Review Article

A State-of-the-Art Review on Mapping and Localization of Mobile Robots Using Omnidirectional Vision Sensors

L. Payá, A. Gil, and O. Reinoso

Universidad Miguel Hernández de Elche, Avda. de la Universidad s/n, Elche, Spain

Correspondence should be addressed to O. Reinoso; o.reinoso@goumh.umh.es

Nowadays, the field of mobile robotics is experiencing a quick evolution, and a variety of autonomous vehicles is available to solve different tasks. The advances in computer vision have led to a substantial increase in the use of cameras as the main sensors in mobile robots. They can be used as the only source of information or in combination with other sensors such as odometry or laser. Among vision systems, omnidirectional sensors stand out due to the richness of the information they provide the robot with, and an increasing number of works about them have been published over the last few years, leading to a wide variety of frameworks. In this review, some of the most important works are analysed. One of the key problems the scientific community is addressing currently is the improvement of the autonomy of mobile robots. To this end, building robust models of the environment and solving the localization and navigation problems are three important abilities that any mobile robot must have. Taking it into account, the review concentrates on these problems; how researchers have addressed them by means of omnidirectional vision; the main frameworks they have proposed; and how they have evolved in recent years.

1. Introduction

Over the last few years, the range of applications of mobile robots has significantly increased and they can be found in diverse environments, such as households, industrial and educational, where they are able to carry out a variety of tasks. The improvements performed both in perception and computation have had an important contribution to increase the range of environments where mobile robots can be used.

In order to be fully functional, a mobile robot must be capable of navigating safely through an unknown environment while simultaneously carrying out the task it has been designed for. With this aim, the robot must be able to build a model or map of this previously unknown environment, to estimate its current position and orientation making use of this model, and to navigate to the target points. Mapping, localization, and navigation are three classical problems in mobile robotics that have received a great deal of attention in the literature and keep on being very active research areas at present. Finding a robust solution to these three key problems is crucial to increase the autonomy and adaptability of mobile

robots to different circumstances and definitely expand their range of applications.

To address the mapping, localization, and navigation problems, it is necessary that the robot has some relevant information about the environment where it moves, which is a priori unknown in most applications. The robots can be equipped with diverse sensorial systems that allow them to extract the necessary information from the environment to be able to carry out their tasks autonomously. This way, from the first works on mobile robotics [1], the control systems have made use of the information collected from the environment, to a greater or lesser extent. The methods used to process this information have evolved as new algorithms have come out and the capabilities of the perception and computation systems have increased. As a result, very different approaches have emerged, considering different kinds of sensory information and processing techniques. Among them, in recent years, the use of omnidirectional vision systems has become popular in many works on mobile robots, and it is worth studying their main properties and applications in mapping, localization, and navigation.

In the light of the above, the purpose of this review is to present the most relevant works carried out in the field of mapping and localization using computer vision, paying special attention to those developments that make use of omnidirectional visual information.

The remainder of the paper is structured as follows. First, Section 2 presents the kinds of sensors that can be used to extract information from the mobile robot and from the environment. Then, Section 3 outlines some preliminary concepts about omnidirectional vision, the possible geometries, and the transformations that can be made with the visual information. After that, Section 4 presents the necessity of describing the information of the scenes and the main approaches that can be used. Later, Section 5 focuses on the problems of map creation, localization, and navigation of mobile robots. To conclude, a final discussion is carried out in Section 6.

2. Sensors in Mobile Robotics

In this section, we analyse first the types of sensors that have been traditionally used in mobile robotics to solve the mapping, localization, and navigation problems (Section 2.1). After that, we focus on vision sensors; we study their advantages (Section 2.2), the possible configurations they offer (Section 2.3), and some current trends in the creation of visual models (Section 2.4).

2.1. Types of Sensors. The solution to the mapping, localization, and navigation of mobile robots depends strongly on the information which is available on the state of the robot and the environment. Many types of sensors have been used so far with this aim, both proprioceptive and exteroceptive. On the one hand, proprioceptive sensors, such as odometry, measure the state of the robot. On the other hand, exteroceptive sensors, such as GPS, SONAR, laser, and infrared sensors, measure some external information from the environment. First, odometry is usually calculated from the measurements of encoders installed in the wheels. It can estimate the displacement of the robot, but the accumulation of errors makes its application infeasible, as a unique source of information, in real applications. This is the reason why it is usually used in combination with other kinds of sensors. Second, GPS (Global Positioning System) constitutes a good choice outdoors, but it loses fiability close to buildings, on narrow streets, and indoors. Third, SONAR (Sound Navigation and Ranging) sensors have a relatively low cost and permit measuring distances to the objects situated around, through the emission of sound pulses and measuring the reception of echoes from these pulses [2, 3]. However, their precision is relatively low, because they tend to present a high angular uncertainty and some noise introduced by the reflection of the sound signals [4]. Fourth, laser sensors determine the distance by measuring the flight time of a laser pulse when it is reflected onto the neighbour objects. Their precision is higher than SONAR's, and they can measure distances from centimeters to dozens of meters with a relatively good precision and angular resolution. Nevertheless, the cost, the weight, and the power consumption of such equipment are considerably high. Lasers have been used to create both 2D [5, 6] and 3D maps [7, 8], combined often with the use of odometry [9]. Finally, it is also possible to find some works that use infrared sensors in navigation tasks, whose range of detection arrives to several tens of centimeters [10].

2.2. Vision Sensors. As an alternative to the perception systems presented in the previous subsection, vision sensors have gained popularity because they present some interesting advantages. The cameras provide a big quantity of information from the environment, and 3D data can be extracted from it. Also, they present a relatively low cost and power consumption, comparing to laser rangefinders, which is specially relevant to the design of autonomous robots which need to work with batteries during long periods of time. Their behaviour is stable both outdoors and indoors, unlike GPS, whose signal tends to degrade in some areas. Finally, the availability of images permits carrying out additional high level tasks, apart from mapping and localization. These tasks include people detection and recognition and identification of the state of some objects which are relevant in robot navigation, such as doors and traffic lights.

Vision systems can be used either as the only perception system of the robot or in conjunction with the information provided by other sensors. For example, Choi et al. [11] present a system that combines SONAR with visual information in a mobile robot navigation application while Hara et al. [12] combine it with laser. Chang and Chuang [13] present a laser-vision system composed of a projector of laser lines and a camera whose information is processed to avoid obstacles and to extract relations between the points identified by the visual sensor and the laser. Some authors have reflected the wide variety of solutions proposed by researchers, depending on the visual techniques used, the combination with other sensors, and the algorithms to process the information. The evolution of the techniques in these fields is widely documented in [14], which shows the developments carried out in map building, localization, and navigation using computer vision until the mid-1990s, and [15], which complements this work with a state of the art from the 1990s.

Traditionally, the solutions based on the use of vision have been applied to Autonomous Ground Vehicles (AGV). However, more recently, these sensors have gained presence in the development of Unmanned Aerial Vehicles (UAV), which have great prospects of use in some applications such as surveillance, search and rescue, inspection of some areas or buildings, risky aerial missions, map creation, or fire detection. The higher quantity of degrees of freedom these kinds of vehicles tend to have makes it necessary a deeper study about how to analyse the sensory information. The aim is to obtain robust data that permit localization with all these necessary degrees of freedom.

2.3. Configuration of Vision Systems. As far as vision sensors are concerned, different configurations have been proposed, depending on the number of cameras used and the field of view they offer. Among them, there are systems based on

monocular configurations [16–18], binocular systems (stereo cameras) [19–21], and even trinocular configurations [22, 23]. Binocular and trinocular configurations permit measuring depth from the images. However, the limited field of view of these systems causes the necessity of employing several images of the environment in order to acquire complete information from it, which is necessary to create a complete model of an unknown environment. In contrast, more recently, the systems that provide omnidirectional visual information [24] have gained popularity thanks, mainly, to the great quantity of information they provide the robot with, as they usually have a field of view of 360 deg around the robot [25]. They can be composed of several cameras pointing towards different directions [26] or a unique camera and a reflective surface (catadioptric vision systems) [27].

Omnidirectional vision systems have further advantages, comparing to conventional vision systems. The features that appear in the images are more stable, since they remain longer in the field of view as the robot moves. Also, they provide information that permits estimating the position of the robot, independently on its orientation, and they permit estimating this orientation too. In general, omnidirectional vision systems have the ability to capture a more complete description of the environment in only one scene, which permits creating exhaustive models of the environment with a reduced number of views. Finally, even if some objects or persons occlude partially the scene, omnidirectional images contain some environment information from other directions.

These systems are usually based on the combination of a conventional camera and a convex mirror which can have various shapes. These structures are known as catadioptric vision systems, and the raw information they provide is known as *omnidirectional image*. However, some transformations can be carried out to obtain other kinds of projections which may be more useful depending on the task to develop and the degrees of freedom of the movement of the robot. They include the spherical, cylindrical, or orthographic projections [25, 28]. This issue will be addressed in Section 3.

2.4. Trends in the Creation of Visual Models. Using any of the configurations and projections presented in the previous subsections, a complete model of the environment can be built. Traditionally, three different approaches have been proposed with this aim: metric, topological, or hybrid. First, a metric map usually defines the position of some relevant landmarks extracted from the scenes with respect to a coordinate system and permits estimating the position of the robot with geometric accuracy. Second, a topological model consists generally of a graph where some representative localizations appear as nodes, along with the connectivity relations that permit navigating between consecutive nodes. Comparing to metric models, they usually permit a rougher localization with a reasonable computational cost. Finally, hybrid maps arrange the information into several layers, with different degrees of detail. They usually combine topological models in the top layers, which permit a rough localization, with metric models in the bottom layers, to refine this localization. This way, they try to combine the advantages of metric and topological maps.

The use of visual information to create models of the environment has an important disadvantage: the visual appearance of the environment changes not only when the robot moves, but also under other circumstances, such as variations in lighting conditions, which are present in all real applications and may produce substantial changes in the appearance of the scenes. The environment may also have some changes after the model has been created, and sometimes, the scenes may be partially occluded by the natural presence of people or other robots moving in the environment which is being modelled. Taking these facts into account, independently on the approach used to create the model, it is necessary to extract some relevant information from the scenes. This information must be useful to identify the environment and the position of the robot when it was captured, independently on any other phenomena that may happen. The extraction and description of this information can be carried out using two different approaches: based on local features or based on global appearance. On the one hand, the approaches based on local features try to extract some landmarks, points, or regions from the scenes and, on the other, global methods create a unique descriptor per scene that contains information on its global appearance. This problem is analysed more deeply in Section 4.

3. Omnidirectional Vision Sensors

As stated in the previous section, the expansion of the field of view that can be achieved with vision sensors is one of the reasons that explains the extensive use of computer vision in mobile robotics applications. Omnidirectional vision sensors stand out because they present a complete field of view around the camera axis. The objective of this section is twofold. On the one hand, some of the configurations that permit capturing omnidirectional images are presented. On the other hand, the different formats that can be used to express this information and their application in robotic tasks are detailed.

There are several configurations that permit capturing omnidirectional visual information. First, an array of cameras can be used to capture information from a variety of directions. The *ladybug* systems [29] constitute an example. They are composed of several vision sensors distributed around the camera and, depending on the model, they can capture a visual field covering more than 90% of a sphere whose centre is situated in the sensor.

Fisheye lenses can also be included within the systems that capture a wide field of view [30, 31]. These lenses can provide a field of view higher than 180 deg. The *Gear 360* camera [32] contains two fisheye lenses, each one capturing a field of view of 180 deg both horizontally and vertically. Combining both images, the camera provides images with a complete 360 deg field of view. Some works have shown how several cameras with such lenses can be combined to obtain omnidirectional information. For example, Li et al. [33] present a vision system equipped with two cameras with fisheye lenses whose field of view is a complete sphere around the sensor. However, using different cameras to create a spherical image can be challenging, taking the different
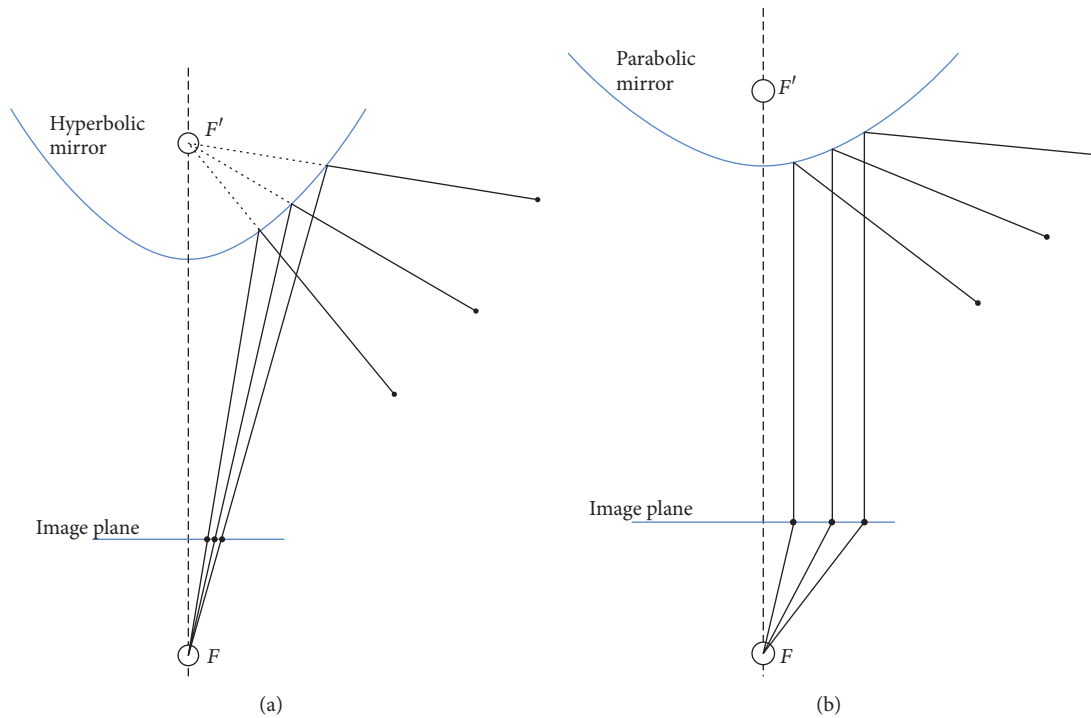
Figure 1: Projection model of central catadioptric systems. (a) Hyperbolic mirror and perspective projection lens and (b) parabolic mirror and orthographic lens.

lighting conditions of each camera into account. This is why Li developed a new system [34] to avoid this problem.

Catadioptric vision sensors are another example [35]. They make use of convex mirrors where the scene is projected. These mirrors may present different geometries, such as spherical, hyperbolic, parabolic, elliptic, or conic, and the camera takes the information from the environment through its reflection onto the mirrors. Due to their importance, the next subsection describes them in depth.

3.1. Catadioptric Vision Sensors. Catadioptric vision systems make use of convex reflective surfaces to expand the field of view of the camera. The vision sensors capture the information through these surfaces. The shape, position, and orientation of the mirror with respect to the camera will define the geometry of the projection of the world onto the camera.

Nowadays, many kinds of catadioptric vision sensors can be found. One of the first developments was done by Rees in 1970 [36]. More recent examples of catadioptric systems can be found in [37–40]. Most of them permit omnidirectional vision, which means having a vision angle equal to 360 deg around the mirror axis. The lateral angle of view depends essentially on the geometry of the mirror and the relative position with respect to the camera.

In the related literature, several kinds of mirrors can be found, as a part of catadioptric vision systems, such as spherical [41], conic [42], parabolic [27], or hyperbolic [43]. Yoshida et al. [44] present a work about the possibility of creating a catadioptric system composed of two mirrors, focusing

on the study on how the changes in the geometry of the system are reflected in the final image captured by the camera. Also, Baker and Nayar [45] present a comparative analysis about the use of different mirror geometries in catadioptric vision systems. According to this work, it is not possible to state, in general, that a specific geometry outperforms the others. Each geometry presents characteristic reflection properties that may be advantageous under some specific circumstances. Anyway, parabolic and hyperbolic mirrors present some particular properties that make them specially useful when perspective projections of the omnidirectional image are needed, as the next paragraph details.

There are two main issues when using catadioptric vision systems for mapping and localization: the single effective viewpoint property and calibration. On the one hand, when using catadioptric vision sensors, it is interesting that the system has a unique projection centre (i.e., that they constitute a central camera). In such systems, all the rays that arrive to the mirror surface converge into a unique point, which is the optical centre. This is advantageous because, thanks to it, it is possible to obtain undistorted perspective images from the scene captured by the catadioptric system [39]. This property appears in [45] referred to as *single effective viewpoint*. According to both works, there are two ways of building a catadioptric vision system that meets this property: with the combination of a hyperbolic mirror and a camera with perspective projection lens (conventional lens or *pin-hole* model), as shown in Figure 1(a) or with a system composed of a parabolic mirror and an orthographic projection lens (Figure 1(b)). In both figures, F and F' are the

Figure 2: Two sample omnidirectional images captured in the same room, from the same position, and using two different catadioptric systems.

foci of the camera and the mirror, respectively. In other cases, the rays that arrive to the camera converge into different focal points, depending on their vertical incidence angle (this is the case of noncentral cameras). Ohte et al. studied this problem with spherical mirrors [41]. In the case of objects that are relatively far from the axis of the mirror, the error produced by the existence of different focal points is relatively small. However, when the object is close to the mirror, the errors in the projection angles become significant and the complete projection model of the camera must be used.

On the other hand, many mapping and localization applications work correctly only if all the parameters of the system are known. With this aim, the catadioptric system needs to go under a calibration process [46]. The result of this process can be either (a) the intrinsic parameters of the camera, the coefficients of the mirror, and the relative position between them or (b) a list of correspondences between each pixel in the image plane and the ray of light of the world that has projected onto it. Many works have been developed on calibration of catadioptric vision systems. For example, Gonçalves and Araújo [47] propose a method to calibrate both central and noncentral catadioptric cameras using an approach based on bundle adjustment. Ying and Hu [48] present a method that uses geometric invariants, extracted from projections of lines or spheres in the world. This method can be used to calibrate central catadioptric cameras. Also, Marcato Jr. et al. [49] develop an approach to calibrate a catadioptric system composed of a wide-angle lens camera and a conic mirror that takes into account the possible misalignment between the camera and mirror axes. Finally, Scaramuzza et al. [50] develop a framework to calibrate central catadioptric cameras using a planar pattern shown at a number of different orientations.

3.2. Projections of the Omnidirectional Information. The raw information captured by the camera in a catadioptric vision system is the omnidirectional image that contains the information of the environment previously reflected onto the mirror. This image can be considered as a polar representation of the world whose origin is the projection of the focus of the mirror onto the image plane. Figure 2 shows two sample catadioptric vision systems, composed of a camera and a hyperbolic mirror, and the omnidirectional images captured with each one. The mirror (a) is the model *Wide 70* of the

manufacturer *Eizoh* and (b) is the model *Super-Wide View Large*, manufactured by *Accowle*.

The omnidirectional scene can be used directly to obtain useful information in robot navigation tasks. For example, Scaramuzza et al. [51] present a description method based on the extraction of radial lines from the omnidirectional image, to characterize omnidirectional scenes captured in real environments. They show how radial lines in omnidirectional images correspond to vertical lines of the environment, while circumferences whose centre is the origin of coordinates of the image correspond to horizontal lines in the world.

From the original omnidirectional image, it is possible to obtain different scene representations through the projection of the visual information onto different planes and surfaces. To make it possible, in general, it is necessary to calibrate the catadioptric system. Using this information, it is possible to project the visual information onto different planes or surfaces that show different perspectives of the original scene. Each projection presents specific properties that can be useful in different navigation tasks.

The next subsections present some of the most important representations and some of the works that have been developed with each of them in the field of mobile robotics. A complete mathematical description of these projections can be found in [39].

3.2.1. Unit Sphere Projection. An omnidirectional scene can be projected onto a unit sphere whose centre is the focus of the mirror. Every pixel of this sphere takes the value of the ray of light that has the same direction with respect to the focus of the mirror. To obtain this projection, the catadioptric vision system has to be previously calibrated. Figure 3 shows the projection model of the unit sphere projection when a hyperbolic mirror and a perspective lens are used. The mirror and the image plane are shown with blue color. F and F' are the foci of the camera and the mirror, respectively. The pixels in the image plane are back-projected onto the mirror (to do it, the calibration of the catadioptric system must be available) and after that, each point of the mirror is projected onto a unit sphere.

This projection has been traditionally useful in those applications that make use of the Spherical Fourier Transform, which permits studying 3D rotations in the space, as Geyer and Daniilidis show [52]. Friedrich et al. [53, 54]
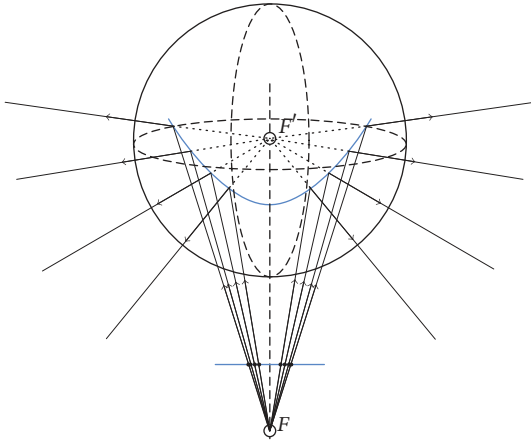
FIGURE 3: Projection model of the unit sphere projection.

FIGURE 4: Projection model of the cylindrical projection.

present an algorithm for robot localization and navigation using spherical harmonics. They use images captured with a hemispheric camera. Makadia et al. address the estimation of 3D rotations from the Spherical Fourier Transform [55–57], using a catadioptric vision system to capture the images. Finally, Schairer et al. present several works about orientation estimation using this transform, such as [58–60], where they develop methods to improve the accuracy in orientation estimation, implement a particle filter to solve this task, and develop a navigation system that combines odometry and the Spherical Fourier Transform applied to low resolution vision information.

3.2.2. Cylindrical Projection. This representation consists in projecting the omnidirectional information onto a cylinder whose axis is parallel to the mirror axis. Conceptually, it can be obtained by changing the polar coordinate system of the omnidirectional image into a rectangular coordinate system. This way, every circumference in the omnidirectional image will be converted into a horizontal line in the panoramic scene. This projection does not require the previous calibration of the catadioptric vision system.

Figure 4 shows the projection model of the cylindrical projection when a hyperbolic mirror and a perspective lens are used. The mirror and the image plane are shown with blue color. F and F' are the foci of the camera and the mirror, respectively. The pixels in the image plane are back-projected onto a cylindrical surface.

The cylindrical projection is commonly known as *panoramic image* and it is one of the most used representations in mobile robotics works, to solve some problems such as map creation and localization [61–63], SLAM [64], and visual servoing and navigation [65, 66]. This is due to the fact that this representation is more easily understandable by human operators and it permits using standard image processing algorithms, which are usually designed to be used with perspective images.

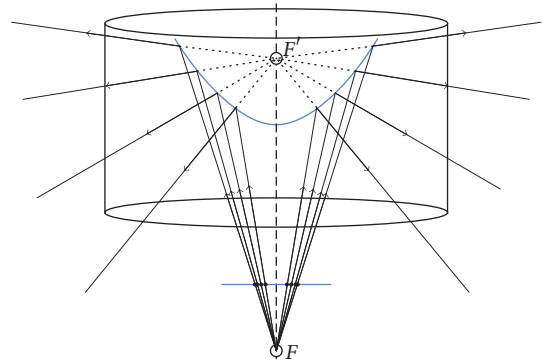*3.2.3. Perspective Projection.* From the omnidirectional image, it is possible to obtain projective images in any direction. They would be equivalent to the images captured by virtual conventional cameras situated in the focus of the mirror. The catadioptric system has to be previously calibrated to obtain such projection. Figure 5 shows the projection model of the perspective projection when a hyperbolic mirror and a perspective lens are used. The mirror and the image plane are shown with blue color. F and F' are the foci of the camera and the mirror, respectively. The pixels in the image plane are back-projected to the mirror (to do it, the calibration of the catadioptric system must be available) and after that, each point of the mirror is projected onto a plane. It is equivalent to capturing an image from the virtual conventional camera placed in the focus F'.

The orthographic projection, also known as *bird's eye view*, can be considered as a specific case of the perspective projection. In this case, the projection plane is situated perpendicularly to the camera axis. If the world reference system is defined in such a way that the floor is the plane xy and the z-axis is vertical, and the camera axis is parallel to the z-axis, then the orthographic projection is equivalent to having a conventional camera situated in the focus of the mirror pointing to the floor plane. Figure 6 shows the projection model of the orthographic view. In this case, the pixels are back-projected onto a horizontal plane.

In the literature, some algorithms which make use of the orthographic projection in map building, localization, and navigation applications can be found. For example, Gaspar et al. [25] make use of this projection to extract parallel lines from the floor of corridors and other rooms to perform navigation indoors. Also, Bonev et al. [67] propose a navigation system that combines the information contained in the omnidirectional image, cylindrical projection, and orthographic projection. Finally, Roebert et al. [68] show how a model of the environment can be created using perspective images and how this model can be used for localization and navigation purposes.

To conclude this section, Figure 7 shows (a) an omnidirectional image captured by the catadioptric system presented in Figure 2(a) and the visual appearance of the projections calculated from this image: (b) unit sphere projection, (c) cylindrical projection, (d) perspective projection onto a vertical plane, and (e) orthographic projection.
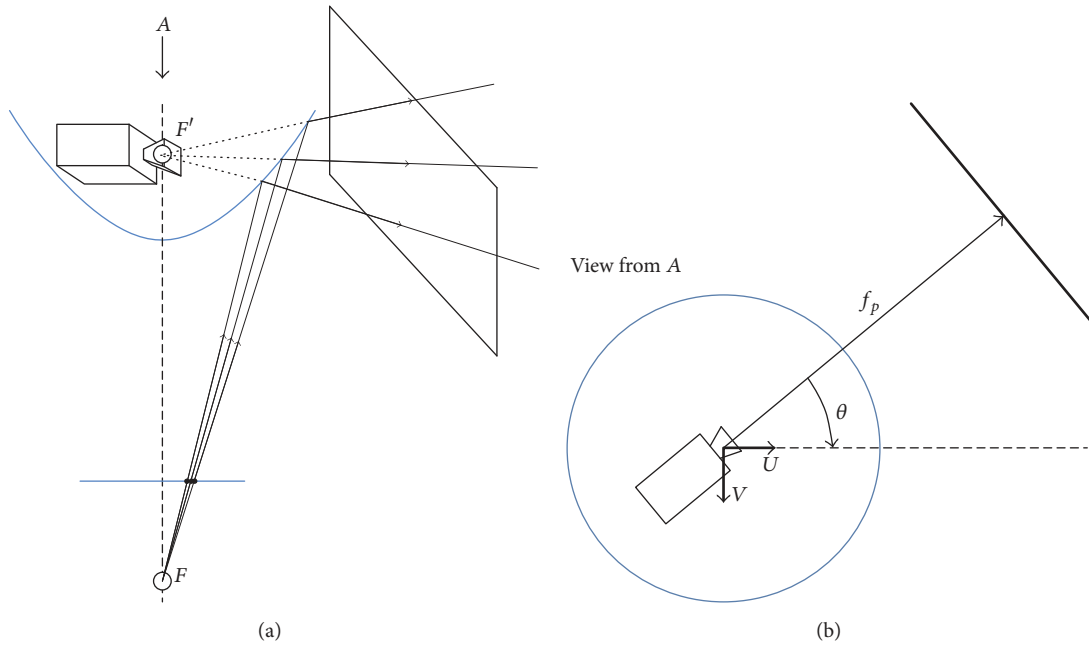
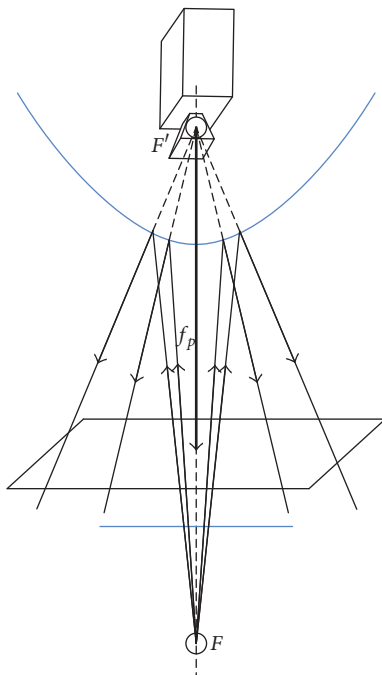FIGURE 5: Projection model of the perspective projection.

FIGURE 6: Projection model of the orthographic projection.

4. Description of the Visual Information

As described in the previous section, numerous authors have studied the use of omnidirectional images or any of their projections both in map building and in localization tasks. The images are highly dimensional data that change not only when the robot moves, but also when there is any change in the environment, such as changes in lighting conditions or in the position of some objects. Taking these facts into account, it is necessary to extract relevant information from the scenes, to be able to solve robustly the mapping and localization tasks. Depending on the method followed to extract this information, the different solutions can be classified into two groups: solutions based on the extraction and description of local features and solutions based on global appearance. Traditionally, researchers have focused on the first family of methods, but more recently some global appearance algorithms have been demonstrated to be also a robust alternative.

Many algorithms can be found in the literature working both with local features and with global appearance of images. All these algorithms imply many parameters that have to be correctly tuned so that the mapping and localization processes are correct. In the next subsections, some of these algorithms and their applications are detailed.

4.1. Methods Based on the Extraction and Description of Local Features

4.1.1. General Issues. Local features approaches are based on the extraction of a set of outstanding points, objects, or regions from each scene. Every feature is described by means of a descriptor, which is usually invariant against changes in the position and orientation of the robot. Once extracted and described, the solution to the localization problem is addressed usually in two steps [69]. First, the extracted features are tracked along a set of scenes, to identify the zones where these features are likely to be in the most recent images. Second, a feature matching process is carried out to identify the features.

Many different philosophies can be found depending on the type of features which are extracted and the procedure followed to carry out the tracking and matching. As an
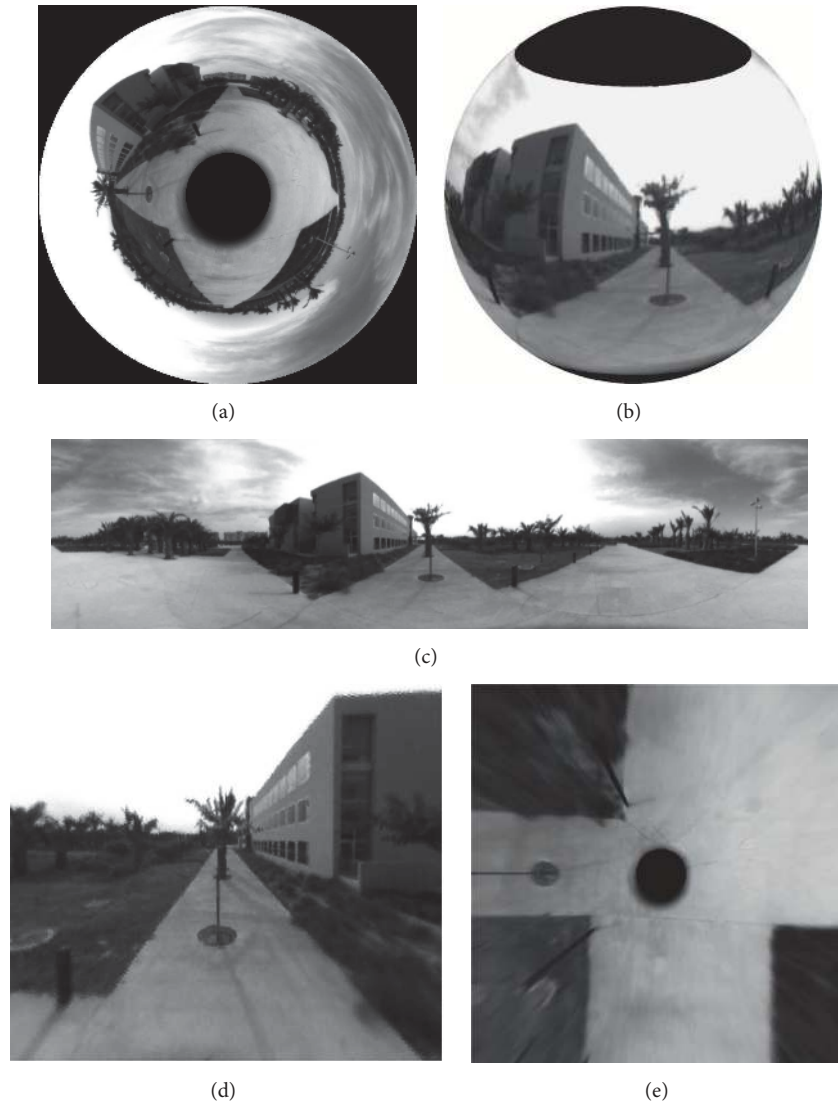
(a)

(b)

(c)

(d)

(e)

FIGURE 7: (a) Sample omnidirectional image and some projections calculated from it: (b) unit sphere projection, (c) cylindrical projection, (d) perspective projection onto a vertical plane, and (e) orthographic projection.

example, Pears and Liang [70] extract corners which are located in the floor plane in indoor environments and use homographies to carry out the tracking. Zhou and Li [71] also use homographies, but they extract features through the Harris corner detector [72]. Sometimes, geometrically more complex features are used, as Saripalli et al. do [73]. They carry out a segmentation process to extract predefined features, such as windows, and a Kalman filtering to carry out the tracking and matching of features.

4.1.2. Local Features Descriptors. Among the methods for feature extraction and description, SIFT and SURF can be highlighted. On the one hand, SIFT (Scale Invariant Feature Transform) was developed by Lowe [74, 75] and provides features which are invariant against scaling, rotation, changes in lighting conditions, and camera viewpoint. On the other hand, SURF (Speeded-Up Robust Features), whose standard version was developed by Bay et al. [76, 77], is inspired in

SIFT but presents a lower computational cost and a higher robustness against image transformations. More recent developments include BRIEF [78], which is designed to be used in real-time applications at the expense of a lower tolerance to image distortions and transformations; ORB [79], which is based on BRIEF, trying to improve its invariance against rotation and resistance to noise, but it is not robust against changes of scale; BRISK [80] and FREAK [81], which try to have the robustness of SURF but with an improved computational cost.

These descriptors have become popular in mapping and localization tasks using mobile robots, as many researchers show, such as Angeli et al. [82] and Se et al. [83], who make use of SIFT descriptors to solve these problems, and the works of Valgren and Lilienthal [84] and Murillo et al. [85], who employ SURF features extracted from omnidirectional scenes to estimate the position of the robot in a previously built map. Also, Pan et al. [86] present a method based on the use of

BRISK descriptors to estimate the position of an unmanned aerial vehicle.

4.1.3. Using Methods Based on Local Features to Build Visual Maps.

Using feature-based approaches in combination with probabilistic techniques, it is possible to build metric maps [87]. However, these methods present some drawbacks; for example, it is necessary that the environment be rich in prominent details (otherwise, artificial landmarks can be inserted in the environment, but this is not always possible); also, the detection of such points is sometimes not robust against changes in the environment and their description is not always fully invariant to changes in robot position and orientation. Besides, camera calibration is crucial in order to incorporate new measurements in the model correctly. This way, small deviations in either the intrinsic or the extrinsic parameters add some error to the measurements. At last, extracting, describing, and comparing landmarks are computationally complex processes that often make it infeasible building the model in real time, as the robot explores the environment.

Feature-based approaches have reached a relative maturity and some comparative evaluations of their performance have been carried out, such as [88] and [89]. These evaluations are useful to choose the most suitable extractor and descriptor to a specific application.

4.2. Methods Based on the Global Appearance of Scenes

4.2.1. General Issues.

The approaches based on the global appearance of scenes work with the images as a whole, without extracting any local information. Each image is represented by means of a unique descriptor, which contains information on its global appearance. This kind of methods presents some advantages in dynamic and/or poorly structured environments, where it is difficult to extract stable characteristic features or regions from a set of scenes. These approaches lead to conceptually simpler algorithms since each scene is described by means of only one descriptor. Map creation and localization can be achieved just storing and comparing pairwise these descriptors. As a drawback, extracting metric relationships from this information is difficult; thus this family of techniques is usually employed to build topological maps (unless the visual information is combined with other sensory data, such as odometry). Despite their simplicity, several difficulties must be faced when using these techniques. Since no local information is extracted from the scenes, it is necessary to use any compression and description method that make the process computationally feasible. Nevertheless, the current image description and compression methods permit optimising the size of the databases to store the necessary information and carrying out the comparisons between scenes with a relative computational efficiency. Occasionally, these descriptors do not present invariance neither to changes in the robot orientation or in the lighting conditions nor to other changes in the environment (position of objects, doors, etc.). They will also experience some problems in environments where *visual aliasing* is present, which is a common

phenomenon in indoor environments with repetitive visual structures.

In spite of these disadvantages, techniques based on global appearance constitute a systematic and intuitive alternative to solve the mapping and localization problems. In the works that make use of this approach, these tasks are usually addressed in two steps. The first step consists in creating a model of the environment. In this step, the robot captures a set of images and describes each of them by means of a unique descriptor and, from the information of these descriptors, a map is created and some relationships are established between images or robot poses. This step is known as *learning phase*. Once the environment has been modelled, in the second step, the robot carries out the localization by capturing an image, describing it, and comparing this descriptor with the descriptors previously stored in the model. This step is also known as *test or auto-localization phase*.

4.2.2. Global Appearance Descriptors.

The key to the proper functioning of this kind of methods is in the global description algorithm used. Different alternatives can be found in the related literature. Some of the pioneer works on this approach were developed by Matsumoto et al. [90] and make a direct comparison between the pixels of a central region of the scenes using a correlation process. However, considering the high computational cost of this method, the authors started to build some descriptors that stored global information from the scenes, using depth information estimated through stereo divergence [91].

Among the techniques that try to compress globally the visual information, Principal Components Analysis (PCA) [92] can be highlighted as one of the first robust alternatives used. PCA considers the images as multidimensional data that can be projected onto a lower dimension space, retaining most of the original information. Some authors, like Kröse et al. [93, 94] and Štimec et al. [95], make use of this method to create robust models of the environment using mobile robots. The first PCA developments presented two main problems. On the one hand, the descriptors were variant against changes in the orientation of the robot in the ground plane and, on the other hand, all the images had to be available to build the model, which means that the map cannot be built online, as the robot explores the environment. If a new image has to be added to the previously built PCA model (e.g., because the robot must update this representation as new relevant images are captured) it is necessary to start the process from the scratch, using again all the images captured so far. Some authors have tried to overcome these drawbacks. First, Jogan and Leonardis [96] proposed a version of PCA that provides descriptors which are invariant against rotations of the robot in the ground plane when omnidirectional images are used, at the expense of a substantial increase of the computational cost. Second, Artač et al. [97] used an incremental version of PCA that permits adding new images to a previously built model.

Other authors have proposed the implementation of descriptors based on the application of the Discrete Fourier Transform (DFT) to the images, with the aim of extracting the most relevant information from the scenes. In this field,

some alternatives can be found. On the one hand, in the case of panoramic images, both the two-dimensional DFT or the Fourier Signature can be implemented, as Payá et al. [28] and Menegatti et al. [98] show, respectively. On the other hand, in the case of omnidirectional images, the Spherical Fourier Transform (SFT) can be used, as Rossi et al. [99] show. In all the cases, the resulting descriptor is able to compress most of the information contained in the original scene in a lower number of components. Also, these methods permit building descriptors which are invariant against rotations of the robot in the ground plane. Apart from this, they contain enough information to estimate not only the position of the robot, but also its relative orientation. At last, the computational cost to describe each scene is relatively low and each descriptor can be calculated independently on the rest of descriptors. Taking these features into account, in this moment DFT outperforms PCA as far as mapping and localization are concerned. As an example, Menegatti et al. [100] show how a probabilistic localization process can be carried out within a visual memory previously created by means of the Fourier Signature of a set of panoramic scenes, as the only source of information.

Other authors have described globally the scenes through approaches based on gradient, either the magnitude or the orientation. As an example, Košecká et al. [101] make use of a histogram of gradient orientation to describe each scene with the goal of creating a map of the environment and carrying out the localization process. However, some comparisons between local areas of the candidate zones are performed to refine these processes. Murillo et al. [102] propose to use a panoramic *gist* descriptor [103] and try to optimise its size while keeping most of the information of the environment. The approaches based on gradients and *gist* tend to present a performance similar to Fourier Transform methods [104].

The information stored in the color channels also constitutes a useful alternative to build global appearance descriptors. The works of Zhou et al. [105] show an example of use of this information. They propose building histograms that contain information on color, gradient, edges density, and texture to represent the images.

4.2.3. Using Methods Based on Global Appearance to Build Visual Maps. Finally, when working with global appearance, it is necessary to consider that the appearance of an image strongly depends on the lighting conditions of the environment represented. This way, the global appearance descriptor should be robust against changes in these conditions. Some researchers have focused on this topic and have proposed different solutions. First, the problem can be addressed by means of considering sets of images captured under different lighting conditions. This way, the model would contain information on the changes that appearance can undergo. As an example, the works developed by Murase and Nayar [106] solved the problem of visual recognition of objects under changing lighting conditions using this approach. The second approach consists in trying to remove or minimise the effects produced by these changes during the creation of the descriptor, to obtain a normalised model. This approach is mainly based on the use of filters, for example, to detect

and extract edges [107], since this information tends to be more insensitive to changes in lighting conditions than the information of intensity, or homomorphic filters [108], which try to separate the luminance and reflectance components and minimise the first component, which is the one that is more prone to change when the lighting conditions do.

5. Mapping, Localization, and Navigation Using Omnidirectional Vision Sensors

This section addresses the problems of mapping, localization, and navigation using the information provided by omnidirectional vision sensors. First, the main approaches to solve these tasks are outlined and then some relevant works developed within these approaches are described.

While the initial works tried to model the geometry of the environment with metric precision, from visual information, and arranging the information through CAD (Computer Assisted Design) models, these approaches gave way to simpler models that represent the environment through occupation grids, topological maps, or even sequences of images. Traditionally, the problems of map building and localization have been addressed using three different approaches:

(i) Map building and subsequent localization: in these approaches, first, the robot goes through the environment to map (usually in a teleoperated way) and collects some useful information from a variety of positions. This information is then processed to build the map. Once the map is available, the robot captures information from its current position and comparing this information with the map, its current pose (position and orientation) is estimated.

(ii) Continuous map building and updating: in this approach, the robot is able to explore autonomously the environment to map and build or update a model while it is moving. The SLAM algorithms (Simultaneous Localization And Mapping) fit within this approach.

(iii) Mapless navigation systems: the robot navigates through the environment by applying some analysis techniques to the last captured scenes, such as optical flow, or through visual memories previously recorded that contain sequences of images and associated actions, but no relation between images. This kind of navigation is associated, basically, with reactive behaviours.

The next subsections present some of the most relevant contributions developed in each of these three frameworks.

5.1. Map Building and Subsequent Localization. Pretty usually, solving the localization problem requires having a previously built model of the environment, in such a way that the robot can estimate its pose by comparing its sensory information with the information captured in the model. In general, depending on how this information is arranged, these models can be classified into three categories: metric, topological, or hybrid [109]. First, metric approaches try to

create a model with geometric precision, including some features of the environment with respect to a reference system. These models are created usually using local features extracted from images [87] and they permit estimating metrically the position of the robot, up to a specific error. Second, topological approaches try to create compact models that include information from several characteristic localizations and the connectivity relations between them. Both local features and global appearance can be used to create such models [104, 110]. They usually permit a rough localization of the robot with a reasonable computational cost. Finally, hybrid maps combine metric and topological information to try to have the advantages of both methods. The information is arranged into several layers, with topological information that permits carrying out an initial rough estimation and metric information to refine this estimation when necessary [111].

The use of information captured by omnidirectional vision sensors has expanded in recent years to solve the mapping, localization, and navigation problems. The next paragraphs outline some of these works.

Initially, a simple way to create a metric model consists in using some visual beacons which can be seen from a variety of positions and used to estimate the pose of the robot. Following this philosophy, Li et al. [112] present a system for the localization of agricultural vehicles using some visual beacons which are perceived using an omnidirectional vision system. These beacons are constituted by four red landmarks situated in the vertices of the environment where these vehicles may move. The omnidirectional vision system makes use of these four landmarks as beacons situated on specific and previously known positions to estimate its pose. In other occasions, a more complete description of the environment is previously available, as in the work of Lu et al. [113]. They use the model provided by the competition RoboCup Middle Size League. Taking this competition into account, the objective consists in carrying out an accurate localization of the robot. With this aim, a Monte Carlo localization method is employed to provide a rough initial localization and, once the algorithm has converged, this result is used as the initial value of a matching optimisation algorithm in order to perform accurate and efficient localization tracking.

The maps composed of local visual features have been extensively used along the last few years. Classical approaches have made use of monocular or stereo vision to extract and track these local features or landmarks. More recently, some researchers have shown that it is feasible to extend these classical approaches to be used with omnidirectional vision. In this line, Choi et al. [114] propose an algorithm to create maps, which is based on an object extraction method, using Lucas-Kanade optical flow motion detection from the images obtained by an omnidirectional vision system. The algorithm uses the outer point of motion vectors as feature points of the environment. They are obtained based on the corner points of an extracted object and using Lucas-Kanade optical flow.

Valgren et al. [115] propose the use of local features that are extracted from images captured in a sequence. These features are used both to cluster the images into nodes and to detect links between the nodes. They employ a variant of SIFT features that are extracted and matched from different viewpoints. Each node of the topological map contains a collection of images which are considered similar enough (i.e., a sufficient number of feature matches must exist among the images contained in the node). Once the nodes are constructed, they are connected taking into account feature matching between the images in the nodes. This way, they build a topological map incrementally, creating new nodes as the environment is explored. In a later work [84], the same authors study how to build topological maps in large environments both indoors and outdoors, using the local features extracted from a set of omnidirectional views, including the epipolar restriction and a clustering method to carry out localization in an efficient way.

Goedemé et al. [116] present an algorithm to build a topological model of complex environments. They also propose algorithms to solve the problems of localization (both global, when the robot has no information about its previous position and local, when this information is available) and navigation. In this approach, the authors propose two local descriptors to extract the most relevant information from the images: a color enhanced version of SIFT and invariant column segments.

The use of bag-of-features approaches has also been considered by some researchers in mobile robotics. As an example, Lu et al. [117] propose the use of a bag-of-features approach to carry out topological localization. The map is built in a previous process, consisting mainly in clustering the visual information captured by an omnidirectional vision system. It is performed in two steps. First, local features are extracted from all the images, and they are used to carry out a K-means clustering, obtaining the k clusters' centres. Second, these centres constitute the visual vocabulary which is used subsequently to solve the localization problem.

Beyond these frameworks, the use of omnidirectional vision systems has contributed to the emergence of new paradigms to create visual maps, using some different fundamentals to those traditionally used with local features. One of these frameworks consists in creating topological maps using a global description of the images captured by the omnidirectional vision system. In this area, Menegatti et al. [98] show how a visual memory can be built using the Fourier Signature, which presents rotational invariance when used with panoramic images. The map is built using global appearance methods and a system of mechanical forces, calculated from the similitude between descriptors. Liu et al. [118] propose a framework that includes a method to describe the visual appearance of the environment. It consists in an adaptive descriptor based on color features and geometric information. Using this descriptor, they create a topological map based on the fact that the vertical edges divide the rooms in regions with a uniform meaning, in the case of indoor environments. These descriptors are used as a basis to create a topological map through an incremental process in which similar descriptors are incorporated into each node. A global threshold is considered to evaluate the similitude between descriptors. Also, other authors have studied the localization problem using color information, as

Ulrich and Nourbakhsh do in [119], where a topological localization framework is proposed, using panoramic color images and color histograms in a voting schema.

Payá et al. [63] propose a method to build topological maps from the global appearance of panoramic scenes captured from different points of the environment. These descriptors are analysed, to create a topological map of the environment. On the one hand, each node in the map contains an omnidirectional image representing the visual appearance of the place where the image was captured. On the other hand, links in the graph are neighbourly relations between nodes so that when two nodes are connected, the environment represented in one node is close to the environment represented by the other node. In this way, a process based on a mass-spring-damper model is developed and the resulting topological map incorporates geometrical relationships that situate the nodes close to the real positions in which the omnidirectional images were captured. The authors develop algorithms to build this model either in a batch or in an incremental way. Subsequently, Payá et al. [120] also propose a topological map building algorithm, and a comparative evaluation of the performance of some global appearance descriptors is included. The authors study the effect of some usual phenomena that may happen in real applications: changes in lighting conditions, noise, and occlusions. Once the map is built, a Monte Carlo localization approach is accomplished to estimate the most probable pose of the vehicle.

Also, Rituerto et al. [121] propose the use of a semantic topological map created from the global appearance description of a set of images captured by an omnidirectional vision system. These authors propose the use of a *gist*-based descriptor to obtain a compact representation of the scene and estimate the similarity between two locations based on the Euclidean distance between descriptors. The localization task is divided into two steps: a first step, or global localization, in which the system does not consider any a priori information about the current localization, and a second step or continuous localization, in which they assume that the image is acquired from the same topological region.

Ranganathan et al. [122] introduce a probabilistic method to perform inference over the space of topological maps. They approximate the posterior distribution over topologies, from the available sensor measurements. To do it, they perform Bayesian inference over the space of all the possible topologies. The application of this method is illustrated using Fourier Signatures of panoramic images obtained from an array of cameras mounted on the robot.

Finally, Štimec et al. [95] present a method based on global appearance to build a trajectory-based map by means of a clustering process with PCA features obtained from a set of panoramic images.

While the previous approaches suppose that the robot trajectory is contained on the ground plane (i.e., they contain information of 3 degrees of freedom), some works have also shown that it is possible to create models that contain information with a higher number of degrees of freedom. As an example, Huhle et al. [123] and Schairer et al. [60] present algorithms for map building and 3D localization, based on the use of the Spherical Fourier Transform and the unit sphere projection of the omnidirectional information.

5.2. Continuous Map Building and Updating. The process to build a map and continuously update it while the robot simultaneously estimates its position with respect to the model is one of the most complex tasks in mobile robotics. Along the last few years, some approaches have been developed to solve it using omnidirectional visual systems and some authors have proposed comparative evaluations to know the performance of omnidirectional imaging comparing to other classic sensors. In this field, Rituerto et al. [124] carry out a comparative analysis between visual SLAM systems using omnidirectional and conventional monocular vision. This approach is based on the use of the Extended Kalman Filter (EKF) and SIFT points extracted from both kinds of scenes. To develop the comparison between both systems, the authors made use of a spherical camera model, whose Jacobian is obtained and used in the EKF algorithm. The results provided show the superiority of the omnidirectional systems in the experiments carried out. Also, Burbridge et al. [125] performed several experiments in order to quantify the advantage of using omnidirectional vision compared to narrow field of view cameras. They also made use of the EKF but, instead of characteristic points, their landmarks are the vertical edges of the environment that appear in the scenes. In this case, the experiments were conducted using simulated data, and they offer results about the accuracy in localization taking some parameters into account, such as the number of landmarks they integrate into the filter, the field of view of the camera, and the distribution of the landmarks. One of the main advantages of the large field of view of the omnidirectional vision systems is that the extracted features remain in the image longer as the robot moves. Thanks to it, the estimation of the pose of the robot and the features can be carried out more precisely. Some other evaluations have compared the efficacy of a SLAM system using an omnidirectional vision sensor with respect to a SLAM system using a laser range finder, such as the work developed by Erturk et al. [126]. In this case, the authors have employed multiple simulated environments to provide the necessary visual data to both systems. Omnidirectional vision proves to be a cost-effective solution that is suitable specially for indoor environments (since outdoor conditions have a negative influence on the visual data).

In the literature, we can also find some approaches which are based on the classical visual SLAM schema, adapted to be used with omnidirectional visual information, and introducing some slight variations. For example, Kim and Oh [127] propose a visual SLAM system in which vertical lines extracted from omnidirectional images and horizontal lines extracted from a range sensor are integrated. Another alternative is presented by Wang et al. [128], who propose a map which is composed of many submaps, each one consisting of all the feature points extracted from an image and the position of the robot with respect to the global coordinate system when this image was captured. Furthermore, some other proposals that have provided good results in the SLAM area, such as Large-Scale Direct- (LSD-) SLAM,

have been proposed and adapted using omnidirectional cameras, computing dense or semidense depth maps in an incremental fashion, and tracking the camera using direct image alignment. Caruso et al. [129] propose an extension of LSD-SLAM to a generic omnidirectional camera model, along with an approach to perform stereo operations directly on such omnidirectional images. In this case, the proposed method is evaluated on images captured with a fisheye lens with a field of view of 185 deg.

More recently, some proposals that go beyond the classical approaches have been made. In these proposals, the features provided by the omnidirectional vision system are used. In this area, Valiente et al. [130] suggest a representation of the environment that tries to optimise the computational cost of the mapping process and to provide a more compact representation. The map is sustained by a reduced set of omnidirectional images, denoted as views, which are acquired from certain poses of the environment. The information gathered by these views permits modelling large environments and, at the same time, they ease the observation process by which the pose of the robot is retrieved. In this work, a view n consists of a single omnidirectional image captured from a certain pose of the robot $x_{ln} = (x_l, y_l, \theta_l)_n$ and a set of interest points extracted from that image. Such arrangement permits exploiting the capability of an omnidirectional image to gather a large amount of information in a simple snapshot, due to its large field of view. So, a view n is constituted by the pose where it was acquired, with $n \in [1, \ldots, N]$, with N being the number total of views constituting the final map. The number of views initialized in the map directly depends on the sort of environment and its visual appearance. In this case, the process of localization is solved by means of an observation model that takes the similarity between the omnidirectional images into account. In a first step, a subset of candidate views from the map is selected, based on the Euclidean distance between the pose of the current view acquired by the robot and the position of each candidate. Once the set of candidates have been extracted, a similarity measure can be evaluated in order to determine the highest similarity with the current image. Finally, the localization of the robot can be accomplished taking into account the correspondences between both views and the epipolar restriction.

An additional step in building topological maps with omnidirectional images is to create a hierarchical map structure that combines metric and topology. In this area, Fernández et al. [131] present a framework to carry out SLAM (Simultaneous Localization And Mapping) using panoramic scenes. The mapping is approached in a hybrid way, constructing simultaneously a model with two layers and checking if a loop closure is produced with a previous node.

5.3. Mapless Navigation Systems. In this case, the model of the environment can be represented as a visual memory that typically contains sequences or sets of images with no other underlying structure among them. In this area, several possibilities can be considered, taking into account if the visual information is described through local features of global appearance.

First, some approaches that use local features can be found. Thompson et al. [132] propose a system that learns places by automatically selecting reliable landmarks from panoramic images and uses them with localization purposes. Also, they employ normalised correlation during the comparisons to minimise the effect of changing lighting conditions. Argyros et al. [133] develop a vision based method for robot homing, using panoramic images too. With this method, the robot can calculate a route that leads it to the position where an image was captured. The method tracks features in panoramic views and exploits only angular information of these features to calculate a control strategy. Long range homing is achieved by organizing the features' trajectories in a visual memory. Furthermore, Lourenço et al. [134] propose an alternative approach to image-based localization which goes beyond, since it takes into account that, in some occasions, it would be possible that the images stored in the database and the query image could have been acquired using different omnidirectional imaging systems. Due to the introduction of nonlinear image distortion in such images, the difference of the appearance between both of them could be significant. In this sense, the authors propose a method that employs SIFT features extracted from the database images and the query image in order to determine the localization of the robot.

Second, as far as the use of global appearance is concerned, Matsumoto et al. [135] propose a navigation method that makes use of a sequence of panoramic images to store information from a route and a template matching approach to carry out localization, steering angle determination, and obstacle detection. This way, the robot can navigate between consecutive views by using global information from the panoramic scenes. Also, Menegatti et al. [100] present a method of image-based localization from the matching of the robot's currently captured view with the previously stored reference view, trying to avoid perceptual aliasing. Their method makes use of the Fourier Signature to represent the image and to facilitate the image-based localization. They make use of the properties that this transform presents when it is used with panoramic images, and they also propose a Monte Carlo algorithm for robust localization. In a similar way, Berenguer et al. [136] propose two methods to estimate the position of the robot by means of the omnidirectional image captured. On the one hand, the first method represents the environment through a sequence of omnidirectional images, and the Radon transform is used to describe the global appearance of the scenes. A rough localization is achieved carrying out the pairwise comparison between the current scene and this sequence of scenes. On the other hand the second method tries to build a local topological map of the area where the robot is located and this map is used to refine the localization of the robot. This way while the first method is a mapless localization, the second one would fit in the methods presented in Section 5.1. Both methods are tested under different lighting conditions and occlusions and the results show their effectiveness and robustness. Murillo et al. [137] propose an alternative approach to solve the localization problem through a pyramidal matching process. The authors indicate that the method can work with any low dimensional

TABLE 1: Main features of the frameworks that use omnidirectional vision systems to solve the mapping and localization problems.

Reference	Approach	Type of image	Type of map	Features	Description
[112]	5.1	Omnidirectional	Metric	Local	Colour
[113]	5.1	Omnidirectional	Metric	Local	FAST-CSLBP
[114]	5.1	Omnidirectional	Metric	Local	LKOF
[84, 115]	5.1	Omnidirectional	Topological	Local	MSIFT
[116]	5.1	Omnidirectional	Topological	Local	SIFT Invariant column segments
[98]	5.1	Panoramic	Topological	Global	Fourier Signature
[118]	5.1	Panoramic	Topological	Global	FACT descriptor
[119]	5.1	Panoramic	Topological	Global	Color histograms
[63]	5.1	Panoramic	Topological	Global	Fourier Signature
[120]	5.1	Panoramic	Topological	Global	Various
[121]	5.1	Omnidirectional	Topological	Global	GIST
[122]	5.1	Panoramic	Topological	Global	Fourier Signature
[95]	5.1	Panoramic	Topological	Global	PCA features
[60, 123]	5.1	Omnidirectional	Topological	Global	Spherical Fourier Transform
[124]	5.1	Omnidirectional	Metric	Local	SIFT
[125]	5.2	Panoramic	Metric	Local	Vertical edges
[126]	5.2	Panoramic	Metric	Local	Contours of objects
[127]	5.2	Omnidirectional	Metric	Local	Vertical + horizontal lines
[128]	5.2	Omnidirectional	Metric	Local	SURF
[129]	5.2	Omnidirectional	Metric	Global	Distance map
[130]	5.2	Omnidirectional	Metric	Local	SURF
[131]	5.2	Panoramic	Hybrid	Global	Fourier Signature
[132]	5.2	Panoramic	Sequence images	Local	Landmarks after reliability test
[133]	5.3	Panoramic	Sequence images	Local	KLT corner detection
[134]	5.3	Omnidirectional	Sequence images	Local	SIFT (improved)
[135]	5.3	Panoramic	Sequence images	Global	Portion of the image
[100]	5.3	Panoramic	Sequence images	Global	Fourier Signature
[136]	5.1, 5.3	Omnidirectional	Topological sequence images	Global	Radon transform
[137]	5.3	Omnidirectional	Sequence images	Global	SURF
[138]	5.3	Unit sphere proj. panor., orthographic	Sequence images	Global	Spherical Fourier Transform Discrete Fourier Transform

features descriptor. They proposed the use of descriptors for the features detected in the images combining topological and metric information in a hierarchical process performed in three steps. In the first step, a descriptor is calculated over all the pixels in the images. All the images in the database with a difference over a threshold are discarded. In a second step, a set of descriptors of each line in the image is used to build several histograms per image. A matching process is performed, obtaining a similarity measurement between the query image and the images contained in the database. Finally in a third step, a more detailed matching algorithm is used taking into account geometric restrictions between the lines.

Finally, not only is it possible to determine the 2D location of the robot based on these approaches, but also 3D estimations can be derived from them considering the description of the omnidirectional images. In this regard, Amorós et al. [138] present a collection of different techniques that provide the relative height between the real pose of the robot and a reference image. In this case, the environment is represented through sequences of images that store the effect of changes in the altitude of the robot and the images are described using their global appearance.

To conclude this section, Table 1 presents an outline of the frameworks presented in this section and their main features: the kind of approach (Sections 5.1, 5.2, or 5.3), the type of image, the type of map, the kind of features (local or global), and the specific description method employed to extract information from the images.

6. Conclusion

During the last decades, vision sensors have become a robust alternative in the field of mobile robotics to capture the necessary information from the environment where the robot moves. Among them, omnidirectional vision systems have experienced a great expansion more recently, mainly owing to the big quantity of data they are able to capture in only one scene. This wide field of view leads to some interesting properties that make them specially useful as the only source of information in mobile robotics applications, which leads to improvements in power consumption and cost.

Consequently, the number of works that make use of such sensors has increased substantially, and many frameworks can be found currently to solve the mapping, localization,

and navigation tasks. This review has revolved around these three problems and some of the approaches that researches have proposed to solve them using omnidirectional visual information.

To this end, this work has started focusing on the geometry of omnidirectional vision systems, specially on catadioptric vision systems, which are the most common structures to obtain omnidirectional images. It has led to the study of the different shapes of mirrors and the projections of the omnidirectional information. This is specially important because the majority of works have made use of perspective projections of the information to build models of the environment, as they can be interpreted more easily by a human operator. After that, the review has homed in on how visual information can be described and handled. Two options are available: methods based on the extraction, description, and tracking of local features and methods based on global appearance description. While local features were the reference option in most initial works, global methods have gained presence more recently as they are able to build intuitive models of the environment in which the localization problem can be solved using more straightforward methods, based on the pairwise comparison between descriptors. Finally, the work has concentrated on the study of the mapping, localization, and navigation of mobile robots using omnidirectional information. The different frameworks have been classified into three approaches: (a) map building and subsequent localization, (b) continuous map building and updating, and (c) mapless navigation systems.

The great quantity of works on these topics shows how omnidirectional vision and mobile robotics are two very active areas, where this quick evolution is expected to continue during the next years. In this regard, finding a relatively robust and computationally feasible solution to some current problems would definitively help to improve the autonomy of mobile robots and expand their range of applications and environments where they can work. Among them, the mapping and localization problems in considerably large and changing environments and the estimation of the position and orientation in the space and under realistic working conditions are worth being addressed to try to arrive at closed solutions.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

References

[1] G. Giralt, R. Sobek, and R. Chatila, "A multi-level planning and navigation system for a mobile robot; a first approach to HILARE," in *Proceedings of the 6th International Joint Conference on Artificial Intellligence*, pp. 335–337, Morgan Kaufmann, Tokyo, Japon, 1979.

[2] O. Wijk and H. Christensen, "Localization and navigation of a mobile robot using natural point landmarks extracted from sonar data," *Robotics and Autonomous Systems*, vol. 31, no. 1-2, pp. 31–42, 2000.

[3] J. D. Tardós, J. Neira, P. M. Newman, and J. J. Leonard, "Robust mapping and localization in indoor environments using sonar data," *International Journal of Robotics Research*, vol. 21, no. 4, pp. 311–330, 2002.

[4] Y. Cha and D. Kim, "Omni-directional image matching for homing navigation based on optical flow algorithm," in *Proceedings of the 12th International Conference on Control, Automation and Systems (ICCAS '12)*, pp. 1446–1451, October 2012.

[5] J. J. Leonard and H. F. Durrant-Whyte, "Mobile robot localization by tracking geometric beacons," *IEEE Transactions on Robotics and Automation*, vol. 7, no. 3, pp. 376–382, 1991.

[6] L. Zhang and B. K. Ghosh, "Line segment based map building and localization using 2D laser rangefinder," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA '00)*, vol. 3, pp. 2538–2543, April 2000.

[7] R. Triebel and W. Burgard, "Improving simultaneous localization and mapping in 3D using global constraints," in *Proceedings of the 20th National Conference on Artificial Intelligence—Volume 3*, pp. 1330–1335, 2005.

[8] C. Yu and D. Zhang, "A new 3D map reconstruction based mobile robot navigation," in *Proceedings of the 8th International Conference on Signal Processing*, vol. 4, IEEE, Beijing, China, November 2006.

[9] A. Y. Hata and D. F. Wolf, "Outdoor mapping using mobile robots and laser range finders," in *Proceedings of the Electronics Robotics and Automotive Mechanics Conference (CERMA '09)*, pp. 209–214, September 2009.

[10] C.-H. Chen and K.-T. Song, "Complete coverage motion control of a cleaning robot using infrared sensors," in *Proceedings of the IEEE International Conference on Mechatronics (ICM '05)*, pp. 543–548, July 2005.

[11] J. Choi, S. Ahn, M. Choi, and W. K. Chung, "Metric SLAM in home environment with visual objects and sonar features," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS '06)*, pp. 4048–4053, October 2006.

[12] K. Hara, S. Maeyama, and A. Gofuku, "Navigation path scanning system for mobile robot by laser beam," in *Proceedings of the SICE Annual Conference of International Conference on Instrumentation, Control and Information Technology*, pp. 2817–2821, August 2008.

[13] W.-C. Chang and C.-Y. Chuang, "Vision-based robot navigation and map building using active laser projection," in *Proceedings of the IEEE/SICE International Symposium on System Integration (SII '11)*, pp. 24–29, IEEE, December 2011.

[14] G. N. DeSouza and A. C. Kak, "Vision for mobile robot navigation: a survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 2, pp. 237–267, 2002.

[15] F. Bonin-Font, A. Ortiz, and G. Oliver, "Visual navigation for mobile robots: a survey," *Journal of Intelligent and Robotic*

Systems: Theory and Applications, vol. 53, no. 3, pp. 263–296, 2008.

[16] R. Sim and G. Dudek, "Effective exploration strategies for the construction of visual maps," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, vol. 3, pp. 3224–3231, Las Vegas, Nev, USA, October 2003.

[17] Q. Zhu, S. Avidan, M.-C. Yeh, and K.-T. Cheng, "Fast human detection using a cascade of histograms of oriented gradients," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '06)*, vol. 2, pp. 1491–1498, IEEE, June 2006.

[18] K. Okuyama, T. Kawasaki, and V. Kroumov, "Localization and position correction for mobile robot using artificial visual landmarks," in *Proceedings of the International Conference on Advanced Mechatronic Systems (ICAMechS '11)*, pp. 414–418, Zhengzhou, China, August 2011.

[19] M. Z. Brown, D. Burschka, and G. D. Hager, "Advances in computational stereo," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 8, pp. 993–1008, 2003.

[20] R. Sim and J. J. Little, "Autonomous vision-based exploration and mapping using hybrid maps and Rao-Blackwellised particle filters," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS '06)*, pp. 2082–2089, Beijing, China, October 2006.

[21] Z. Yong-Guo, C. Wei, and L. Guang-Liang, "The navigation of mobile robot based on stereo vision," in *Proceedings of the 5th International Conference on Intelligent Computation Technology and Automation (ICICTA '12)*, pp. 670–673, January 2012.

[22] Y. Jia, M. Li, L. An, and X. Zhang, "Autonomous navigation of a miniature mobile robot using real-time trinocular stereo machine," in *Proceedings of the IEEE International Conference on Robotics, Intelligent Systems and Signal Processing*, vol. 1, pp. 417–421, Changsha, China, October 2003.

[23] N. Ayache and F. Lustman, "Trinocular stereo vision for robotics," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, no. 1, pp. 73–85, 1991.

[24] N. Winters, J. Gaspar, G. Lacey, and J. Santos-Victor, "Omni-directional vision for robot navigation," in *Proceedings of the IEEE Workshop on Omnidirectional Vision (OMNIVIS '00)*, pp. 21–28, Hilton Head Island, SC, USA, June 2000.

[25] 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.

[26] D. Bradley, A. Brunton, M. Fiala, and G. Roth, "Image-based navigation in real environments using panoramas," in *Proceedings of the IEEE International Workshop on Haptic Audio Visual Environments and their Applications (HAVE '05)*, pp. 57–59, October 2005.

[27] S. Nayar, "Omnidirectional video camera," in *Proceedings of the DARPA Image Understanding Workshop*, pp. 235–241, 1997.

[28] L. Payá, L. Fernandez, O. Reinoso, A. Gil, and D. Ubeda, "Appearance-based dense maps creation. Comparison of compression techniques with panoramic images," in *Proceedings of the International Conference on Informatics in Control, Automation and Robotics*, pp. 238–246, INSTICC, Milan, Italy, 2009.

[29] Point Grey Research Inc, "Ladybug Cameras," https://www.ptgrey.com/360-degree-spherical-camera-systems.

[30] A. Gurtner, D. G. Greer, R. Glassock, L. Mejias, R. A. Walker, and W. W. Boles, "Investigation of fish-eye lenses for small-UAV aerial photography," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, no. 3, pp. 709–721, 2009.

[31] R. Wood, "XXIII. Fish-eye views, and vision under water," *Philosophical Magazine Series 6*, vol. 12, no. 68, pp. 159–162, 1906.

[32] Samsung, Gear 360 camera, http://www.samsung.com/global/galaxy/gear-360.

[33] S. Li, M. Nakano, and N. Chiba, "Acquisition of spherical image by fish-eye conversion lens," in *Proceedings of the IEEE Virtual Reality Conference*, pp. 235–236, IEEE, Chicago, Ill, USA, March 2004.

[34] S. Li, "Full-view spherical image camera," in *Proceedings of the 18th International Conference on Pattern Recognition (ICPR '06)*, vol. 4, pp. 386–390, August 2006.

[35] V. Peri and S. Nayar, "Generation of perspective and panoramic video from omnidirectional video," in *Proceedings of the DARPA Image Understanding Workshop (IUW '97)*, pp. 243–246, 1997.

[36] D. W. Rees, "Panoramic television viewing system," US Patent Application 3505465, 1970.

[37] S. Bogner, "An introduction to panospheric imaging," in *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics, 1995. Intelligent Systems for the 21st Century*, vol. 4, pp. 3099–3106, October 1995.

[38] K. Yamazawa, Y. Yagi, and M. Yachida, "Obstacle detection with omnidirectional image sensor HyperOmni vision," in *Proceedings of the IEEE International Conference on Robotics and Automation*, vol. 1, pp. 1062–1067, May 1995.

[39] V. Grassi Junior and J. Okamoto Junior, "Development of an omnidirectional vision system," *Journal of the Brazilian Society of Mechanical Sciences and Engineering*, vol. 28, no. 1, pp. 58–68, 2006.

[40] G. Krishnan and S. K. Nayar, "Cata-fisheye camera for panoramic imaging," in *Proceedings of the IEEE Workshop on Applications of Computer Vision (WACV 08)*, pp. 1–8, January 2008.

[41] A. Ohte, O. Tsuzuki, and K. Mori, "A practical spherical mirror omnidirectional camera," in *Proceedings of the IEEE International Workshop on Robotic Sensors: Robotic and Sensor Environments (ROSE '05)*, pp. 8–13, IEEE, Ottawa, Canada, October 2005.

[42] Y. Yagi and S. Kawato, "Panorama scene analysis with conic projection," in *Proceedings of the IEEE International Workshop on Intelligent Robots and Systems (IROS '90). Towards a New Frontier of Applications*, vol. 1, pp. 181–187, July 1990.

[43] E. Cabral, J. de Souza, and M. Hunold, "Omnidirectional stereo vision with a hyperbolic double lobed mirror," in *Proceedings of the 17th International Conference on Pattern Recognition (ICPR '04)*, vol. 1, pp. 1–9, August 2004.

[44] K. Yoshida, H. Nagahara, and M. Yachida, "An omnidirectional vision sensor with single viewpoint and constant resolution," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS '06)*, pp. 4792–4797, IEEE, October 2006.

[45] S. Baker and S. Nayar, "A theory of catadioptric image formation," in *Proceedings of the 6th International Conference on Computer Vision*, pp. 35–42, January 1998.

[46] P. Sturm, S. Ramalingam, J.-P. Tardif, S. Gasparini, and J. Barreto, "Camera models and fundamental concepts used in geometric computer vision," *Foundations and Trends in Computer Graphics and Vision*, vol. 6, no. 1-2, pp. 1–183, 2011.

[47] N. Gonçalves and H. Araújo, "Estimating parameters of non-central catadioptric systems using bundle adjustment," *Computer Vision and Image Understanding*, vol. 113, no. 1, pp. 11–28, 2009.
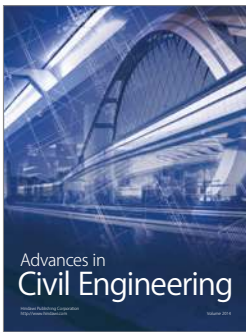
[48] X. Ying and Z. Hu, "Catadioptric camera calibration using geometric invariants," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 10, pp. 1260–1271, 2004.

[49] J. Marcato Jr., A. M. G. Tommaselli, and M. V. A. Moraes, "Calibration of a catadioptric omnidirectional vision system with conic mirror," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 113, pp. 97–105, 2016.

[50] D. Scaramuzza, A. Martinelli, and R. Siegwart, "A flexible technique for accurate omnidirectional camera calibration and structure from motion," in *Proceedings of the 4th IEEE International Conference on Computer Vision Systems (ICVS '06)*, p. 45, IEEE, January 2006.

[51] D. Scaramuzza, R. Siegwart, and A. Martinelli, "A robust descriptor for tracking vertical lines in omnidirectional images and its use in mobile robotics," *The International Journal of Robotics Research*, vol. 28, no. 2, pp. 149–171, 2009.

[52] C. Geyer and K. Daniilidis, "A unifying theory for central panoramic systems and practical implications," in *Computer Vision—ECCV 2000*, D. Vernon, Ed., vol. 1843 of *Lecture Notes in Computer Science*, pp. 445–461, Springer, Berlin, Germany, 2000.

[53] H. Friedrich, D. Dederscheck, K. Krajsek, and R. Mester, "View-based robot localization using spherical harmonics: concept and first experimental results," in *Pattern Recognition*, F. Hamprecht, C. Schnarr, and B. Jahne, Eds., vol. 4713 of *Lecture Notes in Computer Science*, pp. 21–31, Springer, Berlin, Germany, 2007.

[54] H. Friedrich, D. Dederscheck, M. Mutz, and R. Mester, "View-based robot localization using illumination-invariant spherical harmonics descriptors," in *Proceedings of the 3rd International Conference on Computer Vision Theory and Applications (VISIGRAPP '08)*, pp. 543–550, 2008.

[55] A. Makadia and K. Daniilidis, "Direct 3D-rotation estimation from spherical images via a generalized shift theorem," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2, pp. II-217–II-224, IEEE, June 2003.

[56] A. Makadia, L. Sorgi, and K. Daniilidis, "Rotation estimation from spherical images," in *Proceedings of the 17th International Conference on Pattern Recognition (ICPR '04)*, vol. 3, pp. 590–593, IEEE, August 2004.

[57] A. Makadia and K. Daniilidis, "Rotation recovery from spherical images without correspondences," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 7, pp. 1170–1175, 2006.

[58] T. Schairer, B. Huhle, and W. Strasser, "Increased accuracy orientation estimation from omnidirectional images using the spherical Fourier transform," in *Proceedings of the 3DTV Conference: The True Vision—Capture, Transmission and Display of 3D Video*, pp. 1–4, IEEE, Potsdam, Germany, May 2009.

[59] T. Schairer, B. Huhle, and W. Strasser, "Application of particle filters to vision-based orientation estimation using harmonic analysis," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA '10)*, pp. 2556–2561, May 2010.

[60] T. Schairer, B. Huhle, P. Vorst, A. Schilling, and W. Strasser, "Visual mapping with uncertainty for correspondence-free localization using gaussian process regression," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS '11)*, pp. 4229–4235, IEEE, San Francisco, Calif, USA, 2011.

[61] D. Cobzas and H. Zhang, "Cylindrical panoramic image-based model for robot localization," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, vol. 4, pp. 1924–1930, IEEE, November 2001.

[62] L. Gerstmayr-Hillen, O. Schlüter, M. Krzykawski, and R. Möller, "Parsimonious loop-closure detection based on global image-descriptors of panoramic images," in *Proceedings of the IEEE 15th International Conference on Advanced Robotics (ICAR '11)*, pp. 576–581, IEEE, June 2011.

[63] 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. 11468–11497, 2010.

[64] M. Milford and G. Wyeth, "Persistent navigation and mapping using a biologically inspired SLAM system," *The International Journal of Robotics Research*, vol. 29, no. 9, pp. 1131–1153, 2010.

[65] M. Liu, C. Pradalier, and R. Siegwart, "Visual homing from scale with an uncalibrated omnidirectional camera," *IEEE Transactions on Robotics*, vol. 29, no. 6, pp. 1353–1365, 2013.

[66] Y. Yagi, S. Fujimura, and M. Yachida, "Route representation for mobile robot navigation by omnidirectional route panorama Fourier transformation," in *Proceedings of the IEEE International Conference on Robotics and Automation*, vol. 2, pp. 1250–1255, May 1998.

[67] B. Bonev, M. Cazorla, and F. Escolano, "Robot navigation behaviors based on omnidirectional vision and information theory," *Journal of Physical Agents*, vol. 1, no. 1, pp. 27–35, 2007.

[68] S. Roebert, T. Schmits, and A. Visser, "Creating a bird-eye view map using an omnidirectional camera," in *Proceedings of the 20th Belgian-Dutch Conference on Artificial Intelligence, Universiteit Twente, Faculteit Elektrotechniek (BNAIC '08)*, pp. 233–240, Enschede, The Netherlands, October 2008.

[69] E. Trucco and K. Plakas, "Video tracking: a concise survey," *IEEE Journal of Oceanic Engineering*, vol. 31, no. 2, pp. 520–529, 2006.

[70] N. Pears and B. Liang, "Ground plane segmentation for mobile robot visual navigation," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 1513–1518, November 2001.

[71] J. Zhou and B. Li, "Homography-based ground detection for a mobile robot platform using a single camera," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA '06)*, pp. 4100–4105, IEEE, May 2006.

[72] C. Harris and M. Stephens, "A combined corner and edge detector," in *Proceedings of the 4th Alvey Vision Conference*, pp. 147–151, September 1988.

[73] S. Saripalli, G. S. Sukhatme, L. O. Mejías, and P. C. Cervera, "Detection and tracking of external features in an urban environment using an autonomous helicopter," in *Proceedings of the IEEE International Conference on Robotics and Automation*, pp. 3972–3977, IEEE, April 2005.

[74] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proceedings of the 7th IEEE International Conference on Computer Vision (ICCV '99)*, vol. 2, pp. 1150–1157, IEEE, September 1999.

[75] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[76] H. Bay, T. Tuytelaars, and L. Van Gool, "SURF: speeded up robust features," in *Computer Vision—ECCV 2006*, A. Leonardis, H. Bischof, and A. Pinz, Eds., vol. 3951 of *Lecture Notes in Computer Science*, pp. 404–417, Springer, Berlin, Germany, 2006.

[77] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-Up Robust Features (SURF)," *Computer Vision and Image Understanding*, vol. 110, no. 3, pp. 346–359, 2008.

[78] M. Calonder, V. Lepetit, C. Strecha, and P. Fua, "BRIEF: binary robust independent elementary features in," in *Proceedings of the European Conference on Computer Vision*, pp. 778–792, Springer, Crete, Greece, September 2010.

[79] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "ORB: an efficient alternative to SIFT or SURF," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV '11)*, pp. 2564–2571, IEEE, Barcelona, Spain, November 2011.

[80] S. Leutenegger, M. Chli, and R. Y. Siegwart, "BRISK: binary robust invariant scalable keypoints," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV '11)*, pp. 2548–2555, Barcelona, Spain, November 2011.

[81] A. Alahi, R. Ortiz, and P. Vandergheynst, "FREAK: fast retina keypoint," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '12)*, pp. 510–517, Providence, RI, USA, June 2012.

[82] A. Angeli, S. Doncieux, J.-A. Meyer, and D. Filliat, "Visual topological SLAM and global localization," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA '09)*, pp. 4300–4305, IEEE, Kobe, Japan, May 2009.

[83] S. Se, D. G. Lowe, and J. J. Little, "Vision-based global localization and mapping for mobile robots," *IEEE Transactions on Robotics*, vol. 21, no. 3, pp. 364–375, 2005.

[84] C. Valgren and A. J. Lilienthal, "SIFT, SURF & seasons: Appearance-based long-term localization in outdoor environments," *Robotics and Autonomous Systems*, vol. 58, no. 2, pp. 149–156, 2010.

[85] A. C. Murillo, J. J. Guerrero, and C. Sagüés, "SURF features for efficient robot localization with omnidirectional images," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA '07)*, pp. 3901–3907, April 2007.

[86] C. Pan, T. Hu, and L. Shen, "BRISK based target localization for fixed-wing UAV's vision-based autonomous landing," in *Proceedings of the IEEE International Conference on Information and Automation (ICIA '15)*, pp. 2499–2503, IEEE, Lijiang, China, August 2015.

[87] A. Gil, Ó. Reinoso, M. Ballesta, M. Juliá, and L. Payá, "Estimation of visual maps with a robot network equipped with vision sensors," *Sensors*, vol. 10, no. 5, pp. 5209–5232, 2010.

[88] A. Gil, O. M. Mozos, M. Ballesta, and O. Reinoso, "A comparative evaluation of interest point detectors and local descriptors for visual SLAM," *Machine Vision and Applications*, vol. 21, no. 6, pp. 905–920, 2010.

[89] M. Ballesta, A. Gil, O. Reinoso, and D. Úbeda, "Analysis of visual landmark detectors and descriptors in SLAM in indoor and outdoor environments," *Revista Iberoamericana de Automatica e Informatica Industrial*, vol. 7, no. 2, pp. 68–80, 2010.

[90] Y. Matsumoto, K. Ikeda, M. Inaba, and H. Inoue, "Visual navigation using omnidirectional view sequence," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS '99)*, pp. 317–322, October 1999.

[91] Y. Matsumoto, K. Sakai, M. Inaba, and H. Inoue, "Vision-based approach to robot navigation," in *Proceedings of the IEEE International Conference on Intelligent Robots and Systems*, pp. 1702–1708, Takamatsu, Japan, Novemver 2000.

[92] M. Kirby, *Geometric Data Analysis: An Empirical Approach to Dimensionality Reduction and the Study of Patterns*, A Wiley Interscience publication, John Wiley & Sons, New York, NY, USA, 2001.

[93] R. Möller, A. Vardy, S. Kreft, and S. Ruwisch, "Visual homing in environments with anisotropic landmark distribution," *Autonomous Robots*, vol. 23, no. 3, pp. 231–245, 2007.

[94] B. Kröse, R. Bunschoten, S. ten Hagen, B. Terwijn, and N. Vlassis, "Household robots look and learn: environment modeling and localization from an omnidirectional vision system," *IEEE Robotics and Automation Magazine*, vol. 11, no. 4, pp. 45–52, 2004.

[95] A. Štimec, M. Jogan, and A. Leonardis, "Unsupervised learning of a hierarchy of topological maps using omnidirectional images," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 22, no. 4, pp. 639–665, 2008.

[96] M. Jogan and A. Leonardis, "Robust localization using eigenspace of spinning-images," in *Proceedings of the IEEE Workshop on Omnidirectional Vision (OMNIVIS '00)*, pp. 37–44, Hilton Head Island, SC, USA, June 2000.

[97] M. Artač, M. Jogan, and A. Leonardis, "Mobile robot localization using an incremental eigenspace model," in *Proceedings of the IEEE International Conference on Robotics and Automation*, vol. 1, pp. 1025–1030, IEEE, May 2002.

[98] 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.

[99] F. Rossi, A. Ranganathan, F. Dellaert, and E. Menegatti, "Toward topological localization with spherical Fourier transform and uncalibrated camera in," in *Proceedings of the International Conference on Simulation, Modeling, and Programming for Autonomous Robots*, pp. 319–350, Springer, Venice, Italy, November 2008.

[100] 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.

[101] J. Košecká, L. Zhou, P. Barber, and Z. Duric, "Qualitative image based localization in indoors environments," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2, pp. II-3–II-8, June 2003.

[102] A. C. Murillo, G. Singh, J. Kosecká, and J. J. Guerrero, "Localization in urban environments using a panoramic gist descriptor," *IEEE Transactions on Robotics*, vol. 29, no. 1, pp. 146–160, 2013.

[103] A. Oliva and A. Torralba, "Building the gist of ascene: the role of global image features in recognition," in *Progress in Brain Reasearch: Special Issue on Visual Perception*, vol. 155, pp. 23–36, 2006.

[104] L. Payá, O. Reinoso, Y. Berenguer, and D. Úbeda, "Using omnidirectional vision to create a model of the environment: a comparative evaluation of global-appearance descriptors," *Journal of Sensors*, vol. 2016, Article ID 1209507, 21 pages, 2016.

[105] C. Zhou, Y. Wei, and T. Tan, "Mobile robot self-localization based on global visual appearance features," in *Proceedings of the IEEE International Conference on Robotics and Automation*, pp. 1271–1276, Taipei, Taiwan, September 2003.

[106] H. Murase and S. K. Nayar, "Visual learning and recognition of 3-d objects from appearance," *International Journal of Computer Vision*, vol. 14, no. 1, pp. 5–24, 1995.

[107] V. P. de Araújo, R. D. Maia, M. F. S. V. DAngelo, and G. N. R. DAngelo, "Automatic plate detection using genetic algorithm," in *Proceedings of the 6th WSEAS International Conference on Signal, Speech and Image Processing, World Scientific and Engineering Academy and Society (WSEAS '06)*, pp. 43–48, Stevens Point, Wis, USA, 2006.

[108] R. C. González and R. E. Woods, *Digital Image Processing*, Addison-Wesley, Boston, Mass, USA, 3rd edition, 1992.

[109] S. Thrun, "Robotic mapping: a survey, in exploring artificial intelligence in," in *The New Milenium*, pp. 1–35, Morgan Kaufmann, San Francisco, Calif, USA, 2003.

[110] E. Garcia-Fidalgo and A. Ortiz, "Vision-based topological mapping and localization methods: a survey," *Robotics and Autonomous Systems*, vol. 64, pp. 1–20, 2015.

[111] K. Konolige, E. Marder-Eppstein, and B. Marthi, "Navigation in hybrid metric-topological maps," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA '11)*, pp. 3041–3047, May 2011.

[112] M. Li, K. Imou, S. Tani, and S. Yokoyama, "Localization of agricultural vehicle using landmarks based on omni-directional vision," in *Proceedings of the International Conference on Computer Engineering and Applications (ICCEA '09)*, vol. 2, pp. 129–133, IACSIT Press, Singapore, 2009.

[113] H. Lu, X. Li, H. Zhang, and Z. Zheng, "Robust place recognition based on omnidirectional vision and real-time local visual features for mobile robots," *Advanced Robotics*, vol. 27, no. 18, pp. 1439–1453, 2013.

[114] Y.-W. Choi, K.-K. Kwon, S.-I. Lee, J.-W. Choi, and S.-G. Lee, "Multi-robot mapping using omnidirectional-vision SLAM based on fisheye images," *ETRI Journal*, vol. 36, no. 6, pp. 913–923, 2014.

[115] C. Valgren, A. Lilienthal, and T. Duckett, "Incremental topological mapping using omnidirectional vision," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS '06)*, pp. 3441–3447, Beijing, China, October 2006.

[116] 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.

[117] H. Lu, X. Li, H. Zhang, and Z. Zheng, "Robust place recognition based on omnidirectional vision and real-time local visual features for mobile robots," *Advanced Robotics*, vol. 27, no. 18, pp. 1439–1453, 2013.

[118] M. Liu, D. Scaramuzza, C. Pradalier, R. Siegwart, and Q. Chen, "Scene recognition with omnidirectional vision for topological map using lightweight adaptive descriptors," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS '09)*, pp. 116–121, IEEE, St. Louis, Mo, USA, October 2009.

[119] I. Ulrich and I. Nourbakhsh, "Appearance-based place recognition for topological localization," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA '00)*, pp. 1023–1029, April 2000.

[120] L. Payá, F. Amorós, L. Fernández, and O. Reinoso, "Performance of global-appearance descriptors in map building and localization using omnidirectional vision," *Sensors*, vol. 14, no. 2, pp. 3033–3064, 2014.

[121] A. Rituerto, A. C. Murillo, and J. J. Guerrero, "Semantic labeling for indoor topological mapping using a wearable catadioptric system," *Robotics and Autonomous Systems*, vol. 62, no. 5, pp. 685–695, 2014.

[122] 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.

[123] B. Huhle, T. Schairer, A. Schilling, and W. Straßer, "Learning to localize with Gaussian process regression on omnidirectional image data," in *Proceedings of the 23rd IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS '10)*, pp. 5208–5213, October 2010.

[124] A. Rituerto, L. Puig, and J. Guerrero, "Comparison of omni-directional and conventional monocular systems for visual SLAM," in *Proceedings of the Workshop on Omnidirectional Vision, Camera Networks and Non-classical Cameras (OMNIVIS '10)*, Zaragoza, Spain, 2010.

[125] C. Burbridge, L. Spacek, J. Condell, and U. Nehmzow, "Monocular omnidirectional vision based robot localisation and mapping," in *TAROS 2008—Towards Autonomous Robotic Systems 2008, The University of Edinburgh, United Kingdom*, pp. 135–141, IEEE Computer Society Press, 2008, http://uir.ulster.ac.uk/8898/.

[126] H. Erturk, G. Tuna, T. V. Mumcu, and K. Gulez, "A performance analysis of omnidirectional vision based simultaneous localization and mapping," in *Intelligent Computing Technology: 8th International Conference, ICIC 2012, Huangshan, China, July 25–29, 2012. Proceedings*, vol. 7389 of *Lecture Notes in Computer Science*, pp. 407–414, Springer, Berlin, Germany, 2012.

[127] S. Kim and S. Oh, "SLAM in indoor environments using omnidirectional vertical and horizontal line features," *Journal of Intelligent and Robotic Systems*, vol. 51, no. 1, pp. 31–43, 2008.

[128] K. Wang, G. Xia, Q. Zhu, Y. Yu, Y. Wu, and Y. Wang, "The SLAM algorithm of mobile robot with omnidirectional vision based on EKF," in *Proceedings of the IEEE International Conference on Information and Automation (ICIA '12)*, pp. 13–18, IEEE, Shenyang, China, June 2012.

[129] D. Caruso, J. Engel, and D. Cremers, "Large-scale direct SLAM for omnidirectional cameras," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS '15)*, pp. 141–148, Hamburg, Germany, October 2015.

[130] D. Valiente, A. Gil, L. Fernández, and Ó. Reinoso, "A comparison of EKF and SGD applied to a view-based SLAM approach with omnidirectional images," *Robotics and Autonomous Systems*, vol. 62, no. 2, pp. 108–119, 2014.

[131] L. Fernández, L. Payá, O. Reinoso, and L. M. Jimenez, "Appearance-based approach to hybrid metric-topological simultaneous localisation and mapping," *IET Intelligent Transport Systems*, vol. 8, no. 8, pp. 688–699, 2014.

[132] S. Thompson, T. Matsui, and A. Zelinsky, "Localisation using automatically selected landmarks from panoramic images," in *Proceedings of the Australian Conference on Robotics and Automation (ARAA '00)*, pp. 167–172, Melbourne, Australia, 2000.

[133] A. A. Argyros, K. E. Bekris, S. C. Orphanoudakis, and L. E. Kavraki, "Robot homing by exploiting panoramic vision," *Autonomous Robots*, vol. 19, no. 1, pp. 7–25, 2005.

[134] M. Lourenço, V. Pedro, and J. P. Barreto, "Localization in indoor environments by querying omnidirectional visual maps using perspective images," in *Proceedings of the IEEE International Conference on Robotics and Automation*, pp. 2189–2195, St. Paul, Minn, USA, May 2012.

[135] Y. Matsumoto, M. Inaba, and H. Inoue, "Memory-based navigation using omni-view sequence," in *Field and Service Robotics*, pp. 171–178, Springer, London, UK, 1998.

[136] Y. Berenguer, L. Payá, M. Ballesta, and O. Reinoso, "Position estimation and local mapping using omnidirectional images and global appearance descriptors," *Sensors*, vol. 15, no. 10, pp. 26368–26395, 2015.

[137] A. Murillo, J. Guerrero, and C. Sagues, "SURF features for efficient robot localization with omnidirectional images," in

Proceedings of the IEEE International Conference on Robotics and Automation, pp. 3901–3907, Roma, Italy, April 2007.

[138] F. Amorós, L. Payá, O. Reinoso, D. Valiente, and L. Fernández, "Towards relative altitude estimation in topological navigation tasks using the global appearance of visual information," in *Proceedings of the 9th International Conference on Computer Vision Theory and Applications (VISAPP '14)*, vol. 1, pp. 194–201, SciTePress—Science and Technology Publications, January 2014.

International Journal of
Rotating
Machinery

The Scientific
World Journal

Journal of
Sensors

International Journal of
Distributed
Sensor Networks

Journal of
Engineering

Advances in
Civil Engineering

Journal of
Control Science
and Engineering

Journal of
Robotics

Submit your manuscripts at
https://www.hindawi.com

Journal of
Electrical and Computer
Engineering

Advances in
OptoElectronics

VLSI Design

International Journal of
Navigation and
Observation

Modelling &
Simulation
in Engineering

International Journal of
Aerospace
Engineering

International Journal of
Chemical Engineering

International Journal of
Antennas and
Propagation

Active and Passive
Electronic Components

Shock and Vibration

Advances in
Acoustics and Vibration