

Fourier Signature in Log-Polar Images

A. Gasperin, C. Ardito, E. Grisan and E. Menegatti

Intelligent Autonomous Systems Laboratory (IAS-Lab)
Department of Information Engineering
Via Gradenigo 6/b, 35131 Padova, Italy

Abstract. In image-based robot navigation, the robot localises itself by comparing images taken at its current position with a set of reference images stored in its memory. The problem is then reduced to find a suitable metric to compare images, and then to store and compare efficiently a set of images that grows quickly as the environment widens. The coupling of omnidirectional image with Fourier-signature has been previously proved to be a viable framework for image-based localization task, both with regard to data reduction and to image comparison. In this paper, we investigate the possibility of using a space variant camera, with the photosensitive elements organised in a log polar layout, thus resembling the organization of the primate retina. We show that an omnidirectional camera using this retinal camera, provides a further data compression and excellent image comparison capability, even with very few components in the Fourier signature.

1 Introduction

A mobile robot that moves from place to place in a large scale environment needs to know its position in the environment to successfully plan its path and its movements. The general approach to this problem is to provide the robot with a detailed description of the environment (usually a geometrical map) and to use some kind of sensors mounted on the robot to locate in its world representation. Unfortunately, the sensors used by the robots are noisy, and they are easily misled by the complexity of the environment. Nevertheless, several works successfully addressed this solution using high precision sensors like laser range scanners combined with very robust uncertainty management systems [13] [2]. Another solution, very popular in real-life robot applications, is the management of the environment. If artificial landmarks, such as stripes or reflecting dots, are added to the environment, the robot can use these objects, which are easy to spot and locate, to calculate its position on a geometrical map. An example of a successful application of this method is the work of Hu [6]. Unfortunately, these two approaches are not always feasible. There are situations in which an exact map of the environment is either unavailable or useless for example, in old or unexplored buildings or in environments in which the configuration of objects in the space changes frequently. So, the robot needs to build its own representations of the world. This means that in most cases a geometrical map contains more information than that needed by the robot to move in the environment. Often, this adds unnecessary complexity to the map building problem. In addition to the capability of reasoning about the environment

topology and geometry, humans show a capability for recalling memorised scenes that help themselves to navigate. This implies that humans have a sort of visual memory that can help them locate themselves in a large environment. There is also experimental evidence to suggest that very simple animals like bees and ants use visual memory to move in very large environments [3]. From these considerations, a new approach to the navigation and localization problem developed, namely, image-based navigation. The robotic agent is provided with a set of views of the environment taken at various locations. These locations are called reference locations because the robot will refer to them to locate itself in the environment. The corresponding images are called reference images. When the robot moves in the environment, it can compare the current view with the reference images stored in its visual memory. When the robot finds which one of the reference images is more similar to the current view, it can infer its position in the environment. If the reference positions are organised in a metrical map, an approximate geometrical localization can be derived. With this technique, the problem of finding the position of the robot in the environment is reduced to the problem of finding the best match for the current image among the reference images. The problem now is how to store and to compare the reference images, which for a wide environment can be a large number. In order to store and match a large number of images efficiently, it has been shown in [9] the transformation of omnidirectional views into a compact representation by expanding it into its Fourier series. The agent memorises each view by storing the Fourier coefficients of the low frequency components. This drastically reduces the amount of memory required to store a view at a reference location. Matching the current view against the visual memory is computationally inexpensive with this approach.

We show that a further reduction in memory requirements and computations can be met by using log-polar images, obtained by a retina-like sensor, without any loss in the discriminatory power of the methods.

2 Materials

2.1 Omnidirectional Retinal Sensor

The retina-like sensor used in this work is the Giotto camera developed by Lira-Lab at the University of Genova [11] [12] and by the Unitek Consortium [4]. It is built using the $35\mu\text{m}$ CMOS technology, and arranging the photosensitive elements in a log-polar geometry. A constant number of elements is placed on concentric rings, so that the size of these elements necessarily decreases from the periphery toward the center. This kind of geometric arrangement has a singularity in the origin, where the element dimension would shrink to zero. Since this dimension is constrained by the building technology used, there is a ring from which no dimension decrement is possible for accomodating a constant number sensitive elements. Hence, the area inside this limiting ring does not show a log-polar geometry in the arrangement of the elements, but is nevertheless designed to preserve the polar structure of the sensor and at the same time tessellate the area with pixels of the same size. This internal region will be called the *fovea* of the sensor for its analogy with the fovea in the animal retina, whereas the region with constant number of pixels per ring will be called *periphery*.

The periphery is composed by $N_{per} = 110$ rings with $M = 252$ pixels each, and the

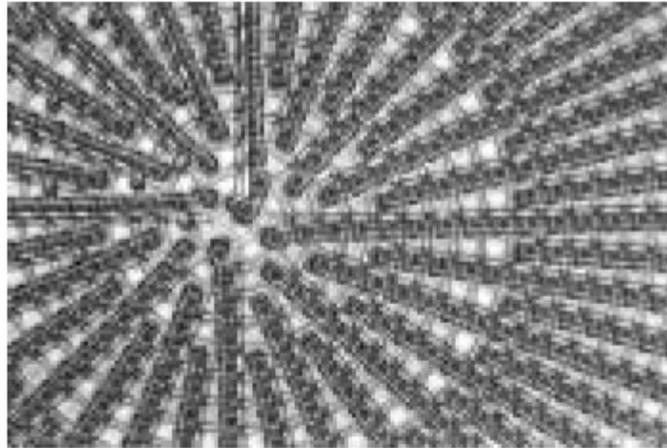


Fig. 1. The central part of the electronic layout of the retinal sensor (from [4]).

fovea is composed by $N_{fov} = 42$ rings (see Fig. 1). This lead to a log-polar image having size of $M \times N = 252 \times 152$, where $N = (N_{per} + N_{fov})$, and the image is obtained from a sensor with 38.304 photosensitive elements. It is claimed in [4] that given its resolution, the log polar sensor yields an image equivalent to a 1090x1090 image acquired with a usual CCD: a sample image acquired with this camera is shown in Fig. 2, together with its cartesianig:retina remapping in Fig. 3

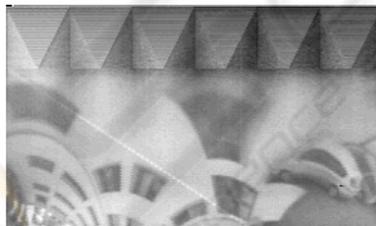


Fig. 2. A sample 252x152 image acquired with the retina-like camera.

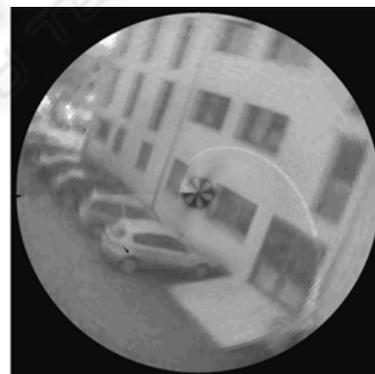


Fig. 3. The sample image of Fig. 2 transformed in a 1090x1090 cartesian image.

To obtain the omnidirectional sensor, the retina-like camera is coupled with an hyperbolic mirror with a black needle at the apex of the mirror to avoid internal reflections

on the glass cylinder [7]: the sensor can be seen in Fig. 4(a). A single omnidirectional image gives a 360° view of the environment, as can be seen in Fig. 4(b).

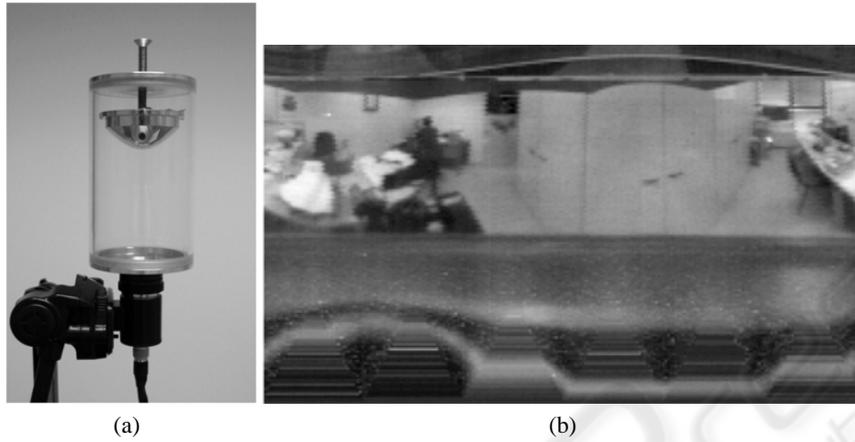


Fig. 4. (a) The omnidirectional sensor composed by the retina-like camera and the hyperbolic mirror. (b) A sample image acquired with the omnidirectional retinal sensor.

3 Methods

3.1 Log-Polar Omnidirectional Image

The pixel coordinates of the output image of the retinal sensor are polar coordinates (ρ, ϑ) , that are related to the usual cartesian coordinates (x, y) via:

$$\begin{aligned}\rho &= \log\left(\sqrt{x^2 + y^2}\right) \\ \vartheta &= \arctan\left(\frac{y}{x}\right)\end{aligned}\quad (1)$$

There are two main issues to be considered while dealing with log-polar images. The first is that there is a singularity in the transformation near the origin, where the pixel dimension tend to zero. The transformation can thus be considered exact only in the region outside the fovea, whereas inside the fovea the mapping depends on the particular arrangement of the retinal sensor.

The second point is the consideration that given the sampling in polar coordinates induced by the sensor, moving from the center toward the periphery of the image, the mapping is not bijective from $(\rho_i, \vartheta_i) \rightarrow (x_i, y_i)$, but rather one point in the log polar image correspond to a sector of annular ring:

$$(\rho_i, \vartheta_i) \rightarrow \{(x, y) | \rho \in [\rho_i, \rho_{i+1}] \cap \vartheta \in [\vartheta_i, \vartheta_{i+1}]\} \quad (2)$$

This means that from the center of the image toward its outer boundary, the resolution decreases, as a pixel in the log-polar image gather information from a bigger area than

a pixel, e.g., in the fovea.

An interesting property of the retinal sensor appears when it is coupled with an hyperbolic mirror, so to provide an omnidirectional sensor. In fact the space-variant resolution of the sensor, if matched with the hyperbolic projection provide an omnidirectional image of nearly constant resolution. Moreover, the image acquired by this omnidirectional sensor is already in the form of a panoramic cylinder, without need of further transformations [12, 10].

3.2 Fourier Signature

In image-based navigation the main problem is the storage of reference images and the comparison of these images with those acquired during the localization. In [9] was shown the effectiveness of using a small number of Fourier coefficients to characterize an image: that method both reduce drastically the dimension of the information to be stored and proved to be enough to discriminate different images, without the need of image alignment as in [1] [5] [8].

The Fourier signature is computed in two steps. First, we calculate the 1-D Fourier transform of every line of the log-polar image and we store in a matrix the Fourier coefficients line by line. Then, we keep only a subset of the Fourier coefficients, those corresponding to the lower spatial frequencies, as signature for the image.

To fully exploit the further dimensionality reduction imposed by the retina-like sensor, we have to recall that in the fovea the effective physical pixels (and therefore the amount of information) is 1 in the center, that is mapped in first line of the log-polar image, 4 in the second innermost ring, mapped in the second line, and so on until the number of pixels in the ring match that of the periphery, where the amount of pixels per ring is constant. A number of physical pixels smaller than the number of image pixels induces a smaller band on the signal than it would be possible given the image dimension. This leads to the consideration that in the foveal region we need to retain less Fourier coefficients than in the periphery to achieve a storage efficiency without loosing any information. The choice is therefore to decrease linearly the number of coefficients used to build the signature: from the k_{max} per line in the periphery (rows $N_{fov} + 1$ to N in the log-polar image), to the 1 coefficient of the first line of the image. Hence for the row y :

$$k(y) = \begin{cases} \lceil \frac{k_{max}-1}{N_{fov}-1} \cdot y + \frac{k_{max}-N_{fov}}{k_{max}-1} \rceil & \text{if } y \leq N_{fov} \\ k_{max} & \text{if } y > N_{fov} \end{cases} \quad (3)$$

with $\lceil x \rceil$ meaning the ceiling of x .

3.3 Dissimilarity Measure

Given an image I , and the discrete set of its Fourier coefficients for the line y , $a_{y,k}$, with $y = 1, \dots, N$, we can define the Fourier signature as the vector \mathbf{F} containing the juxtaposition of all Fourier coefficients of the signature for each line:

$$\mathbf{F}(I) = [a_{1,1}, \dots, a_{1,k(1)}, \dots, a_{N,1}, \dots, a_{N,k(N)}] \quad (4)$$

A distance between two images I_i and I_j can be evaluated as the L^1 norm between the two vectors of their Fourier signature:

$$d(I_i, I_j) = |\mathbf{F}(I_i) - \mathbf{F}(I_j)|_1 \quad (5)$$

When a database of images is available, and a new image have to be compared with those in the database to find the best match, is often more intuitive to use a measure of relative distance of the image under examination from one in the database, given all the images in the database:

$$p(I_i, I_j) = 1 - \frac{d(I_i, I_j)}{\max_{i,j} (d(I_i, I_j))} \quad (6)$$

This is a normalized distance in the database, assuming values in the interval $[0, 1]$, and can therefore be viewed as a probability *a posteriori* of an image I_i to be equal to image I_j , with $j = 1, \dots, N$, and N the number of images in the database.

4 Results and Discussion

To test the proposed measure, an image database was built by acquiring a frame from different positions in an indoor environment, using the retinal omnidirectional camera described previously. The acquisition sites were 15 locations 20cm apart.

First of all, we made experimentations to evaluate which is the minimum number of Fourier coefficients necessary to construct a Fourier signature that retains all and only the necessary information. Hence, we calculated the similarity of each input image against all the reference image of the dataset varying the number of coefficients per row (k_{max}) of the Fourier signature. Since the Nyquist frequency of each row is $f_{Ny} = \frac{N}{2}$, the maximum number of coefficients of the DFT which yield effective information is $\frac{N}{2}$. Therefore, we made $k_{max} \in K = [1, \dots, \frac{N}{2}]$.

For each $k_{max} \in K$ we first evaluated the similarity measure Eq. (6) of each image in the reference database from every other image in the reference database. By this mean, we show that Eq. (6) is an effective measure to distinguish different images, and can therefore be used to provide a good localization performance in autonomous robot navigation tasks. In Fig. 7 we show three successive sample images (relative distance equal to 15cm) from the reference database, and the similarity value of an input image taken at a location corresponding to the second reference image. The similarity value yields a correct match between input and reference.

In Fig. 5 and Fig. 6, it is shown the values of the similarity value for different values of k_{max} of an image in the reference database with every other image. In both figures, the similarity peak corresponds to the correct image, and the similarity values decrease around the peak, the higher k_{max} , the sharper the decrease.

The choice of k_{max} influences the trade off between dissimilarity accuracy and image storage efficiency. A good measure of the accuracy of the proposed measure is the minimum difference between $1 - d(I_i, I_j)$ for $i \neq j$. This is equivalent to evaluate a *classifier*

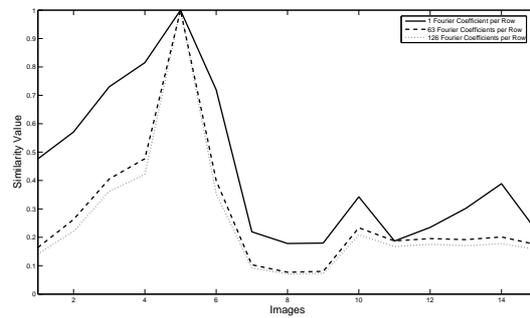


Fig. 5. Similarity measure $p(I_5, I_j)$ for $j = 1, \dots, N$, for $k_{max} = 1, 63, 126$. It is clear that the correct image always yields a similarity measure of 1, whereas the decreasing in the similarity is sharper for high values of k_{max} .

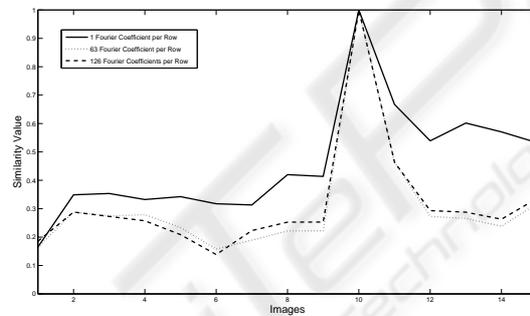


Fig. 6. Similarity measure $p(I_{10}, I_j)$ for $j = 1, \dots, N$, for $k_{max} = 1, 63, 126$. It is clear that the correct image always yields a similarity measure of 1, whereas the decreasing in the similarity is sharper for high values of k_{max} .

margin, in the separation between two different images. In Fig. 8, we show that after a monotonic increase in this margin with the number of coefficients, it reaches a kind of plateau after $k_{max} \simeq 20$.

The storage efficiency achieved is clear when comparing the number of Fourier coefficients needed to form the Fourier signature of a log-polar image with the number of the equivalent cartesian image, which has dimension 1090x1090 pixels. In Fig. 9 we show for different k_{max} the dimension of the Fourier signature for the equivalent cartesian image, for the log-polar image with k_{max} coefficients per row, and for the log-polar image with $k(y)$ coefficients per row, meaning that we have a reduced number of coefficients in the foveal rings. It is well apparent the storage reduction that can be achieved using a retina-like sensor.

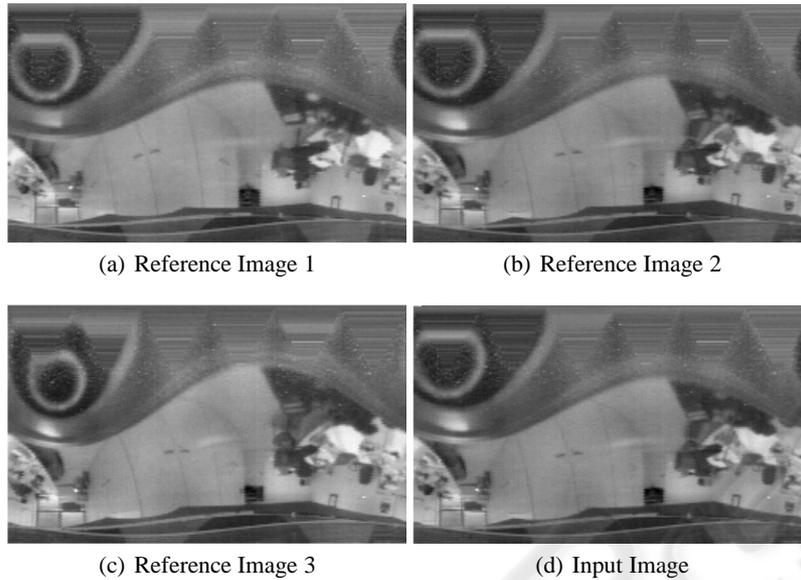


Fig. 7. Three reference image taken 15 cm apart, to be confronted with an input image acquired at location (b). With $k_{max} = 10$, the similarity value of the input image with image (a) is 0.59, with (b) is 0.96, and with (c) is 0.54: the correct match has the highest similarity value.

5 Conclusions

In this paper we show that retinal omnidirectional images can be successfully used to localize an autonomous robot with the image-based navigation approach. Within this approach, the direct comparison of images is not robust, is too computationally cumbersome, and the storage of the whole images requires an excessive memory space. Representing the images with their Fourier signature has been proved a viable way to overcome these problems. In this paper, we showed that coupling this technique with log-polar sensor yields a further dimensionality reduction with sufficient accuracy. The reduction is achieved by exploiting the different bandwidth of each ring of the retina-like sensor with respect to the constant bandwidth of a cartesian sensor, where each row contains the same number of photosensitive element. This allows to keep a decreasing number of Fourier coefficients in the signature, moving from the periphery toward the center of the sensor.

Despite the storage requirement reduction, we show that using a simple L_1 norm on difference of signature vectors have an excellent discriminatory power in distinguishing images taken at different sites.

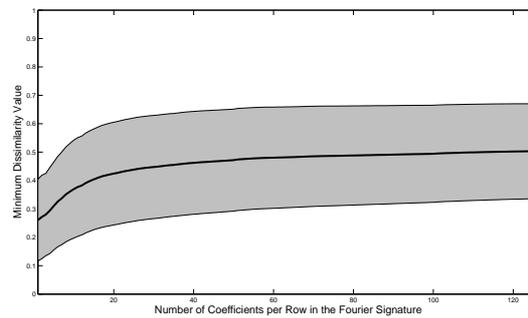


Fig. 8. Minimum difference in the proposed similarity measure $p(I_i, I_j)$ for every different image in the database, and for $k_{max} = 1, \dots, 126$. The solid line represent the mean minimum difference μ , and the gray area represent the variability of this value $\mu \pm \sigma$.

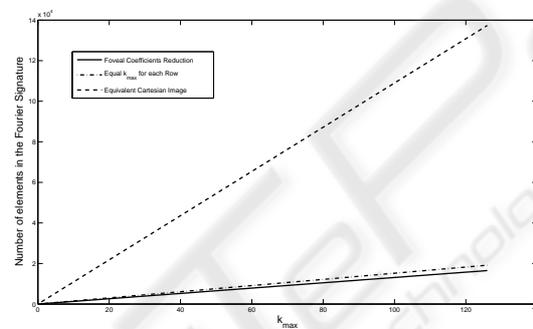


Fig. 9. Total number of coefficients needed to form the proposed Fourier Signature, with respect to k_{max} .

Acknowledgements

We would like to thank Prof. G. Sandini and coworkers for having kindly provided the retinal camera, of which only few prototypes exist.

This research has been partially supported by the Italian Ministry for Education and Research (MIUR) and by the University of Padova (Italy).

References

1. H. A. and N. Iwasa, N. Yokoya, and H. Takemura. Memory-based self-localisation using omnidirectional images. In *Proceedings of the 14th International Conference on Pattern Recognition*, volume 1, pages 1799–1803, 1998.
2. W. Burgard, D. Fox, M. Moors, R. Simmons, and S. Thrun. Collaborative multi-robot exploration. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, 2000.

3. T. Collett, E. Dillmann, A. Giger, and R. Wehner. Visual landmarks and route following in desert ants. *Journal of Comparative Physiology A*, 170:435–442, 1992.
4. Consorzio Unitek. Giotto : retina-like camera.
5. J. Gaspar, N. Winters, and J. Santos-Victor. Vision-based navigation and environmental representations with an omnidirectional camera. *IEEE Transactions on robotics and automation*, 16(6):890–898, December 2000.
6. H. Hu and D. Gu. Landmark based localisation of industrial mobile robots. *International Journal of Industrial Robot*, 27(6):458–467, November 2000.
7. H. Ishiguro. Development of low-cost compact omnidirectional vision. In R. Benosman and S. B. Kang, editors, *Panoramic Vision*, chapter 3. Springer, 2001.
8. B. J. A. Kroese, N. Vlassis, R. Bunschoten, and Y. Motomura. A probabilistic model for appearance-based robot localization. *Image and Vision Computing*, 19(6):381–391, April 2001.
9. E. Menegatti, T. Maeda, and H. Ishiguro. Image-based memory for robot navigation using properties of the omnidirectional images. *Robotics and Autonomous Systems*, 47(4):251–267, July 2004.
10. T. Pajdla and H. Roth. Panoramic imaging with SVAVISCA camera - simulations and reality. Ctu-cmp-2000-16, Center for machine perception - Czech Technical University, October 2000.
11. G. Sandini and G. Metta. *Retina-like sensors: motivations, technology and applications*. 2002.
12. G. Sandini, J. Santos-Victor, T. Pajdla, and F. Berton. OMNIVIEWS: direct omnidirectional imaging based on a retina-like sensor. In *Proceedings of IEEE Sensors 2002*, June 12-14 2002.
13. S. Thrun, M. Beetz, M. Bennwitz, W. Burgard, A. B. Cremers, F. D. Fox, D. Haehnel, C. Rosenberg, N. Roy, J. Schulte, and D. Schulz. Probabilistic algorithms and the interactive museum tour-guide robot Minerva. *International Journal of Robotics Research*, 19:972–999, 2000.

