t}_{map,i}$ of the collection of samples that models the posterior distribution $P(G^{t}_{map}|h^{0:t}_{env})$ at $t$, map graph candidates $\left\{\widetilde{G}^{t+1}_{map,i,k}\right\}^{|V_{map,i}|}_{k=0}$ are generated where $k$ refers to the vertex which the new observation is predicted to correspond to. Gaussian white noise $\mathcal{N}(0, \sigma_g)$ is added artificially to the labels of the map graph candidates to model the inherent uncertainties associated with robot perception. If an additional vertex is introduced it is labelled with the new observation. If the current location and the place which is predicted to be occupied next by the robot are not connected in the topological map yet, an edge is introduced to connect the corresponding vertices.

The proposal distribution on topological maps is updated using the new observation. In Monte-Carlo approaches this is done by weighting the samples using the data. We compute the weight of a topological map graph candidate $\widetilde{G}^{t+1}_{map,i,k}$ using the current observation $l^{t+1}_{env}$ and the set of $n$-grams $\Gamma_{env}$, so

$$w_i = P(l^{t+1}_{env}, \Gamma_{env}|\widetilde{G}^{t+1}_{map,i,k}) = P(l^{t+1}_{env}|\widetilde{G}^{t+1}_{map,i,k})P(\Gamma_{env}|\widetilde{G}^{t+1}_{map,i,k})P(\widetilde{G}^{t+1}_{map,i,k}). \quad (6)$$

The term $P(l^{t+1}_{env}|\widetilde{G}^{t+1}_{map,i,k})$ computes the probability of the measured label and the label of the vertex $k$ to be identical, so

$$P(l^{t+1}_{env}|\widetilde{G}^{t+1}_{map,i,k}) = \exp\left(-\left(\frac{l^{t+1}_{env} - \tilde{l}_k}{\sigma_l}\right)^2\right) \quad (7)$$

where $\tilde{l}_k \in \widetilde{L}^{t+1}_{map,i,k}$ and $\sigma_l$ denotes a weighting factor.

The second term in Equation 6 considers the probability that map graph $\widetilde{G}^{t+1}_{map,i,k}$ is $n$-consistent with the information given in the history. Using Equation 1 the consistency probability is computed with

$$P(\Gamma_{env}|\widetilde{G}^{t+1}_{map,i,k}) = \exp\left(-\left(\frac{d_H(\Gamma_{env}, \widetilde{\Gamma}^{t+1}_{map,i,k})}{\sigma_c}\right)^2\right) \quad (8)$$

where $\widetilde{\Gamma}_{map,i,k} = Grams(\tilde{h}^{t+1}_{map,i,k}, n)$ is generated from $\tilde{h}^{t+1}_{map,i,k} \in \widetilde{L}^{*,t+1}_{map,i,k}$.

In parametric methods the probability of the model to represent the data is increased when the number of parameters in the model is increased. In our case, a map graph which consists of one component for each $n$-gram in $Grams(h_{env}, n)$ would be consistent with the information from the history but is inappropriate for navigation, containing too many vertices. Hence, the prior should favour small topological maps. Consequently, the last term in Equation 6 penalises map graphs which contain vertices with similar labels

$$P(\widetilde{G}^{t+1}_{map,i,k}) = \prod_{a=0}^{|\widetilde{V}^{t+1}_{map,i,k}|} \prod_{b=0}^{|\widetilde{V}^{t+1}_{map,i,k}|} \left(1 - \phi\left(\exp - \left(\frac{\tilde{l}_a - \tilde{l}_b}{\sigma_l}\right)^2\right)\right) \quad (9)$$

where $\tilde{l}_a, \tilde{l}_b \in \widetilde{L}^{t+1}_{map,i,k}$ and $\phi$ weights the influence of the penalty.

The posterior distribution on topological maps is computed by drawing $N$ samples from the proposal distribution.

### 3.2 Localisation

While building the map, the place the robot occupies is implicitly estimated whenever the map graph is updated with a new observation. The vertex whose label is updated or additionally introduced using the observation indicates the estimated location of the robot. If a new vertex is introduced the robot is hence located at the place which corresponds to that vertex. The location, in turn, is used to guide the mapping process by introducing adjacencies between the current and the previous place occupied.

## 4 Experiments

Our experimental set up environment covers an indoor office environment area of about 20,000 square meters (Wean Hall at Carnegie Mellon University), see Figure 1. The robot uses a panoramic camera to acquire information about the environment.

The experimental platform traverses the environment using the generalised Voronoi graph (GVG) strategy developed by Choset and Nagatani [16]. It is based on the Voronoi diagram which is a special kind of decomposition of a metric space into segments and nodes determined by distances to a specified discrete set of objects in the space. Our robot measures distances using sonar readings.

Segments capture the points in the plane that are equidistant to two sites. Travelling along the Voronoi segments, the robot can keep in the middle of corridors while exploring the environment. The Voronoi nodes are the points equidistant to three (or more) obstacles. In indoor environments this corresponds naturally to T-junctions or intersections of corridors as shown in Figure 1. Once such a locus point is identified, a panoramic image is taken.

Note, we use the GVG strategy only for exploring the environment and the identification of places. Mapping and localisation is performed using purely visual information. The GVG of our experimental environment is displayed in Figure 1.

In some areas of our experimental environment bright ceiling lights are installed whereas some other areas have wall lights. As a result, the images the robot takes at each place suffer from loss of clarity of visual information within shadows or near strong lights (over and under exposed regions). We use a method

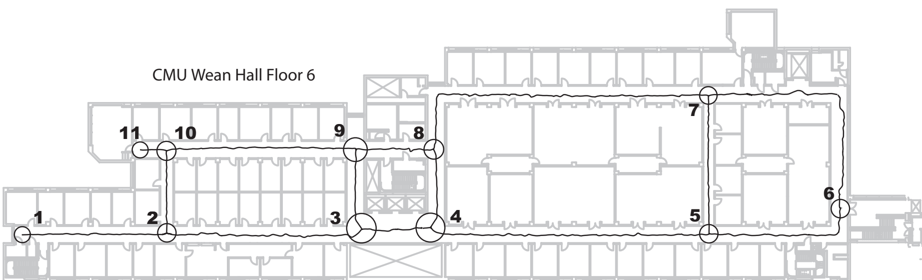


**Fig. 1.** The floor plan of Wean Hall Floor 6 at Carnegie Mellon University (CMU). Embedded is the topological graph (vertices and their connectivity) that reflects the ground truth of the topological map that we wish the SLAM algorithm to infer.

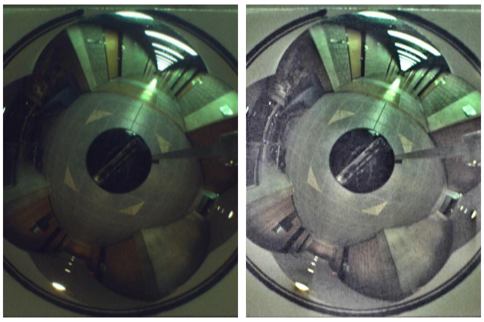

**Fig. 2.** An under exposed (left) image measured at place 10 in Wean Hall (see Figure 1) and the enhancement (right) using Vonikakis and Andreadis method [18].

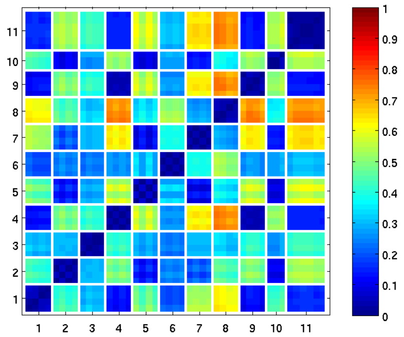


**Fig. 3.** Distance matrix of the fingerprints of places (see Figure 1). Blue small distance, red big distance.

proposed by Vonikakis and Andreadis to enhance the acquired images by lightening under exposed regions and darkening over exposed regions without affecting the correctly exposed ones [18]. Figure 2 shows the application of the enhancement method for an under exposed image.

In our system, the visual appearance of a place is measured using colour histograms. Usually a colour histogram is created by calculating an $N$-bin histogram for each of the $R, G$ and $B$ colour bands and so looses the 3D spatial information of the $RGB$ tuples in colour space. To retain the 3D spatial information of the $RGB$ tuples in colour space, we use 3D histograms in $RGB$ space where the histogram consists of $N^3$ equally sized bins.

The number of $n$-grams we can extract at most from a history $h$ of lenth $m$ is at most $m - n + 1$ whereby the maximum number of unique $n$-grams which are derivable from a strongly connected graph is $O(|L|^n)$ [3]. However, using fast vision techniques such as 3D colour histograms with $5^3$ bins keeps the system fast despite potentially big sets of $n$-grams. In comparison, a single standard SIFT feature as is represented through a 128 dimensional vector, whereby thousands of such features may be identified in a single image [7].

We have conducted several exploration runs and recorded a total of 50 images of places and stored the fingerprints of the enhanced images in a data base. Figure 3 displays the similarity matrix of the fingerprints from the enhanced images. It is apparent that the environment contains numerous topological ambiguities when using colour histograms as fingerprints of the places.

Given the data base and the ground truth environment graph, we can simulate arbitrary traversals of the environment. The robot starts at an initial vertex and selects an arbitrary adjacent vertex as next place. According to the vertex the robot occupies, a random observation from the data base is sampled. For the following evaluations, $2{\times}200$ paths of length 70 were generated. One set of paths uses fingerprints from the enhanced images and the other set uses fingerprints from the original images. Each path represents an exhaustive exploration of the

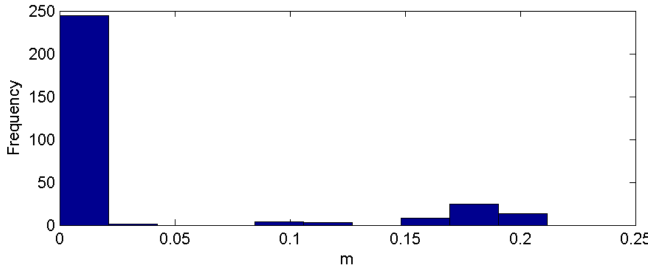


**Fig. 4.** The consistency measured using the Hausdorff distance of the inferred topological maps and the corresponding simulated histories is displayed. Most of the inferred maps using the enhanced visual perceptions are highly consistency respective the consistency measure.

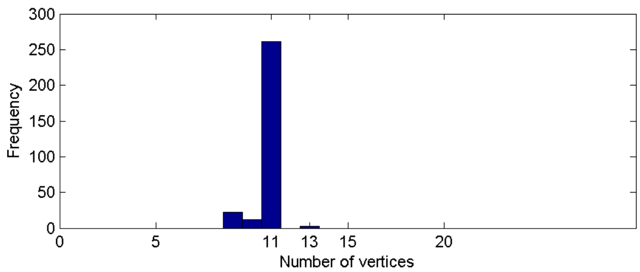


**Fig. 5.** The number of vertices of the inferred topological maps are histogrammed. Most of the inferred maps have only small consistency errors which occur due to measurement noise. A clear peak at 11 vertices is to see what corresponds to the environment graph (see Figure 1).

environment. For each path the set $\Gamma_{env}$ of $n$-grams is derived before starting the algorithm. Note, it is actually not necessary to assume an exhaustive exploration of the environment as the inferred map is a representation of the environment which is consistent to the measurements at a certain time.

We have applied the proposed SLAM method using 30 map candidates to model the posterior distribution on topological maps and 3-consistency mapping. The fingerprints of places are represented through 3D colour histograms with $5^3$ bins from the enhanced images.

Figure 4 shows a histogram of the 3-consistency of the inferred maps of the simulated random traversals. It can be seen that most of the inferred maps using the enhanced images are very consistent with the information from the history, whereby little divergence occurs as a result of measurement noise inherent to sensor perception. Rare outliers may occur when the mapping process is misled due its probabilistic nature so the inferred map is inconsistent with the observations.

The overall goal in topological mapping is to build an internal representation which is isomorphic to the environment. Here, we investigate whether the inferred map graphs are isomorphic to the environment graph in order to measure the quality of the proposed approach for topological mapping. In the case of applying the image enhancement method, we found all map graphs (193, or 97%) with the same number of vertices as the environment graph to be isomorphic to the environment graph (see Figure 5). The results support the strategy of the proposed algorithm to use the current position estimation with the new observation to map the connectivity of the environment.

## 5 Discussion

In this paper we proposed a Bayesian approach for topological SLAM that does not rely on any motion model or metric information, but uses a history of noisy visual measurements from visited places only. Using colour histograms as fingerprints of places makes our system fast but entails physically different places to appear similar to the robot's senses. In order to deal with this problem, the sequential Monte-Carlo SLAM technique embeds a method to reliably disambiguate places which appear to be the same but in fact are different. The method aims to maintain consistency with the observed data while minimising the number of vertices contained in the map. The consistency between a topological map and the observations is measured using the Hausdorff distance.

Experiments in an indoor environment which is subject to severe ambiguities due to repeated structures demonstrate the capability of the idea to use neighbourhood clues in order to disambiguate otherwise identical vertices. Our approach mostly infers topological maps with only small inconsistencies with respect to the data. Moreover, most of the resulting maps are isomorphic to the environment graph what supports reliable topologial navigation despite severe perceptual aliasing.

## Acknowledgements

NICTA is funded by the Australian Government as represented by the Department of Broadband, Communications and the Digital Economy and the Australian Research Council through the ICT Centre of Excellence program.

Computational resources and services used in this work were provided by the HPC and Research Support Group, Queensland University of Technology, Brisbane, Australia.

The authors would like to thank Stephen Tully for the help with the experiments and Howie Choset for valuable discussions.

## References

1. Filliat, D., Meyer, J.A.: Map-based navigation in mobile robots: I. a review of localization strategies. Cog. Sys. Res. 4(4), 243–282 (2003)
2. Thrun, S., Burgard, W., Fox, D.: Probabilistic Robotics. MIT Press, Cambridge (2005)

3. Werner, F., Gretton, C., Maire, F., Sitte, J.: Induction of topological environment maps from sequences of visited places. In: IROS, Nice, France, pp. 2890–2895. IEEE, Los Alamitos (2008)
4. Angeli, A., Doncieux, S., Meyer, J.A., Filliat, D.: Incremental vision-based topological slam. In: IROS, September 2008, pp. 1031–1036 (2008)
5. Valgren, C., Duckett, T., Lilienthal, A.: Incremental spectral clustering and its application to topological mapping. In: ICRA, Roma, Italy, pp. 4283–4288. IEEE, Los Alamitos (2007)
6. Cummins, M., Newman, P.: Fab-map: Probabilistic localization and mapping in the space of appearance. Int. J. Rob. Res. 27(6), 647–665 (2008)
7. Lowe, D.: Distinctive image features from scale-invariant keypoints. In: IJCV, vol. 60, pp. 91–110. Kluwer Academic Publishers, Dordrecht (2003)
8. Bay, H., Tuytelaars, T., Van Gool, L.: SURF: Speeded up robust features. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006. LNCS, vol. 3951, pp. 404–417. Springer, Heidelberg (2006)
9. Tuytelaars, T., Mikolajczyk, K.: Local invariant feature detectors: A survey. FnT Comp. Graphics and Vision 3(3), 177–280 (2008)
10. Werner, F., Sitte, J., Maire, F.: Visual topological localisation and mapping using colour histograms. In: ICARCV, Hanoi, Vietnam. IEEE, Los Alamitos (2008)
11. Werner, F., Sitte, J., Maire, F.: Automatic place determination using colour histograms and self-organising maps. In: ICAR, Jeju, Korea Republic, August 2007, pp. 111–115 (2007)
12. Werner, F., Sitte, J., Maire, F.: On the induction of topological maps from sequences of colour histograms. In: DICTA, Adelaide, Australia, pp. 167–174. IEEE, Los Alamitos (2007)
13. Ulrich, I., Nourbakhsh, I.R.: Appearance-based place recognition for topological localization. In: ICRA, pp. 1023–1029 (2000)
14. Rañó, I., Lazkano, E., Sierra, B.: On the application of colour histograms for mobile robot localisation. In: EMCR, vol. 1, pp. 189–193 (2005)
15. Ranganathan, A., Dellaert, F.: Inference in the space of topological maps: An mcmc-based approach. In: IROS, Sendai, Japan, vol. 2, pp. 1518–1523. IEEE, Los Alamitos (2004)
16. Choset, H., Nagatani, K.: Topological simultaneous localization and mapping (slam): Toward exact localization without explicit localization. IEEE Trans. Robot. Automat. 17(2), 125–137 (2001)
17. De, A., Lee, J., Cowan, N.J.: Toward slam on graphs. In: WAFR, Guanajuato, México (December 2008)
18. Vonikakis, V., Andreadis, I.T.: Fast automatic compensation of under/Over- exposured image regions. In: Mery, D., Rueda, L. (eds.) PSIVT 2007. LNCS, vol. 4872, pp. 510–521. Springer, Heidelberg (2007)