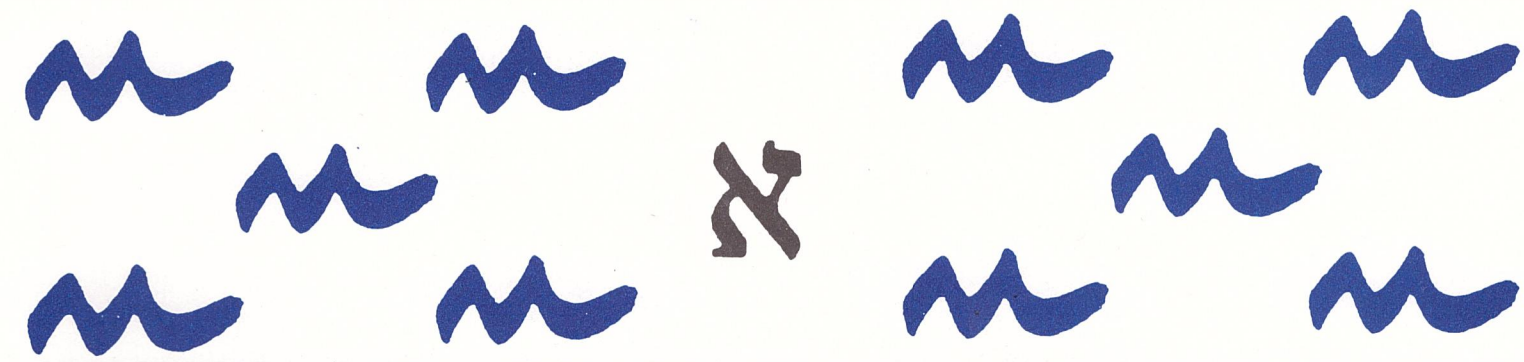
Universitat de les Illes Balears

Departament de Ciències
Matemàtiques i Informàtica

State-of-the-Art in Vision-Based Topological Mapping and Localization Methods

EMILIO GARCIA-FIDALGO and ALBERTO ORTIZ



April / 2014 TECHNICAL REPORT A-01-2014

State-of-the-Art in Vision-Based Topological Mapping and Localization Methods*

Emilio Garcia-Fidalgo and Alberto Ortiz
Department of Mathematics and Computer Science,
University of the Balearic Islands, Spain
{emilio.garcia,alberto.ortiz}@uib.es

Abstract

Topological maps model the environment as a graph, where nodes are distinctive places of the environment and edges indicate the relationships between them. They present an interesting alternative to the classic metric maps, due to their simplicity and storage needs, which convert them in an active research area. Several kinds of sensors have been used during years for topological mapping and localization. However, in the last decades, vision approaches have emerged because of the technology improvements and the amount of useful information that a camera can provide. In this paper, we review the main solutions presented in the last 15 years, and classify them in accordance to the kind of image descriptor employed. Advantages and disadvantages of each approach are thoroughly reviewed and discussed.

*This work is supported by the European Social Fund through the grant FPI11-43123621R (Conselleria d'Educacio, Cultura i Universitats, Govern de les Illes Balears) and by FP7 project INCASS (GA 605200).

1 Introduction

Mapping and localization are essential problems in mobile robotics. As a result of the mapping process, a representative map of the environment is generated while the localization process computes the pose of the robot within the map according to the sensor data perceived from the environment. Both processes can be used for navigation-related tasks, such as path planning or obstacle avoidance. This is of special interest for autonomous vehicles, which need to be able to operate without any human intervention.

Localization is sometimes solved using external structures, such as beacons at fixed, know positions or the Global Positioning System (GPS). However, the former implies modifications in the environment and the latter is not available in places such as indoor, underground or underwater scenarios. In these situations, the localization must be solved internally by the robot, using its own sensor suite. Ultrasonic and laser sensors have been used for years to this end. Nevertheless, recently there has been a significant increase in the number of visual solutions because of the low cost of cameras and the richness of the sensor data provided.

As far as robotic mapping is concerned, two main paradigms are generally accepted: metric and topological mapping. Metric maps represent the world as accurate as possible, maintaining a lot of information about environment details, such as distances, measures or sizes, and they are usually referenced according to a global coordinate system. This representation is most appropriate for vehicle localization and guidance, as well as for obstacle avoidance. However, metric maps are more difficult to build and maintain, and are computationally demanding. Conversely, topological maps represent the environment in an abstract manner by means of a graph, where nodes represent distinctive places in the environment and arcs model the relations between them. These maps are simple and compact, scale better and require much less space to be stored than metric maps. They are not useful for tasks with accuracy needs, for example obstacle avoidance, but simplifies others, like path planning. There exists another paradigm, called hybrid maps, that tries to maximize the advantages and minimize the problems of each kind of map alone for combining them in a different mapping technique.

Despite mapping and localization can be performed as independent tasks, they are closely re-

lated. In order to build a map, the pose of the structures and the obstacles of the environment needs to be known. On the other hand, during localization, the pose of the agent is computed against a reference map. In this case the map of the working scenario must be available before starting the navigation, which limits the autonomy of the vehicle. To solve this egg-and-hen problem, several approaches have been proposed where both tasks take place at the same time, creating an incremental map of an unknown environment while localizing the robot within this map. These techniques are generically known as Simultaneous Localization and Mapping (SLAM) [1]. In SLAM, loop closure detection is a key challenge to overcome which entails the correct detection of previously visited places from sensor data. This allows generating consistent maps and reducing their uncertainty. Although the dominant theme in the literature is the metric SLAM approach, a map can be constructed satisfying much less requirements in methods that fall into the category of topological SLAM.

In this paper we review the main approaches published in the last 15 years with regard to topological mapping and localization by visual means. Despite we focus on topological maps, we also consider hybrid solutions that perform some kind of topological processes. Other publications can be found in the related literature, although they are more focused on navigation [2] and visual SLAM [3].

Although there is no clear consensus about what a topological map is, we consider a classic definition of the problem [4, 5, 6], where nodes represent distinct places of the environment where the characteristics of the environment change significantly, and the edges denote the ability to navigate from one node to another. For this reason, we are not interested in the pose-graph SLAM problem, since the nodes in this kind of solutions represent poses reached by the agent and not distinctive places of the environment. The position in pose-graph SLAM is a metric position of the vehicle and not a qualitative estimation in a discrete model of the appearance of the world. These approaches are then out of the scope of this survey.

The loop closure detection is an important component in topological schemes. When using vision as a source, this problem is usually solved comparing images directly, resulting into appearance-based approaches. In this regard, a related research field is scene categorization or visual place categorization (VPC) [7]. The main goal of this area is to find the class of a place in a

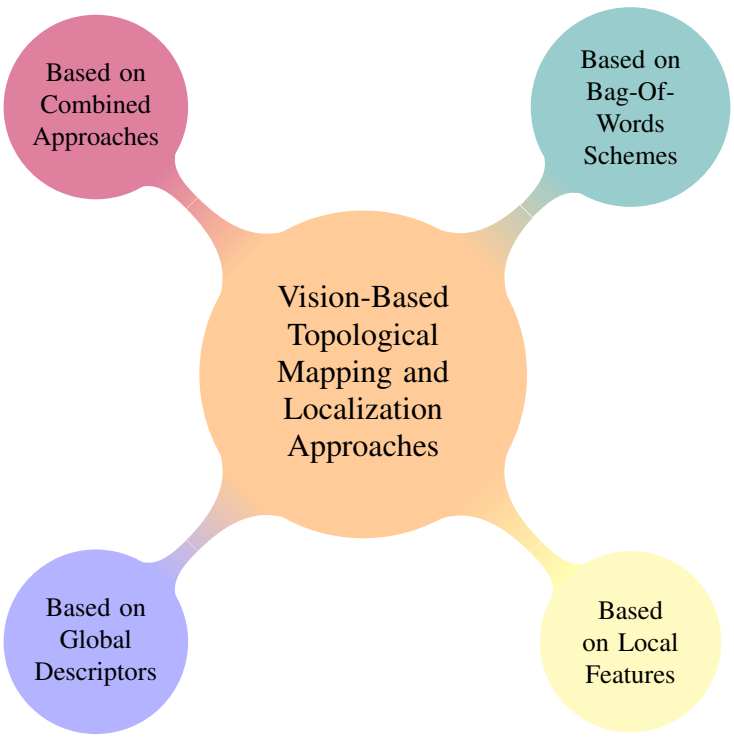


Figure 1: Taxonomy for classifying vision-based topological schemes according to their image representation method.

rough manner. For instance, given the current image, the objective is to conclude that the current place is a kitchen. Some authors create topological maps using this frameworks, forming a graph of known places. However, VPC can be considered as a different research line and these works are also out of the scope of this paper.

In order to perform mapping and localization tasks using vision, it is necessary to describe the acquired images and be able to compare these descriptions. Consequently, the quality of the map and the posterior localization will directly rely on the method used for visually describing the different environment locations. For this reason, we believe that the different approaches can be classified according to the description method employed as approaches based on global descriptors, approaches based on local features and approaches based on Bag-Of-Words (BoW) schemes. We also identify that these methods can be combined. See Fig. 1 for a graphical description of this classification.

Given the taxonomy of the problem, the rest of the paper is organized as follows: Section 2 enumerates fundamental works based on global descriptors; approaches based on local features are

presented in Section 3; Section 4 introduces main solutions built under BoW schemes; Section 5 enumerates principal works that represent the image as a combination of the other ones; and Section 6 concludes the report, including a discussion and proposing some open research issues.

2 Methods based on Global Descriptors

Global descriptors describe the image in a holistic manner, using the full image in the process. These descriptors are normally very fast to compute, what simplifies the matching process between images and reduces the computational needs of mapping and localization tasks. This kind of descriptor has been used in several applications comprising scene classification, giving good results in all cases.

A summary of global descriptors used in some approaches is shown in Table 1. There exist other global descriptors that have not been included in the table because, to the best of our knowledge, they have not been employed in topological mapping and localization solutions, although they could be interesting for the reader. Examples of them include descriptors for scene categorization (Census Transform Histogram (CENTRIST [8]), Pyramid Histogram of Oriented Gradients (PHOG [9]), Histogram of Oriented Uniform Patterns (HOUP [10]), Multi-Resolution BoW [11]) and for pedestrian detection (Histogram of Oriented Gradients (HOG [12])).

Many authors have proposed different solutions for topological mapping and localization using global image representations, which are summarized in Table 2. This table indicates, for each solution, the imaging configuration adopted, whether the resulting map is a pure topological map or otherwise is a hybrid representation, the intended tasks, the environments where the approach was assessed and the image descriptor used.

2.1 Histograms

Histograms provide a compact way of representing an image and have been used for topological mapping and localization in different forms. An example of that is the work of Ulrich and Nourbakhsh [15]. They proposed a topological localization method based on appearance. Each image is represented by six one-dimensional colour histograms, three extracted from the HLS colour space

Table 1: Summary of global image descriptors.

Name	References
Principal Components	[13, 14]
Colour Histograms	[15]
Gradient Orientation Histograms	[16]
WGOH	[17]
WGII	[18]
OACH	[19]
Receptive Field Histograms	[20]
Gist	[21]
Omni-Gist	[22]
BRIEF-Gist	[23]
Spherical Harmonics	[24]
Fingerprints	[25]
FACT	[26]
DP-FACT	[27]
Fourier Signatures	[28, 29]
Colour Segmented Images	[30]
Scanline Intensity Profile	[31]
Normalized Patches	[32]
2D Haar Wavelet Decomposition	[33, 34]
WI-SURF	[35]
DIRD	[36]
OFM	[37]
OFSC	[37]

and other three extracted from the RGB colour space. Given a query image, they retrieved reference images from the map using a nearest neighbour learning scheme in their topological map. The Jeffrey divergence was used as a distance measure between two histograms. They assessed their system in several environments, obtaining at least 87.5% of correctly classified images in all of them. Werner et al. [38] also employed colour histograms combined with a Bayes filter for providing a topological SLAM solution. They used the Hausdorff distance to compare the topological map and the visual observations received by the robotic platform. They argued that colour histograms are not distinctive enough, and that the Bayes filter helps to disambiguate places with similar appearance.

Kosecka et al. [16] proposed a navigation strategy using gradient orientation histograms as image descriptor. In an exploration phase, a topological map was built by comparing successive frame descriptors. For each node, a set of representative views was computed using Learning Vec-

Table 2: Summary of topological mapping and localization solutions based on global image descriptors.

References	Camera	Map	Tasks	Environment	Descriptor
Winters [13]	Omnidir	Topo	Map + Loc	Indoors	PCA
Gaspar [14]	Omnidir	Topo	Map + Loc	Indoors	PCA
Ulrich [15]	Omnidir	Topo	Map + Loc	In + Out	Colour Hist.
Werner [38]	Omnidir	Topo	SLAM	Indoors	Colour Hist.
Kosecka [16]	Mono	Topo	Map + Loc	Indoors	Gradient Orien. Hist.
Bradley [17]	Mono	Topo	Map + Loc	Outdoors	WGOH
Weiss [18]	Mono	Topo	Map + Loc	Outdoors	WGII
Wang [19]	Mono	Topo	Map + Loc	In + Out	OACH
Pronobis [20]	Mono	Topo	Loc	Indoors	Receptive Field Hist.
Singh [39]	Omnidir	Topo	Map + Loc	Outdoors	Gist
Murillo [22]	Omnidir	Hybrid	Map + Loc	In + Out	Omni-Gist
Rituerto [40]	Omnidir	Topo	Mapping	Indoors	Omni-Gist
Sunderhauf [23]	Mono	Topo	SLAM	Outdoors	BRIEF-Gist
Liu [41]	Mono	Topo	SLAM	Outdoors	Gist
Chapoulie [42]	Sphere	Topo	Map + Loc	In + Out	Gist
Chapoulie [24]	Sphere	Topo	Map + Loc	In + Out	Spherical Harmonics
Lamon [25]	Omnidir	Topo	Loc	Indoors	Fingerprints
Tapus [43, 44]	Omnidir	Topo	Map + Loc	Indoors	Fingerprints
Liu [26]	Omnidir	Topo	Mapping	Indoors	FACT
Liu [27]	Omnidir	Topo	Mapping	Indoors	DP-FACT
Menegatti [28, 29]	Omnidir	Topo	Map + Loc	Indoors	Fourier Signatures
Paya [45]	Omnidir	Topo	Map + Loc	Indoors	Fourier Signatures
Ranganathan [46]	Omnidir	Topo	Mapping	Indoors	Fourier Signatures
Milford [47]	Mono	Hybrid	SLAM	Indoors	Colour Segmentation
Prasser [48]	Omnidir	Hybrid	SLAM	Outdoors	Colour Hist.
Milford [31]	Mono	Hybrid	SLAM	Outdoors	Scan Intensity Prof.
Glover [49]	Mono	Hybrid	SLAM	Outdoors	Scan Intensity Prof.
Milford [32, 50]	Mono	Topo	SLAM	Outdoors	Normalized Patches
Lui [33, 34]	Omnidir	Hybrid	SLAM	In + Out	2D Haar Wavelet Dec.
Badino [35]	Mono	Hybrid	Map + Loc	Outdoors	WI-SURF
Lategahn [36]	Mono	Hybrid	SLAM	Outdoors	DIRD
Nourani [37]	Mono	Topo	Map + Loc	In + Out	OFM/OFSC

tor Quantization (LVQ). During the navigation, the current frame's histogram was extracted and compared with each node representatives using the Euclidean distance to determine the most similar location. Inspired by Kosecka's work, Bradley et al. [17] introduced a topological localization approach in large outdoor environments using Weighted Gradient Orientation Histogram (WGOH) features. These features were computed partitioning the image into a grid, and extracting an 8-bin histogram of the gradient orientations for each part of the grid, weighted by the magnitude of the

gradient at each point and the distance from the center to the region. A WGOH descriptor was formed concatenating each histogram and normalizing it to the unit length. In order to avoid a dependence of the feature vector to any particular component, values higher than 0.2 were capped to 0.2 and the final descriptor was re-normalized again. Their experiments covered over 100.000 images and 67 km of traverse with a high success. Similarly, Weiss et al. [18] also split each image into a grid, but computing an 8×8 histogram of integral invariants using two relational kernels. These integral invariant features are features which are invariant to some Euclidean motions, such as rotations or translations. The main idea is to apply all possible transformations to each sub-image and obtain and averaged version of these image transformations. They called these approach Weighted Grid Integral Invariant (WGII) features. These features were combined with a particle filter for outdoor mobile robot localization. Wang et al. introduced Orientation Adjacency Coherence Histograms (OACH) [19] to solve the coarse part of a topological localization process. OACH is an extension of the traditional gradient orientation histograms where two Orientation Adjacency Histograms (OAH) are computed respectively in the edge and corner regions of the image according to the Harris detector response and concatenated to form the final descriptor. In an OAH, the gradient orientations of the center pixel's 4-neighbourhood are accumulated and then normalized by the number of center pixels of each orientation. The Jeffrey divergence between OACH descriptors was used to compare the images in the framework.

Pronobis et al. [20] showed that receptive field responses summarized into histograms can be used for place recognition. In a training phase, several histograms were acquired from the environment and used to train Support Vector Machines (SVM) as classifiers which served as a basis of a topological localization process.

2.2 The Gist Descriptor

Recently, several approaches have proposed to use the Gist global descriptor [21]. Initially developed for scene recognition, it is based on the observation that humans are able to classify images at a single glance under certain conditions. Their authors concluded that humans are receptive to what they called the *spatial envelope* of the scene, defined as a set of perceptual properties related to the shape of the space. They demonstrated that this spatial envelope is closely correlated with

second-order statistics (Discriminant Spectral Template) and with the spatial arrangement of structures in the scene (Windowed Discriminant Spectral Template). A bank of filters (such as Gabor filters [51]) can be used to infer a global descriptor of the scene. Principal Component Analysis (PCA) can be used in order to reduce the final dimension of the descriptor.

Singh and Kosecka [39] computed a Gist descriptor for panoramas applying the algorithm to each of the four views that the omnidirectional image consisted of. They introduced a novel similarity measure between image panoramas for these descriptors and evaluated its efficiency for loop closure detection in urban environments. Murillo et al. [22] extended this proposal and introduced *omni-gist*, an adapted version of the descriptor to be used with omnidirectional images extracted from catadioptric cameras, instead of multi-camera systems. They improved the similarity measure for these descriptors and proposed a hierarchical topological localization and map building algorithm based on them. In a more recent work [40], omni-gist was used in a semantic labelling process for building indoor topological maps. The images were classified as *places* or *transitions*, which corresponds to, respectively, the nodes and the edges of the topological map. This place classification module was integrated with a Hidden Markov Model (HMM) to ensure the temporal consistency.

Motivated by the success of Gist and the BRIEF [52] binary feature descriptor, Sunderhauf and Protzel [23] adapted the former to be used as a global descriptor, introducing the *BRIEF-Gist* descriptor. The implementation is very straightforward: the image is downsampled to a patch size and the BRIEF descriptor is computed from the patch. Other possible implementation consists of partitioning the image into a grid, compute BRIEF descriptor for each patch and concatenate them to form the final descriptor. They used this simple descriptor for loop closing, presenting a SLAM system that can be used in a large-scale scenario, as is shown in their experiments.

Liu and Zhang [41] employed PCA to reduce the dimensionality of a Gist descriptor for improving the efficiency and the discriminative power of the descriptor. Then, they presented a particle filter for detecting loop closures in a SLAM system. These descriptors were taken into account in the update step of the filter. As a result, they showed that a high recall can be obtained at 100% precision with only a few particles.

Chapoulie et al. [42] presented an approach for segmenting the environment into topological

places using spherical images. This segmentation approach was based on detecting changes in the environment and an adapted version of Gist for spherical images. In a more recent work [24], they argued that Gist is not well adapted to represent this kind of images because the sphere spatial periodicity is partially lost. Then, they introduced a new global image representation based on spherical harmonics adapted for spherical views.

2.3 Vertical Regions

Extracting vertical lines in order to define globally omnidirectional images has also been used for topological mapping and localization, specially for indoor environments because of the nature of their structures. In this regard, Lamon et al. [25] presented the concept of *fingerprints* of places. A fingerprint is a circular list of features extracted using different algorithms. In their case, they used two detectors: a vertical edge detector based on histograms and a colour patch detector. They also presented an algorithm for matching these sequences of features based on a minimum energy algorithm, and employed this framework for global localization. Tapus et al. [43] demonstrated that this fingerprint representation combined with an uncertainty model of the features can improve the localization results. After this work, Tapus and Siegwart [44] expanded the fingerprint concept incorporating information from a laser range finder in an incremental topological mapping approach for multi-room indoor environments.

Liu et al. [26] avoided the use of a laser presenting their Fast Adaptive Color Tags (FACT) descriptor for a topological mapping approach. It is based on the fact that, in indoor environments, the important vertical edges (windows, columns, etc.) naturally divide the indoor environment into several meaningful cuts. For each cut, the average colour value in the U-V space is computed. This U-V average value and the width of the region form a region descriptor called *tag*. A scene descriptor is formed concatenating each region descriptor in a vector. Scene matching between new scenes and existing nodes was performed computing the 2D Euclidean distance between colour descriptors, and recursively comparing the widths of the regions according to an empirically determined inequality. In order to take into the account the main drawbacks that this solution presented, they improved their descriptor publishing another version called DP-FACT [27], where a Dirichlet Process Mixture Model is used to combine colour and geometry features extracted

from omnidirectional images.

2.4 Discrete Fourier Transform

Several authors have proposed to use the Discrete Fourier Transform (DFT) as a global image representation method. Menegatti et al. [28] unwarped omnidirectional images over a panoramic cylinder. These panoramic cylinders were expanded row by row into their Fourier series. An image was represented by the first 15 Fourier coefficients i.e. the 15 lowest frequency components, reducing the storage needs for each reference view. The set of these selected coefficients was called by their authors as *Fourier signatures*. They also proposed a method for an automatic organization of a set of reference images obtained in an exploration phase into a *visual memory* and a navigation approach using this framework. To overcome the perceptual aliasing problem that the original approach presented, in a following work [29], they improved their localization system fusing this image representation with a particle filter. Based on these works, Paya et al. [45] contributed with an incremental mapping process, creating the map while the robot is traversing the environment and Ranganathan et al. [46] introduced the concept of Probabilistic Topological Maps (PTM), where a particle filter was employed for approximating the posterior distribution over the possible topologies given the available sensor measurements and an odometry source.

2.5 Biologically-Inspired Approaches

Biologically-inspired solutions try to emulate the information processing methods and problem resolution abilities of the biological systems, simulating the behaviour of living organisms. Several topological mapping and localization solutions fall under this category.

Gaspar et al. [14] mapped an indoor environment emulating the vision-based navigation capabilities of insects using an omnidirectional camera. The images of the topological map were encoded as a manifold in a low-dimensional eigenspace obtained from PCA. In an offline phase, they created a representation of the environment resulting into a topological map, which was later used to navigate using a visual following approach.

Milford et al. [47] introduced RatSLAM, a single-camera SLAM system derived from models of the hippocampal complex in rodents. According to the authors, the operation of these models

appears to be related with some topological and metric properties to its advantage, so it can be considered as a hybrid approach. The environment representation was built using a competitive attractor network structure called *pose cells*, which was used to concurrently represent the belief about the location and orientation of the robot. The system performed a colour segmentation process [30] to detect some coloured cylinders spread around the experimental area in order to update these pose cells. This approach was later adapted by Prasser et al. [48] to be used in outdoor environments and using an omnidirectional camera as a main input sensor. Images were described using histograms of the hue and saturation colour bands and compared using the χ^2 statistic. Later, Milford and Wyeth [31] mapped a path of 66 km along an entire suburb using RatSLAM, showing that it can be used in a long-term operation. A scanline intensity profile is employed as image descriptor, which is a one-dimensional vector formed by summing the intensity values in each pixel column, and then normalizing the final vector. Glover et al. [49] combined RatSLAM with other approaches in order to address the challenging problem of producing coherent maps across several times of the day.

2.6 Other Approaches

Winters et al. [13] utilized an omnidirectional camera to create a topological map from the environment during a training phase. Nodes were sets of images with common properties, and links were sequences of consecutive views between two nodes. The large image set obtained was compressed using PCA, resulting in a low-dimensional eigenspace from which the robot could determine its global topological position using an appearance-based method.

Badino [35] presented an outdoor localization approach based in a descriptor called Whole Image SURF (WI-SURF), where a Speeded Up Robust Feature (SURF) descriptor of the entire image is computed according to [53]. Each node of the map is associated with the GPS coordinates where it was acquired, and a Bayesian filter is used to compute the probability of being in each discrete place of the map. They showed experiments in a long term localization and solved the global localization problem.

Lategahn et al. [36] studied how to generate robust descriptors for changing environments. They proposed to use building blocks which can be used to construct millions of descriptors. In

that work, an evaluation function to evaluate the performance of these descriptors was presented, as well as a search algorithm for them. Results for loop closure detection were also presented. The experiments were carried on using the best combination of these building blocks found and was called *Dird is an Illumination Robust Descriptor* (DIRD).

A complete loop closing system for autonomous mobile robots was proposed by Lui and Jarvis, where omnidirectional images was described employing a GPU-based 2D Haar Wavelet decomposition. These images are used to create a database of signatures. A relaxation algorithm is executed to adjust the topology each time the vehicle revisits a previously seen place.

Nourani-Vatani et al. [37] proposed to use optical flow information to detect changes in the environment, using the Optical Flow Moment (OFM) and the Optical Flow Shape Context (OFSC) descriptors. Then, statistical attributes from the flow were extracted in order to define each location. Once a database of nodes was generated, where a node was defined as a detected scene change, the most likely location was obtained using the Mahalanobis and χ^2 distances. They assessed their approach in indoor and outdoor environments, showing that it could be used in several kinds of scenarios.

In a more recent research line, Milford and Wyeth presented SeqSLAM [32], where instead of searching for a single previously seen image given the current frame, they performed the localization process recognizing coherent sequences of local consecutive images. They showed that this approach could be used for visual navigation under weather or season changes. They employed normalized patches in a cropped version of the original image, and Sum of Absolute Differences (SAD) to compare these patches. They have also showed in recent works that route recognition can be accomplished even with a few bits per image [50] and studied the effect of the length of the sequences onto the SeqSLAM algorithm performance [54].

3 Methods based on Local Features

In the previous section we have reviewed solutions based on global representations, where the description is performed using the entire image content. Such descriptions work well for capturing the general structure of the scene, but they are not able to cope well with several visual problems

like partial occlusions or camera rotations. These problems have been addressed more intensively through the recent development of local features.

During the *extraction* step, a set of distinctive local features, which capture the essence of the image, are detected. These features can be derived from the application of a neighbourhood operation or searching for specific structures within the image, such as corners, blobs or regions. Then, a *description* step is performed, where some measurements are taken from the vicinity of each local feature to form a descriptor. Initially, descriptors were formed as a multi-dimensional floating-point vectors. Recently, several authors have proposed binary descriptors, where local features are defined as bit strings, reducing the storage and computational needs.

In order to identify the same local features in other images, they need to be invariant to certain properties, such as camera rotations or affine transformations. According to [55], a good feature detector should have the following properties: repeatability, distinctiveness, locality, quantity, accuracy and efficiency. The most important property is repeatability, that can be achieved either by invariance, when large deformations are expected because of relevant view changes, or by robustness, in case of relatively small deformations.

Tables 3 and 4 collect relevant information about main feature detectors and descriptors. Formal an detailed descriptions of them are, however, considered out of the scope of this survey. The interested reader is referred to [55, 56, 57, 58]. In the tables, detectors are classified based on the type of the feature extracted following the guidelines of [55], where they distinguished between corner, blob and region detectors. The descriptors are classified according to their type (floating-point or binary). The descriptor size, in number of components, is also showed in the table. These tables do not intend to be complete, but a summary of the most important facts about local feature detection and description. The main topological solutions based on local features can be found in Table 5, following the same guidelines than the previous section.

Several authors have used local features to perform topological mapping and localization tasks, specially since the release of the Lowe's Scale-Invariant Feature Transformt (SIFT) algorithm. Kosecka and Yang [84, 85] used SIFT features for describing images in indoor environments and performed a global localization process based on a simple voting scheme. In order to overcome the problems resulting from dynamic changes in the environment, they proposed to incorporate addi-

Table 3: Summary of local feature detectors. Check marks between parentheses indicate that there exist versions that are invariant to scale or affine transformations.

			Invariant		
Name	References	Type of detector	Rotation	Scale	Affine
Harris	[59]	Corners	✓	(✓)	(✓)
Shi and Tomasi	[60]	Corners	✓		
SUSAN	[61]	Corners	✓		
FAST	[62]	Corners	✓	(✓)	
FAST-ER	[63]	Corners	✓	(✓)	
ORB	[64]	Corners	✓	✓	
AGAST	[65]	Corners	✓	(✓)	
BRISK	[66]	Corners	✓	✓	
SIFT	[67]	Blobs	✓	✓	
SURF	[68]	Blobs	✓	✓	
CenSure	[53]	Blobs	✓	✓	
Star	[69]	Blobs	✓	✓	
SUSurE	[70]	Blobs	✓	✓	
KAZE	[71]	Blobs	✓	✓	
AKAZE	[72]	Blobs	✓	✓	
ASIFT	[73]	Blobs	✓	✓	✓
MSER	[74]	Regions	✓	✓	✓

tional knowledge about neighbourhood relationships between individual locations using a Hidden Markov Model. The likelihood function was based on the number of correspondences between the current image and the past locations. Following this work, in [86] they presented a feature selection strategy in order to reduce the number of keypoints per location. This strategy was carried on measuring the discriminability of the individual features to describe each topological location. Zhang [87] also presented a method for selecting a subset of visual features from an image called Bag-of-Raw-Features (BoRF). The features are selected according to the scale where they are found. A location was represented by the set of features that can be matched consecutively in several images, applying a keyframe selection policy based on his previous work [125]. The main problem that BoRF presents is that the number of features to manage increases while new images are added, and a linear search for matching becomes intractable. This drawback was overcome in [88] by indexing features through kd-tree structures.

Using the idea of maintaining only persistent features, several authors have proposed vari-

Table 4: Summary of local feature descriptors.

Name	References	Component type	Number of components	Invariant: Rotation	Invariant: Scale	Invariant: Affine
SIFT	[67]	Float	128	✓	✓	
SURF	[68]	Float	32, 64, 128	✓	✓	
U-SURF	[68]	Float	32, 64, 128		✓	
GLOH	[58]	Float	64, 128	✓	✓	
PCA-SIFT	[75]	Float	36	✓	✓	
M-SIFT	[76]	Float	128	✓	✓	
DAISY	[77]	Float	200	✓	✓	
LESH	[78]	Float	128	✓	✓	
ASIFT	[73]	Float	128	✓	✓	✓
KAZE	[71]	Float	64	✓	✓	
BRIEF	[52]	Bit	128, 256, 512			
ORB	[64]	Bit	256	✓	✓	
BRISK	[66]	Bit	512	✓	✓	
FREAK	[79]	Bit	512	✓	✓	
AKAZE	[72]	Bit	488	✓	✓	
D-BRIEF	[80]	Bit	32	✓	✓	
LDAHash	[81]	Bit	128	✓	✓	
BinBoost	[82]	Bit	64	✓	✓	
LDB	[83]	Bit	256, 512	✓	✓	

ous solutions to the community. Rybski et al. [89] used Kanade-Lucas-Tomasi (KLT) feature tracker for matching persistent features in a sequence of omnidirectional images and constructed a topological map incrementally. He et al. [90] proposed to use manifold constraints to find representative feature prototypes, which are useful to represent any image within the environment in an efficient manner. Sabatta [91] introduced a mapping and localization algorithm that exploits the persistence of SIFT features within consecutive omnidirectional images to improve data association. He also modified the SIFT algorithm in order to include colour information in the descriptor. More recently, Johns and Yang [92] introduced an approach where the map is composed by a set of landmarks detected across multiple images, spanning the continuous space between nodal images. Given a query image, matches are then made to landmarks instead of individual images, resulting into a dense continuous topological map without sacrificing the speed of the solution. They presented a probabilistic localization approach using the learned discriminative properties of each landmark.

Table 5: Summary of topological mapping and localization solutions based on local features.

References	Camera	Map	Tasks	Environment	Feature
Kosecka [84, 85, 86]	Mono	Topo	Map + Loc	Indoors	SIFT
Zhang [87]	Mono	Topo	Map + Loc	Indoors	SIFT
Zhang [88]	Mono	Topo	SLAM	Indoors	SIFT
Rybski [89]	Omnidir	Topo	Map + Loc	Indoors	KLT
He [90]	Mono	Topo	Map + Loc	Outdoors	SIFT
Sabatta [91]	Omnidir	Topo	Map + Loc	Indoors	SIFT
Johns [92]	Mono	Topo	Map + Loc	Indoors	SIFT
Kawewong [93, 94]	Omnidir	Topo	SLAM	In + Out	PIRF (SIFT)
Tongprasit [95]	Omnidir	Topo	SLAM	In + Out	PIRF (SURF)
Morioka [96]	Omnidir	Hybrid	SLAM	Indoors	3D-PIRF (SURF)
Andreasson [76]	Omnidir	Topo	Map + Loc	Indoors	KLT/M-SIFT
Valgren [97]	Omnidir	Topo	Mapping	Indoors	KLT/M-SIFT
Valgren [98]	Omnidir	Topo	Mapping	In + Out	SIFT
Valgren [99]	Omnidir	Topo	Loc	Outdoors	SIFT/SURF
Ascani [100]	Omnidir	Topo	Loc	In + Out	SIFT/SURF
Anati [101]	Omnidir	Topo	Map + Loc	In + Out	SIFT
Zivkovic [102]	Omnidir	Hybrid	Map + Loc	Indoors	SIFT
Booij [103]	Omnidir	Hybrid	Map + Loc	Indoors	SIFT
Booij [104]	Omnidir	Hybrid	Map + Loc	In + Out	SIFT
Dayoub [105]	Omnidir	Hybrid	Map + Loc	Indoors	SURF
Blanco [106, 107]	Stereo	Hybrid	SLAM	Indoors	SIFT
Tully [108]	Omnidir	Hybrid	Map + Loc	Indoors	SIFT
Tully [109]	Omnidir	Hybrid	SLAM	Indoors	SIFT
Segvic [110]	Mono	Hybrid	Map + Loc	Outdoors	SIFT/Harris/MSER
Ramisa [111]	Omnidir	Topo	Map + Loc	Indoors	MSER/SIFT/GLOH
Badino [112]	Mono	Hybrid	Map + Loc	Outdoors	SURF/U-SURF
Dayoub [113]	Omnidir	Topo	Map + Loc	Indoors	SURF
Bacca [114, 115]	Omnidir	Topo	Map + Loc	Indoors	SIFT/SURF
Bacca [116]	Omnidir	Topo	SLAM	Indoors	Edges
Romero [117, 118]	Omnidir	Topo	SLAM	Outdoors	MSER
Majdik [119]	Mono	Topo	Loc	Outdoors	ASIFT
Saedan [120]	Omnidir	Hybrid	SLAM	Indoors	Wavelets
Kessler [121]	Omnidir	Topo	SLAM	Indoors	SIFT
Maohai [122]	Omnidir	Topo	Map + Loc	Indoors	ASIFT
Garcia-Fidalgo [123]	Mono	Topo	SLAM	In + Out	SURF
Garcia-Fidalgo [124]	Mono	Topo	SLAM	In + Out	SIFT

Kawewong et al. presented Position-Invariant Robust Features (PIRFs) [93, 94], a method for generating averaged features from SIFT descriptors that can be matched along several consecutive frames in a temporal window given the input sequence of images. Each place was represented by a dictionary of these representative PIRFs, whose variation of appearance was assumed relatively

small with regard to robot motion. These features were then used in an incremental appearance-based SLAM algorithm called PIRF-Nav, which was based on a majority voting scheme. Despite they showed several improvements in terms of recall regarding other common solutions, the main problem of this approach was the computational cost, since some images took long time to be processed. In order to improve this performance, Tongprasit et al. [95] modified the original PIRF algorithm and added a new dictionary management in a SLAM approach called PIRF-Nav 2. This method was 12 times faster than the original PIRF-Nav sacrificing only a small percentage of recall. Morioka et al. [96] presented a method for mapping PIRFs in three-dimensional space combining them with an odometry source. Their method, called 3D-PIRF, was validated navigating in crowded indoor environments.

Andreasson and Duckett [76] presented a simplified version of the SIFT algorithm (M-SIFT) adapted to omnidirectional images, where the descriptors are only found in one resolution, because full invariance to scale and translation is not required in their case. Interest points are selected using the Shi and Tomasi method. Several methods for topological localization were presented, showing the M-SIFT approach the best performance with regard to the other ones. Using the M-SIFT descriptor, Valgren et al. [97] represented the environment by means of an image similarity matrix. They avoided exhaustively computing the affinity matrix by searching for cells which are more likely to describe existing loop closures. Later, in [98], they employed exhaustive search, but introduced an incremental spectral clustering algorithm to reduce the search space incrementally when new images are processed. They also addressed the topological localization problem for outdoor environments over time [99], comparing SIFT and SURF for these purposes and concluding that SURF performs better for topological localization in outdoor scenarios. Moreover, Ascani et al. [100] found that SIFT performs better in indoor environments for topological localization tasks. Other authors that created a topological map from a similarity matrix are Anati and Daniilidis [101]. In their work, they introduced a novel image similarity measure for panoramas which involves dynamic programming to match images using both the appearance and the relative positions of local features simultaneously. The probability of loop closures is modelled using a Markov Random Field (MRF) over the image similarity matrix.

Some researchers construct hierarchical maps of the environment from a set of input images.

These approaches combine higher level conceptual maps (usually topological) with lower level and geometrically accurate maps, trying to maximize the advantages and minimize the problems of each kind of map alone and combine them in a different mapping technique. Zivkovic et al. [102] presented an algorithm for automatically generating hierarchical maps from images. A low-level map is built using SIFT features and geometrical constraints. They then use the graph-cuts algorithm to cluster nodes to construct a high-level representation. This hierarchical representation was later employed in [103], where they showed a navigation system based on a topological space which used the epipolar geometry and a planar floor constraint to obtain a heading estimation. This work was further improved in [104] proposing a incremental data association scheme based on the concept of Connected Dominating Set (CDS) of a graph. Given a new image, this method is used to find a subset of past images that represents the complete image set, enabling an efficient loop closure detection during the trajectory of the robot. Dayoub et al. [105] presented a solution where an initial dense pose-graph map of the environment were generated using a graph-based SLAM algorithm. This map is then used to infer a sparse hybrid map with two levels, global and local. The global level is represented by a topological map built using a dual clustering approach. On the local level, each node stores a spherical view representation of the features extracted from images recorded at the position of the node, which is used for estimating the robot's heading using a multiple-view geometry approach.

Instead of inferring a high-level topological map from a set of geometric relations, other authors have proposed an alternative hybrid representation where each node of a global topological map includes its own metric sub-map. Blanco el al. [106] presented an approach called Hybrid Metric-Topological SLAM (HTM-SLAM). The sequence of areas traversed by the robot is modelled as a graph whose nodes are annotated with metric sub-maps and whose arcs include the coordinate transformation between these areas. They also proposed a unified Bayesian approach using these maps in order to estimate the robot's path while traversing the environment. This work was improved in [107] using spectral techniques to efficiently partition the map into sub-maps and deriving expressions for applying their ideas to other sensors, such as a stereo camera. In the same line, Tully et al. [108] proposed a hybrid localization solution based on the *hierarchical atlas* map [126], a structure specially created for robots operating in large environments. In this

framework, a global topological map decomposes the space into regions within which a feature-based map is built. The localization process is separated in two steps. First, a discrete probability distribution is computed using a recursive Bayesian filter in order to determine the most probable map. Next, a metric position is estimated within the correspondent sub-map using a Kalman filter. Later, in [109], they investigated SLAM as a multi-hypothesis topological loop closing problem. Both works were combined in a more complete solution recently in [127].

Segvic et al. [110] created a hybrid visual navigation framework for large-scale mapping and localization combining several features extracted from monocular perspective images. Despite the approach supported navigation based exclusively on 2D image measurements, it relied in 3D reconstruction procedures. Ramisa et al. [111] also tried to combine several local feature region detectors in order to create a signature of a place for localization purposes. They showed that these combinations increase notably the performance compared with the use of one descriptor alone. Badino et al. [112] integrated metric data directly into a topological map in their hybrid approach called *topometric* localization. Each node of the graph is stored together with its GPS position. They grab images at a constant Euclidean distance, and for each one, visual local features are extracted. A feature database is generated next, where each feature is stored with a reference to the node corresponding to its real location. This database is then used by a Bayes filter to estimate the probability density function of the position of the observer as the vehicle moves along the route.

The multi-store model of human memory proposed by Atkinson and Shiffrin [128] has inspired several approaches. This model divides the human memory into three stores: Sensory Memory (SM), Short-Term Memory (STM) and Long-Term Memory (LTM). Input information is stored in the SM. A selective attention process determines which information can be moved to the STM. Information stored in this memory can be forgotten as soon as it is no longer attended to. Through a rehearsal process, information is moved from the STM to the LTM in order to be retained for longer periods. Dayoub and Duckett [113] used these concepts in order to keep up to date the appearance of a particular place in a map in response to the dynamic changes of the environment during a long-term operation. Bacca et al. [114, 115] adapted this human memory model considering a weighted voting scheme. This allows to pass to the STM only strong features present in the

environment. The memory model is implemented using a Feature Stability Histogram (FSH), which stores information about the number of times each feature has been observed in each node. A more complete FSH approach was presented in [116], adapting the initial solution to operate in SLAM conditions.

Romero and Cazorla [117, 118] proposed an approach to construct topological maps matching graphs of invariant features. Each image is segmented into regions in order to group the extracted invariant features in a graph so that each graph defines a single region of the image. The matching process takes into the account the features and their structure using the Graph Transformation Matching (GTM) algorithm.

Recently, Majdik et al. [119] dealt with the *air-ground* matching localization problem, where images taken by a camera mounted on a Micro Aerial Vehicle (MAV) need to be matched with a set of images stored in a database of geotagged pictures obtained from Google Street View. To overcome the severe viewpoint changes presented, they proposed to generate virtual views of each scene, exploiting the air-ground geometry of the system. The best image correspondences are obtained using a histogram-voting scheme. They compared their solution with several state-of-the-art approaches, outperforming them in computational terms and precision-recall rates.

Other solutions based on local features [120, 121] included particle filters as a method to estimate the probability distribution of the location over the topological map. More recently, Maohai et al. [122] combined a particle filter with a GPU-based image description and matching algorithm to define a complete topological autonomous navigation system for indoor environments.

In a previous work [123], we proposed an appearance-based approach for visual mapping and localization. On the one hand, a new image similarity measure between images based on number of matchings and their associated distances was introduced. On the other hand, to optimize running times, matchings between the current image and previous visited places were determined using an index based on a set of randomized KD-trees. Further, a discrete Bayes filter was used for predicting loop candidates, taking into account the previous relationships between visual locations. The approach was validated using image sequences from several environments. In order to avoid redundant information in the resulting maps, we recently presented a map refinement framework [124], which takes into account the visual information stored in the map for refining the final

topology of the environment. These refined maps save storage space and improve the execution times of the localization tasks.

4 Methods based on Bag-Of-Words Algorithm

The Bag-of-Words (BoW) algorithm was initially developed for text retrieval, where a BoW is a sparse vector representation of a document counting the number of occurrences of each word given a predefined vocabulary. Documents with more words in common are likely to describe the same topic. Exporting these concepts to the computer vision field [129], the idea is to treat local features as visual words and quantize them according to a set of representative features, known as *codebook* or *visual vocabulary*. This quantization is performed by mapping each descriptor of the image to the nearest image word in the dictionary. Then, the image is represented by a histogram of occurrences of each reference local feature presented in the image, reducing the total set of feature descriptors found to a vector of integers. Since some words are more discriminative than others when identifying an image, the BoW vector is normally weighted by some scoring algorithm such as the Term Frequency-Inverse Document Frequency (TF-IDF). The most common way of generating a visual dictionary is to cluster the descriptors extracted from a set of training images using some clustering algorithm, such as k-means, where the learned centroids are considered as the reference visual words.

As will be seen in Section 6, generating the visual dictionary in an offline phase presents several problems. In order to overcome these drawbacks, some authors have proposed to build it in an incremental fashion, adapting the codewords to the appearance of the operating scenario. In this section, the BoW-based works are classified according to this criterion. The main approaches based on the BoW algorithm are summarized in Table 6 specifying the same features than in previous sections.

4.1 Offline Visual Vocabulary Approaches

Despite the BoW algorithm has been used in other areas, such as for internet search engines or for scene categorization [163, 164], it was first applied to visual search techniques in the semi-

Table 6: Summary of topological mapping and localization solutions based on the BoW algorithm.

References	Camera	Map	Tasks	Environment	Quantized Feature
Wang [130, 131]	Mono	Topo	Map + Loc	In + Out	HARRIS/SIFT
Fraundorfer [132]	Mono	Topo	Map + Loc	Indoors	MSER/SIFT
Konolige [69]	Stereo	Hybrid	SLAM	In + Out	STAR/FAST/SAD
Cummins [133, 134]	Mono	Topo	SLAM	Outdoors	SIFT/SURF
Cummins [135, 136]	Mono	Topo	SLAM	Outdoors	SURF
Cummins [137, 138]	Omnidir	Topo	SLAM	Outdoors	SURF
Newman [139]	Omnidir	Hybrid	SLAM	Outdoors	SURF
Maddern [140, 141]	Omnidir	Hybrid	SLAM	Outdoors	SURF
Maddern [142]	Omnidir	Hybrid	SLAM	Indoors	SURF
Paul [143]	Mono	Topo	SLAM	Outdoors	SURF
Johns [144, 145]	Mono	Topo	SLAM	Outdoors	SIFT
Galvez [146, 147]	Mono	Topo	SLAM	In + Out	FAST/BRIEF
Ranganathan [148]	Mono	Hybrid	SLAM	Indoors	SIFT
Cadena [149]	Stereo	Topo	SLAM	In + Out	SURF
Ciarfuglia [150]	Mono	Topo	SLAM	In + Out	SURF
Majdik [151]	Mon/Ste	Topo	SLAM	Outdoors	SURF
Schindler [152]	Mono	Topo	Map + Loc	Outdoors	SIFT
Achar [153]	Mono	Topo	Map + Loc	Outdoors	SIFT
Lee [154]	Mono	Topo	SLAM	Indoors	MSLD
Filliat [155]	Mono	Topo	Map + Loc	Indoors	SIFT
Angeli [156]	Mono	Topo	SLAM	Indoors	SIFT
Angeli [157]	Mono	Topo	SLAM	In + Out	SIFT/Color Hist.
Angeli [158]	Mono	Topo	SLAM	Indoors	SIFT/Color Hist.
Labbe [159, 160]	Mono	Topo	SLAM	In + Out	SURF
Nicosevici [161, 162]	Mono	Topo	SLAM	Underwater	SURF

nal work of Sivic an Zisserman [129], where this model was employed in order to find similar scenes in video sequences. SIFT features were extracted from each frame and then quantized as BoW vectors, creating a database of BoW image representations. They presented an interactive application where the user could query the image database to find similar frames , i.e. with enough features in common. A lookup table called inverted file, which mapped image words to the video frames where they were found, was also used to speed up the retrieval process. Wang et al. [130, 131] presented a coarse-to-fine global localization system based on the BoW model, where interest points detected with the Harris-Laplace detector were described using the SIFT algorithm. In an offline phase, the vocabulary and the inverted index were created, and then used for localization. An epipolar geometry step was incorporated in order to verify whether the loop candidate obtained from the BoW stage was plausible.

The size of a dictionary can vary within a large range, which has an impact on the performance of the retrieval process. The larger the size, the more discriminative the vocabulary is, but at a higher computational cost for finding the nearest reference descriptor. The hierarchical visual vocabulary has been proposed as a relevant improvement towards alleviating this problem [165], where the original training set of descriptors is clustered in a small number of clusters, and then each cluster is recursively clustered again until achieving the desired number of words. Given a query descriptor, finding its closest word consists in traversing the tree from the root until reaching a leaf node. This hierarchical representation, in addition to the inverted index, makes the BoW algorithm an ideal and scalable approach for searching millions of images in an efficient way and it is a good option to consider when mapping large environments. Fraundorfer et al. [132] applied this hierarchical dictionary to the visual navigation problem, presenting a highly scalable vision-based localization and mapping method using image collections. For each frame captured by the camera, they used the dictionary structure and the inverted file to retrieve the most likely images. Using a RANSAC procedure, they performed a geometry verification step against these candidates, which can be used to determine if the image closes a loop or otherwise is a new place to be added to the map. They used the local geometric information to navigate within the generated topological map. Konolige et al. [69] proposed a SLAM solution based on an adapted scheme of this hierarchical codebook using a stereo camera. As shown in their results, the approach, which was assessed in indoor and outdoors environments, was able to find loop closures in paths of several kilometers. A strong geometric filter was used to eliminate false positives when detecting loop closures.

Probably the most well-known solution that falls into this category is the Cummins and Newman's Fast Appearance-Based Mapping (FAB-MAP) [133, 134] approach, proposed under the assumption that modeling the probabilities that the visual words appear simultaneously can help in the localization process. These probabilities were approximated by a Chow Liu tree, computed from a set of training data as the maximum-weight spanning tree of a directed graph of co-occurrences between visual words. This approximation permitted the authors to compute efficiently an observation likelihood which was used in a Bayes filter for predicting loop closure candidates. The main drawback presented by the original FAB-MAP algorithm was the high

computational cost, since every time the robot collected an observation, the likelihood needed to be computed for each location existent in the map. To solve this problem, Cummins and Newman [135, 136] introduced a probabilistic bail-out test based on the use of concentration inequalities for rapidly identifying promising loop closure hypotheses and then avoid to compute the likelihood for all locations. Later, an even faster version called FAB-MAP 2.0 [137, 138] was presented adapting the probabilistic model to be used with an inverted index architecture similar to image typical search engines. This scheme was assessed using a dataset of 1000 km composed by omnidirectional images and GPS coordinates to be used as ground truth. FAB-MAP was combined with a laser in the work of Newman et al. [139], where it was used as a component to detect loop closures to describe urban scenes.

Initially, the authors only published FAB-MAP as binaries to the community. For this reason, Glover et al. developed OpenFABMAP [166], a fully open-source implementation of the algorithm. OpenFABMAP was a key component in the solution proposed by Maddern et al. called Continuous Appearance-based Trajectory SLAM (CAT-SLAM) [140, 141], where an appearance-based SLAM system was improved with odometric information using a particle filter in order to obtain an estimation of the position of the vehicle. An extension of CAT-SLAM called CAT-Graph was introduced in [142] combining multiple visits to the same place to build a topological graph-based representation of indoor environments. These graphs were used in the mapping and localization processes according to the loop closures detected by the appearance-based module.

Since the BoW model used in FAB-MAP does not take into account the spatial arrangement of features, Paul and Newman introduced FAB-MAP 3D [143], where they demonstrated that integrating this kind of information in the algorithm improved the localization accuracy. Using a random graph, they modeled the word co-occurrences as well as their pairwise distances and showed how to accelerate the inference process with a Delaunay tessellation of this graph. Another attempt to include spatial information within the BoW model for localization is the recent work by Johns and Yang, where they presented the Feature Co-occurrence Maps (Cooc-Map) [144], where local features are quantized in both feature and image space and a set of statistics regarding their co-occurrence at different times of the day are calculated. They also introduce a new geometric feature matching algorithm for this kind of representation and showed that sequential matching can

be incorporated into their solution. They also showed that learning the properties of local features observed during long periods of time can be more accurate for localization than representing a location using a single image [145].

An attempt to create a visual dictionary from binary features can be found in the work of Galvez-Lopez and Tardos [146, 147]. They adapted the hierarchical BoW model of Nister to be used with keypoints detected with FAST and described with the BRIEF algorithm. Other novelties of their work included a direct index to obtain correspondences between images in a efficient manner and matching images in groups to increase the accuracy of the loop closure detection process. Using this framework, they are able to detect loop closures in sequences of 19000 images spending an average time of 16 ms per image, presenting an interesting improvement in performance in comparison to other solutions.

Ranganathan and Dellaert presented Online Probabilistic Topological Mapping (OPTM) [148], an online loop-closing algorithm based on a Rao-Blackwellized particle filter which was used for updating incrementally the posterior on the space of all possible topologies whenever a new measurement arrived. Since OPTM was sensor independent, it was assessed with a laser range finder, an odometry source and visual input in indoor environments. A BoW model based on a multivariate Polya distribution was used for quantizing SIFT descriptors. OPTM improves a previous framework called Probabilistic Topological Maps (PTM) [167] by enhancing the inference process so that it can be used online.

Cadena et al. [149] introduced a place recognition framework based on stereo vision which combined a BoW model for obtaining loop closure candidates and an algorithm based on Conditional Random Fields (CRF-Matching) in order to verify these candidates. This matching method, according to the authors, was more robust than using only epipolar geometry, since it used 3D information provided by the stereo images. This module was later used in [168], where a method for removing past incorrect loop closures using the Realizing, Reversing, Recovering (RRR) algorithm was presented.

Some authors have proposed weighing strategies different to the one typically used in BoW approaches, i.e. The TF-IDF. For a start, Ciarfuglia et al. [150] showed a discriminative criterion to assign weights to the visual words in a training phase. The weights are learnt in an approach

based on the large margin paradigm and can be applied to several similarity functions in order to compare images. This weighing scheme was assessed as of a loop closure detector within a SLAM framework for navigating in indoor and outdoor environments. Another case is Majdik et al. [151] who proposed an adaptive loop closure algorithm based on the hierarchical BoW model that was able to update the weights of the visual words according to their importance when detecting loop closures. They assessed their approach using both single and stereo cameras in outdoor environments.

While in outdoor environments GPS can be used for estimating the location of a robot, urban environments present more challenging situations since buildings can block the satellite signals. Clearly, vision becomes an option as exteroceptive sensor in these cases. Nevertheless, indexing images from a city can be very difficult in computational terms, reason why the BoW model can be of help for this kind of situations. In line with this scenario, Schindler et al. [152] presented a localization system for recognizing scenes in cities, where they were able to index 30000 images from a city using a BoW scheme. They showed that this huge amount of information can be more efficiently retrieved by selecting the most informative features from the training dataset, understanding these features as the ones that occur in all images of some specific location but not in other places. This concept was measured using the information gain formula. They also proposed an alternative search algorithm called Greedy N-Best Paths (GNP) improving the image retrieval performance. A more recent solution for urban localization can be found in the work by Achar et al. [153], where geometric inferencing was used to identify features corresponding to moving objects in the scene. These features are then used for global localization.

Recently, Lee et al. [154] proposed a place recognition system that, instead of quantizing interest points, they processed lines using Mean Standard-Deviation Line descriptors (MSLD). A hierarchical visual dictionary was trained using these vectors, which was employed in combination with a Bayes filter for detecting loop closures in indoor environments. They integrated this loop closure detection module into a SLAM solution.

4.2 Online Visual Vocabulary Approaches

An alternative in order to maintain the dictionary adapted to the operating environment is to generate it online, at the same time that the robot explores the world. In this regard, Filliat [155] introduced an approach to construct dynamically a visual dictionary. The closest visual word to a given local feature was selected performing a simple linear search algorithm. If these features were very far in distance, the query local feature was added as a new word to the dictionary. This scheme was assessed using different feature spaces and employed for mapping and localization tasks, but it was limited to small distances due to the inefficiency of the linear search algorithm. This model was extended by Angeli et al. [156] to incremental conditions to be used in a place recognition module. Their approach relied on a discrete Bayes filter to estimate the probability of loop closures and to ensure temporal coherency between predictions. During the calculation of the likelihood, the TF-IDF coefficients are extracted according to the distinctiveness of each word given the current image. This work was improved in [157], where two visual vocabularies were trained and used together as input to the Bayes filter, and further expanded in [158] by constructing a complete topological SLAM system.

Inspired by the work of Angeli, Labbe and Michaud presented Real-Time Appearance-Based Mapping (RTAB-Map) [159, 160] a loop closure detection approach for large-scale and long-term SLAM. The main contribution of this solution was that they provided memory management mechanisms for caching a subset of the online learnt visual words in the main memory (called Working Memory), and this subset was used for detecting loop closures. The rest were stored in a database stored in an external memory called Long Term Memory. The transition of words between memories was ruled by the time taken for processing images in an adaptive way. This scheme allowed to obtain high recall rates at 100% of precision while maintaining the real time performance of the solution.

Nicosevici and Garcia [161, 162] introduced Online Visual Vocabulary (OVV), where the words were generated at the same time that the robot was exploring the environment using a modified version of an agglomerative clustering algorithm. The elementary clusters were created from features that can be tracked along the images of the sequence, represented by the mean descriptor of a feature and the covariance matrix of the observed descriptors at the current point. In order to

Table 7: Summary of topological mapping and localization solutions based on combined approaches.

References	Camera	Map	Tasks	Environment	Combination
Goedeme [172, 173]	Omnidir	Topo	Map + Loc	Indoors	SIFT/Columns
Murillo [174, 175]	Omnidir	Hybrid	Map + Loc	In + Out	SURF/Color Hist.
Wang [176]	Mono	Topo	Map + Loc	In + Out	OACH/SIFT
Weiss [177, 178]	Mono	Topo	Map + Loc	Outdoors	WGOH/WGII/SIFT
Siagian [179]	Mono	Topo	Map + Loc	Outdoors	Gist/SIFT
Chapoulie [180]	Sphere	Topo	SLAM	Outdoors	SIFT/Spatial Hists.
Wang [181]	Omnidir	Topo	Map + Loc	Indoors	SURF/Convex Hull
Lin [182]	Omnidir	Topo	Map + Loc	In + Out	SURF/Convex Hull
Wang [183]	Mono	Topo	Map + Loc	Outdoors	Harris/Color Hist.
Korrapati [184, 185]	Omnidir	Topo	Mapping	Outdoors	SURF/BoW

merge these clusters, they provided a novel criterion based on the Fisher's linear discriminant that took into account the global distribution of the data, resulting into more distinctive visual words. A method for efficiently reindexing the images when the vocabulary changes is also proposed. An interesting aspect of their experimental results is that, in addition to outdoor scenarios, the approach was assessed in underwater environments.

Despite they are more related to the pose-graph SLAM field, there exists other solutions that used a BoW scheme built in an online manner that can be interesting for the reader, such as the works of Eade and Drummond [169], Botterill et al. [170] and Pradeep et al. [171].

5 Methods based on Combined Approaches

In order to maximize the benefits of each approach, several authors have proposed solutions based on combinations of different image descriptors for topological mapping and localization. The main approaches that fall into this category are summarized in Table 7 specifying the same features as in previous sections.

A common approach is to use a global descriptor to perform a fast selection of similar images during an image search and then use a more accurate process in order to confirm the association, such as matching local features. Goedeme et al. [172] presented a localization system for omni-directional cameras where, for each acquired image, they extracted vertical column segments and described them with ten different descriptors. After a clustering process, these local descriptors

were inserted into a kd-tree structure that was used by the localization process. When a query image arrived, the same local descriptors applied to the vertical structures were computed over the entire image and used to rapidly retrieve possible loop candidates. Next, a matching distance based on the column segments was applied between the image and each of the candidates in order to ensure a correct image matching. The localization process was supported by a Bayes filter, which allowed them to deal with noisy measurements. Their work was improved in [173], presenting a complete navigation system, adding SIFT features to the framework and applying the Dempster-Shafer probabilistic theory to the topological map construction.

Murillo et al. [174] proposed a three-step hierarchical localization method for omnidirectional images. A global color descriptor was applied to obtain a set of susceptible loop candidates, and then line features described by their line support regions were matched using pyramidal matching in order to find the most similar image given a predefined visual memory. The 1D radial trifocal tensor was employed to obtain a metric localization. Their work was expanded incorporating SURF local invariant features to the framework [175].

Wang and Yagi [176] combined recently their OACH global descriptor with local features extracted with the Harris-Laplace detector and described by the SIFT descriptor. They created two databases: one for OACH descriptors for coarse localization and a SIFT database for fine localization. During the global localization stage, a set of candidate images was extracted and then a fine localization step against this subset was performed. A RANSAC-based fundamental matrix estimation strategy was employed in order to verify if the image association was correct.

Weiss et al. [177] performed outdoor localization using a particle filter where particle weights were updated according to the similarities computed using two global descriptors: WGOH and WGII. To calculate the similarity between two images, each descriptor is compared independently using normalized histogram intersection and the final distance is the product of the previous results. This method was compared with SIFT, presenting a slightly minor recall, but four times faster. Later in [178], SIFT was incorporated into their framework as an alternative to compute the position of the robot in those cases where it can not be inferred using the combined global descriptors method.

Another localization approach based on particle filters and inspired in biological concepts can

be found in the work proposed by Siagian and Itti [179], which is based in Gist and saliency features, implemented in parallel using shared raw feature channels.

Chapoulie et al. [180] introduced a loop closing algorithm to be used with spherical images. SIFT features were extracted as local features, while histograms of their distribution over the features space were used as global features. These representations were combined in a Bayes filter in order to detect loop closure candidates under outdoor environments.

Wang and Lin presented a combined local and global descriptor for omnidirectional images called Hull Census Transform (HCT) [181], which consisted of repeatedly generating the convex hull from the extracted SURF features and computing the relative magnitude between these features that compose the convex hull, resulting into a set of binary vectors. This representation was then used for detecting scene changes, generating a set of topological node lists. This work was recently expanded by Lin et al. [182] in a new combined descriptor called Extended-HTC, where they included color information from the environment, encoded as color histograms, as well as the structure information of the convex hulls, computed by means of the centroid of the features and the total distance between any two feature point locations.

A location recognition system which combined edges, local features and color histograms was proposed by Wang and Yagi [183]. The image description process was computed in an integrated way: the Harris detector was used to obtain both edges and interests points, while SIFT algorithm was used for describing interest points.

Recently, Korrapati et al. [184] presented a hierarchical mapping model which organized images into a topological map using the Vector of Locally Aggregated Descriptors (VLAD), where the quantization residues of the local features descriptors, such as SURF, were combined into a single descriptor. This allowed them to create maps containing over 11000 images and a decent amount of frames per second. In a more recent work [185], they also proposed a hierarchical topological mapping algorithm using a sparse node representation where Hierarchical Inverted Files (HIF) were employed for an efficient two-level map storage.

6 Discussion and Conclusions

In the last decades, there has been a significant increase in the number of visual solutions for topological mapping and localization because of the low cost of cameras and the richness of the sensor data provided. This paper surveyed the main approaches emerged in the last 15 years. We identified that these works can be classified, according to the method used for representing the image, into four main categories:

- methods based on global descriptors, where the image is represented by a general descriptor computed using the entire visual information as input;
- methods based on local descriptors, where interest points are found in the image and then a patch around this point is described in order to identify them in other similar images;
- methods based on the BoW algorithm, where local features are quantized according to a set of feature models called visual dictionary, representing images as histograms of occurrences of each word in the image; and
- methods based on combined descriptors, where several techniques described above are used together as a new solution.

The main advantages and disadvantages of each method are summarized in Table 8. All these methods are active research areas and authors publish continuously solutions for mapping, localization or SLAM facing the problem from the point of view of these approaches.

In this work, we consider that a topological node of a map is a unique place of the robot's environment that can be represented by its appearance. For this reason, pose-graph SLAM solutions, that consider the environment as a graph of poses, have been considered out of the scope of this paper. Nevertheless, we do have included works that make use of hybrid metric-topological maps.

Regarding the different categories of methods enumerated above, global descriptors are normally very fast to compute, favouring the matching process between the images and reducing the computational needs of mapping and localization tasks. As a main disadvantages, they offer less robustness to occlusion and illumination effects, what results in a lower discriminative power and

Table 8: Advantages and disadvantages of each method.

Feature	Global Descriptors	Local Features	BoW Schemes
CPU Needs	*	* * *	**
Storage Needs	*	* * *	**
Matching Complexity	**	* * *	*
Discrimination Power	*	* * *	**
Perceptual Aliasing Effect	* * *	*	**
Large-Scale Operation	**	*	* * *
Spatial Loss Information	**	*	* * *
Pose Recovery Complexity	* * *	*	**

an increment of the perceptual aliasing effect, where different places can be perceived as the same. They have been used intensively in other related research areas, such as scene categorization.

Local features are usually more robust to occlusions and changes in scale, rotation and illumination. These methods start with a detection phase, where interest points are found in the image, and are followed by a description phase, where some measures are extracted from the surroundings of these keypoints. Local features present a better discrimination capacity, resulting into higher recognition rates and less detection errors. Furthermore, the recovery of relative poses between images, which can be used for confirming if two images come from the same scene, can be performed easily. However, the storage requirements and the computational cost are higher than for global descriptors and the matching process is also more complex, since sometimes each query descriptor requires to find their closest neighbour within a large set of features. According to the surveyed works, the most used feature is SIFT, followed by SURF, both representing features as vectors of floating point numbers. Recently, a number of binary descriptors have been proposed in the literature, providing an interesting research line to explore regarding topological mapping and localization, because they are cheaper to compute, compact to store and faster to compare.

While global descriptors and local features demonstrate useful approaches for robot mapping and localization, they do not result to be satisfactory when the number of images to process is high. Matching hundreds of images using local features can take a long time when trying to associate the current frame with every previously seen location. Indexing structures can be used to accelerate the search. However, with a high number of descriptors, memory problems and computational bottlenecks appear. Global descriptors are easier to compute and save storage space, but sacrificing

Table 9: Advantages and disadvantages of methods for generating visual dictionaries in BoW schemes.

Feature	Offline	Online
Training Phase	Yes	No
Scenario Specific	No	Yes
Incremental Memory	No	Yes
Feature Management	No	Yes
High Sizes	Yes	No

discriminative power which reduces the performance of the solution. In this case, an alternative approach for describing and matching images is the Bag-Of-Words (BoW) algorithm, which can efficiently index a huge amount of images incorporating a hierarchical scheme and an inverted index structure. Due to this fast image retrieval, works classified in this category are mainly SLAM approaches. As main limitation, it can be mentioned the fact that the effect of perceptual aliasing worsens due to the quantization process, the presence of noisy words due to the coarseness of the vocabulary construction method and the loss of the spatial relations between the words. Some authors have proposed several improvements in order to overcome this last drawback [186, 144].

The visual dictionaries can be generated offline or online. As a main shortcoming, the offline approaches need a training phase, where sometimes millions of descriptors have to be clustered. This can take hours, depending on the number of images and the clustering technique used. Furthermore, the robot can operate in an environment with an appearance totally different to the training set employed for generating the dictionary, which implies that it is not representative of the scenario, augmenting false detections. An alternative is to build the codebook online in an incremental manner, while the robot is navigating across the environment. However, this implies inserting and deleting features to/from the dictionary, limiting its possible size. An interesting study about the reuse of visual dictionaries and their universality is presented by Hou et al. [187]. Nowadays, despite several approaches have been proposed, managing efficiently online visual dictionaries for BoW schemes can be considered as a topic of interest. Another interesting issue is long-term mapping, in order to manage maps during long periods of time under changes in the appearance of the environment. The main advantages and limitations of each dictionary-generation approach are summarized in Table 9.

Although there is no consensus on how to evaluate the performance of the different approaches, a number of datasets have been made public for algorithm benchmarking purposes. Some of them are enumerated below.

Datasets

- RawSeed:

  http://www.rawseeds.org/home/

- Lip6:

  http://cogrob.ensta-paristech.fr/loopclosure.html

- Oxford:

  http://www.robots.ox.ac.uk/~mobile/IJRR_2008_Dataset/

  http://www.robots.ox.ac.uk/NewCollegeData/

- Crowded Canteen:

  http://haselab.info/papers/crowded_canteen_dataset_31-05-2011.zip

- University of Sherbrooke:

  https://introlab.3it.usherbrooke.ca/mediawiki-introlab/index.php/RTAB-Map

- University of Alberta:

  http://webdocs.cs.ualberta.ca/~hajebi/datasets/

- Radish:

  http://radish.sourceforge.net/

- COLD COsy Localization Database:

  http://www.cas.kth.se/COLD/

- KTH-IDOL:

  http://www.cas.kth.se/IDOL/

- LIBVISO2:

  http://www.cvlibs.net/datasets/karlsruhe_sequences/

- KITTI Dataset:

  http://www.cvlibs.net/datasets/kitti/

- St. Lucia:

  https://wiki.qut.edu.au/display/cyphy/UQ+St+Lucia

- Ford Campus:

  http://robots.engin.umich.edu/SoftwareData/Ford

- Malaga Parking Dataset:

  http://www.mrpt.org/downloads/dataset2009/

- Malaga Urban Dataset:

  http://www.mrpt.org/MalagaUrbanDataset

- Omni Zaragoza:

  http://robots.unizar.es/omnicam/

References

[1] H. Durrant-Whyte and T. Bailey, "Simultaneous localisation and mapping (slam): Part i the essential algorithms," *IEEE Robotics and Automation Magazine*, vol. 2, pp. 99–110, 2006.

[2] F. Bonin-Font, A. Ortiz, and G. Oliver, "Visual navigation for mobile robots: A survey," *Journal of Intelligent and Robotic Systems*, vol. 53, no. 3, pp. 263–296, 2008.

[3] J. Fuentes-Pacheco, J. Ruiz-Ascencio, and J. M. Rendón-Mancha, "Visual Simultaneous Localization and Mapping: A Survey," *Artificial Intelligence Review*, 2012.

[4] B. Kuipers and Y.-T. Byun, "A Robot Exploration and Mapping Strategy Based on a Semantic Hierarchy of Spatial Representations," *Robotics and Autonomous Systems*, vol. 8, pp. 47–63, 1991.

[5] H. Choset and K. Nagatani, "Topological Simultaneous Localization and Mapping (SLAM): Toward Exact Localization Without Explicit Localization," *IEEE Transactions on Robotics and Automation*, vol. 17, no. 2, pp. 125–137, 2001.

[6] E. Remolina and B. Kuipers, "Towards a General Theory of Topological Maps," *Artificial Intelligence*, vol. 152, no. 1, pp. 47–104, 2004.

[7] J. Wu, H. Christensen, and J. Rehg, "Visual Place Categorization: Problem, Dataset, and Algorithm," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 4763–4770, 2009.

[8] J. Wu and J. M. Rehg, "CENTRIST: A Visual Descriptor for Scene Categorization," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 8, pp. 1489–1501, 2011.

[9] A. Bosch, A. Zisserman, and X. Munoz, "Representing Shape with a Spatial Pyramid Kernel," *Image Processing*, vol. 5, no. 2, pp. 401–408, 2007.

[10] E. Fazl-Ersi and J. K. Tsotsos, "Histogram of Oriented Uniform Patterns for Robust Place Recognition and Categorization," *International Journal of Robotics Research*, vol. 31, no. 4, pp. 468–483, 2012.

[11] L. Zhou, Z. Zhou, and D. Hu, "Scene Classification using a Multi-Resolution Bag-of-Features Model," *Pattern Recognition*, vol. 46, no. 1, pp. 424–433, 2013.

[12] N. Dalal and B. Triggs, "Histograms of Oriented Gradients for Human Detection," in *International Conference on Computer Vision and Pattern Recognition*, pp. 886–893, 2005.

[13] N. Winters, J. Gaspar, G. Lacey, and J. Santos-Victor, "Omni-directional Vision for Robot Navigation," in *IEEE Workshop on Omnidirectional Vision*, pp. 21–28, 2000.

[14] J. Gaspar, N. Winters, and J. Santos-Victor, "Vision-Based Navigation and Environmental Representations with an Omnidirectional Camera," *IEEE Transactions on Robotics and Automation*, vol. 16, no. 6, pp. 890–898, 2000.

[15] I. Ulrich and I. Nourbakhsh, "Appearance-Based Place Recognition for Topological Localization," in *IEEE International Conference on Robotics and Automation*, vol. 2, pp. 1023–1029, 2000.

[16] J. Kosecka, L. Zhou, P. Barber, and Z. Duric, "Qualitative Image Based Localization in Indoors Environments," in *International Conference on Computer Vision and Pattern Recognition*, vol. 2, pp. II–3–II–8, 2003.

[17] D. Bradley, R. Patel, N. Vandapel, and S. Thayer, "Real-Time Image-Based Topological Localization in Large Outdoor Environments," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 3670–3677, 2005.

[18] C. Weiss and A. Masselli, "Fast Outdoor Robot Localization using Integral Invariants," in *International Conference on Computer Vision*, 2007.

[19] J. Wang, H. Zha, and R. Cipolla, "Efficient Topological Localization Using Orientation Adjacency Coherence Histograms," in *International Conference on Pattern Recognition*, pp. 271–274, 2006.

[20] A. Pronobis, B. Caputo, P. Jensfelt, and H. Christensen, "A Discriminative Approach to Robust Visual Place Recognition," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 3829–3836, 2006.

[21] A. Oliva and A. Torralba, "Modeling the Shape of the Scene : A Holistic Representation of the Spatial Envelope," *International Journal of Computer Vision*, vol. 42, no. 3, pp. 145–175, 2001.

[22] A. C. Murillo, P. Campos, J. Kosecka, and J. Guerrero, "Gist Vocabularies in Omnidirectional Images for Appearance Based Mapping and Localization," in *IEEE Workshop on Omnidirectional Vision, Camera Networks and Non-classical Cameras*, 2010.

[23] N. Sunderhauf and P. Protzel, "BRIEF-Gist - Closing the Loop by Simple Means," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 1234–1241, 2011.

[24] A. Chapoulie, P. Rives, and D. Filliat, "Appearance-Based Segmentation of Indoors and Outdoors Sequences of Spherical Views," in *IEEE International Conference on Robotics and Automation*, pp. 1946–1951, 2013.

[25] P. Lamon, I. Nourbakhsh, B. Jensen, and R. Siegwart, "Deriving and Matching Image Fingerprint Sequences for Mobile Robot Localization," in *IEEE International Conference on Robotics and Automation*, vol. 2, pp. 1609–1614, 2001.

[26] M. Liu, D. Scaramuzza, C. Pradalier, R. Siegwart, and Q. Chen, "Scene Recognition with Omnidirectional Vision for Topological Map using Lightweight Adaptive Descriptors," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 116–121, 2009.

[27] M. Liu and R. Siegwart, "DP-FACT: Towards Topological Mapping and Scene Recognition With Color for Omnidirectional Camera," in *IEEE International Conference on Robotics and Automation*, pp. 3503–3508, 2012.

[28] E. Menegatti, T. Maeda, and H. Ishiguro, "Image-Based Memory for Robot Navigation using Properties of Omnidirectional Images," *Robotics and Autonomous Systems*, vol. 47, no. 4, pp. 251–267, 2004.

[29] E. Menegatti, M. Zoccarato, E. Pagello, and H. Ishiguro, "Image-Based Monte Carlo Localisation with Omnidirectional Images," *Robotics and Autonomous Systems*, vol. 48, no. 1, pp. 17–30, 2004.

[30] D. Prasser and G. Wyeth, "Probabilistic Visual Recognition of Artificial Landmarks for Simultaneous Localization and Mapping," in *IEEE International Conference on Robotics and Automation*, vol. 1, pp. 1291–1296, 2003.

[31] M. Milford and G. Wyeth, "Mapping a Suburb With a Single Camera Using a Biologically Inspired SLAM System," *IEEE Transactions on Robotics*, vol. 24, no. 5, pp. 1038–1053, 2008.

[32] M. Milford and G. Wyeth, "SeqSLAM: Visual Route-Based Navigation for Sunny Summer Days and Stormy Winter Nights," in *IEEE International Conference on Robotics and Automation*, pp. 1643–1649, 2012.

[33] W. L. D. Lui and R. Jarvis, "A Pure Vision-Based Approach to Topological SLAM," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 3784–3791, 2010.

[34] W. L. D. Lui and R. Jarvis, "A Pure Vision-based Topological SLAM System," *International Journal of Robotics Research*, vol. 31, no. 4, pp. 403–428, 2012.

[35] H. Badino, D. Huber, and T. Kanade, "Real-Time Topometric Localization," in *IEEE International Conference on Robotics and Automation*, pp. 1635–1642, 2012.

[36] H. Lategahn, J. Beck, B. Kitt, and C. Stiller, "How to Learn an Illumination Robust Image Feature for Place Recognition," in *IEEE Intelligent Vehicles Symposium*, 2013.

[37] N. Nourani-Vatani, P. Borges, J. Roberts, and M. Srinivasan, "On the Use of Optical Flow for Scene Change Detection and Description," *Journal of Intelligent and Robotic Systems*, 2013.

[38] F. Werner, F. Maire, and J. Sitte, "Topological SLAM using Fast Vision Techniques," in *Advances in Robotics*, pp. 187–196, 2009.

[39] G. Singh and J. Kosecka, "Visual Loop Closing using Gist Descriptors in Manhattan World," in *Workshop on Omnidirectional Robot Vision*, 2010.

[40] A. Rituerto, A. C. Murillo, and J. Guerrero, "Semantic Labeling for Indoor Topological Mapping using a Wearable Catadioptric System," *Robotics and Autonomous Systems*, 2013.

[41] Y. Liu and H. Zhang, "Visual Loop Closure Detection with a Compact Image Descriptor," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 1051–1056, 2012.

[42] A. Chapoulie, P. Rives, and D. Filliat, "Topological Segmentation of Indoors/Outdoors Sequences of Spherical Views," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 4288–4295, 2012.

[43] A. Tapus, N. Tomatis, and R. Siegwart, "Topological Global Localization and Mapping with Fingerprints and Uncertainty," in *International Symposium on Experimental Robotics*, pp. 18–21, 2004.

[44] A. Tapus and R. Siegwart, "Incremental Robot Mapping with Fingerprints of Places," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 2429–2434, 2005.

[45] L. Payá, L. Fernández, A. Gil, and O. Reinoso, "Map Building and Monte Carlo Localization using Global Appearance of Omnidirectional Images.," *Sensors*, vol. 10, no. 12, pp. 11468–97, 2010.

[46] A. Ranganathan, E. Menegatti, and F. Dellaert, "Bayesian Inference in the Space of Topological Maps," *IEEE Transactions on Robotics*, vol. 22, no. 1, pp. 92–107, 2006.

[47] M. Milford, G. Wyeth, and D. Prasser, "RatSLAM: A Hippocampal Model for Simultaneous Localization and Mapping," in *IEEE International Conference on Robotics and Automation*, pp. 403–408, 2004.

[48] D. Prasser, M. Milford, and G. Wyeth, "Outdoor Simultaneous Localisation and Mapping using RatSLAM," in *International Conference on Field and Service Robots*, 2005.

[49] A. Glover, W. Maddern, M. Milford, and G. Wyeth, "FAB-MAP + RatSLAM: Appearance-based SLAM for Multiple Times of Day," in *IEEE International Conference on Robotics and Automation*, pp. 3507–3512, 2010.

[50] M. Milford, "Visual Route Recognition with a Handful of Bits," in *Robotics: Systems and Science*, 2013.

[51] C. Siagian and L. Itti, "Rapid Biologically-Inspired Scene Classification using Features Shared with Visual Attention," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 2, pp. 300–12, 2007.

[52] M. Calonder, V. Lepetit, C. Strecha, and P. Fua, "BRIEF : Binary Robust Independent Elementary Features," in *European Conference on Computer Vision*, vol. 6314 of *Lecture Notes in Computer Science*, pp. 778–792, 2010.

[53] M. Agrawal, K. Konolige, and M. R. Blas, "CenSurE: Center Surround Extremas for Realtime Feature Detection and Matching," in *European Conference on Computer Vision*, vol. 5305, pp. 102–115, 2008.

[54] M. Milford, "Vision-Based Place Recognition: How Low Can You Go?," *International Journal of Robotics Research*, vol. 32, no. 7, pp. 766–789, 2013.

[55] T. Tuytelaars and K. Mikolajczyk, "Local Invariant Feature Detectors: A Survey," *Foundations and Trends® in Computer Graphics and Vision*, vol. 3, no. 3, pp. 177–280, 2007.

[56] A. Schmidt, M. Kraft, and A. Kasinski, "An Evaluation of Image Feature Detectors and Descriptors for Robot Navigation," in *International Conference on Computer Vision and Graphics*, vol. 6375 of *Lecture Notes in Computer Science*, pp. 251–259, 2010.

[57] O. Miksik and K. Mikolajczyk, "Evaluation of Local Detectors and Descriptors for Fast Feature Matching," in *International Conference on Pattern Recognition*, pp. 2681–2684, 2012.

[58] K. Mikolajczyk and C. Schmid, "A Performance Evaluation of Local Descriptors," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 10, pp. 1615–1630, 2005.

[59] C. Harris and M. Stephens, "A Combined Corner and Edge Detector," in *Alvey Vision Conference*, pp. 147–151, 1988.

[60] J. Shi and C. Tomasi, "Good Features to Track," in *International Conference on Computer Vision and Pattern Recognition*, pp. 593–600, 1994.

[61] S. Smith and M. Brady, "SUSAN — A New Approach to Low Level Image Processing," *International Journal of Computer Vision*, vol. 23, no. 1, pp. 45–78, 1997.

[62] E. Rosten and T. Drummond, "Machine Learning for High-Speed Corner Detection," in *European Conference on Computer Vision*, no. 1 in Lecture Notes in Computer Science, pp. 430–443, 2006.

[63] E. Rosten, R. Porter, and T. Drummond, "Faster and Better: A Machine Learning Approach to Corner Detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, pp. 105–19, Jan. 2010.

[64] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "ORB: An Efficient Alternative to SIFT or SURF," in *International Conference on Computer Vision*, vol. 95, pp. 2564–2571, 2011.

[65] E. Mair, G. D. Hager, D. Burschka, M. Suppa, and G. Hirzinger, "Adaptive and Generic Corner Detection Based on the Accelerated Segment Test," in *European Conference on Computer Vision*, vol. 6312 of *Lecture Notes in Computer Science*, pp. 183–196, 2010.

[66] S. Leutenegger, M. Chli, and R. Siegwart, "BRISK: Binary Robust Invariant Scalable Keypoints," in *International Conference on Computer Vision*, pp. 2548–2555, 2011.

[67] D. G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[68] H. Bay, T. Tuytelaars, and L. Van Gool, "SURF: Speeded Up Robust Features," in *European Conference on Computer Vision*, vol. 3951 of *Lecture Notes in Computer Science*, pp. 404–417, 2006.

[69] K. Konolige, J. Bowman, J. Chen, P. Mihelich, M. Calonder, V. Lepetit, and P. Fua, "View-Based Maps," *International Journal of Robotics Research*, vol. 29, no. 8, pp. 941–957, 2010.

[70] M. Ebrahimi and W. Mayol-Cuevas, "SUSurE: Speeded Up Surround Extrema Feature Detector and Descriptor for Realtime Applications," in *International Conference on Computer Vision and Pattern Recognition*, pp. 9–14, 2009.

[71] P. F. Alcantarilla, A. Bartoli, and A. J. Davison, "KAZE Features," in *European Conference on Computer Vision*, pp. 214–227, 2012.

[72] P. F. Alcantarilla, J. Nuevo, and A. Bartoli, "Fast Explicit Diffusion for Accelerated Features in Nonlinear Scale Spaces," in *British Machine Vision Conference*, 2013.

[73] J.-M. Morel and G. Yu, "ASIFT: A New Framework for Fully Affine Invariant Image Comparison," *SIAM Journal on Imaging Sciences*, vol. 2, no. 2, pp. 438–469, 2009.

[74] J. Matas, O. Chum, M. Urban, and T. Pajdla, "Robust Wide Baseline Stereo from Maximally Stable Extremal Regions," in *British Machine Vision Conference*, pp. 1–10, 2002.

[75] Y. Ke and R. Sukthankar, "PCA-SIFT: A More Distinctive Representation for Local Image Descriptors," in *International Conference on Computer Vision and Pattern Recognition*, pp. 506–513, 2004.

[76] H. Andreasson and T. Duckett, "Topological Localization for Mobile Robots using Omnidirectional Vision and Local Features," in *IFAC Intelligent Autonomous Vehicles Symposium*, 2008.

[77] E. Tola, V. Lepetit, and P. Fua, "DAISY: An Efficient Dense Descriptor Applied to Wide Baseline Stereo," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 5, pp. 815–830, 2010.

[78] M. S. Sarfraz and O. Hellwich, "Head Pose Estimation in Face Recognition Across Pose Scenarios," in *International Conference on Computer Vision Theory and Applications*, pp. 235–242, 2008.

[79] A. Alahi, R. Ortiz, and P. Vandergheynst, "FREAK : Fast Retina Keypoint," in *International Conference on Computer Vision and Pattern Recognition*, pp. 510–517, 2012.

[80] T. Trzcinski and V. Lepetit, "Efficient Discriminative Projections for Compact Binary Descriptors," in *European Conference on Computer Vision*, vol. 7572 of *Lecture Notes in Computer Science*, pp. 228–242, 2012.

[81] C. Strecha, A. M. Bronstein, M. M. Bronstein, and P. Fua, "LDAHash: Improved Matching with Smaller Descriptors," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 1, 2012.

[82] T. Trzcinski, C. Christoudias, P. Fua, and V. Lepetit, "Boosting Binary Keypoint Descriptors," in *International Conference on Computer Vision and Pattern Recognition*, pp. 2874–2881, 2013.

[83] X. Yang and K.-T. Cheng, "Local Difference Binary for Ultrafast and Distinctive Feature Description," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 1, pp. 188–94, 2014.

[84] J. Kosecka and X. Yang, "Location Recognition and Global Localization Based on Scale-Invariant Keypoints," in *European Conference on Computer Vision*, 2004.

[85] J. Kosecka and F. Li, "Vision Based Topological Markov Localization," in *IEEE International Conference on Robotics and Automation*, vol. 2, pp. 1481–1486, 2004.

[86] F. Li and J. Kosecka, "Probabilistic Location Recognition Using Reduced Feature Set," in *IEEE International Conference on Robotics and Automation*, pp. 3405–3410, 2006.

[87] H. Zhang, "BoRF: Loop-Closure Detection with Scale Invariant Visual Features," in *IEEE International Conference on Robotics and Automation*, pp. 3125–3130, 2011.

[88] H. Zhang, "Indexing Visual Features: Real-Time Loop Closure Detection Using a Tree Structure," in *IEEE International Conference on Robotics and Automation*, pp. 3613–3618, 2012.

[89] P. Rybski, F. Zacharias, J.-F. Lett, O. Masoud, M. Gini, and N. Papanikolopoulos, "Using Visual Features to Build Topological Maps of Indoor Environments," in *IEEE International Conference on Robotics and Automation*, vol. 1, pp. 850–855, 2003.

[90] X. He, R. Zemel, and V. Mnih, "Topological Map Learning from Outdoor Image Sequences," *Journal of Field Robotics*, vol. 23, no. 11-12, pp. 1091–1104, 2006.

[91] D. G. Sabatta, "Vision-Based Topological Map Building and Localisation using Persistent Features," in *Robotics and Mechatronics Symposium*, pp. 1–6, 2008.

[92] E. Johns and G.-Z. Yang, "Global Localization in a Dense Continuous Topological Map," in *IEEE International Conference on Robotics and Automation*, pp. 1032–1037, 2011.

[93] A. Kawewong, S. Tangruamsub, and O. Hasegawa, "Position-Invariant Robust Features for Long-Term Recognition of Dynamic Outdoor Scenes," *IEICE Transactions on Information and Systems*, vol. E93-D, no. 9, pp. 2587–2601, 2010.

[94] A. Kawewong, N. Tongprasit, S. Tangruamsub, and O. Hasegawa, "Online and Incremental Appearance-based SLAM in Highly Dynamic Environments," *International Journal of Robotics Research*, vol. 30, no. 1, pp. 33–55, 2011.

[95] N. Tongprasit, A. Kawewong, and O. Hasegawa, "PIRF-Nav 2: Speeded-Up Online and Incremental Appearance-Based SLAM in an Indoor Environment," in *IEEE Workshop on Applications of Computer Vision*, pp. 145–152, 2011.

[96] H. Morioka, S. Yi, and O. Hasegawa, "Vision-Based Mobile Robot's SLAM and Navigation in Crowded Environments," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 3998–4005, 2011.

[97] C. Valgren, A. Lilienthal, and T. Duckett, "Incremental Topological Mapping Using Omnidirectional Vision," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 3441–3447, 2006.

[98] C. Valgren, T. Duckett, and A. Lilienthal, "Incremental Spectral Clustering and Its Application to Topological Mapping," in *IEEE International Conference on Robotics and Automation*, pp. 10–14, 2007.

[99] C. Valgren and A. Lilienthal, "SIFT, SURF and Seasons: Long-term Outdoor Localization Using Local Features," in *European Conference on Mobile Robotics*, vol. 128, pp. 1–6, 2007.

[100] A. Ascani, E. Frontoni, A. Mancini, and P. Zingaretti, "Feature Group Matching for Appearance-Based Localization," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 3933–3938, 2008.

[101] R. Anati and K. Daniilidis, "Constructing Topological Maps using Markov Random Fields and Loop-Closure Detection," in *Advances in Neural Information Processing Systems*, pp. 37–45, 2009.

[102] Z. Zivkovic, B. Bakker, and B. Krose, "Hierarchical Map Building Using Visual Landmarks and Geometric Constraints," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 2480–2485, IEEE/RSJ, 2005.

[103] O. Booij, B. Terwijn, Z. Zivkovic, and B. Krose, "Navigation Using an Appearance Based Topological Map," in *IEEE International Conference on Robotics and Automation*, pp. 3927–3932, 2007.

[104] O. Booij, Z. Zivkovic, and B. Krose, "Efficient Data Association for View Based SLAM using Connected Dominating Sets," *Robotics and Autonomous Systems*, vol. 57, no. 12, pp. 1225–1234, 2009.

[105] F. Dayoub, G. Cielniak, and T. Duckett, "A Sparse Hybrid Map for Vision-Guided Mobile Robots," in *European Conference on Mobile Robotics*, pp. 213–218, 2011.

[106] J. L. Blanco, J. A. Fernandez-Madrigal, and J. Gonzalez, "Towards a Unified Bayesian Approach to Hybrid Metric-Topological SLAM," *IEEE Transactions on Robotics*, vol. 24, no. 2, pp. 259–270, 2008.

[107] J. L. Blanco, J. Gonzalez, and J. A. Fernandez-Madrigal, "Subjective Local Maps for Hybrid Metric-Topological SLAM," *Robotics and Autonomous Systems*, vol. 57, no. 1, pp. 64–74, 2009.

[108] S. Tully, H. Moon, D. Morales, G. Kantor, and H. Choset, "Hybrid Localization using The Hierarchical Atlas," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 2857–2864, 2007.

[109] S. Tully, G. Kantor, H. Choset, and F. Werner, "A Multi-Hypothesis Topological SLAM Approach for Loop Closing on Edge-Ordered Graphs," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 4943–4948, 2009.

[110] S. Segvic, A. Remazeilles, A. Diosi, and F. Chaumette, "A Mapping and Localization Framework for Scalable Appearance-Based Navigation," *Computer Vision and Image Understanding*, vol. 113, no. 2, pp. 172–187, 2009.

[111] A. Ramisa, A. Tapus, D. Aldavert, R. Toledo, and R. Lopez de Mantaras, "Robust Vision-Based Robot Localization using Combinations of Local Feature Region Detectors," *Autonomous Robots*, vol. 27, no. 4, pp. 373–385, 2009.

[112] H. Badino, D. Huber, and T. Kanade, "Visual Topometric Localization," in *IEEE Intelligent Vehicles Symposium*, pp. 794–799, 2011.

[113] F. Dayoub and T. Duckett, "An Adaptive Appearance-Based Map for Long-Term Topological Localization of Mobile Robots," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 3364 – 3369, 2008.

[114] B. Bacca, J. Salvi, J. Batlle, and X. Cufi, "Appearance-Based Mapping and Localisation Using Feature Stability Histograms," *Electronics Letters*, vol. 46, no. 16, p. 1120, 2010.

[115] B. Bacca, J. Salvi, and X. Cufi, "Appearance-Based Mapping and Localization for Mobile Robots using a Feature Stability Histogram," *Robotics and Autonomous Systems*, vol. 59, no. 10, pp. 840–857, 2011.

[116] B. Bacca, J. Salvi, and X. Cufi, "Long-term mapping and localization using feature stability histograms," *Robotics and Autonomous Systems*, vol. 61, no. 12, pp. 1539–1558, 2013.

[117] A. Romero and M. Cazorla, "Topological SLAM Using Omnidirectional Images: Merging Feature Detectors and Graph-Matching," in *Advanced Concepts for Intelligent Vision Systems*, vol. 6474 of *Lecture Notes in Computer Science*, pp. 464–475, 2010.

[118] A. Romero and M. Cazorla, "Topological Visual Mapping in Robotics," *Cognitive Processing*, vol. 13, no. 1, pp. 305–308, 2012.

[119] A. Majdik, Y. Albers-Schoenberg, and D. Scaramuzza, "MAV Urban Localization from Google Street View Data," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2013.

[120] M. Saedan, C. W. Lim, and M. Ang, "Appearance-Based SLAM with Map Loop Closing using an Omnidirectional Camera," in *International Conference on Advanced Intelligent Mechatronics*, pp. 1–6, 2007.

[121] J. Kessler, A. König, and H.-M. Gross, "An Improved Sensor Model on Appearance Based SLAM," in *Autonome Mobile Systeme*, vol. 216487, pp. 153–160, 2009.

[122] L. Maohai, W. Han, S. Lining, and C. Zesu, "Robust Omnidirectional Mobile Robot Topological Navigation System using Omnidirectional Vision," *Engineering Applications of Artificial Intelligence*, vol. 26, no. 8, pp. 1942–1952, 2013.

[123] E. Garcia-Fidalgo and A. Ortiz, "Probabilistic Appearance-Based Mapping and Localization Using Visual Features," in *Iberian Conference on Pattern Recognition and Image Analysis*, (Funchal (Portugal)), pp. 277–285, 2013.

[124] E. Garcia-Fidalgo and A. Ortiz, "Vision-Based Topological Mapping and Localization by means of Local Invariant Features and Map Refinement," *Robotica*, 4 2014.

[125] H. Zhang, B. Li, and D. Yang, "Keyframe Detection for Appearance-Based Visual SLAM," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 2071–2076, 2010.

[126] B. Lisien, D. Morales, D. Silver, G. Kantor, I. M. Rekleitis, and H. Choset, "The Hierarchical Atlas," *IEEE Transactions on Robotics*, vol. 21, no. 3, pp. 473–481, 2005.

[127] S. Tully, G. Kantor, and H. Choset, "A Unified Bayesian Framework for Global Localization and SLAM in Hybrid Metric/Topological Maps," *International Journal of Robotics Research*, vol. 31, no. 3, pp. 271—-288, 2012.

[128] R. C. Atkinson and R. M. Shiffrin, "Human Memory: A Proposed System and Its Control Processes," *The Psychology of Learning and Motivation: Advances in Research and Theory*, vol. 2, pp. 89–105, 1968.

[129] J. Sivic and A. Zisserman, "Video Google: A Text Retrieval Approach to Object Matching in Videos," in *International Conference on Computer Vision*, pp. 1470–1477, 2003.

[130] J. Wang, R. Cipolla, and H. Zha, "Vision-Based Global Localization using a Visual Vocabulary," in *IEEE International Conference on Robotics and Automation*, pp. 4230–4235, 2005.

[131] J. Wang, H. Zha, and R. Cipolla, "Coarse-to-Fine Vision-Based Localization by Indexing Scale-Invariant Features," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 36, no. 2, pp. 413–422, 2006.

[132] F. Fraundorfer, C. Engels, and D. Nister, "Topological Mapping, Localization and Navigation Using Image Collections," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 3872–3877, 2007.

[133] M. Cummins and P. Newman, "Probabilistic Appearance Based Navigation and Loop Closing," in *IEEE International Conference on Robotics and Automation*, pp. 2042–2048, 2007.

[134] M. Cummins and P. Newman, "FAB-MAP: Probabilistic Localization and Mapping in the Space of Appearance," *International Journal of Robotics Research*, vol. 27, no. 6, pp. 647–665, 2008.

[135] M. Cummins and P. Newman, "Accelerated Appearance-Only SLAM," in *IEEE International Conference on Robotics and Automation*, pp. 1828–1833, 2008.

[136] M. Cummins and P. Newman, "Accelerating FAB-MAP With Concentration Inequalities," *IEEE Transactions on Robotics*, vol. 26, no. 6, pp. 1042–1050, 2010.

[137] M. Cummins and P. Newman, "Highly Scalable Appearance-Only SLAM - FAB-MAP 2.0.," in *Robotics: Systems and Science*, 2009.

[138] M. Cummins and P. Newman, "Appearance-Only SLAM at Large Scale with FAB-MAP 2.0," *International Journal of Robotics Research*, vol. 30, no. 9, pp. 1100–1123, 2011.

[139] P. Newman, G. Sibley, M. Smith, M. Cummins, A. Harrison, C. Mei, I. Posner, R. Shade, D. Schroeter, L. Murphy, W. Churchill, D. Cole, and I. Reid, "Navigating, Recognizing and Describing Urban Spaces With Vision and Lasers," *International Journal of Robotics Research*, vol. 28, pp. 1406–1433, July 2009.

[140] W. Maddern, M. Milford, and G. Wyeth, "Continuous Appearance-Based Trajectory SLAM," in *IEEE International Conference on Robotics and Automation*, pp. 3595–3600, 2011.

[141] W. Maddern, M. Milford, and G. Wyeth, "Cat-slam: Probabilistic localisation and mapping using a continuous appearance-based trajectory," *International Journal of Robotics Research*, vol. 31, no. 4, pp. 429–451, 2012.

[142] W. Maddern, M. Milford, and G. Wyeth, "Towards Persistent Indoor Appearance-based Localization, Mapping and Navigation using CAT-Graph," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 4224–4230, 2012.

[143] R. Paul and P. Newman, "FAB-MAP 3D: Topological Mapping with Spatial and Visual Appearance," in *IEEE International Conference on Robotics and Automation*, pp. 2649–2656, 2010.

[144] E. Johns and G.-Z. Yang, "Feature Co-occurrence Maps: Appearance-based Localisation Throughout the Day," in *IEEE International Conference on Robotics and Automation*, pp. 3212–3218, 2013.

[145] E. Johns and G.-Z. Yang, "Dynamic Scene Models for Incremental, Long Term, Appearance Based Localisation," in *IEEE International Conference on Robotics and Automation*, pp. 2731–2736, 2013.

[146] D. Galvez-Lopez and J. Tardos, "Real-Time Loop Detection with Bags of Binary Words," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 51–58, 2011.

[147] D. Galvez-Lopez and J. Tardos, "Bags of Binary Words for Fast Place Recognition in Image Sequences," *IEEE Transactions on Robotics*, vol. 28, no. 5, pp. 1188–1197, 2012.

[148] A. Ranganathan and F. Dellaert, "Online Probabilistic Topological Mapping," *International Journal of Robotics Research*, vol. 30, no. 6, pp. 755–771, 2011.

[149] C. Cadena, D. Galvez-Lopez, F. Ramos, J. Tardos, and J. Neira, "Robust Place Recognition with Stereo Cameras," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 5182–5189, 2010.

[150] T. Ciarfuglia, G. Costante, P. Valigi, and E. Ricci, "A Discriminative Approach for Appearance Based Loop Closing," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 3837–3843, 2012.

[151] A. Majdik, D. Galvez-Lopez, G. Lazea, and J. Castellanos, "Adaptive Appearance Based Loop-Closing in Heterogeneous Environments," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 1256–1263, 2011.

[152] G. Schindler, M. Brown, and R. Szeliski, "City-Scale Location Recognition," in *International Conference on Computer Vision and Pattern Recognition*, pp. 1–7, 2007.

[153] S. Achar, C. Jawahar, and K. Madhava Krishna, "Large Scale Visual Localization in Urban Environments," in *IEEE International Conference on Robotics and Automation*, pp. 5642–5648, 2011.

[154] J. H. Lee, G. Zhang, J. Lim, and I. H. Suh, "Place Recognition using Straight Lines for Vision-Based SLAM," in *IEEE International Conference on Robotics and Automation*, pp. 3799–3806, 2013.

[155] D. Filliat, "A Visual Bag of Words Method for Interactive Qualitative Localization and Mapping," in *IEEE International Conference on Robotics and Automation*, pp. 3921–3926, 2007.

[156] A. Angeli, S. Doncieux, J.-A. Meyer, and D. Filliat, "Real-Time Visual Loop-Closure Detection," in *IEEE International Conference on Robotics and Automation*, pp. 1842–1847, 2008.

[157] 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 Transactions on Robotics*, vol. 24, no. 5, pp. 1027–1037, 2008.

[158] A. Angeli, S. Doncieux, J.-A. Meyer, and D. Filliat, "Incremental Vision-Based Topological SLAM," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 1031–1036, 2008.

[159] M. Labbe and F. Michaud, "Memory Management for Real-Time Appearance-Based Loop Closure Detection," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 1271–1276, 2011.

[160] M. Labbe and F. Michaud, "Appearance-Based Loop Closure Detection for Online Large-Scale and Long-Term Operation," *IEEE Transactions on Robotics*, vol. 29, no. 3, pp. 734–745, 2013.

[161] T. Nicosevici and R. Garcia, "On-line Visual Vocabularies for Robot Navigation and Mapping," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 205–212, 2009.

[162] T. Nicosevici and R. Garcia, "Automatic Visual Bag-of-Words for Online Robot Navigation and Mapping," *IEEE Transactions on Robotics*, vol. 28, no. 4, pp. 886–898, 2012.

[163] G. Csurka, C. Dance, L. Fan, J. Willamowski, and C. Bray, "Visual Categorization with Bags of Keypoints," in *European Conference on Computer Vision*, vol. 1, pp. 1–22, 2004.

[164] F.-F. Li and P. Perona, "A Bayesian Hierarchical Model for Learning Natural Scene Categories," in *International Conference on Computer Vision and Pattern Recognition*, pp. 524–531, 2005.

[165] D. Nister and H. Stewenius, "Scalable Recognition with a Vocabulary Tree," in *International Conference on Computer Vision and Pattern Recognition*, vol. 2, pp. 2161–2168, 2006.

[166] A. Glover, W. Maddern, M. Warren, S. Reid, M. Milford, and G. Wyeth, "OpenFABMAP: An Open Source Toolbox for Appearance-based Loop Closure Detection," in *IEEE International Conference on Robotics and Automation*, pp. 4730 – 4735, 2012.

[167] A. Ranganathan and F. Dellaert, "A Rao-Blackwellized Particle Filter for Topological Mapping," in *IEEE International Conference on Robotics and Automation*, pp. 810–817, 2006.

[168] Y. Latif, C. Cadena, and J. Neira, "Realizing, Reversing, Recovering : Incremental Robust Loop Closing over Time Using the iRRR Algorithm," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 4211–4217, 2012.

[169] E. Eade and T. Drummond, "Unified Loop Closing and Recovery for Real Time Monocular SLAM," in *British Machine Vision Conference*, 2008.

[170] T. Botterill, S. Mills, and R. Green, "Bag-of-Words-Driven, Single-Camera Simultaneous Localization and Mapping," *Journal of Field Robotics*, vol. 28, no. 2, pp. 204–226, 2011.

[171] V. Pradeep, G. Medioni, and J. Weiland, "Visual Loop Closing using Multi-Resolution SIFT Grids in Metric-Topological SLAM," in *International Conference on Computer Vision and Pattern Recognition*, pp. 1438–1445, 2009.

[172] T. Goedemé, M. Nuttin, T. Tuytelaars, and L. Van Gool, "Markerless Computer Vision Based Localization using Automatically Generated Topological Maps," in *European Navigation Conference*, pp. 235–243, 2004.

[173] T. Goedemé, M. Nuttin, T. Tuytelaars, and L. Van Gool, "Omnidirectional Vision Based Topological Navigation," *International Journal of Computer Vision*, vol. 74, no. 3, pp. 219–236, 2007.

[174] A. C. Murillo, C. Sagues, and J. Guerrero, "From Omnidirectional Images to Hierarchical Localization," *Robotics and Autonomous Systems*, vol. 55, no. 5, pp. 372–382, 2007.

[175] A. C. Murillo, J. Guerrero, and C. Sagues, "SURF Features for Efficient Robot Localization with Omnidirectional Images," in *IEEE International Conference on Robotics and Automation*, pp. 3901–3907, 2007.

[176] J. Wang and Y. Yagi, "Efficient Topological Localization Using Global and Local Feature Matching," *International Journal of Advanced Robotic Systems*, 2013.

[177] C. Weiss, A. Masselli, and A. Zell, "Fast Vision-Based Localization for Outdoor Robots using a Combination of Global Image Features," in *IFAC Intelligent Autonomous Vehicles Symposium*, pp. 119–124, 2007.

[178] C. Weiss, H. Tamimi, A. Masselli, and A. Zell, "A Hybrid Approach for Vision-Based Outdoor Robot Localization using Global and Local Image Features," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 1047–1052, 2007.

[179] C. Siagian and L. Itti, "Biologically Inspired Mobile Robot Vision Localization," *IEEE Transactions on Robotics*, vol. 25, no. 4, pp. 861–873, 2009.

[180] A. Chapoulie, P. Rives, and D. Filliat, "A Spherical Representation for Efficient Visual Loop Closing," in *International Conference on Computer Vision*, pp. 335–342, 2011.

[181] M.-L. Wang and H.-Y. Lin, "A Hull Census Transform for Scene Change Detection and Recognition Towards Topological Map Building," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 548–553, 2010.

[182] H.-Y. Lin, Y.-H. Lin, and J.-W. Yao, "Scene Change Detection and Topological Map Construction Using Omnidirectional Image Sequences," in *International Conference on Machine Vision and Applications*, pp. 4–7, 2013.

[183] J. Wang and Y. Yagi, "Robust Location Recognition based on Efficient Feature Integration," in *IEEE International Conference on Robotics and Biomimetics*, pp. 97–101, 2012.

[184] H. Korrapati, F. Uzer, and Y. Mezouar, "Hierarchical Visual Mapping with Omnidirectional Images," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 3684–3690, 2013.

[185] H. Korrapati and Y. Mezouar, "Vision-Based Sparse Topological Mapping," *Robotics and Autonomous Systems*, 2014.

[186] S. Lazebnik, C. Schmid, and J. Ponce, "Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories," in *International Conference on Computer Vision and Pattern Recognition*, vol. 2, pp. 2169–2178, 2006.

[187] J. Hou, W.-X. Liu, X. E, Q. Xia, and N.-M. Qi, "An Experimental Study on the Universality of Visual Vocabularies," *Journal of Visual Communication and Image Representation*, vol. 24, no. 7, pp. 1204–1211, 2013.