APPEARANCE-BASED VISUAL ODOMETRY WITH OMNIDIRECTIONAL IMAGES

A Practical Application to Topological Mapping

Lorenzo Fernández, Luis Payá, Oscar Reinoso and Francisco Amorós
Departamento de Ingeniera de Sistemas y Automtica, Miguel Hernández University
Avda. de la Universidad s/n, Alicante, Elche, Spain

Keywords: Visual odometry, Global appearance, Omnidirectional images, Fourier descriptor, Visual compass.

Abstract: In this paper we deal with the problem of map creation and localization of a mobile robot using omnidirectional images. We describe a real-time algorithm for topological mapping, using as input data only a set of images captured by a single omnidirectional camera mounted at a fixed position on the mobile robot. To compute the topological relationships between locations in the map, we use techniques based in the global appearance of the images. When using these methods, it is important to remove redundant information to get an acceptable computational cost when comparing locations. With this aim, we describe each omnidirectional image by a single Fourier descriptor that represents the appearance as well as the relative orientation between images. This algorithm permits computing the relative topological position of a location with respect to the previous one, acting as a visual compass. We have carried out a complete set of experiments to study the validity of the proposed visual odometry and topological mapping and to perform an objective comparison between the results obtained using the robot odometry, our visual odometry and the ground truth. We have also checked the time consumption to carry out the process and the geometrical accuracy obtained comparing to the ground truth.

1 INTRODUCTION

The design of algorithms to perform some tasks autonomously in a real environment is a key point in mobile robotics. In this field, an essential problem to consider is the computation of the localization of the autonomous vehicle in the environment. It is important that the mobile robot has a map or an internal representation of the environment, so that the robot can make decisions about its localization and about the path to follow to move from its current position to the target points. Omnidirectional vision systems are commonly used at this kind of applications due to their relative low cost and the richness of the information they provide. Different representations of the visual information can be used when working with these catadioptric systems, such as the omnidirectional, panoramic and bird eye view images. In this paper, we use the panoramic representation of the scenes as it can offer invariance to ground-plane rotations, and it allows us to use only the information provided by the vision sensor to perform the process.

Different authors have studied how to use the omnidirectional images both to solve the mapping and the localization problems. We can categorize these solutions into two main groups:

- Feature-based solutions, in which a number of significant points or landmarks from each image are extracted and each point is described using an invariant descriptor. For example, (Murillo et al., 2007) and (Valgren and Lilienthal, 2010) use SURF features (Bay et al., 2008) extracted from a set of omnidirectional images to find the localization of the robot in a given map.
- Appearance-based solutions, in which the whole appearance of the omnidirectional image is represented by a single descriptor, with no local feature extraction. For example (Menegatti et al., 2004b) present a method to build a topological map using a Fourier descriptor of omnidirectional images and (Payá et al., 2010) perform a probabilistic localization in some environments.

Using appearance-based techniques is useful when working in unstructured environments and offer an intuitive way to construct the map and to get

Fernández L., Payá L., Reinoso O. and Amorós F..
APPEARANCE-BASED VISUAL ODOMETRY WITH OMNIDIRECTIONAL IMAGES - A Practical Application to Topological Mapping.
DOI: 10.5220/0003533602050210
In *Proceedings of the 8th International Conference on Informatics in Control, Automation and Robotics* (ICINCO-2011), pages 205-210
ISBN: 978-989-8425-75-1
Copyright © 2011 SCITEPRESS (Science and Technology Publications, Lda.)

the position. However, as no relevant information is extracted from the images, it is necessary to apply a compression method to reduce the computational cost of the mapping and localization processes. Several researchers have developed DFT (Discrete Fourier Transform) methods to get the most relevant information from the images (Menegatti et al., 2004a). These descriptors present rotational ground-plane invariance, concentrate the most relevant information in the low frequency components of the transformed images and each image descriptor is computed independently of the rest of images.

For these reasons, and based on some prior works (Payá et al., 2009; Payá et al., 2010; Fernández et al., 2010), we have decided to describe each omnidirectional image by means of a Fourier descriptor. We use the Fourier Signature (Menegatti et al., 2004a) to compress each image captured. The processing time needed to compute the Fourier Signature is noticeably lower than in other common feature extraction algorithms (Payá et al., 2009), and it permits a fast comparison between the images in the map by means of a vector distance measurement. Also, when using the Fourier Signature, we exploit better the invariance to ground-plane rotations in panoramic images, property that will be of utmost importance when deploying our visual compass.

In this paper we present a methodology to build a topological map using the global appearance of the panoramic images to model the topological relationships between successive nodes in the map. We use the Fourier Signature to get a robust descriptor that allow us to work in real-time. However, the methods described here are independent of the descriptor used to represent the images, and other appearance-based descriptors may also be applied. To represent the distance between two consecutive poses we have used the normalized Euclidean distance between their Fourier Signatures and to get the relative angle between them we have implemented a Fourier-based visual compass. Our main objective consists in evaluating the feasibility of using purely global-appearance methods in these tasks and how the main features of the descriptor influence the final result.

The paper is organized as follows. Section 2 presents the fundamentals of topological mapping approaches. In section 3, we describe the Fourier descriptor and how to use it with omnidirectional images to implement the visual compass. Section 4 deals with the problem of localization and map creation using visual odometry. Next, Section 5 presents the experimental setting and the results obtained. Finally, we present the conclusions and future work in Section 6.

2 TOPOLOGICAL MAP BUILDING. STATE OF THE ART

With respect to the mapping problem we can establish two approaches: *metric* and *topological*. The first one consists in modeling the environment using a metric map obtained with geometrical accuracy when representing the position of the robot in it. For example (Gil et al., 2010) present an approach to carry out the mapping process with a team of mobile robots and visual information. On the other side, topological mapping consists in the creation of maps that represent graphical models of the environment that capture places and their connectivity in a compact form. An example of this approach is presented in (Payá et al., 2010) where a topological representation of the environment is obtained by applying a method based on the physics of harmonic oscillators. Also, (Tully et al., 2009) describe a probabilistic method for topological SLAM (Simultaneous Localization and Mapping), solving the topological graph loop-closing problem. At last, (Fernández et al., 2010) describe a Monte-Carlo Localization using the robot odometry and the appearance of omnidirectional images to localize in a topological map.

Recovering relative robot poses from a set of camera images has been a largely studied problem in recent years. For example (Nistér et al., 2006) present a system that estimates the motion of a stereo head using a feature tracker or (Scaramuzza and Siegwart, 2008) describe a real-time algorithm for computing the ego-motion of a vehicle using as input only omnidirectional images and (Scaramuzza et al., 2010) study how to close the loop by using the omnidirectional visual odometry and a vocabulary tree. This work shows how it is possible to carry out a process of robot localization and mapping simultaneously using as input data only omnidirectional images and a loop-closing process.

In this paper, we face the mapping problem as a *relative camera pose recovering* problem, using the overall appearance of the panoramic images, without any feature extraction process. We describe a real-time algorithm for computing an appearance-based topological map through visual odometry. The main contributions of the work are the development of a visual compass that permits computing the position and orientation of each new location in the map, with a low computational cost.

3 FOURIER SIGNATURE

3.1 Fourier Signature

The Fourier Signature presents several advantages among other Fourier-based methods. It is simple to compute, it presents a low computational cost in terms both of computation time and memory required, and it exploits well the invariance against ground-plane rotations using panoramic images (Payá et al., 2009).

The Fourier Signature presents the same properties as the 2D Discrete Fourier Transform. The most important information is concentrated in the low frequency components of each row, so we can work only with the information from the k_1 first columns in the signature ($k_1 < N_y$), and its presents rotational invariance when working with panoramic images. It is possible to prove that if each row of the original image is represented by the sequence $\{a_n\}$ and each row of the rotated image by $\{a_{n-q}\}$ (being q the amount of shift), when the Fourier Transform of the shifted sequence is computed, we obtain the same amplitudes A_k than in the non-shifted sequence, and there is only a phase change, proportional to the amount of shift q, (Eq. 1).

$$F[\{a_{n-q}\}] = A_k \exp\left(-j\frac{2\pi ql}{N_y}\right);\ l = 0,\ldots,N_y - 1 \tag{1}$$

Thanks to this shift theorem we can separate the computation of the robot position and the orientation. With this aim, we decompose the Fourier Signature of the image $I^j \in \Re^{N_x,N_y}$ in two matrices, one containing the modules $d^j \in \Re^{N_x,k_1}$ and the other the phases $p^j \in \Re^{N_x,k_2}$ of this signature. Finally, it is interesting to highlight also that the Fourier Signature is an inherently incremental method.

3.2 Visual Compass

Once we have studied the kind of information to store in the database, we have to establish some relationships between the stored poses to carry out the relative camera pose recovering. When we have the Fourier Signature of two panoramic images that have been captured in two points that are geometrically close in the environment, it is possible to compute their relative orientation using the shift theorem (eq. 1).

Since the Fourier Signature is invariant to ground-plane rotations and there is a relationship between phases of the Fourier Signature of a panoramic image taken at one position and the phases of the Fourier Signature of another panoramic image taken at the same point but with different orientation (eq. 1), we can expand this property and calculate the approximate rotation $\phi^{t+1,t}$ between two panoramic images taken on two consecutive poses. In fig. 1, v^t is the velocity of the robot at time t, v^{t+1} at time $t+1$ and the relative orientation $\phi^{t+1,t}$. The angle obtained corresponds to the rotation the robot has performed in the ground-plane when going from the first to the second point as shown in fig. 1.

To obtain $\phi^{t+1,t}$ we have implemented a convolution operation between the phases of the Fourier Signatures of the panoramic images of the two poses, by applying eq. 1.

4 TOPOLOGICAL MAP CREATION

We build a graph-based map where when a new image is captured, a new node is added to the map, and the topological relationships with the previous node are computed using the global appearance information of the scenes. With our procedure, this computation is made online, as the robot is going through the environment, in a simple and robust way.

We consider that our map is composed of a set of nodes $L = \{l^1, l^2, \ldots, l^N\}$. Each node l^j is represented by an omnidirectional image $I^j \in \Re^{N_x,N_y}$ associated and a Fourier descriptor that describes the global appearance of the omnidirectional image, composed of a modules matrix $d^j \in \Re^{N_x,k_1}$ and a phases matrix $p^j \in \Re^{N_x,k_2}$. Also, with the algorithm implemented, we can compute the position (l^j_x, l^j_y) and the orientation l^j_θ of each node in the map thus $l^j = \{(l^j_x, l^j_y, l^j_\theta), d^j, p^j, I^j\}$ (fig. 1).

We consider that the robot captures a new image at time $t+1$ and then, the Fourier descriptors d^{t+1} and p^{t+1} are computed. Comparing it with the descriptors of the previously captured image d^t and p^t we can find the topological relationships between these two nodes. We can separate the computation of the robot position and the orientation at time $t+1$ $(l^{t+1}_x, l^{t+1}_y, l^{t+1}_\theta)$ thanks to the shift theorem (eq. 1).

In the surroundings of the point where one image is taken, the distance between Fourier Signatures is approximately proportional to the actual geometrical distance (Fernández et al., 2010). To compute the difference between the appearance of two scenes, we use the Euclidean distance between the modules of the Fourier signature. If d^i is the Fourier signature of the image I^i and d^j is the Fourier signature of the image I^j, then the distance between scenes i and j is:

$$D^{i,j} = \sqrt{\sum_{u=0}^{N_x}\sum_{v=0}^{k_1}(d^i(u,v) - d^j(u,v))^2} \tag{2}$$

On the other hand, thanks to the visual compass implemented, we can estimate the relative orientation between images. After this process, the position of the current node is computed from the previous node as:

$$l_x^{t+1} = l_x^t + D^{t+1,t} \cdot cos(\theta^{t+1,t}) \qquad (3)$$

$$l_y^{t+1} = l_y^t + D^{t+1,t} \cdot sin(\theta^{t+1,t}) \qquad (4)$$

$$l_\theta^{t+1} = l_\theta^t + (\phi^{t+1,t}) \qquad (5)$$

These relationships are shown graphically in fig. 1.

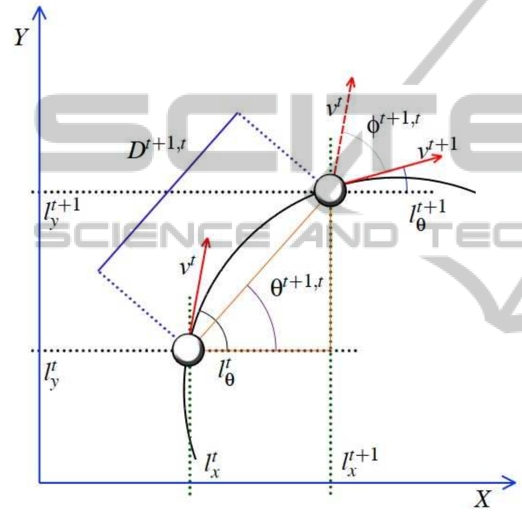


Figure 1: Position and orientation of the new node in the map computed incrementally from the previous node.

Once the topological map is built from the Visual Odometry data, we need a mechanism to test the performance of our approach. We have decided to evaluate how similar is the layout of the resulting map comparing to the actual map (*Ground Truth* or *Real Map*) of the images captured. With this aim, we use a method based on a shape analysis, as in (Payá et al., 2010). As a result of this analysis, a parameter $\mu \in [0,1]$ can be obtained. μ is a measure of the shape correspondence between the sets of points where original images were taken and the set of points in the map. The lower is μ, the more similar are these sets. We name this parameter *shape difference* along the paper. We use this difference with the only purpose to know the feasibility of our appearance-based Visual Odometry, and its use is possible due to the fact that we know the coordinates of the points in the original map (ground truth).

5 EXPERIMENTS

To carry out the experiments, we have used two different sets of omnidirectional images taken from a catadioptric system consisting of a CCD camera and a hyperbolic mirror. The first set has been captured in an office environment, when the robot performs the trajectory shown in fig. 2 (ground truth), which includes a loop closing. This set is composed by 200 images with a distance of 10cm between images. The second set (fig. 3) has been captured in a laboratory environment. It is composed by 150 images and the image acquisition has been automated so that a new image is captured when the difference with the previous one goes over a threshold.

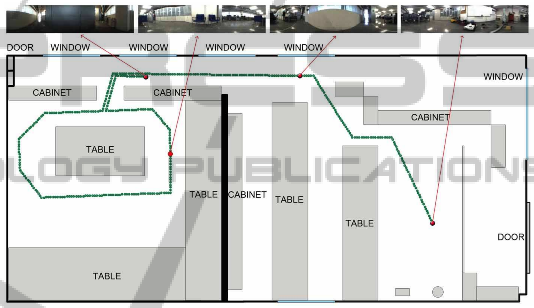


Figure 2: Trajectory followed by the robot when capturing the first set of images.

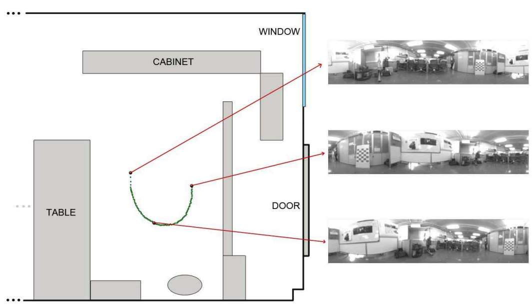


Figure 3: Trajectory followed by the robot when capturing the second set of images.

We have designed a complete set of experiments in order to test the validity of the global appearance-based approach in topological map building. With this aim, we study some features that define the feasibility of the procedures, such as the accuracy of the map built (how similar it is comparing to the actual map) and the computational cost of the method.

Fig. 4 shows an example of the map computed with the first set of images and the actual map (ground truth). This map has been built using all the images of the set (geometrical distance between images equals

SCITEPRESS SCIENCE AND TECHNOLOGY PUBLICATIONS

10 cm.), $k_1 = 64$ module components to compute distance between Fourier descriptors and $k_2 = 64$ phase components used to compute relative orientations in the visual compass. Fig. 5 shows an example with the second set of images. The visual odometry map has been built with all the images in the set, and $k_1 = k_2 = 64$. In this case, we compare it with the actual map and with the map computed using the odometry of the robot. The map obtained with our visual odometry algorithm clearly outperforms the map obtained with the odometry of the robot.

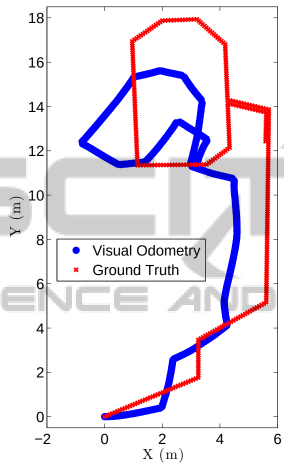


Figure 4: Example of map built with the visual odometry and the first set of images and ground truth.

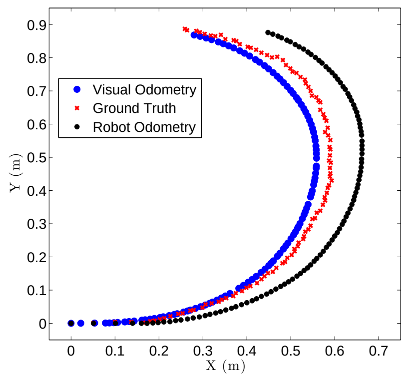


Figure 5: Example of map built with the visual odometry and the second set of images, ground truth and map built using the robot odometry.

In the set of experiments we have implemented to study the performance of our algorithm, we have tested the influence of the number of module components to compute the distance $D^{t+1,t}$ (k_1) and the number of components to compute the phase difference (k_2). We also test the influence of the geometrical distance between the points where the images were captured.

Fig. 6 (a) shows the shape difference of the resulting map comparing to the actual map when using the first set of images. Fig. 6 (b) shows the same result when building the map with the second set of images. When we use all the images of set 1, the minimum shape distance is around 0.04 when $k_1 = 26$ and $k_2 = 20$. In set 2, this factor is 0.015 when $k_1 = 4$ and $k_2 = 8$. The shape distance tends to increase when k_2 does, due to the fact that the first components contain the main information so, the high frequency component may be adding noise to the computation. As far as k_1 is concerned, the tendency is not clear but, in general, the shape distance is quite insensitive to this parameter.

To test the feasibility of the method presented to work in real time, we performed a series of experiments in which we obtained the average time needed at each step depending on the number of components both in magnitude (k_1) and phase (k_2). When we use all the images of set 1, the average time needed at each step to place the new location in the map is around 0.153 seconds when $k_1 = 26$ and $k_2 = 20$. In set 2, this factor is 0.054 seconds when $k_1 = 4$ and $k_2 = 8$.

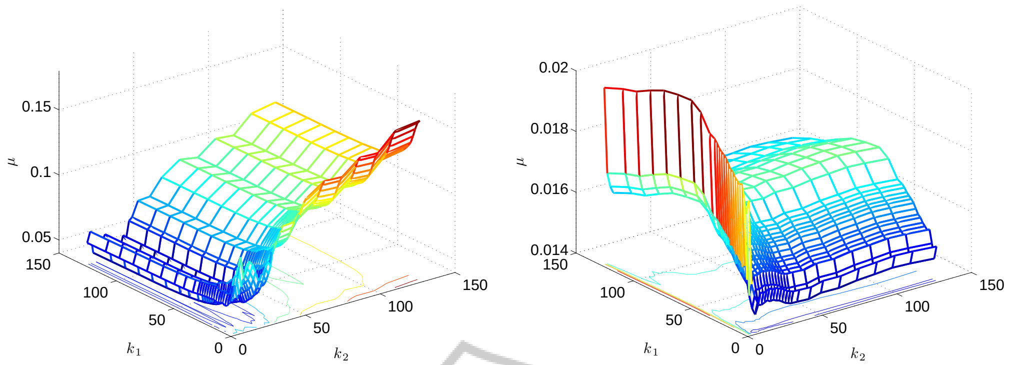
6 CONCLUSIONS

In this paper we have studied the applicability of the approaches based on the global appearance of omnidirectional images in topological mapping, using a set of images a robot has captured when traversing a trajectory in an environment. The main contributions of the paper include the development of a visual compass that allows building a map of the environment online, while the robot is going through the environment, the development of a method to compare the accuracy of the layout of the map computed and the study of the influence of the parameters of the process both in the layout of the resulting map and in the processing time.

All the experiments have been carried out with two sets of omnidirectional images captured by a catadioptric system mounted on the mobile robot. Each scene is described through a Fourier-based signature that presents a good performance in terms of amount of memory and processing time, and it is also invariant to ground-plane rotations and an inherently incremental method.

We present a methodology to build graph-based maps of the environment. As we use a topological approach, these maps represent the real world except for a scale factor and a rotation. To make an homogeneous comparison between the map computed and

SCITEPRESS SCIENCE AND TECHNOLOGY PUBLICATIONS



(a) (b)

Figure 6: Shape difference versus number of module components (k_1) and phase components (k_2) for (a) the set of images 1 and (b) the set of images 2.

the real map, we have developed a method based in the Procrustes analysis. As shown in the results, when the parameters of the system are correctly tuned, accurate results can be obtained, maintaining a reasonable computational cost.

We are now working in this approach to build a topological SLAM algorithm (Simultaneous Localization and Map Building) using just the global appearance of omnidirectional images and different map topologies.

ACKNOWLEDGEMENTS

This work has been supported by the Spanish government through the project DPI2010-15308. ”Exploración Integrada de Entornos Mediante Robots Cooperativos para la Creación de Mapas 3D Visuales y Topológicos que Puedan ser Usados en Navegación con 6 Grados de Libertad”.

REFERENCES

Bay, H., Ess, A., Tuytelaars, T., and Gool, L. V. (2008). Speeded-up robust features (surf). In *Comput. Vis. Image Underst.*

Fernández, L., Gil, A., Payá, L., and Reinoso, O. (2010). An evaluation of weighting methods for appearance-based monte carlo localization using omnidirectional images. In *Proc. of Workshop on Omnidirectional Robot Vision ICRA*.

Gil, A., Reinoso, O., Ballesta, M., Juliá, M., and Payá, L. (2010). Estimation of visual maps with a robot network equipped with vision sensors. *Sensors*.

Menegatti, E., Maeda, T., and Ishiguro, H. (2004a). Image-based memory for robot navigation using properties of omnidirectional images. *Robot. Auton. Syst.*

Menegatti, E., Zoccarato, M., Pagello, E., and Ishiguro, H. (2004b). Image-based monte-carlo localisation with omnidirectional images. In *Robot. Auton. Sys.*

Murillo, A. C., Guerrero, J. J., and Sagues, C. (2007). Surf features for efficient robot localization with omnidirectional images. In *Proc. of ICRA*.

Nistér, D., Naroditsky, O., and Bergen, J. (2006). Visual odometry for ground vehicle applications. *Journal of Field Robot.*

Payá, L., Fernández, L., Gil, A., and Reinoso, O. (2010). Map building and monte carlo localization using global appearance of omnidirectional images. *Sensors*.

Payá, L., Fernández, L., Reinoso, Ó., Gil, A., and Úbeda, D. (2009). Appearance-based dense maps creation. In *Proc. of ICINCO*.

Scaramuzza, D., Fraundorfer, F., and Pollefeys, M. (2010). Closing the loop in appearance-guided omnidirectional visual odometry by using vocabulary trees. *Robot. Auton. Syst.*

Scaramuzza, D. and Siegwart, R. (2008). Appearance-guided monocular omnidirectional visual odometry for outdoor ground vehicles. *IEEE Trans. on Robot.*

Tully, S., Kantor, G., Choset, H., and Werner, F. (2009). A multi-hypothesis topological slam approach for loop closing on edge-ordered graphs. In *Proc. of IEEE/RSJ*.

Valgren, C. and Lilienthal, A. (2010). Sift, surf and seasons: Appearance-based long-term localization in outdoor environments. In *Robot. Auton. Sys.*

SCITEPRESS SCIENCE AND TECHNOLOGY PUBLICATIONS