Robotica (2014) volume 32, pp. 803–821. © Cambridge University Press 2013
doi:10.1017/S0263574713001070

Topological simultaneous localization and mapping: a survey

Jaime Boal*, Álvaro Sánchez-Miralles and Álvaro Arranz

Institute for Research in Technology (IIT), ICAI School of Engineering, Comillas Pontifical University, Madrid, Spain

(Accepted October 29, 2013. First published online: December 3, 2013)

SUMMARY
One of the main challenges in robotics is navigating autonomously through large, unknown, and unstructured environments. Simultaneous localization and mapping (SLAM) is currently regarded as a viable solution for this problem. As the traditional metric approach to SLAM is experiencing computational difficulties when exploring large areas, increasing attention is being paid to topological SLAM, which is bound to provide sufficiently accurate location estimates, while being significantly less computationally demanding. This paper intends to provide an introductory overview of the most prominent techniques that have been applied to topological SLAM in terms of feature detection, map matching, and map fusion.

KEYWORDS: Mobile robots; SLAM; Topological modeling of robots; Feature detection; Robot localization

1. Introduction
Mobile robotics' ultimate aim is to develop fully autonomous entities capable of performing rather complicated tasks, without the need for human intervention, during extended periods of time. Over the past three decades, this objective has constantly faced harsh difficulties, which have hindered progress. The most recurrent issues in the literature, which are yet to be completely resolved, are stated below.

A mobile robot must be able to navigate through the environment in order to achieve its goals. According to Leonard and Durrant-Whyte,[63] this general problem can be summarized in three questions: "Where am I?," "Where am I going?," and "How should I get there?" The first question addresses the *localization problem*, which intends to estimate the robot's pose (i.e., location and orientation) using data gathered by distinct sensors and knowledge of previous locations. However, the presence of noisy sensor measurements makes this problem harder than it may seem at first sight. The precision with which this problem is solved decisively affects the answer to the other two questions, as it is necessary to localize oneself in the environment to safely interact with it, decide what the following step should be, and how to accomplish it.

During the localization process, a robot must resort to some kind of reference system; in other words, it requires a map. The extensive research survey carried out by Thrun[110] collects the main open issues concerning *robotic mapping*, which are succinctly presented henceforth. Currently, there are robust methods for mapping structured, static, and bounded environments, whereas mapping unstructured, dynamic, or large-scale unknown environments remains largely an unsolved problem.

According to Thrun,[110] the robotic mapping problem is "that of acquiring a spatial model of a robot's environment." To this end, robots must be equipped with sensors that enable them to perceive the outside world. Once again, sensor errors and range limitations pose a great difficulty.

The first challenge in robotic mapping develops from the measurement noise. Usually, this issue can be overcome if the noise is statistically independent, as it can be canceled out performing enough measurements. Unfortunately, this does not always occur in robotic mapping because, whenever

* Corresponding author. E-mail: jaime.boal@iit.upcomillas.es

incremental sensors (e.g., encoders) are used, errors in navigation control accumulate progressively and condition the way in which subsequent measurements are interpreted. As a result, if a robot does not rely on the layout of the environment whatever it infers about its surroundings is plagued by systematic, correlated errors. Leonard and Durrant-Whyte[64] state the *correlation problem* as follows:

> If a mobile robot uses an observation of an imprecisely known target to update its position, the resulting vehicle position estimate becomes correlated with the feature location estimate. Likewise, correlations are introduced if an observation taken from an imprecisely known position is used to update the location estimate of a feature in the map.

The second difficulty of the robot mapping problem derives from the amount and complexity of the features required to describe the objects that are being mapped, as the computational burden grows exponentially as the map becomes more detailed. Obviously, it is absolutely different to restrict to the description of corridors, intersections, and doors, than to build a 3D visual map.

A third, and perhaps the hardest, issue is the *correspondence problem*, which attempts to determine if sensor measurements taken at different times correspond to the same physical entity. A specific instance of this problem occurs when returning to an already visited area, because the robot has to realize that it has arrived at a previously mapped location. This is known as the *loop-closing problem*. Another particular case is the so-called *first location problem* or *kidnapped robot problem*,[53] which occurs when a robot is placed in an unknown position of an environment for which it has a map.

Fourth, the vast majority of environments are dynamic. Doh *et al.*[28] further classify the concept of dynamic environments in *temporary* dynamics, which are instantaneous changes that can be discarded by consecutive sensor measurements (e.g., moving objects like walking people), and *semi-permanent* dynamics or *scene variability*,[58] which are changes that persist for a prolonged period of time. This second type of dynamics makes the correspondence problem even more difficult to solve, as it provides another manner in which apparently inconsistent sensor measurements can be interpreted. Suppose a robot perceives a closed door that was previously modeled as open. This observation may be explained by two equally plausible hypotheses: either the door position has changed, or the robot is in error about its current location. At present, there are almost no mapping algorithms capable of coping with this difficulty. On the contrary, most approaches assume a static world and, as a consequence of this simplification, anything that moves apart from the robot is regarded as noise. In fact, the majority of the experimental tests in the literature are carried out in rather controlled environments and never mention how to deal with these troublesome dynamics. Doh *et al.*[28] are an exception to this trend due to the fact they take door position changes into consideration.

Finally, robots must navigate through the environment while mapping on account of sensor range limitations. The operation of generating navigation commands with the aim of building a map is known as *robotic exploration*. Although the commands issued during the exploration of the environment provide relevant information about the locations at which different sensor measurements were obtained, motion is also subject to errors (e.g., wheel slippage). Therefore, these controls alone are insufficient to determine a robot's pose.

2. The Simultaneous Localization and Mapping (SLAM) Problem

As mentioned by Thrun,[110] the localization and mapping problems are often tackled together in the literature. Essentially, both problems are uncertain and, when trying to solve them individually, the other introduces systematic error. By contrast, estimating both at the same time makes the measurement and control noises independent. Notice, nevertheless, that robot mapping is like the *chicken and egg problem*: "A robot needs to know its position to build a map, and it requires a map in order to determine its position.[120]"

The immediate question inferred from this idea is if it is possible for a mobile robot to be placed at an unknown location in an unknown environment and, despite this, incrementally build a consistent map of the environment using local information while simultaneously determining its location within this map. This is known as the *simultaneous localization and mapping (SLAM) problem*.[8,33] During more than a decade, a solution to this issue has been regarded as a key milestone in the pursuit for truly autonomous robots. At present, it can be safely asserted that the SLAM problem has been solved in different manners, at least, from a theoretical point of view. Notwithstanding, substantial issues remain open concerning the implementation of these SLAM solutions.

The majority of the problems that researchers are currently facing are those of computational nature.[8] In order to overcome the correspondence problem, each location in the environment must be unequivocally distinguishable from all the rest. This implies gathering either plenty of similar features or a more restricted number with richer information in every place analyzed. In any case, the computational burden rapidly increases to intractable levels in large environments. Therefore, most approaches make a trade-off between computation times and precision or global distinctiveness, that is, they either limit the number of locations considered or reduce the number of features analyzed in each place.

3. Types of Maps

This paper has so far referred to mapping as a whole. However, there exist several types of maps that require diverse data acquisition techniques and present different associated problems. In general, maps can be divided into the four groups listed below:

- *Metric maps* represent the environment as a set of object or obstacle coordinates with the aid of raw data and geometric features (e.g., lines, edges). Although localization and mapping with this approach can be very accurate and result in very high precision representations of the environment, the required data volume grows at a much higher rate than the size of the region being mapped and, therefore, involves complicated calculations.[36,64]
- Conversely, *topological maps* model the environment as a graph. They are based on the discretization of the continuous world into a finite set of places (nodes) connected according to their relative position in the environment. These maps provide a compact representation, since only distinctive places within the environment are encoded. Consequently, they are much less computationally demanding, as there is no need for a precise localization, and navigation commands follow naturally from the graph. Nevertheless, the main problem of this method is *perceptual aliasing*, in other words, that there is always a risk that two distinct locations appear identical to the robot's sensors.[18,36,56]
- *Hybrid maps* are a combination of the previous two that intend to compensate the drawbacks of both approaches when applied alone. On the one hand, reduce the computational burden of metric maps and, on the other hand, increase topological distinctiveness. To this end, they use a global topological map to move between places, and rely on a metric representation in bounded local spaces for precise navigation.[11,83,121] It is important to bear in mind that these maps are often referred to in the literature as *hierarchical*. However, this term should be avoided, as it can be easily confused with topological graph representations that involve several abstraction processes (i.e., create an *atlas* with progressively detailed sub-maps).[66]
- Finally, *semantic maps* contain, in addition to spatial information about the environment, assignments of the mapped features to entities of known classes. This means that they hold data on objects, functionalities, events, or relations in the robot's environment whose knowledge permits a high-level goal-directed behavior, enables reasoning, and helps to resolve location ambiguities.[84]

According to the previous definitions, maps can be sorted in increasing level of abstraction in metric, hybrid, topological, and semantic (Fig. 1).

4. Why Choose a Topological Approach?

In principle, two classical opposite approaches exist to address the SLAM problem. The first one models the environment using a metric map, enabling an accurate estimation of the robot's position. It provides a dense representation of the environment, which has large storage requirements, and is particularly well suited to precise trajectory planning.

In the second approach, the environment is segmented into distinctive places using a topological map, which relies on a higher level of representation than metric mapping, making symbolic goal-directed planning and navigation possible. It also provides a more compact representation that is more in accordance with the size of the environment[4] in spite of requiring more complex sensory information that often implies more processing. The largely cited papers by Kuipers and Levitt[60] and Kuipers and Byun[59] can be regarded as the seminal work that triggered a paradigm shift from a metric to a topological approach in robotic map building. Contrary to previous developments, which extracted

https://doi.org/10.1017/S0263574713001070 Published online by Cambridge University Press

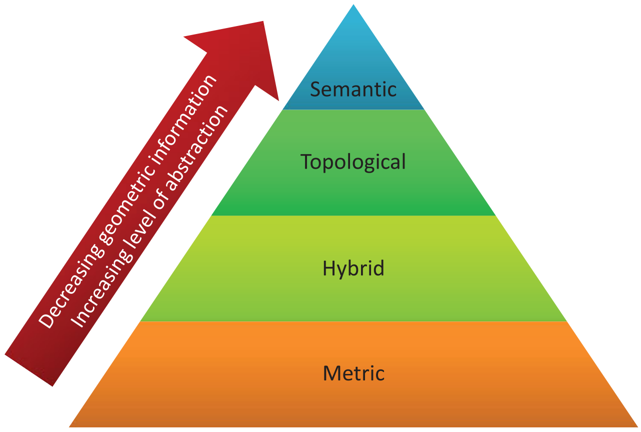

Fig. 1. (Colour online) Level of abstraction hierarchy for maps.

the geometry of the environment from sensor measurements and then inferred the topology from it (see Chatila and Laumond, 1985[17] for instance), they proposed constructing a topological description based on simple control strategies in the first place, and incorporate local metric information in each of the identified nodes afterwards.

Albeit, considering that metric maps are more accurate and that a hybrid approach helps to overcome storage problems, why should purely topological maps be used? To begin with, topological navigation is a behavior employed by a variety of different animal species, including human beings. We do not need to answer the question "Where am I?" in millimeters and degrees in order to safely move through the environment.[15] On the contrary, rather than navigating using coordinates, we have an abstract notion of distance but are still able to recognize where we are in space.[89] Moreover, Brooks[14] supports the belief that topological maps are a means of coping with uncertainty in mobile robot navigation. The absence of metric and geometric information, which is replaced by notions of proximity and order, eliminates dead-reckoning error issues, which no longer accumulate.

In conclusion, topological representation resembles human intuitive navigation system, which has been proven to deal efficiently with uncertainty, and results in a straightforward map from which path planning follows naturally.

5. Topological SLAM

5.1. Breaking up the problem

Implementing a topological SLAM algorithm is a four step process. First of all, it is essential to define what is going to be considered a landmark in the environment and choose the appropriate sensor technologies to perceive them. Once this decision is made, the following step is to determine which feature extraction algorithms are going to be applied.

Afterwards, the gathered data must be compared with the stored nodes, starting with the last observed. Due to the fact that it is almost impossible to extract exactly the same features when revisiting a place, and that several locations may look alike, the most common situation is that the robot is uncertain about its position after performing this comparison. This is depicted in Fig. 2 with gray nodes. The robot may either be in various known positions or, alternatively, have reached a new node (illustrated in Fig. 2 by a discontinuous line).

Consequently, the robot is forced to keep record of the probability of being in each node until the uncertainty is somehow resolved. At this point, both the robot's location and the map become simultaneously unambiguous. Should it happen to be no match, the system must determine if the current location is susceptible of being labeled as a distinctive place according to the adopted criteria. Otherwise, data should be added to the current node definition in order to enhance its distinctiveness for future revisiting.

https://doi.org/10.1017/S0263574713001070 Published online by Cambridge University Press

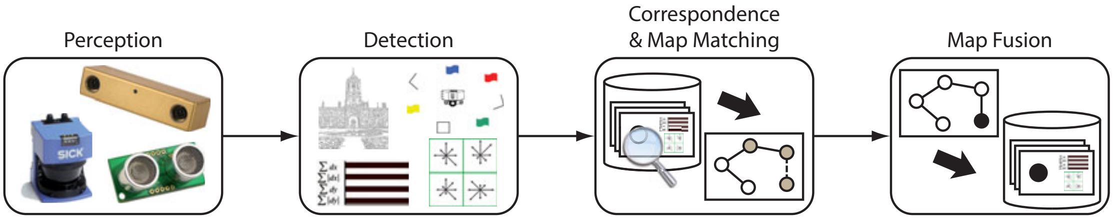

Fig. 2. (Colour online) Topological SLAM overview. From left to right: the system acquires sensory information from one or several sources; selected features are extracted and encoded; the current location is compared with a database of previously visited nodes resulting in a belief state (i.e., the robot could be in several locations with different probabilities); finally, once the uncertainty has been resolved, either a new node is added to the database or the information of an existing one is updated.

5.2. Perception and detection

In previous sections, it has been asserted that place definition plays a key role in topological map building. The underlying problem concerning this matter could be stated as that of deciding whether a given location should be encoded as a place.

The most common solution involves looking for places that are locally distinguishable and store the distinctive features or landmarks. According to Stankiewicz and Kalia,[106] the use of landmarks implicitly assumes three properties: persistence, saliency, and informativeness. To begin with, a landmark should be *persistent*, that is, the features should still be present when the robot returns to the location anytime in the future. Furthermore, it ought to be *perceptually salient*, which means that the landmark must be easily detectable and identifiable. Finally, a landmark needs to be *informative*. In other words, it should provide evidence about the robot's pose or the action it should take when observing it.

Following with the reasoning by Stankiewicz and Kalia,[106] there exist two different types of landmarks: structural and object landmarks. The former are defined as geometric features that can serve as cues, like intersections or entrances, named *gateways* by Kortenkamp and Weymouth,[56] and corners or edges.[108,111] The latter are objects in the environment that are independent of its structure, such as signs. These are often identified using computer vision by means of interest points or regions. From these definitions, it is intuitively obvious that object landmarks typically provide more information concerning spatial coordinates, as two intersections look alike but a poster on a wall is probably unique. Unfortunately, it is more than somewhat unlikely to find a single type of cue that combines all of the previous properties.

This section concentrates on the sensors and algorithms applied in the literature to detect and extract landmarks from an environment. Although not all references cited henceforth correspond strictly to SLAM implementations, they are relevant insofar as they present several techniques for topological feature detection. Table I collects the different sensor technologies that have been applied to extract topological data over the past two decades. Papers that build on earlier work are grouped together. As stated by Ranganathan and Dellaert,[93] laser range scanners are currently *de facto* standard in robotics, due to their ability to provide precise depth estimates and form dense point clouds that resemble the scene structure, although substantial research is currently being carried out on computer vision due to recent progress in image processing, and because cameras are typically less expensive and provide more distinctive features, which is fundamental for topological SLAM.[72,88]

The use of visual data as the primary source of information in SLAM systems has not had time to converge to generally efficient and robust solutions yet, hence leaving much room for experimentation and improvement. Notwithstanding, albeit perceiving the world through a camera lens can be less accurate than laser range sensing, the richness of the information encoded has already proved to be sufficient to obtain reliable estimates of camera motion and scene structure. However, it is important to point out that the vast majority of the articles reviewed in Table I opt for omnidirectional cameras. This can be easily explained by the fact that omnidirectional cameras are the only ones that guarantee *rotational invariance* (i.e., no matter what orientation a robot has in a given location, the image captured is always the same).

Table II shows the numerous methods employed in the references presented in Table I. At early stages, due to the fact that the only widespread sensor technology was sonar, feature detection reduced

https://doi.org/10.1017/S0263574713001070 Published online by Cambridge University Press

Table I. Sensors used in the literature to identify topological landmarks.

					Cameras		
Reference	Sonar	Laser	Encoder	Compass	Monocular	Omnidir.	Stereo
Kuipers & Byun, 1991[59]	✓			✓			
Kortenkamp & Weymouth, 1994[56]	✓				✓		
Owen & Nehmzow, 1998[85]	✓			✓			
Gutmann & Konolige, 1999[45]		✓	✓				
Hafner, 2000[46]				✓		✓	
Ulrich & Nourbakhsh, 2000[113]						✓	
Choset & Nagatani, 2001[18]	✓	✓	✓				
Tomatis *et al.*, 2002[111]		✓					
Anguelov *et al.*, 2004[7]		✓				✓	
Kuipers *et al.*, 2004[61]	✓	✓					
Modayil *et al.*, 2004[78]		✓	✓				
Andreasson *et al.*, 2005[3]			✓			✓	
Goedemé *et al.*, 2005,[43] 2007[41]						✓	
Stachniss *et al.*, 2005[105]		✓			✓		
Tapus, 2005[108,109]		✓	✓			✓	
Zivkovic *et al.*, 2005[121]						✓	
Fraundorfer *et al.*, 2007[39]					✓		
Vasudevan *et al.*, 2007[114]		✓	✓				✓
Angeli *et al.*, 2008[4–6]					✓		
Cummins & Newman, 2008,[22] 2011[25]					✓		
Koenig *et al.*, 2008[52]			✓			✓	
Nüchter & Hertzberg, 2008[84]		✓					
Ranganathan & Dellaert, 2008,[93] 2011[94]		✓			✓		
Sabatta *et al.*, 2008,[100] 2010[99]						✓	
Liu *et al.*, 2009,[67] Liu & Siegwart, 2012[68]						✓	✓
Doh *et al.*, 2009[28]		✓	✓				
Tully *et al.*, 2009[112]	✓		✓				
Werner *et al.*, 2009,[116] 2012[119]						✓	
Werner *et al.*, 2009[117]	✓						
Lui & Jarvis, 2010[72]					✓	✓	✓
Romero & Cazorla, 2010,[97] 2012[98]						✓	
Maddern *et al.*, 2011,[73] 2012[74]			✓			✓	
Total	7	12	10	3	7	14	3

to what has been called geometric features in Table II (i.e., distances to different obstacles that allow to identify simple topological landmarks such as corners or dead ends) and gateways, which are an extension of the previous to detect openings.

With the rise of laser range scanners, these approaches became more precise owing to the acquisition of dense point clouds and, more recently, with the introduction of computer vision techniques, simple methods like color histograms were applied. For instance, Ulrich and Nourbakhsh[113] extract histograms in the RGB and HSL color spaces from omnidirectional images. However, it was soon widely accepted that the information obtained from histograms was not sufficiently distinctive and reliable—they can be potentially identical for two images with different content, and are very sensitive to illumination changes—to use them as a sole characteristic detector. Thus, this approach has now become a part of, or a complement for, other more consistent and

https://doi.org/10.1017/S0263574713001070 Published online by Cambridge University Press

Table II. Landmark extraction techniques for topological navigation grouped according to sensor technologies: range sensors (i.e., sonar and laser), cameras, and both; and the type of features obtained: distances, lines, frequency-based, edges, keypoints, affine regions, and probabilistic.

	Range sensors						Cameras									R/C
	Distances		Lines				Frequency		Edges		Keypoints			Regions		Prob.
Reference	Geometric features	Gateways	Douglas-Peucker	EM	RANSAC	Hough trans.	Hist.	Haar wavelets	Sobel operator	Invariant columns	SIFT	SURF	KLT	Harris-affine	MSER	Bayesian surprise
Kuipers & Byun, 1991[59]	✓															
Kortenkamp & Weymouth, 1994[56]		✓							✓							
Owen & Nehmzow, 1998[85]	✓															
Gutmann & Konolige, 1999[45]	✓															
Hafner, 2000[46]							✓									
Ulrich & Nourbakhsh, 2000[113]							✓									
Choset & Nagatani, 2001[18]	✓															
Tomatis *et al.*, 2002[111]	✓	✓														
Anguelov *et al.*, 2004[7]	✓	✓		✓		✓	✓									
Kuipers *et al.*, 2004[61]		✓														
Modayil *et al.*, 2004[78]	✓															
Andreasson *et al.*, 2005[3]							✓				✓					
Goedemé *et al.*, 2005,[43] 2007[41]										✓	✓					
Stachniss *et al.*, 2005[105]	✓															
Tapus, 2005[108,109]			✓				✓		✓							
Zivkovic *et al.*, 2005[121]											✓					
Fraundorfer *et al.*, 2007[39]															✓	
Vasudevan *et al.*, 2007[114]			✓								✓					
Angeli *et al.*, 2008[4–6]											✓					
Cummins & Newman, 2008,[22] 2011[25]												✓				

Table II. Continued.

	Range sensors						Cameras									R/C
	Distances		Lines				Frequency		Edges		Keypoints			Regions		Prob.
Reference	Geometric features	Gateways	Douglas-Peucker	EM	RANSAC	Hough trans.	Hist.	Haar wavelets	Sobel operator	Invariant columns	SIFT	SURF	KLT	Harris-affine	MSER	Bayesian surprise
Koenig *et al.*, 2008[52]							✓									
Nüchter & Hertzberg, 2008[84]		✓			✓											
Ranganathan & Dellaert, 2008,[93] 2011[94]											✓			✓	✓	✓
Sabatta *et al.*, 2008,[100] 2010[99]											✓					
Doh *et al.*, 2009[28]	✓	✓														
Liu *et al.*, 2009,[67] Liu & Siegwart, 2012[68]							✓		✓							
Tully *et al.*, 2009[112]	✓															
Werner *et al.*, 2009,[116] 2012[119]							✓									
Werner *et al.*, 2009[117]	✓															
Lui & Jarvis, 2010[72]								✓				✓	✓			
Romero & Cazorla, 2010,[97] 2012[98]															✓	
Maddern *et al.*, 2011,[73] 2012[74]												✓				
Total	11	6	2	1	1	1	8	1	3	1	7	3	1	1	3	1

https://doi.org/10.1017/S0263574713001070 Published online by Cambridge University Press

informative methods. In addition, other procedures like line extractors, Haar wavelets, edge, keypoint, and affine covariant region detectors, and Bayesian surprise saw the light of day.

The rest of this section concentrates on the detection methods found in the literature. For those techniques that are common knowledge in the field, only references to surveys or seminal papers are put forward. Emphasis is put on the more recent and original techniques.

5.2.1. Line extractors. Human-made environments are full of vertical and horizontal lines and, therefore, constitute an invaluable source of topological information. Line extraction techniques are usually employed in conjunction with laser range scanners. There exist many approaches for line extraction, some of which are compared by Nguyen *et al.*[82] As far as topological feature detection is concerned, the Douglas-Peucker algorithm[108] (also known as *split-and-merge*), EM (Expectation-Maximization) applied to line fitting,[87] the Hough transform,[38] and RANSAC (RANdom SAmple Consensus)[37] have been employed. Finally, it is worth mentioning that the latter is a general algorithm for model adjustment in the presence of many data outliers, which has further applications, for instance, Nüchter and Hertzberg,[84] adopt this technique for plane extraction.

5.2.2. Haar wavelets. Yet another attempt to topological feature extraction is that of Lui and Jarvis[72] who use a feature extraction method for unwarped stereo panoramic images based on the standard 2D Haar wavelet decomposition proposed by Jacobs *et al.*[49] and adapted for mobile robotics by Ho and Jarvis.[47] Similarly to the Fourier transform, which is used to decompose complex signals into a series of sine waves, Haar wavelets are applied to obtain a summation of simpler images that can be used to extract a discriminative and robust to occlusions and light changes signature, although rotation variant.

5.2.3. Edge-based detectors. They are used to obtain outlines in the context of computer vision. In particular, Tapus[108] utilizes the Sobel operator as an intermediate step to obtain segments of vertical edges, whereas Goedemé *et al.*[42] employs this operator to apply the so-called *invariant column segments* method, which is not an edge detector strictly speaking but a specialization of the so-called affine invariant regions that are commented below. For further reference, a comparison of several edge detectors can be found in Maini and Aggarwal.[76]

5.2.4. Keypoint detectors. In the context of feature detection using computer vision, blobs are points in the image that are either significantly brighter or darker than its neighbors. An initial comment is required before proceeding with the most remarkable algorithms. Although the title alludes to detectors, most of the methods cited below also include a *descriptor* to encode the distinguishing data that can be extracted from the features localized using the *detector*. For the sake of simplicity, they will be treated as a whole because they are usually presented together. Nevertheless, it is important to bear in mind that detectors and descriptors are interchangeable.

The most pre-eminent blob detector algorithm is Scale Invariant Feature Transform (SIFT),[69,70,102] which is the current standard for vision-based SLAM. Later on, Bay *et al.*[9] developed Speeded-Up Robust Features (SURF) with the aim of reducing the computational burden of SIFT. This fact makes it a better candidate for real-time applications. Last but not least, it is worth mentioning another promising feature detector named Center Surround Extremas (CenSurE),[1] which is much faster than the previous two methods at the expense of a slight increase in rotation sensitivity. More recently, Ebrahimi and Mayol-Cuevas[34] presented SUSurE, an interest point detector and descriptor based on CenSurE, which is capable of executing two to three times faster with only a slight loss in repeatability.

However, there exist other type of interest points apart from blobs. An example of these is Kanade-Lucas-Tomasi (KLT), included within the OpenCV library,[13] which is a corner detector that has also been applied in topological SLAM systems to perform visual odometry.[72]

5.2.5. Affine covariant region detectors. Affine covariant region detectors emerged with the idea of extracting features from images that were robust to perspective transformations. It is unclear which is the best among them, as they are often complementary and well suited for extracting regions with different properties. Mikolajczyk *et al.*[77] carried out a survey comparing the most common detectors, among which Harris-affine and Maximally Stable Extremal Regions (MSER) can be found.

It is also interesting to point out that Romero and Cazorla[97] run the JSEG segmentation algorithm[27] prior to applying MSER described with SIFT with the aim of grouping features according to the image region to which they belong and produce a graph with them.

https://doi.org/10.1017/S0263574713001070 Published online by Cambridge University Press

5.2.6. Bayesian surprise. Mainly based on the concept of saliency, it states that relevant stimuli represent statistical outliers or, in other words, sudden or unexpected changes in the environment.[48,93] Thus, at least one of its attributes needs to be unique or rare over the entire scene (e.g., a red coat is perceptually salient among black suits but not among many other red coats). This method, which claims to fire at almost all locations that would be regarded as landmarks by a human being, as well as at some others that would not, can be implemented for different sensor technologies, predominantly laser and cameras, and applied to several elementary features such as color, intensity, orientation, or motion. For example, Ranganathan and Dellaert[93] illustrate this technique using laser range scanners and, in the context of computer vision, by simultaneously applying this method to SIFT descriptors computed over Harris-affine and MSER features.

5.2.7. A hybrid approach: fingerprint of places. Once set forth the most common feature extraction methods, it is clear that they all have advantages and disadvantages that make them suitable for specific applications. Thus, in the pursuit of a more generally applicable method, some authors have tried to combine several of the aforementioned techniques.

An interesting approach has its origin in the paper by Lamon *et al.*[62] where the term *fingerprint* of places was coined to refer to a circular list of complementary simple features (color patches and vertical edges), obtained from omnidirectional images, whose order matches their relative position around the robot. This idea led to the publication of a series of pieces of work that further developed on the concept of fingerprint. Of special relevance is that of Tapus and Siegwart[109] where, thanks to the information provided by two 180° laser range scanners, corners and empty areas (i.e., when there are more than 20° of free space between two features) are additionally detected.

More recently, Liu *et al.*[67] proposed a much simpler fingerprint procedure, exclusively based on panoramic images, which extracts vertical edges under the belief that the prevailing lines naturally segment a structured environment into meaningful areas, and uses the distance among those lines and the mean U-V chrominance of the defined regions as a lightweight descriptor called FACT, which was later granted with statistical meaning and renamed DP-FACT[68].

5.3. Correspondence and map matching

After detecting the distinguishing features in the environment with any of the hitherto presented algorithms, the subsequent step in traditional metric SLAM implementations is to track the features detected between two consecutive sensor samples. The distance between equal features is then used to compute how much the robot has moved and, if there is an encoder available, both measurements are merged with the aim of minimizing errors. Afterwards, according to the movement, the current location in the map is calculated.

By contrast, in pure topological SLAM systems, correspondence and map matching are the same. In general, there is no need to know how much the robot has moved, but only to identify if it has returned to an already visited place. Thus, it forces to repeatedly solve a loop-closing problem because correspondence is computed among the already encoded nodes instead of with the previous sample.[117]

It is important to remember that it is almost impossible to obtain two identical samples because of measurement noise, changes in the environment, and, in addition, because when revisiting a place the robot performs the measurements in a slightly different location or with another orientation. For these reasons, correspondence and map matching are usually carried out by means of *dissimilarity measurements*, like the Mahalanobis,[40,100] Euclidean,[3,40] and χ^2 distances[35], or the Jeffrey divergence,[113] whereas Tapus and Siegwart[109] employ a modified version of the *global alignment* algorithm, proposed by Needleman and Wunsch[80] to compare DNA sequences, which takes the uncertainty of the detected features into consideration. The latter opted for this approach—which accounts for an average 83.82% of correct classifications in indoor and outdoor environments—after comparing it with Bayesian programming and a hybrid technique which merges the global alignment with uncertainty and Bayesian programming methods.

Moreover, in the context of visual topological SLAM Angeli *et al.*[4] and Romero and Cazorla[97] utilize the *relative position of the features* within the images as a matching criteria. However, while the former uses RANSAC to ensure that epipolar geometry constraints are met, the latter applies the Graph Transformation Matching (GTM) algorithm by Aguilar *et al.*[2] In addition, Li and Olson[65] proposes the Incremental Posterior Joint Compatibility (IPJC) test to match constellations of features together rather than considering them individually. Although its formulation is equivalent to the well know Joint Compatibility Branch and Bound (JCBB) test,[81] it is faster and more accurate, and performs better on non-linear problems.

Finally, because map matching becomes more demanding as the mapped area grows, some authors like Goedemé *et al.*[42] or Romero and Cazorla[97] propose applying *clustering techniques* like *k*d-trees to reduce the dimensionality of the features in order to optimize the search and comparison processes. Cummins and Newman[22] employ a Chow-Liu tree. Notice, nevertheless, that both Goedemé *et al.*[42] and Cummins and Newman[22] perform tree building offline due to time constraints.

5.4. Map fusion: dealing with loop-closing uncertainty

The final stage in topological SLAM involves updating the recorded map. If the current location does not correspond to any node known in advance, then the robot is in an unexplored area and, therefore, if the measurements meet the requirements to be considered a distinctive place, it should be added to the map. A more complex situation occurs when there is a positive match. Remember that for topological SLAM one of the most awkward problems is perceptual aliasing, and suppose that for map matching only sensory information is used. Consequently, there may be several nodes in the map that coincide with the measurements. Notwithstanding, this by no means signifies that it is an already visited place. This section concentrates on the different manners in which loop-closing uncertainty in topological maps has been tackled in the literature.

5.4.1. The consistent pose estimation paradigm. Some of the early developments on map fusion are inspired by the concept of consistent pose estimation (CPE) introduced by Lu and Milios,[71] which attempts to globally optimize the recorded set of poses based on how well neighboring sensor scans match. Gutmann and Konolige[45] presented the Local Registration/Global Correlation (LRGC) algorithm that is based on building local metric maps (named local patches) from the last few measurements in order keep the accumulated odometric error low and ensure topological correctness. The global metric map is then incrementally updated by comparing the topological structure of the latest patch with older portions of the map. A high match score with low ambiguity and variance indicates a loop closure. The experiments, carried out with robots equipped with laser sensors and encoders in four different environments of up to 80 by 25 m, yield fairly good metric maps under the assumption that local patches are accurate enough. Later on, Konolige[55] presented an efficient algorithm for multiple-loop maps that allows to extend the CPE method to map much larger areas (i.e., around 10^5 distinct locations).

5.4.2. Spatial semantic hierarchy. The Spatial Semantic Hierarchy (SSH) is a model of knowledge for large-space introduced by Kuipers.[57] It involves four qualitative and quantitative representations. At the *control* level, the agent continuously seeks *distinctive states* with a combination of trajectory-following and hill-climbing strategies. The *causal* level abstracts this pattern of behavior into a discrete model described in terms of states, sensory views, actions, and the causal relations among them. The *topological* level introduces the concepts of places, paths, and regions, and links them through turn and travel actions in order to explain the regularities observed among views in the control level. Finally, the *metrical* level represents a global geometric map of the environment in a single frame of reference. This framework was subsequently formalized using non-monotonic logic by Remolina and Kuipers.[96]

Kuipers *et al.*[61] extended the basic SSH with *local perceptual maps* (LPMs), a bounded occupancy grid. In this work, they identify *gateways* in corridors as the locations where the distance between the medial axis edge and the obstacles is a local minimum close to a larger maximum. However, they believe that other alternatives are possible. In addition, they include *path fragments* associated to the gateways. This information, along with travel control laws, is employed to obtain a *local topology* of a place in terms of distinctive states and directed paths.

In order to obtain the global topological map, a tree whose nodes are topological map-distinctive state pairs is maintained and pruned over time by matching local topologies, and LPMs if necessary. Instead of pruning, Johnson and Kuipers[50] proposed expanding only the most probable hypothesis to ensure that you can always backtrack in case of error and find the correct map. Further developments of this research line include improvements to loop-closing with the incorporation of the planarity constraint,[101] and the construction of accurate global metric maps from the topological skeleton obtained.[78]

5.4.3. Markov decision processes. Markov decision processes—Hidden Markov Models (HMMs)[44,103] and their extension, Partially Observable Markov Decision Processes (POMDPs)[16,51,54]—have also been employed to determine the navigation policy that the robot should follow in order to reduce uncertainty. Subsequently, Tomatis *et al.*[111] and Tapus and Siegwart[109] extended POMDPs to perform multi-hypothesis tracking and determine a pose distribution. However, as computing an optimal policy is intractable in large environments, Tomatis *et al.*[111] suggested using the *most likely state* (MLS) criterion to choose the following action, whereas Tapus and Siegwart[109] opted for another heuristic, the entropy of the current location probability distribution, to decide the control commands. In the latter case, whenever the entropy falls below an experimentally determined threshold, the robot's location is assumed certain and the map is updated accordingly, either by adding a new node or by merging the latest fingerprint information with the node representative.

Loop closures are also identified by means of the POMDP. Whenever the robot returns to a previously visited location, the probability distribution should split in two. One hypothesis would correspond to a new location and the other to a node already present in the map. If both divergent peaks evolve similarly over time, a loop closure is assumed.[108]

5.4.4. Probabilistic topological maps. A Bayesian inference framework has also been explored for topological mapping. Ranganathan and Dellaert coined the term *Probabilistic Topological Map* (PTM), a sample-based representation that estimates the posterior distribution over all the possible topologies that can be built given a set of sensor measurements.[90,91] Due to the fact that this is a problem of a combinatorial nature, they proposed approximating the solution by drawing samples from the distribution using Markov-Chain Monte Carlo (MCMC) sampling.[91,95] In principle, this technique is applicable to any landmark detection scheme as long as the landmark detection algorithm does not provide false negatives (i.e., the robot's sensors do not fail to recognize landmarks).

Afterwards, they presented *Rao-Blackwellized Particle Filters (RBPFs)*[29,79] as an alternative to MCMC sampling for PTMs.[90,92,94] Particle filters is yet another Monte Carlo localization technique used to probabilistically estimate the state of a system under noisy measurement conditions. They claim that this technique permits incremental inference in the space of topologies—conversely to MCMC, which is a batch algorithm—and can therefore be computed in real time. In order to overcome the samples degeneracy problem over time,[30] that can lead to convergence issues, they suggest integrating odometric data to draw more likely particles with higher probability. However, the selection of the appropriate number of particles still remains an open issue, as particle filtering inherently has the risk of disposing the correct map. Koenig *et al.*[52] also employ a RBPF. Each particle incrementally constructs its own graph of the environment using color histograms and odometry information. Local graphs are compared with the global graph to determine the best matches and, simultaneously, the resampling weights for each particle.

The main advantage of PTMs is that all decisions are reversible and the algorithm is therefore capable of recovering from incorrect loop closures. In the end, only a small set of similar topologies have non-negligible probabilities. The experiments conducted suggest that, if the environment is unambiguous, the ground-truth topology is assigned a much higher posterior probability mass than the other alternatives.

5.4.5. Voronoi graphs and neighboring information. Choset and Nagatani[18] represent the environment by means of a generalized Voronoi graph (GVG). A GVG is a one-dimensional set of points equidistant to n obstacles in n dimensions. When used in the plane, it reduces to the set of points equidistant to two (or more) obstacles, and define a roadmap of the robot's free space. Voronoi nodes, which are locations equidistant to $n + 1$ obstacles, are used as natural landmarks because they provide topologically meaningful information that can be extracted online (e.g., junctions, dead-ends, etc.). The main problem with Voronoi nodes is that they are very sensitive to changes in the configuration of the environment. If non-structural obstacles are moved, Voronoi vertices may appear or vanish.

In order to achieve SLAM, the robot follows simple control commands looking for these nodes in the environment. Loop-closing is carried out by comparing the subgraph built from the latest observed nodes to the already encoded map. Ambiguity is resolved by following a candidate path and ruling out inconsistent matches based on the new visited places. This method as is assumes that the robot is equipped with infinite range sonar sensors, and is only suitable for static and planar environments with plenty of obstacles. Based on this idea, Beeson *et al.*[10] introduced extended Voronoi graphs (EVGs) to address the problems of GVGs derived from limited sensory horizons by means of local perceptual maps (LPMs).

The research path initiated by Werner *et al.*[115,118] is also remarkable. They apply Bayesian inference to obtain a topological map in ambiguous environments that explains the set of observations without the need for motion knowledge. The method is based on guaranteeing consistency between the local neighboring information extracted from the latest n images and the constructed map while keeping the number of topological vertices as low as possible, following the Occam's razor principle. Topological places, where captures are acquired, are identified by means of a GVG using sonar readings. The algorithm assumes that there exist some prior information about the connectivity but not about the number of distinct locations in the environment.

Initially, a sequential Monte Carlo technique was employed to maintain a series of candidate maps,[116] which was later replaced by a particle filter.[117] In order to be able to recover from incorrect loop closures, Tully *et al.*[112] introduced a multi-hypothesis approach based on a tree expansion algorithm specifically conceived for edge-ordered graphs,[32] as well as a series of pruning rules to keep the number of hypothesis under control. Recently, Tao *et al.*[107] discussed the benefits of saturated generalized Voronoi graphs (S-GVG), that employ a wall-following behavior to navigate within sensor range limits, and performed SLAM using a similar hypothesis tree. Finally, Werner *et al.*[119] suggested applying *stochastic local search* (SLS) to produce the topological map.

Before concluding this section, it is worth mentioning the work by Doh *et al.*,[28] who deal with semi-permanent dynamics induced by door opening and closing. They classify GVG nodes in invariant (i.e., junctions, corners, and ends of corridors) and variant (i.e., doors). Nodes are told apart using the areas between two local minimums of a sensor scan (which identify doors), and looking for a vanishing point from a range scan or in an image (for invariant nodes).

5.4.6. Appearance-based topological SLAM. Most early approaches to inferring topological maps based only on visual information—they discard employing odometric data because it is prone to cumulative errors, especially on slippery surfaces—rely on SIFT keypoints extracted from omnidirectional images. Some examples include the work by Zivkovic *et al.*,[121] who solve the map building process using graph-cuts, and Goedemé *et al.*,[41] who resort to Dempster-Shafer theory of evidence[26] for loop-closing. Unfortunately, these solutions require offline computation.

Later on, Fraundorfer *et al.*[39] presented a real-time framework based on the *bag-of-words* paradigm,[19] where images are quantized in terms of unordered elementary features taken from an offline-built dictionary. Loop-closing is identified by visual word comparison following a voting scheme. Romero and Cazorla[97,98] take a similar approach but without the need for a dictionary. They build graphs from homogeneous regions using MSER features described with SIFT and use the GTM algorithm for matching. They then compare the graphs from newly acquired images with the latest visited topological node representative. If the matching score is below a threshold, it is then compared—using another threshold—with the rest of the encoded vertices in order to identify loop closures. If no match is found, a new node is added to the map. The main drawback of this algorithm is that it is extremely sensitive to the two thresholds. The value of these parameters has a decisive impact on the final topology obtained.

Angeli *et al.*[4–6] proposed a method that builds the vocabulary online, following the procedure developed by Filliat.[35] The problem of loop-closing is addressed following a Bayesian approach. The probability of transition between locations is modeled using a sum of Gaussians to assign higher probability to adjacent states, whereas the correspondence likelihood is computed by means of voting using the *tf-idf* coefficient.[104]

Furthermore, *Fast Appearance-Based Mapping* (FAB-MAP), which is a Bayesian framework for navigation and mapping exclusively based on appearance information developed by Cummins and Newman as a solution to loop closure detection,[20–22,24] has attracted a great deal of attention. It relies on a vocabulary model built offline from the clustering of SURF features extracted from a large collection of independent images. The words obtained are then organized using a Chow-Liu tree to capture the dependencies among them (i.e., car wheels and car doors are likely to appear together). This vocabulary model is used to approximate the partition function in the Bayesian formulation, which provides a natural probabilistic measure of when an observation should be labeled as a new location.

The experiments conducted outdoors suggest that it performs well in repetitive environments and is fast enough for online loop-closing. The fact that it requires offline training is an drawback, although tests carried out indoors with the bag-of-words model built for outdoor environments produce surprisingly good results according to the authors.

Some improvements have been introduced to the original algorithm since its presentation. First, speed was increased by more than 25 times, with only a slight degradation in accuracy, thanks to the usage of concentration inequalities to reduce the number of hypothesis considered.[23] The formulation of the algorithm was also modified to operate on very large environments (over trajectories of around 1000 km).[25] Finally, Paul and Newman[86] incorporated the spatial arrangement of visual words to improve distinctiveness.

5.4.7. Continuous appearance-based trajectory SLAM. Continuous Appearance-based Trajectory SLAM (CAT-SLAM)[73] incorporates odometry, following the approach of FastSLAM,[79] to appearance-based SLAM using FAB-MAP. The current location is modeled as a probability distribution over a trajectory and appearance is treated as a continuous variable. The evaluation of the distribution is carried out using a RBPF. Compared to FAB-MAP, it identifies three times as many loop closures at 100% precision (i.e., with no false positives). By contrast, FAB-MAP is capable of recognizing places when approached from a different direction, whereas CAT-SLAM cannot because it relies on odometric information. Enhancements to computational and memory storage requirements, like pruning those nodes in the trajectory that are locally uninformative once a preset maximum number of nodes is reached, were subsequently introduced to allow continuous operation on much larger environments.[74,75]

5.4.8. Closing the loop with visual odometry. Lui and Jarvis[72] have implemented a different correction algorithm for loop closure detection, which relies on *visual odometry*. They employ the Kanade-Lucas-Tomasi (KLT) features present in the OpenCV computer vision library[13] to estimate the distance traveled and column image comparison using the Sum of Absolute Differences (SAD) for the front 180° field of view (FOV) of the robot to estimate the bearing. These are then used to reduce the matches retrieved from the database, using a Haar wavelet-based signature, utilizing the relaxation algorithm proposed by Duckett *et al.*[31] The current location is then told apart by means of SURF.[9] This system has been proven effective in indoor and semi-outdoor environments. However, its main drawback lies in the complexity of the robot infrastructure, which includes an omnidirectional stereovision system and a web camera to perform visual odometry, as well as a stereo camera for obstacle avoidance.

5.4.9. The final stage: updating the map. Finally, once the uncertainty has been resolved, the new information gathered should be incorporated to the map for future reference. On the one hand, some authors suggest removing any unobserved nodes, features, and relations or, better, implementing a gradual "forgetting" process that could take into account changes in the environment (e.g., an open door appears closed when revisiting a place).[114] On the other hand, Kuipers and Beeson[58] and Tapus and Siegwart[109] propose applying *clustering techniques* to create a mean node representative with a view to reducing the impact of scene variability.

6. Conclusion

There is still a long way to go as far as topological SLAM is concerned. Although there exist plenty of partial implementations and ideas for some of the phases, a robust and globally applicable method is yet to be developed.

In detection, it seems that the best results have been achieved with a wisely chosen collection of features. Nevertheless, the selection of these features is crucially affected by the sensory technology used. Computer vision is rising as a promising alternative because, even though processing the data gathered can be more difficult and computationally expensive, it provides richer information than other sensors, like laser range scanners, and can be easily installed in any mobile entity.

In map matching and updating, the probabilistic approach seems to be the most consolidated research line. However, in spite of the topological representation being less computationally demanding, there are still some open issues in unbounded, dynamic, and ambiguous unknown environments. Constantly solving a loop-closing problem can be cumbersome in large maps as the robot can simultaneously believe to be in several locations, which results in having to deal with

https://doi.org/10.1017/S0263574713001070 Published online by Cambridge University Press

a huge pose distribution that multiply the calculations required. Because of this, new tracks are being explored in the pursuit of scalability in metric SLAM. The work by Blanco *et al.*,[12] which demonstrates that an appropriate formulation of the SLAM problem can pose an upper limit on loop closure complexity in unbounded environments, is a remarkable example.

References

1. M. Agrawal, K. Konolige and M. R. Blas, "CenSurE: Center Surround Extremas for Realtime Feature Detection and Matching," *LNCN* **5305**, 102–115 (2008).
2. W. Aguilar, Y. Frauel, F. Escolano, M. E. Martínez-Pérez, A. Espinosa-Romero and M. Á. Lozano, "A robust graph transformation matching for non-rigid registration," *Image Vis. Comput.* **27**, 897–910 (2009).
3. H. Andreasson, A. Treptow and T. Duckett, "Localization for Mobile Robots Using Panoramic Vision, Local Features and Particle Filter," *Proceedings of the IEEE International Conference Robotics and Automation*, Barcelona, Spain (2005) pp. 3348–3353.
4. A. Angeli, S. Doncieux, J.-A. Meyer and D. Filliat, "Incremental Vision-Based Topological SLAM," *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, Nice, France (2008a) pp. 1031–1036.
5. A. Angeli, D. Filliat, S. Doncieux and J.-A. Meyer, "A fast and incremental method for loop-closure detection using bags of visual words," *IEEE Trans. Robot.* **24**(5), 1027–1037 (2008b).
6. A. Angeli, D. Filliat, S. Doncieux and J.-A. Meyer, "Real-Time Visual Loop-Closure Detection," *Proceedings of the IEEE International Conference on Robotics and Automation*, Pasadena, CA (2008c) pp. 1842–1847.
7. D. Anguelov, D. Koller, E. Parker and S. Thrun, "Detecting and modeling doors with mobile robots," *Proc. IEEE Int. Conf. Robot. Autom.* **4**, 3777–3784 (2004).
8. T. Bailey and H. F. Durrant-Whyte, "Simultaneous localization and mapping (SLAM): Part II," *IEEE Robot. Autom. Mag.* **13**(3), 108–117 (2006).
9. H. Bay, A. Ess, T. Tuytelaars and L. van Gool, "SURF: Speeded up robust features," *Comput. Vis. Image Underst.* **110**(3), 346–359 (2008).
10. P. Beeson, N. K. Jong and B. Kuipers, "Towards Autonomous Topological Place Detection Using the Extended Voronoi Graph," *Proceedings of the IEEE International Conference on Robotics and Automation*, Barcelona, Spain (2005) pp. 4373–4379.
11. J. L. Blanco, J. A. Fernández-Madrigal and J. González, "Toward a unified Bayesian approach to hybrid metric-topological SLAM," *IEEE Trans. Robot.* **24**(2), 259–270 (2008).
12. J. L. Blanco, J. González-Jiménez and J. A. Fernández-Madrigal, "Sparser Relative Bundle Adjustment (SRBA): Constant-Time Maintenance and Local Optimization of Arbitrarily Large Maps," *Proceedings of the IEEE International Conference Robotics and Automation*, Karlsruhe, Germany (2013) pp. 70–77.
13. G. Bradski, "The OpenCV library" (2000). `http://opencv.willowgarage.com/`.
14. R. A. Brooks, "Visual map making for a mobile robot," *Proc. IEEE Int. Conf. Robot. Autom.* **2**, 824–829 (1985).
15. R. A. Brooks, "Elephants don't play chess," *Robot. Auton. Syst.* **6**, 3–15 (1990).
16. A. R. Cassandra, L. P. Kaelbling and J. A. Kurien, "Acting Under Uncertainty: Discrete Bayesian Models for Mobile-Robot Navigation," *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, Vol. 2, Osaka, Japan (1996) pp. 963–972.
17. R. Chatila and J.-P. Laumond, "Position referencing and consistent world modeling for mobile robots," *Proc. IEEE Int. Conf. Robot. Autom.* **2**, 138–145 (1985).
18. H. Choset and K. Nagatani, "Topological simultaneous localization and mapping (SLAM): Toward exact localization without explicit localization," *IEEE Trans. Robot. Autom.* **17**(2), 125–137 (2001).
19. G. Csurka, C. Dance, L. Fan, J. Williamowski and C. Bray, "Visual Categorization with Bags of Keypoints," *ECCV International Workshop on Statistical Learning in Computer Vision*, Prague, Czech Republic (2004) pp. 59–74.
20. M. Cummins, *Probabilistic Localization and Mapping in Appearance Space Ph.D. Thesis* (University of Oxford, 2009).
21. M. Cummins and P. Newman, "Probabilistic Appearance Based Navigation and Loop Closing," *Proceedings of the IEEE International Conference on Robotics and Automation*, Rome, Italy (2007) pp. 2042–2048.
22. M. Cummins and P. Newman, "FAB-MAP: Probabilistic localization and mapping in the space of appearance," *Int. J. Robot. Res.* **27**(6), 647–665 (2008).
23. M. Cummins and P. Newman, "Accelerating FAB-MAP with concentration inequalities," *IEEE Trans. Robot.* **26**(6), 1042–1050 (2010a).
24. M. Cummins and P. Newman, "FAB-MAP: Appearance-Based Place Recognition and Mapping Using a Learned Visual Vocabulary Model," *Proceedings of the International Conference on Machine Learning*, Haifa, Israel (2010b) pp. 3–10.
25. M. Cummins and P. Newman, "Appearance-only SLAM at large scale with FAB-MAP 2.0," *Int. J. Robot. Res.* **30**(9), 1100–1123 (2011).
26. A. P. Dempster, "Upper and lower probabilities induced by a multivalued mapping," *Ann. Math. Stat.* **38**(2), 325–339 (1967).

27. Y. Deng and B. S. Manjunath, "Unsupervised segmentation of color-texture regions in images and video," *IEEE Trans. Pattern Anal. Mach. Intell.* **23**(8), 800–810 (2001).
28. N. L. Doh, K. Lee, W. K. Chung and H. Cho, "Simultaneous localisation and mapping algorithm for topological maps with dynamics," *IET Control Theor. Appl.* **3**(9), 1249–1260 (2009).
29. A. Doucet, N. de Freitas, K. Murphy and S. Russell, "Rao-Blackwellised Particle Filtering for Dynamic Bayesian Networks," *Proceedings of the 16th Conference on Uncertainty in Artificial Intelligence*, Stanford, CA (2000a) pp. 176–173.
30. A. Doucet, S. Godsill and C. Andrieu, "On sequential Monte Carlo sampling methods for Bayesian filtering," *Stat. Comput.* **10**, 197–208 (2000b).
31. T. Duckett, S. Marsland and J. Shapiro, "Learning Globally Consistent Maps by Relaxation," *Proceedings of the IEEE International Conference on Robotics and Automation*, San Francisco, CA, USA (2000) pp. 3841–3846.
32. G. Dudek, P. Freedman and S. Hadjres, "Using Local Information in a Non-Local Way for Mapping Graph-Like Worlds," *International Joint Conference on Artificial Intelligence*, Chambery, France (1993) pp. 1639–1647.
33. H. F. Durrant-Whyte and T. Bailey, "Simultaneous localization and mapping: Part I," *IEEE Robot. Autom. Mag.* **13**(2), 99–110 (2006).
34. M. Ebrahimi and W. W. Mayol-Cuevas, "SUSurE: Speeded up Surround Extrema Feature Detector and Descriptor for Realtime Applications," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Miami, FL (2009) pp. 9–14.
35. D. Filliat, "A Visual Bag of Words Method for Interactive Qualitative Localization and Mapping," *Proceedings of the IEEE International Conference on Robotics and Automation*, Rome, Italy (2007) pp. 3921–3926.
36. D. Filliat and J.-A. Meyer, "Map-based navigation in mobile robots: I. A review of localization strategies," *Cogn. Syst. Res.* **4**(4), 243–282 (2003).
37. M. A. Fischler and R. C. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM* **24**(6), 381–395 (1981).
38. D. A. Forsyth and J. Ponce, *Computer Vision: A Modern Approach* (Prentice Hall, Upper Saddle River, NJ, 2003).
39. F. Fraundorfer, C. Engels and D. Nistér, "Topological Mapping, Localization and Navigation Using Image Collections," *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, San Diego, CA, USA (2007) pp. 3872–3877.
40. T. Goedemé and L. van Gool, "Robust Vision-Only Mobile Robot Navigation with Topological Maps," **In**: *Mobile Robots Motion Planning, New Challenges* (X.-J. Jing, ed.) (InTech, Austria, 2008) chap. 4, pp. 63–88.
41. T. Goedemé, M. Nuttin, T. Tuytelaars and L. van Gool, "Omnidirectional vision based topological navigation," *Int. J. Comput. Vis.* **74**(3), 219–236 (2007).
42. T. Goedemé, T. Tuytelaars and L. van Gool, "Fast wide baseline matching for visual navigation," *Proc. IEEE Comp. Soc. Conf. Comp. Vision Pattern Recogn.* **1**, 24–29 (2004).
43. T. Goedemé, T. Tuytelaars, L. van Gool, G. Vanacker and M. Nuttin, "Feature Based Omnidirectional Sparse Visual Path Following," *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, Edmonton, Canada (2005) pp. 1806–1811.
44. R. Gutiérrez-Osuna and R. C. Luo, "LOLA Probabilistic navigation for topological maps," *AI Mag.* **17**(1), 55–62 (1996).
45. J.-S. Gutmann and K. Konolige, "Incremental Mapping of Large Cyclic Environments," *Proceedings of the International Symposium on Computational Intelligence in Robotics and Automation* (1999), pp. 318–325.
46. H. H. Hafner, "Learning Places in Newly Explored Environments," **In**: *Proceedings of the International Conference on Simulation of Adaptive Behavior* (Meyer, Berthoz, Floreano, Roitblat and Wilson, eds.) (International Society for Adaptive Behavior, Honolulu, HI, USA, 2000) pp. 111–120.
47. N. Ho and R. Jarvis, "Vision Based Global Localisation Using a 3D Environmental Model Created by a Laser Range Scanner," *Proceedings of the IEEE/RSJ International Conference Intelligent Robots and Systems*, Nice, France (2008) pp. 2964–2969.
48. L. Itti and P. Baldi, "A principled approach to detecting surprising events in video," *Proc. IEEE Comp. Soc. Conf. Comp. Vision Pattern Recogn.* **1**, 631–637 (2005).
49. C. E. Jacobs, A. Finkelstein and D. H. Salesin, "Fast Multiresolution Image Querying," *Proceedings of the Annual Conference on Computer Graphics and Interactive Techniques*, Los Angeles, CA (1995) pp. 277–286.
50. C. Johnson and B. Kuipers, "Efficient Search for Correct and Useful Topological Maps," *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, Vilamoura, Algarve, Portugal (2012) pp. 5277–5282.
51. L. P. Kaelbling, M. L. Littman and A. R. Cassandra, "Planning and acting in partially observable stochastic domains," *Artif. Intell.* **101**(1–2), 99–134 (1998).
52. A. Koenig, J. Kessler and H.-M. Gross, "A Graph Matching Technique for an Appearance-Based, Visual SLAM-Approach Using Rao-Blackwellized Particle Filters," *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, Nice, France (2008) pp. 1576–1581.

53. S. Koenig, A. Mudgal and C. Tovey, "A Near-Tight Approximation Lower Bound and Algorithm for the Kidnapped Robot Problem," *Proceedings of the Symposium on Discrete Algorithms*, Miami, FL (2006) pp. 133–142.
54. S. Koenig and R. G. Simmons, "Unsupervised Learning of Probabilistic Models for Robot Navigation," *Proceedings of the IEEE International Conference on Robotics and Automation*, Minneapolis, MN, USA 1996) pp. 2301–2308.
55. K. Konolige, "Large-Scale Map-Making," *Proceedings of the National Conference on Artificial Intelligence*, San Jose, CA (2004) pp. 457–463.
56. D. Kortenkamp and T. Weymouth, "Topological Mapping for Mobile Robots Using a Combination of Sonar and Vision Sensing," *Proceedings of the American Association for Artificial Intelligence (AAAI) Conference*, Seattle, WA, USA (1994).
57. B. Kuipers, "The spatial semantic hierarchy," *Artif. Intell.* **119**, 191–233 (2000).
58. B. Kuipers and P. Beeson, "Bootstrap Learning for Place Recognition," *Proceedings of the 18th National Conference on Artificial Intelligence*, Edmonton, Alberta, Canada (2002) pp. 174–180.
59. B. Kuipers and Y.-T. Byun, "A robot exploration and mapping strategy based on a semantic hierarchy of spatial representations," *Robot. Auton. Syst.* **8**(1), 47–63 (1991).
60. B. Kuipers and T. Levitt, "Navigation and mapping in large-scale space," *AI Mag.*, **9**(2), 25–43 (1988).
61. B. Kuipers, J. Modayil, P. Beeson, M. MacMahon and F. Savelli, "Local metrical and global topological maps in the hybrid spatial semantic hierarchy," *Proc. IEEE Int.Conf. Robot. Autom.* **5**, 4845–4851 (2004).
62. P. Lamon, I. Nourbakhsh, B. Jensen and R. Siegwart, "Deriving and Matching Image Fingerprint Sequences for Mobile Robot Localization," *Proceedings of the IEEE International Conference on Robotics and Automation*, Seoul, Korea (2001).
63. J. J. Leonard and H. F. Durrant-Whyte, "Mobile robot localization by tracking geometric beacons," *IEEE Trans. Robot. Autom.* **7**(3), 376–382 (1991a).
64. J. J. Leonard and H. F. Durrant-Whyte, "Simultaneous Map Building and Localization for an Autonomous Mobile Robot," *Proceedings of the IEEE/RSJ International Workshop on Intelligent Robots and Systems*, Osaka, Japan (1991b) pp. 1442–1447.
65. Y. Li and E. B. Olson, "IPJC: The Incremental Posterior Joint Compatibility Test for Fast Feature Cloud Matching," *Proceedings of the IEEE/RSJ International Conference Intelligent Robots and Systems*, Vilamoura, Algarve, Portugal (2012) pp. 3467–3474.
66. B. Lisien, D. Morales, D. Silver, G. Kantor, I. Rekleitis and H. Choset, "Hierarchical Simultaneous Localization and Mapping," *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, Las Vegas, NV (2003) pp. 448–453.
67. M. Liu, D. Scaramuzza, C. Pradalier, R. Siegwart and Q. Chen, "Scene Recognition with Omnidirectional Vision for Topological Map Using Lightweight Adaptive Descriptors," *Proceedings of the IEEE/RSJ International Conference Intelligent Robots and Systems*, St. Louis, MO, USA (2009) pp. 116–121.
68. M. Liu and R. Siegwart, "DP-FACT: Towards Topological Mapping and Scene Recognition with Color for Omnidirectional Camera," *IEEE International Conference on Robotics and Automation*, Saint Paul, MN, USA (2012).
69. D. G. Lowe, "Object recognition from local scale-invariant features," *Proc. IEEE Int. Conf. Computer Vision* **2**, 1150–1157.
70. D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.* **60**(2), 91–110 (2004).
71. F. Lu and E. Milios, "Globally consistent range scan alignment for environment mapping," *Auton. Robots* **4**, 333–349 (1997).
72. W. L. D. Lui and R. Jarvis, "A Pure Vision-Based Approach to Topological SLAM," *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, Taipei, Taiwan (2010).
73. W. Maddern, M. Milford and G. Wyeth, "Continuous Appearance-Based Trajectory SLAM," *Proceedings of the IEEE International Conference on Robotics and Automation*, Shanghai, China (2011) pp. 3595–3600.
74. W. Maddern, M. Milford and G. Wyeth, "Capping Computation Time and Storage Requirements for Appearance-Based Localization with CAT-SLAM," *Proceedings of the IEEE International Conference on Robotics and Automation*, Saint Paul, MN, USA (2012a) pp. 822–827.
75. W. Maddern, M. Milford and G. Wyeth, "Towards Persistent Indoor Appearance-Based Localization, Mapping and Navigation Using CAT-SLAM," *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, Vilamoura, Algarve, Portugal (2012b) pp. 4224–4230.
76. R. Maini and H. Aggarwal, "Study and comparison of various image edge detection techniques," *Int. J. Image Process.* **3**(1), 1–11 (2009).
77. K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir and L. van Gool, "A comparison of affine region detectors," *Int. J. Comput. Vis.* **65**(1), 43–72 (2006).
78. J. Modayil, P. Beeson and B. Kuipers, "Using the Topological Skeleton for Scalable Global Metrical Map-Building," *Proceedings of the IEEE/RSJ International Conference on lntelligent Robots and Systems*, Sendai, Japan (2004) pp. 1530–1536.
79. M. Montemerlo, S. Thrun, D. Koller and B. Wegbreit, "FastSLAM: A Factored Solution to the Simultaneous Localization and Mapping Problem," *Proceedings of the AAAI National Conference on Artificial Intelligence*, Edmonton, Canada (2002).

80. S. B. Needleman and C. D. Wunsch, "A general method applicable to the search for similarities in the amino acid sequence of two proteins," *J. Mol. Biol.* **48**, 443–453 (1970).
81. J. Neira and J. D. Tardós, "Data association in stochastic mapping using the joint compatibility test," *IEEE Trans. Robot. Autom.* **17**(6), 890–897 (2001).
82. V. Nguyen, A. Martinelli, N. Tomatis and R. Siegwart, "A Comparison of Line Extraction Algorithms Using 2D Laser Rangefinder for Indoor Mobile Robotics," *Proceedings of the IEEE/RSJ International Conference Intelligent Robots and Systems*, Edmonton, Canada (2005) pp. 1929–1934.
83. J. I. Nieto, J. E. Guivant and E. M. Nebot, "The Hybrid Metric Maps (HYMMs): A novel map representation for DenseSLAM," *Proc. IEEE Int. Conf. Robot. Autom.* **1**, 391–396 (2004).
84. A. Nüchter and J. Hertzberg, "Towards semantic maps for mobile robots," *Robot. Auton. Syst.* **56**, 915–926 (2008).
85. C. Owen and U. Nehmzow, "Landmark-Based Navigation for a Mobile Robot," **In**: *Proceedings of the Simulation of Adaptive Behaviour* (MIT Press, 1998) pp. 240–245.
86. R. Paul and P. Newman, "FAB-MAP 3D: Topological Mapping with Spatial and Visual Appearance," *Proceedings of the IEEE International Conference on Robotics and Automation*, Anchorage, AK, USA (2010) pp. 2649–2656.
87. S. T. Pfister, S. I. Roumeliotis and J. W. Burdick, "Weighted line fitting algorithms for mobile robot map building and efficient data representation," *Proc. IEEE Int. Conf. Robot. Autom.* **1**, 1304–1311 (2003).
88. P. Pirjanian, N. Karlsson, L. Goncalves and E. di Bernardo, "Low-cost visual localization and mapping for consumer robotics," *Ind. Robot Int. J.* **30**(2), 139–144 (2003).
89. F. T. Ramos, B. Upcroft, S. Kumar and H. F. Durrant-Whyte, "A Bayesian Approach for Place Recognition," *Proceedings of the IJCAI Workshop on Reasoning with Uncertainty in Robotics*, Edinburgh, Scotland (2005).
90. A. Ranganathan, *Probabilistic Topological Maps Ph.D. Thesis* (Georgia Institute of Technology, 2008).
91. A. Ranganathan and F. Dellaert, "Inference in the Space of Topological Maps: An MCMC-Based Approach," *Proceedings of the IEEE/RSJ International Conference Intelligent Robots and Systems*, Sendai, Japan (2004) pp. 1518–1523.
92. A. Ranganathan and F. Dellaert, "A Rao-Blackwellized Particle Filter for Topological Mapping," *Proceedings of the IEEE International Conference on Robotics and Automation*, Orlando, FL, USA (2006) pp. 810–817.
93. A. Ranganathan and F. Dellaert, Automatic Landmark Detection for Topological Mapping Using Bayesian Surprise *Technical Report* (Georgia Institute of Technology, 2008).
94. A. Ranganathan and F. Dellaert, "Online probabilistic topological mapping," *Int. J. Robot. Res.* **30**(6), 755–771 (2011).
95. A. Ranganathan, E. Menegatti and F. Dellaert, "Bayesian inference in the space of topological maps," *IEEE Trans. Robot.* **22**(1), 92–107 (2006).
96. E. Remolina and B. Kuipers, "Towards a general theory of topological maps," *Artif. Intell.* **152**, 47–104 (2004).
97. A. Romero and M. Cazorla, "Topological SLAM Using Omnidirectional Images: Merging Feature Detectors and Graph-matching," *Proceedings of the Advanced Concepts for Intelligent Vision Systems*, Sydney, Australia (2010) pp. 464–475.
98. A. Romero and M. Cazorla, "Topological visual mapping in robotics," *Cogn. Process.* **13**(1 Supplement), 305–308 (2012).
99. D. Sabatta, D. Scaramuzza and R. Siegwart, "Improved Appearance-Based Matching in Similar and Dynamic Environments Using a Vocabulary Tree," *Proceedings of the IEEE International Conference on Robotics and Automation*, Anchorage, AK (2010).
100. D. G. Sabatta, "Vision-Based Topological Map Building and Localisation Using Persistent Features," *Robotics and Mechatronics Symposium*, Bloemfontein, South Africa (2008) pp. 1–6.
101. F. Savelli and B. Kuipers, "Loop-Closing and Planarity in Topological Map-Building," *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, Sendai, Japan (2004) pp. 1511–1517.
102. S. Se, D. G. Lowe and J. J. Little, "Vision-based global localization and mapping for mobile robots," *IEEE Trans. Robot.* **21**(3), 364–375 (2005).
103. H. Shatkay and L. P. Kaelbling, "Learning Topological Maps with Weak Local Odometric Information," *Proceedings of the International Conference on Artificial Intelligence*, Nagoya, Japan (1997) pp. 920–929.
104. J. Sivic and A. Zisserman, "Video Google: A text retrieval approach to object matching in videos," *Proc. IEEE Int. Conf. Comput. Vis.* **2**, 1470–1477 (2003).
105. C. Stachniss, O. Martínez-Mozos, A. Rottmann and W. Burgard, "Semantic Labeling of Places," *Proceedings of the International Symposium of Robotics Research*, San Francisco, CA, USA (2005).
106. B. J. Stankiewicz and A. A. Kalia, "Acquisition of structural versus object landmark knowledge," *J. Exp. Psychol. Hum. Perception Perform.* **33**(2), 378–390 (2007).
107. T. Tao, S. Tully, G. Kantor and H. Choset, "Incremental Construction of the Saturated-GVG for Multi-Hypothesis Topological SLAM," *Proceedings of the IEEE International Conference on Robotics and Automation*, Shanghai, China (2011) pp. 3072–3077.
108. A. Tapus, *Topological SLAM – Simultaneous Localization and Mapping with Fingerprints of Places Ph.D. Thesis* (École Polytechnique Fédérale de Lausanne, Switzerland, 2005).

109. A. Tapus and R. Siegwart, "Incremental Robot Mapping with Fingerprints of Places," *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, Edmonton, Canada (2005) pp. 2429–2434.
110. S. Thrun, "Robotic Mapping: A Survey," **In**: *Exploring Artificial Intelligence in the New Millenium* (G. Lakemeyer and B. Nebel, eds.) (Morgan Kaufmann, San Francisco, CA, 2002) pp. 1–35.
111. N. Tomatis, I. Nourbakhsh and R. Siegwart, "Hybrid Simultaneous Localization and Map Building: Closing the Loop with Multi-Hypotheses Tracking," *Proceedings of the IEEE International Conference Robotics and Automation*, Washington, DC (2002) pp. 2749–2754.
112. S. Tully, G. Kantor, H. Choset and F. Werner, "A Multi-Hypothesis Topological SLAM Approach for Loop Closing on Edge-Ordered Graphs," *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, St. Louis, MO, USA (2009) pp. 4943–4948.
113. I. Ulrich and I. Nourbakhsh, "Appearance-Based Place Recognition for Topological Localization," *Proceedings of the IEEE International Conference on Robotics and Automation*, San Francisco, CA, USA (2000) pp. 1023–1029.
114. S. Vasudevan, S. Gáchter, V. Nguyen and R. Siegwart, "Cognitive maps for mobile robots: an object based approach," *Robot. Auton. Syst.* **55**, 359–371 (2007).
115. F. Werner, C. Gretton, F. Maire and J. Sitte, "Induction of Topological Environment Maps from Sequences of Visited Places," *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, Nice, France (2008a) pp. 2890–2895.
116. F. Werner, F. Maire and J. Sitte, "Topological SLAM Using Fast Vision Techniques," *Proceedings of the Advances in Robotics: FIRA RoboWorld Congress*, Incheon, Korea (2009a) pp. 187–198.
117. F. Werner, F. Maire, J. Sitte, H. Choset, S. Tully and G. Kantor, "Topological SLAM Using Neighbourhood Information of Places," *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, St. Louis, MO, USA (2009b).
118. F. Werner, J. Sitte and F. Maire, "Visual Topological Mapping and Localisation Using Colour Histograms," *Proceedings of the International Conference on Control, Automation, Robotics and Vision*, Hanoi, Vietnam (2008b) pp. 341–346.
119. F. Werner, J. Sitte and F. Maire, "Topological map induction using neighbourhood information of places," *Auton. Robots* **32**(4), 405–418 (2012).
120. B. Yamauchi, A. Schultz and W. Adams, "Mobile Robot Exploration and Map-Building with Continuous Localization," *Proceedings of the International Conference on Robotics and Automation*, Leuven, Belgium (1998).
121. Z. Zivkovic, B. Bakker and B. Kröse, "Hierarchical Map Building Using Visual Landmarks and Geometric Constraints," *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, Edmonton, Canada (2005) pp. 2480–2485.