SLAM Algorithm by using Global Appearance of Omnidirectional Images

Yerai Berenguer, Luis Payá, Adrián Peidró and Oscar Reinoso
Departamento de Ingeniería de Sistemas y Automática, Miguel Hernández University, Spain

Keywords: SLAM, Global Appearance, Omnidirectional Images, Vision systems, Image description, Loop Closure

Abstract: This work presents a SLAM algorithm to estimate the position and orientation of a mobile robot while simultaneously creating the map of the environment. It uses only visual information provided by a catadioptric system mounted on the robot formed by a camera pointing towards a convex mirror. It provides the robot with omnidirectional images that contain information with a field of view of 360 degrees around the camera-mirror axis. Each omnidirectional scene acquired by the robot is described using global appearance descriptors. Thanks to their compactness, this kind of descriptors permits running the algorithm in real time. The method consists of three different steps. First, the robot calculates the pose of the robot (location and orientation) and creates a new node in the map. This map is formed by connected nodes between them. Second, it detects loop closures between the new node and the nodes of the map. Finally, the map is optimized by using an optimization algorithm and the detected loop closures. Two different sets of images have been used to test the effectiveness of the method. They were captured in two real environments, while the robot traversed two paths. The results of the experiments show the effectiveness of our method.

1 INTRODUCTION

The Simultaneous Localization and Mapping (SLAM) has been regarded as one of the most important technologies in mobile robot research during the last few years (Munguia et al., 2016; Whelan et al., 2016). Many of these works are focused on the use of visual information to carry out SLAM algorithms, due to the rich information provided by the visual sensors.

In addition to this, visual sensors can be configured in different ways like conventional cameras, stereo systems, array of cameras and catadioptric systems. This last configuration is formed by a single-viewpoint camera pointing towards a convex mirror (Nagahara et al., 2007). The resulting image captured by this last kind of camera contains information on a field of view of 360 degrees around the camera-mirror axis. They are named omnidirectional images.

In the field of SLAM, omnidirectional images have many advantages because they contain information with a field of view of 360 degrees around the mirror axis and the robot does not need to rotate. We can find many previous works that use omnidirectional images in mapping and localization tasks. For example, Valiente et al. (Valiente et al., 2014) present a comparison between two different visual SLAM methods using omnidirectional images and Garcia et al. (Garcia-Fidalgo and Ortiz, 2015) make a survey of vision-based topological mapping and localization methods.

Traditionally, the developments in mobile robots using visual sensors are based on the extraction and description of some landmarks from the scenes, such as SIFT (Scale-Invariant Feature Transform) (Lowe, 1999) and SURF (Speeded-Up Robust Features) (Bay et al., 2006) descriptors. This approach presents some disadvantages: the computational time to calculate and compare the descriptors is usually high, and it leads to relatively complex mapping and localization algorithms. As an advantage, only few positions need to be stored in the map to make the localization process possible.

More recently, some works propose using the global information to describe the scenes, creating a unique descriptor per image. These techniques have demonstrated to be a good option to solve the localization and navigation problems when the movement of the robot is contained in the floor plane. For example, Chang et al. (Chang et al., 2010) presents a vision-based navigation and localization system using the gist descriptor, Payá et al. (Payá et al., 2010) use

Berenguer, Y., Payá, L., Peidró, A. and Reinoso, O.
SLAM Algorithm by using Global Appearance of Omnidirectional Images.
DOI: 10.5220/0006434503820388
In *Proceedings of the 14th International Conference on Informatics in Control, Automation and Robotics (ICINCO 2017) - Volume 2*, pages 382-388
ISBN: Not Available

a descriptor based on the Fourier signature in Monte Carlo localization tasks, and Wu et al. (Wu et al., 2014) propose an efficient visual loop closure detection method. In (Payá et al., 2014), several methods to obtain global descriptors from panoramic scenes are analyzed and compared to demonstrate their validity in map building and localization. The majority of these global appearance descriptors can be used in real time because the computational time to calculate and handle them is low, and they usually lead to more straightforward mapping and localization algorithms.

Sometimes, the mapping process produces an error measure in each map position due to the iterative calculation of new poses of the robot. This can be a big problem in extensive environments when the robot has to calculate many new poses because the error is increasing in each iteration. This uncertainty can be reduced by detecting loop closures and using optimization algorithms to relocate each previous pose. This problem is thoroughly studied in this work.

The contribution of this work is the creation of a method to carry out SLAM tasks by only using visual information of the environment and global appearance descriptors. Each omnidirectional scene acquired by the robot is described using global appearance descriptors. The method consists of three different steps: calculating the pose of the robot (location and orientation), detecting loop closures (by comparing global appearance descriptors) and optimizing the map (by using the G2O optimization algorithm). The optimization algorithm used in the presented paper is named G2O and it was presented by Kümerle et al. (Kümmerle et al., 2011).

The experiments have been carried out with two different sets of images captured while the robot traversed two real working environments. The first one has been taken following a rectangular path indoors. The second one has been captured following a real path including several rooms in a building.

The remainder of this paper is structured as follows. Section 2 introduces some preliminary concepts about image description and graph optimization. Section 3 presents the SLAM algorithm we have implemented to solve the simultaneous localization and mapping problem. Section 4 describes our databases used to carry out the experiments and presents the experiments and results. At last, section 5 outlines the conclusions.

2 PRELIMINARIES

Along the paper, two methods are used to describe the global appearance of scenes: the Radon transform and the Histogram of Oriented Gradients (HOG). This section includes some information on them. Also, we present the fundamentals of the methods used to calculate the difference between two images captured from different locations. At last, we describe the optimization algorithm used to recalculate the previous map positions after detecting loop closures.

2.1 Global Appearance Descriptors

Methods based on the global appearance of the scenes constitute a robust alternative to methods based on landmark extraction and description. This is because global appearance descriptors represent the environment through high level features that can be interpreted and handled easily.

This subsection presents the image descriptors we have used to describe the omnidirectional images. Both of them are based on global appearance, without any segmentation or local landmark extraction.

2.1.1 Radon Transform

The Radon transform was described initially in (Radon, 1917). Previous research demonstrates the efficacy of this descriptor in shape description and segmentation such as (Hoang and Tabbone, 2010) and (Hasegawa and Tabbone, 2011). Hoang et al. (Hoang and Tabbone, 2010) present a shape descriptor, invariant to geometric transformations, based on the Radon, Fourier and Mellin transforms, and Hasegawa et al. (Hasegawa and Tabbone, 2011) describe a shape descriptor combining the histogram of the Radon transform, the logarithmic-scale histogram and the phase-only correlation. Berenguer et al. (Berenguer et al., 2015) present a 2D localization method using a global appearance descriptor based on the Radon transform. They demonstrate the effectiveness and robustness of this descriptor.

Mathematically, the Radon transform of an image $im(i,j) \in \mathbb{R}^{KxL}$ along the line $c_1(\phi,d)$ (Figure 1) can be obtained as:

$$\begin{aligned}\mathcal{R}\{im(i,j)\} &= \lambda_f(\phi,d) = \\ &= \int_{\mathbb{R}} im(d\cos\phi - j'\sin\phi, d\sin\phi + j'\cos\phi)\, dj' \quad (1)\end{aligned}$$

where $\mathcal{R}$ is the Radon transform operator. $im(i,j)$ is the image to transform. λ_f is the transformed function, which depends on two new variables: the distance from the line c_1 to the origin d and the angle between the x axis and the i' axis, ϕ (Figure 1). j' axis is parallel to the c_1 line.

By considering different values for d and ϕ in Equation (1), the transformed function $\lambda_f(\phi,d)$ will become a matrix with M rows and N columns. M is the number of orientations considered (normally chosen to cover the whole circumference), and N is the number of parallel lines considered at each orientation (to cover the whole image). The distance between each pair of consecutive lines is considered constant.

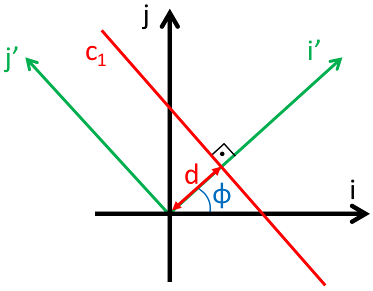


Figure 1: Line parametrization through the distance origin d and the angle between the normal line and the i axis, ϕ.

When the Radon transform is applied to an image, it calculates the image projections along the specified directions through a cluster of line integrals along parallel lines in this direction. The distance between the parallel lines is usually one pixel.

2.1.2 Histogram of Oriented Gradients (HOG)

HOG has been used traditionally as a description method in the field of object detection. It was initially described by (Dalal and Triggs, 2005). They used it in people detection tasks. However, there are several researches in which this description method has been improved, such as (Zhu et al., 2006), who improve the accuracy and the computational cost.

The basic implementation consists in dividing the image in small connected cells and the histogram of gradient orientations is calculated in each cell. Then, the descriptor is composed of these histograms arranged in a single vector.

Fernandez et al. (Fernández et al., 2016) analyze this kind of descriptor in outdoor localization tasks. Furthermore, they make a comparative analysis between several methods to describe outdoor panoramic images.

2.2 Phase Only Correlation (POC)

In this subsection we present the method used to compare the Radon transform of two images. It provides a measurement of difference between the visual appearance of two locations and estimation of the change of orientation of the robot between these locations.

POC (Phase Only Correlation), proposed in (Kuglin and Hines, 1975), is an operation made in the frequency domain that provides a correlation coefficient between two images (Kobayashi et al., 2004). In our case we compare two Radon transforms but this does not affect the POC performance because the Radon transform can be interpreted as an image. In general, it permits obtaining both the relative orientation between two different Radon transforms and a similitude coefficient between them, as shown in (Berenguer et al., 2015).

The correspondence between two images $im_1(i,j)$ and $im_2(i,j)$ calculated by POC is given by the following equation:

$$C(i,j) = \mathcal{F}^{-1}\left\{\frac{\mathbf{IM}_1(u,v)\cdot\mathbf{IM}_2^*(u,v)}{|\mathbf{IM}_1(u,v)\cdot\mathbf{IM}_2^*(u,v)|}\right\} \quad (2)$$

Where $\mathbf{IM}_1$ is the Fourier transform of the image 1 and $\mathbf{IM}_2^*$ is the conjugate of the Fourier transform of the image 2. $\mathcal{F}^{-1}$ is the inverse Fourier transform operator.

To estimate the distance between the two images (im_1 and im_2) we have used the following expression:

$$dist(im_1, im_2) = 1 - max\{C(i,j)\} \quad (3)$$

$max\{(C(i,j)\}$ is a coefficient that takes values in the interval $[0,1]$ and it measures the similitude between the two images $im_1 and im_2$.

This operation is invariant against shifts of the images along the i and j axes. Furthermore, it is possible to estimate these shifts Δ_x and Δ_y along both axes by:

$$(\Delta_x, \Delta_y) = argmax_{(i,j)}\{C(i,j)\} \quad (4)$$

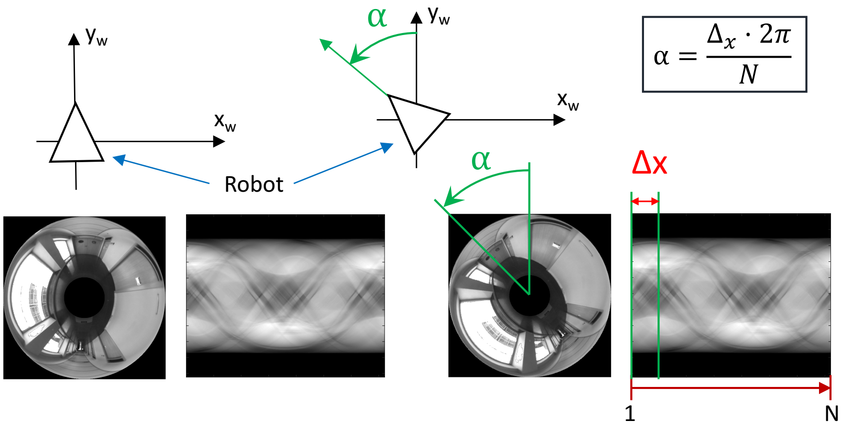
If we compare the Radon transforms of two omnidirectional images using POC, the value Δ_x is proportional to the relative orientation α of the robot when capturing the images according to Equation (5). The Figure 2 shows the Radon transforms of two different omnidirectional images captured from the same point (x_w, y_w, z_w) but with different robot orientation with respect to the z_w axis, θ (Figure 2).

$$\alpha = \frac{\Delta_x \cdot 2\pi}{N} \quad (5)$$

This way, POC is able to compare two images independently on the orientation and it is also able to estimate this change in orientation.

2.3 Optimization Algorithm: G2O

G2O is an optimization algorithm described in (Kümmerle et al., 2011). This method was created for optimizing graph-based nonlinear error functions.



(a) (b)

Figure 2: (a) Omnidirectional image captured from a specific position of a virtual environment and its Radon transform. (b) Omnidirectional image taken from the same location changing only the robot orientation around the z_w axis, and its Radon transform. A change in the robot orientation around the z_w axis produces a shift in the columns of the Radon transform, Δx.

In the field of SLAM, the robot has to calculate its pose when it takes every new image with respect to the previous poses included in the map. This operation has an error associated that increments in each pose calculation, so we need to correct the poses to decrease this deviation. G2O is able to recalculate each pose of the map using new pose restrictions. One of these restrictions can be obtained when loop closures between one pose of the existing map and the new pose of the robot occur. Then, G2O relocates each pose of the map gradually modifying them to fulfill the loop closure restriction.Then, the new pose is located the same position than the equivalent pose stored in the map.

3 SLAM METHOD

In this section, we introduce our visual SLAM approach. The robot goes through the environment and captures images from some positions. Every time a new image arrives, the robot includes a new node inside the map. This map is formed by nodes. Then, the SLAM problem is solved, following these three steps:

First, the robot calculates two descriptors of the image: the Radon transform and the HOG descriptor; and stores them in the node. Then, the robot creates a new node and locates it inside the map calculating the position and orientation of the new node with respect to the previously added node, and both nodes are connected. This localization process is carried out by using only visual information.

Second, the robot checks the existence of possible loop closures comparing the new scene with the previous scenes stored in the map.

Finally, the map is optimized by using the G2O algorithm with the loop closures detected. This process is repeated in each new location.

3.1 Creating the Map

This subsection presents the method proposed to calculate the coordinates of each new node location. These new poses are estimated by calculating the distance and the angle between scenes. For each new node, the robot stores the Radon descriptor and the HOG descriptor of the new omnidirectional image taken to do possible the localization of it.

Figure 3 shows a scheme of the mapping process. It consist of the calculation of the (x_k, y_k) coordinates of each new node. These coordinates are calculated from the distance and angle between poses.

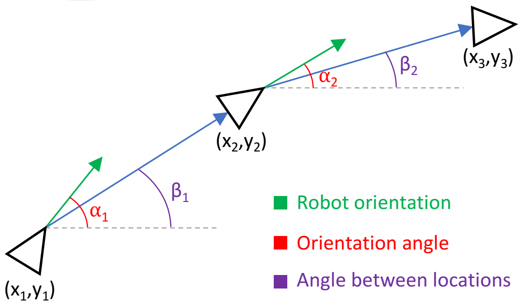


Figure 3: Mapping process scheme.

The distance between locations is calculated using the Equation 3. It is an image distance and is not a metric distance, i.e. this distance is not an actual measurement unit, but it is proportional to the metric distance.

The calculation of the angle between node positions is an approximation, because we consider that the angle between locations, β_k, is approximately the orientation angle, α_j, computed using Equation 5:

$$\beta_k \approx \alpha_k \tag{6}$$

The closer the poses are, the more accurate the approximation is. It is because the robot orientation change is smaller, and the error made in the approximation is reduced. However, it will be reduced in the optimization process.

Finally, the node coordinates (x_k, y_k) are calculated by these equations:

$$x_k = dist(im_{k-1}, im_k) \cdot cos(\alpha_k) \tag{7}$$

$$y_k = dist(im_{k-1}, im_k) \cdot sin(\alpha_k) \tag{8}$$

where $dist(im_{k-1}, im_k)$ is the POC distance between two consecutive images, calculated by using the Equation 3. And α_k is the orientation angle of the k node, calculated by using Equation 5.

3.2 Loop Closures

The next step of the algorithm consists in detect loop closures, i.e. comparing the HOG descriptor of the new image taken by the robot with the HOG descriptors stored in the map. To calculate the distance between HOG descriptors we use the cosine similarity between them to calculate the distance:

$$dist(\overrightarrow{d_1}, \overrightarrow{d_2}) = 1 - \frac{\overrightarrow{d_1} \cdot \overrightarrow{d_2}'}{\sqrt{(\overrightarrow{d_1} \cdot \overrightarrow{d_1}')(\overrightarrow{d_2} \cdot \overrightarrow{d_2}')}} \tag{9}$$

where $\overrightarrow{d_1}$ and $\overrightarrow{d_2}$ are the HOG descriptors of two different images.

The loop closures have to be determined by defining a maximum threshold of distance,W (Equation 10). This threshold is defined as a constant in the beginning of the SLAM process. If the distance is lower than this threshold, the two poses compared will be considered as the same location (x,y), but the orientation of the robot can be different.

$$if(HOG_{distance} < W) \rightarrow loop\ closure \tag{10}$$

3.3 Optimization of the Map

Taking the detected loop closures into account, the robot uses this information to optimize the stored map. This optimization is made by using the G2O optimization algorithm.

When the robot detects a loop closure, it has to relocate all previous nodes to reduce the error associated in each node position. This process modifies all the node positions in the map to accomplish the new restriction calculated by the loop closure detection.

The nodes location modification is made by the G2O algorithm. It receive as input all the node positions of the map and the loop closure restriction. Then, G2O gives as an output the new recalculated node positions.

Therefore, the two nodes of the loop closure are localized in the same position and the coordinates of the rest of the map nodes are modified.

4 EXPERIMENTS

This subsection presents the different sets of omnidirectional images used to test our method and the results obtained in these experiments.

4.1 Databases

In order to check the performance of the proposed technique, two sets of images captured by ourselves are used. To capture the first set, the robot was teleoperated to follow a rectangular path. The second set was captured while the robot followed a more complicated path through several rooms inside a building. Figure 4 shows a sample omnidirectional image of each environment.

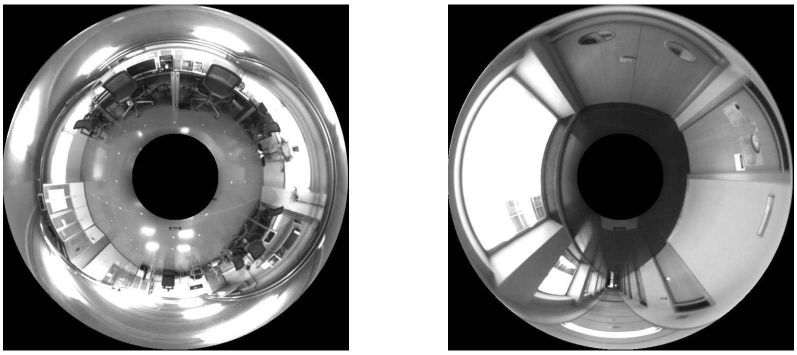
These two databases have been created taking one new omnidirectional image every 40 cm approximately. The Figure 5 shows the omnidirectional acquisition system used to capture the omnidirectional images, formed by the camera (model: DFK-41BF02) and the hyperbolic mirror (model: Eizo Wide70).

4.2 Results

In this section the results of the experiments with our SLAM algorithm are shown. The two databases described in section 4.1 have been used to carry out these experiments.

The maximum threshold of distance between HOG descriptors is an important parameter to tune. To do that, we have made some tests and chosen the best value to detect loop closures. After these tests we consider a threshold equal to 0.006 as a good value of distance between HOG descriptors.

Figure 6 shows the results of the SLAM algorithm after incorporating the final position of the first path. The blue line is the map created without optimization and the green line is the same map optimized. This



(a) (b)

Figure 4: (a) Sample omnidirectional image of the rectangular path. (b) Sample image of the second path.

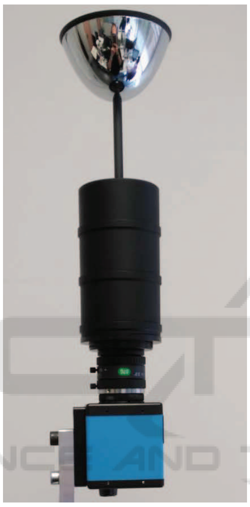


Figure 5: Omnidirectional acquisition system.

optimization is made in each iteration but the map without any optimization is shown to view the difference. As we can see, the green line is a square path.

Figure 7 shows the same results than in Figure 6 but using the second path. The blue line is the map created without optimization, the green line is the same map optimized and the yellow line is the ground truth.

As for the computational time, the robot spends an average of 0.65 seconds in each iteration of the SLAM process. This time is increasing in each iteration because the map is formed by larger amount of nodes and the loop closure detection needs to compare an higher number of HOG descriptors.

5 CONCLUSIONS

In this paper we have presented a SLAM method to estimate the position and orientation of a mobile robot in an environment while the robot is creating the map. We use two different global appearance descriptors to carry out the SLAM process and the map is formed by these two descriptors of each image. At last, the algorithm has been tested with two sets of images captured in two different indoor environments.

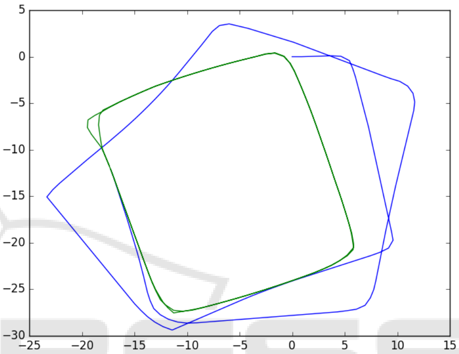


Figure 6: Map created using the first path. The blue line is the map created without optimization and the green line is the same map optimized.

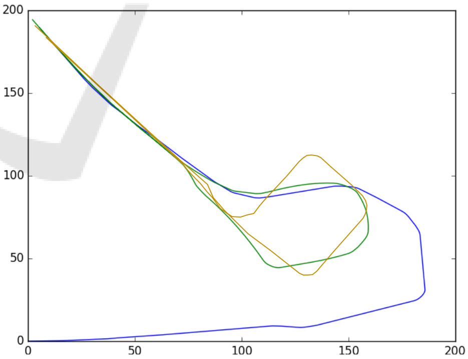


Figure 7: Map created using the real path. The blue line is the map created without optimization, the green line is the same map optimized and the yellow line is the ground truth.

The results have demonstrated the accuracy of the method. As for the values of the parameters, the maximum threshold of distance between HOG descriptors is the main tuning parameter in this method.

The results presented in this paper show the effectiveness of the global appearance descriptors of omnidirectional images to do SLAM thanks to the richness of the information they contain. We are now working to improve this method and we are trying to estimate more accurately the relative orientation between nodes. Furthermore, we are implementing a clustering method to reduce the computational time to detect loop closures when the number of nodes is increased.

ACKNOWLEDGEMENTS

This work has been supported by the Spanish government through the project DPI2016-78361-R (AEI/FEDER, UE) "*Creación de Mapas Mediante Métodos de Apariencia Visual para la Navegación de Robots*'."

REFERENCES

Bay, H., Tuytelaars, T., and Gool, L. (2006). Surf: Speeded up robust features. *Computer Vision at ECCV*, 3951:404–417.

Berenguer, Y., Payá, L., Ballesta, M., and Reinoso, O. (2015). Position estimation and local mapping using omnidirectional images and global appearance descriptors. *Sensors*, 15(10):26368.

Chang, C., Siagian, C., and Itti, L. (2010). Mobile robot vision navigation and localization using gist and saliency. In *IROS 2010, International Conference on Intelligent Robots and Systems*, pages 4147–4154.

Dalal, N. and Triggs, B. (2005). Histograms of oriented gradients for human detection. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, volume 1, pages 886–893 vol. 1.

Fernández, L., Payá, L., Reinoso, O., Jiménez, L., and Ballesta, M. (2016). A study of visual descriptors for outdoor navigation using google street view images. *Journal of Sensors*, 2016.

Garcia-Fidalgo, E. and Ortiz, A. (2015). Vision-based topological mapping and localization methods: A survey. *Robotics and Autonomous Systems*, 64:1 – 20.

Hasegawa, M. and Tabbone, S. (2011). A shape descriptor combining logarithmic-scale histogram of radon transform and phase-only correlation function. In *2011 International Conference on Document Analysis and Recognition (ICDAR)*, pages 182–186.

Hoang, T. and Tabbone, S. (2010). A geometric invariant shape descriptor based on the radon, fourier, and mellin transforms. In *20th International Conference on Pattern Recognition (ICPR)*, pages 2085–2088.

Kobayashi, K., Aoki, T., Ito, K., Nakajima, H., and Higuchi, T. (2004). A fingerprint matching algorithm using phase-only correlation. *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, pages 682–691.

Kuglin, C. and Hines, D. (1975). The phase correlation image alignment method. In *In Proceedings of the IEEE, International Conference on Cybernetics and Society*, pages 163–165.

Kümmerle, R., Grisetti, G., Strasdat, H., Konolige, K., and Burgard, W. (2011). G2o: A general framework for graph optimization. In *2011 IEEE International Conference on Robotics and Automation*, pages 3607–3613.

Lowe, D. (1999). Object recognition from local scale-invariant features. In *ICCV 1999, International Conference on Computer Vision*, volume 2, pages 1150–1157.

Munguia, R., Urzua, S., and Grau, A. (2016). Delayed monocular slam approach applied to unmanned aerial vehicles. *PLOS ONE*, 11(12):1–24.

Nagahara, H., Yoshida, K., and Yachida, M. (2007). An omnidirectional vision sensor with single view and constant resolution. In *2007 IEEE 11th International Conference on Computer Vision*, pages 1–8.

Payá, L., Amorós, F., Fernández, L., and Reinoso, O. (2014). Performance of global-appearance descriptors in map building and localization using omnidirectional vision. *Sensors*, 14(2):3033–3064.

Payá, L., Fernández, L., Gil, L., and Reinoso, O. (2010). Map building and monte carlo localization using global appearance of omnidirectional images. *Sensors*, 10(12):11468–11497.

Radon, J. (1917). Uber die bestimmung von funktionen durch ihre integralwerte langs gewisser mannigfaltigkeiten. *Berichte Sachsische Akademie der Wissenschaften*, 69(1):262–277.

Valiente, D., Gil, A., Fernández, L., and Reinoso, O. (2014). A comparison of ekf and sgd applied to a view-based slam approach with omnidirectional images. *Robotics and Autonomous Systems*, 62(2):108 – 119.

Whelan, T., Salas-Moreno, R. F., Glocker, B., Davison, A. J., and Leutenegger, S. (2016). Elasticfusion: Real-time dense slam and light source estimation. *The International Journal of Robotics Research*, 35(14):1697–1716.

Wu, J., Zhang, H., and Guan, Y. (2014). An efficient visual loop closure detection method in a map of 20 million key locations. In *2014 IEEE International Conference on Robotics and Automation (ICRA)*, pages 861–866.

Zhu, Q., Yeh, M.-C., Cheng, K.-T., and Avidan, S. (2006). Fast human detection using a cascade of histograms of oriented gradients. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, volume 2, pages 1491–1498.