

Academy Publish

Volume 2

Issue 1

ISSN: 2161-7112

Journal of Computer and Information Technology

Table of Contents

NUMERICAL SIMULATION OF DIRECT INJECTION ENGINE WITH USING POROUS MEDIUM <i>Arash Mohammadi, Ali Jazayeri, Masoud Ziabasharhagh</i>	4
VIRTUAL REALITY TECHNOLOGY APPLIED TO BUILDING MAINTENANCE: COATING OF INTERIOR AND EXTERIOR WALLS OF BUILDINGS <i>Alcínia Zita Sampaio, Daniel P. Rosário and Ana Rita Gomes</i>	19
INTERFERENCE SUPPRESSION STRATEGY BASED ON ADAPTIVE FILTERING FOR MANET <i>Danyang Qin, Ming Qin, Lin Ma, Hongbin Ma, Qun Ding</i>	34
ON-LINE HEAD POSE ESTIMATION WITH BINOCULAR HAND-EYE ROBOT BASED ON EVOLUTIONARY MODEL-BASED MATCHING <i>Fujia Yu, Mamoru Minami, Wei Song, Jianing Zhu, Akira Yanou</i>	43
MAP BUILDING AND LOCALIZATION USING GLOBAL-APPEARANCE DESCRIPTORS APPLIED TO PANORAMIC IMAGES <i>Francisco Amorós, Luis Payá, Oscar Reinoso, Lorenzo Fernández</i>	55
A NOVEL ARTIFICIAL IMMUNE - ANT COLONY OPTIMIZATION ALGORITHM <i>Fang Liu</i>	72
DESIGN OF WEB-BASED E-LEARNIG SYSTEM BASED ON MONITORING FACIAL EXPRESSIONS: DETECTION MECHANISM OF LEARNERS' ATTENTION <i>Hong-Ren Chen</i>	78
BEETLE-INSPIRED BEHAVIOR OF MOBILE ROBOT WITH A FINITE STATE MACHINE DESIGN <i>Lejla Banjanovic-Mehmedovic, Darmin Blazevic, Fahrudin Mehmedovic, Dusko Lukac, Rheinische Fachhochschule</i>	83
COMPUTER MODELING OF RADIAL-DIRECT EXTRUSION OF COMPLEX-SHAPED DETAILS <i>Lyudmila Ryabicheva, Dmytro Usatyuk, Nikolaj Beloshitskij</i>	93
NUMERICAL SIMULATION OF DOUBLE-REFLECTOR ANTENNA WITH FLAT SUBDISH <i>Oleg Yurtsev and Mohamed Dghali</i>	104

Table of Contents

ADAPTIVITY TO KNOWLEDGE ACQUISITION WITH BAYESIAN NETWORK

Samuel Fontes Lima, Sandra Maria Dotto Stump, Nizam Omar 112

IMPLEMENTATION TECHNIQUES FOR THE PROPOSED DYNAMIC SERVING SYSTEMS CAPACITY PLANNING METHOD, THEIR LIMITATIONS, DISADVANTAGES AND SOLUTIONS

Saša Klampfer, Bojan Kotnik, Jože Mohorko, Amor Chowdhury, Žarko Čučej 122

FUNCTION OPTIMIZATION BASED ON CULTURAL ALGORITHMS

Xuesong Yan, Qinghua Wu..... 143

VIRTUAL DOCTOR SYSTEM-VDS: REASONING ISSUES BASED ON FUZZY CONCEPTS

Hamido Fujita, Masaki Kurematsu, Jun Hakura..... 150

**MAP BUILDING AND LOCALIZATION USING
GLOBAL-APPEARANCE DESCRIPTORS APPLIED TO PANORAMIC IMAGES**

**MAP BUILDING AND LOCALIZATION USING
GLOBAL-APPEARANCE DESCRIPTORS APPLIED TO PANORAMIC IMAGES**

Francisco Amorós, Luis Payá, Oscar Reinoso and Lorenzo Fernández

lpaya@umh.es, o.reinoso@umh.es

Departamento de Ingeniería de Sistemas y Automática, Miguel Hernández University
Avda. de la Universidad, s/n. 03202, Elche (Alicante), Spain

Global-Appearance techniques represent a very promising alternative in image processing applied to the tasks of robotic map building and localization. Their inherent characteristics present advantages over classical features descriptors, especially in unstructured environments where landmarks can be difficult to extract. However, when we apply them to real-time navigation, we have to take into account some restrictions that might limit their application since mapping and robot navigation requires specific advisable characteristics when building a descriptor.

This article makes a review and comparison of different methods based on global-appearance to create descriptors of panoramic scenes in order to extract the most relevant information, studying several characteristics and parameters, as invariance against rotations of the robot on the ground plane, computational requirements and accuracy in localization. For this purpose, we have carried out a set of experiments with panoramic images captured in real lighting conditions in indoor environments to demonstrate the applicability of the different descriptors to robotic navigation tasks and to measure their goodness in localization recovery and their computational requirements. The experimental part lets us to validate them and to make an analysis and comparison of each technique.

The autonomous navigation of a robot or a team of robot through an environment usually implies an internal representation of the workspace. This representation has to permit us to build a map where the robot can estimate its pose, i.e., its position and orientation, using the information provided by the different sensors the robot is provided with. Although there are different sensors the robot can be equipped with to carry out the task, omnidirectional visual systems can be stood out due to the richness of the information they provide and its relatively low cost. We can find several researches into mobile robots using vision systems focused on local features descriptors, which are based on the extraction of natural or artificial landmarks from the image to build the map and carry out the localization of the robot, as we can see in (Thrun, 2003).

However, it is not necessary to extract such kind of

landmarks to recognize the pose of the robot. More recent approaches propose processing the image as a whole, with no local feature extraction. These methods, known as global-appearance techniques, become an interesting option when dealing with unstructured environments where it may be difficult to find patterns in order to recognize the scene.

As a disadvantage, they have to deal with a large quantity of information, which means a high computational cost. For that reason, it is necessary to study compression techniques to make feasible the creation of low-dimensional descriptors that concentrate the information of the scene, preserving low-level features such as color or spatial frequencies. These low-dimensional representations let us to build dense maps. With this maps, the problem of finding the position of the robot is reduced to the problem of getting the best match for the current image among

the map images. However, when working with the global appearance of images, we have to deal with inherent problems such as visual aliasing, changing lighting conditions or small changes in the environment among other possible situation there the robustness of the compression technique is proved.

The descriptors must be able to recognize the location of the robot into the map regardless of its orientation, i.e., rotational invariance in the ground plane. Moreover, some orientation information is also necessary to estimate the pose of the robot. It is also advisable incremental methods, since some navigation tasks require modifying the map created during the process, as Simultaneous Localization and Mapping (SLAM).

In the literature we can find several approaches to compress the visual information with global-appearance methods. As an example, PCA (Principal Components Analysis) is a widely used method that has demonstrated being robust in image processing applications, as (Krose et al., 2007). Due to the fact that conventional PCA is not a rotational invariant method, other authors introduced an approach that takes into account different orientations of each image in order to build a more complete map ((Jogan and Leonardis, 2000)), although that improvement makes the algorithm more complex and computationally heavier, fact that might limit its application in real experiments. There are authors that use the Fourier Transform as a generic method to extract the most relevant information of an image. In this field, (Menegatti et al., 2004) extend the Fourier concept defining the Fourier Signature, based on the Discrete Fourier Transform of the image's rows. On the other hand, (Dalal and Triggs, 2005) present a method based on the Histogram of Oriented Gradients (HOG) to pedestrian detection tasks, proving that it could be a useful descriptor for computer vision and image processing using the scenes' appearance, since it avoids the extraction of characteristics points in the image.

(Paya et al., 2009) present a comparative study about

global-appearance techniques. We extend that work studying and comparing three global-appearance methods in robot navigation applications: Fourier Signature, Rotational PCA and Gist-Gabor. The last technique has shown previous promising results, although not in localization and mapping tasks.

The remainder of the article is organised as follows: in section 2 the compared techniques are outlined. In the next section, the process of building the map is presented, studying the computational cost for each method. Section 4 explains the way we gauge the pose of the robot and the time and memory requirements for each algorithm. To finish, there is a summary with the main conclusions.

REVIEW OF COMPRESSION TECHNIQUES

In this section we present the techniques to extract the most relevant information from a database made up of panoramic images using global-appearance approaches.

Fourier-based Technique

As shown in (Menegatti et al., 2004), it is possible to represent an image using the Discrete Fourier transform of each row. So, we can expand each row of the panoramic image $\{a_n\} = \{a_0, a_1, \dots, a_{N-1}\}$ into the sequence of complex numbers

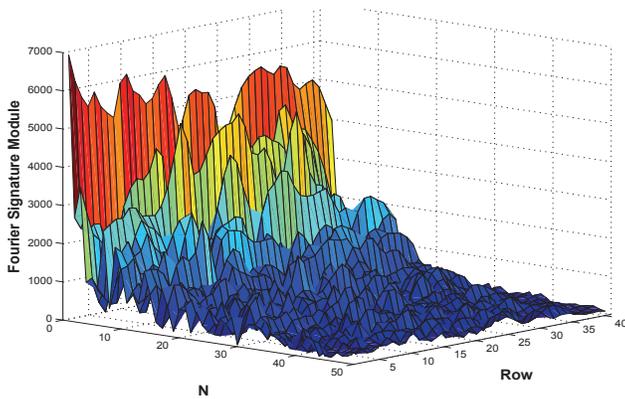
$$\{A_n\} = \{A_0, A_1, \dots, A_{N-1}\} :$$

$$\{A_n\} = \mathfrak{F}[\{a_n\}] = \sum_{n=0}^{N-1} a_n e^{-j \frac{2\pi}{N} kn} ; k = 0, \dots, N-1.$$

The Fourier Transform properties make possible to represent each row with just the first coefficients since the most relevant information is concentrated in the low frequency components of the sequence.

MAP BUILDING AND LOCALIZATION USING GLOBAL-APPEARANCE DESCRIPTORS APPLIED TO PANORAMIC IMAGES

Figure 1. Representation of the 50 first Coefficient Modulus of the Fourier Signature of an image.



We can also take profit of the fact that we are working with omnidirectional images, as this kind of images has the same pixels per row although it is rotated. Due to the fact that the rotation of a panoramic image is represented as a shift of its columns, i.e., a shift along the horizontal axis of each row, the Fourier Transform component module will be the same. So the Fourier Signature presents rotational invariance. Representing each row of the original image as $\mathfrak{F}\{a_n\}$ and $\mathfrak{F}\{a_{n-q}\}$ the same row shifted q pixels, being q proportional to the relative rotation between images, the rotational invariance can be seen in the shift theorem:

$$\mathfrak{F}\{a_{n-q}\} = \sum_{n=0}^{N-1} A_k e^{-j \frac{2\pi}{N} qk}; k = 0, \dots, N-1 \tag{2}$$

where $\mathfrak{F}\{a_{n-q}\}$ is the Fourier Transform of the shifted sequence, and A_k are the components of the

Fourier Transform of the non-shifted sequence. According to this expression, the amplitude of the transform is the same as the original sequence, and just the phase changes. This phase variation lets us to find the phase lag between two rotated images. Therefore, the modulus of the Fourier Signature let the position estimation, whereas we can find out the phase of the robot by comparing the Fourier phase coefficients, as we can see in eq. 2.

PCA-Based Technique

The data compression with PCA techniques has demonstrated being very efficient (Krose et al., 2004). The Principal Components Analysis makes possible the representation of a set of P images with M pixels each, $\vec{I}^j \in \mathfrak{R}^{M \times 1}, j = 1, \dots, P$, in a new subspace where the images are transformed in a feature vector, known as projection of the image, $\vec{p}^j \in \mathfrak{R}^{k \times 1}, j = 1, \dots, P$, being k the PCA features that preserve the most of the variance of the database (Turk and Pentland1991). But if we apply PCA directly over the set of images we obtain a map without rotational invariance, since it keeps information of one orientation for each scene, but no for other possible orientations. In (Leonardis and Jogan, 2000) a possible solution is proposed with the use of the *Eigenspace of Spinning-Images*.

This technique uses the property of panoramic images to create a set of spinning images from the original, obtaining a database that contains a set of N in plane rotations of each point. The phase lag between consecutive rotated images will be constant and proportional to the number of rotations, i.e., $2\pi/N$. After that, the database is compressed by means of PCA analysis. Therefore we have not only a representation that deals with the orientation problem but also a more robust location algorithm since it stores different views of a same point that contains the same visual information.

When working with a set of rotated images, the model presents specific properties that we can use to make the PCA transform easier. Being

MAP BUILDING AND LOCALIZATION USING GLOBAL-APPEARANCE DESCRIPTORS APPLIED TO PANORAMIC IMAGES

$X = [x_0 | x_1 \dots | x_{N-1}] \in \mathfrak{R}^{M \times N}$ a matrix where each column contains a scene of a set of rotations of an image, the covariance matrix $Q = X^T X \in \mathfrak{R}^{N \times N}$ forms a circulant matrix. This property appears regardless of the number of rotations provided that the phase lag between consecutive rotations is constant, as we can see in fig. 1. As (Uenohara and Kanade, 1998) show, the eigenvectors v_i of every circulant matrix are not dependent on its information, and correspond to the basis vectors from the Fourier matrix $F = [v_0 \ v_1 \ \dots \ v_{N-1}]$ where:

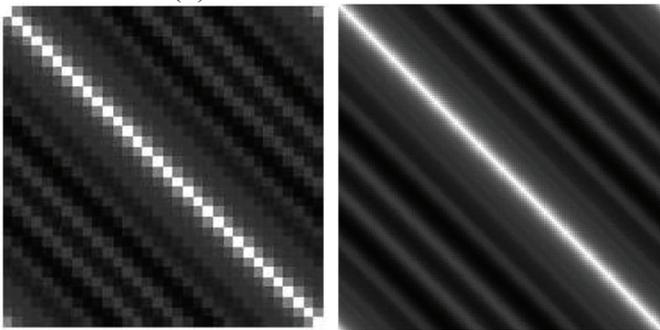
$$v_k = [1 \ \omega^k \ \omega^{2k} \ \dots \ \omega^{(N-1)k}]^T$$

$$k = 0, \dots, n-1, \quad \omega = e^{-2\pi j / N}, \quad j = \sqrt{-1}$$

(3)

Taking profit of this property, we can compute the eigenvectors without computing the SVD decomposition of the covariance matrix. Moreover, once we have calculated the respective eigenvalues, the eigenvectors are sorted by decreasing order of its eigenvalue in order to obtain a subspace that retains the most of variance in the first coefficients of the projections.

Figure 2. Covariance matrix of a set of (a) 32 rotations and (b) 128 rotations.



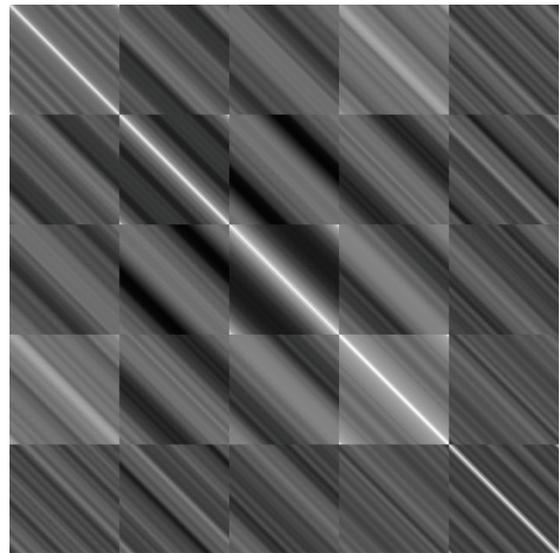
In our case, because we have multiple scenes, the problem is extended to P sets of N rotated images. The inner product $C = X^T X$ is composed of $P \times P$ circulant blocks whose size is $N \times N$ (fig. 2). Moreover,

the blocks are symmetric with respect to the diagonal of the matrix C . Because of that characteristic, the initial problem can be reduced to the problem of solving the SVD of C to solving N decompositions of order P .

$$C = \begin{bmatrix} Q_{11} & Q_{12} & \dots & Q_{1P} \\ Q_{21} & Q_{22} & \dots & Q_{2P} \\ \vdots & \vdots & \ddots & \vdots \\ Q_{P1} & Q_{P1} & \dots & Q_{PP} \end{bmatrix}, \text{ being } Q_{ij} \text{ circulant blocks.}$$

(4)

Figure 3. Inner product of a matrix of a set of images with $P=5$ locations and $N=128$ rotations.



We can calculate the eigenvectors of every Q_{ij} using the vectors of the Fourier Matrix since they are circulant matrices. So all the blocks have the same set of eigenvectors v'_k .

The new eigenvalue problem can be expressed as:

MAP BUILDING AND LOCALIZATION USING GLOBAL-APPEARANCE DESCRIPTORS APPLIED TO PANORAMIC IMAGES

$$Cw' = \mu w'. \tag{5}$$

The eigenvectors w' shall be found among the vectors of the form:

$$w'_k = [\alpha_{k1}v'_k \quad \alpha_{k2}v'_k \quad \dots \quad \alpha_{kP}v'_k], k = 1, \dots, P. \tag{6}$$

The eq. 6 can be also expressed by blocks as:

$$\sum_{j=1}^P Q_{ij}(\alpha_{kj}v'_k) = \mu \alpha_{ki}v'_k, i = 1, \dots, P. \tag{7}$$

Moreover, since v'_k is an eigenvector of every block Q_{ij} , the eq. (7) can be simplified as:

$$\sum_{j=1}^P \alpha_{ij} \lambda^{kj}_{ij} v'_k = \mu \alpha_{ki} v'_k, i = 1, \dots, P. \tag{9}$$

where λ^{kj}_{ij} is the eigenvalue of Q_{ij} corresponding with v'_k .

That implies a new eigenvalue problem:

$$\Lambda \alpha_k = \mu \alpha_k, \tag{10}$$

where $\Lambda = \begin{bmatrix} \lambda^{1K}_{11} & \lambda^{1K}_{12} & \dots & \lambda^{1K}_{1P} \\ \lambda^{1K}_{21} & \lambda^{1K}_{22} & \dots & \lambda^{1K}_{2P} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda^{1K}_{P1} & \lambda^{1K}_{P2} & \dots & \lambda^{1K}_{PP} \end{bmatrix}$ and

$$\alpha_k = [\alpha_{k1}, \alpha_{k2}, \dots, \alpha_{kP}]^T.$$

Since C is symmetric by blocks, Λ is symmetric, having P independent eigenvectors α_k . We apply the

same method for each v'_k , obtaining $N \cdot P$ independent eigenvectors. So instead of performing the SVD decomposition of C (what it would be a computationally expensive process), we solve N decompositions of order P .

Again, we can select just a subset of the eigenvectors with the highest characteristic values to reduce the dimensionality of the projection space (Jogan and Leonardis, 2003).

The representation of the database in the new subspace is in the complex plane. It can be proved that the coefficients of an image and its rotations have the same modulus, with only a change in its phase. Moreover, the phase lag between consecutive rotations of an image is constant (fig. 4). For that reason, we only need to store the projection of one orientation for each point, and the phase lag between the components of consecutive rotations. The angle between the coefficients of the projection of two consecutive rotations of an image can be calculated as:

$$\Delta\varphi_{ij} = \arctan \frac{\text{Re}(q_{(i+1)j} - q_{ij})}{\text{Im}(q_{(i+1)j} - q_{ij})} \tag{11}$$

being q_{ij} a coefficient of the projection, where $i = 0, \dots, N-1$ denotes the rotation, and j is the number of the coefficient of the image projection.

With this information, the projections of all the others orientations q_{rj} can be artificially simulated.

$$q_{rj} = q_{ij} e^{2\pi\sqrt{-1}(r-1)\Delta\varphi_{ij}}, r = i + 1, \dots, (i + N - 1), j = 1, \dots, k \tag{12}$$

with K the number of selected coefficients of each projection.

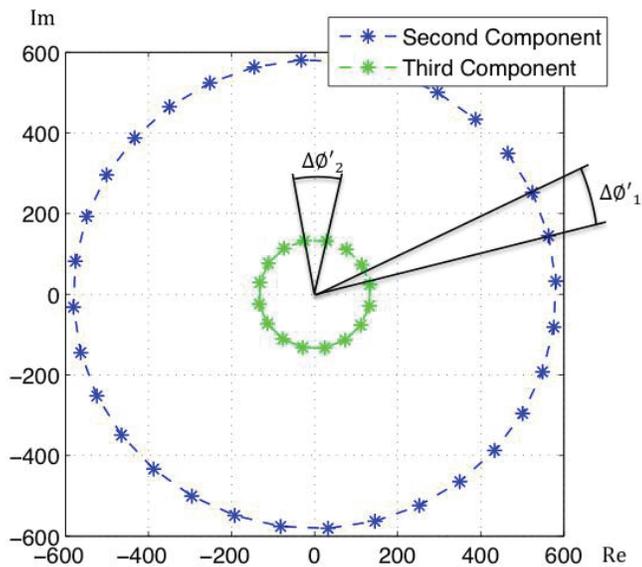
MAP BUILDING AND LOCALIZATION USING GLOBAL-APPEARANCE DESCRIPTORS APPLIED TO PANORAMIC IMAGES

Hence, the localization of the robot in the map is handled by comparison of the modules. To estimate the phase, we simulate the projections of all the rotations of the image of the map selected. The angular resolution depends on the number of siblings of each image included in the database:

$$\text{Min. Angle}(\text{°}) = \frac{360}{N}$$

(13)

Figure 4. Projections of two components of a set of $N=16$ rotations of an image.



Hence, the localization of the robot in the map is handled by comparison of the modules. To estimate the phase, we simulate the projections of all the rotations of the image of the map selected. The angular resolution depends on the number of siblings of each image included in the database:

$$\text{Min. Angle}(\text{°}) = \frac{360}{N}$$

(13)

Gist-Based Techniques

In (Friedman, 1979) we can find the concept of the *gist* of a scene. It can be defined as an abstract representation that activates the memory of scenes' categories. The descriptors based on Gist try to obtain the essential information of the image simulating the human perception system through its ability to recognise a scene through the identification of colour or remarkable structures, avoiding the representation of specific objects.

Some authors have studied the problem of image categorization considering features based on the human ability to recognise images. In (Oliva and Schyns, 1997) it is presented a descriptor that classifies blurred images where only the spatial distribution is seen under the name of *shape of the scene*. (Oliva and Torralba, 2001) develop this idea with the name of *holistic representation of the spatial envelope* to create a descriptor. In (Torralba, 2003) we can see this model computed using global scene features, such as spatial frequencies and different scales based on Gabor filtering. Although it has demonstrated its capacity for scene recognition and classification, we have not found any reference of applications in robotic mapping and localization tests, so we have adapted this technique to be used in such applications.

The descriptor we propose is called Gist-Gabor because it is based in Gabor filtering masks in order to obtain frequency and orientation information. The different masks cover the entire image, without any segmentation.

A Gabor filter is a linear filter whose impulse response is a sinusoid modulated with a Gaussian function (Gabor, 1946). Therefore, a Gabor mask is localized both in the spatial and in the frequency domains (fig. 6). Thanks to its properties regarding textures treatment, Gabor filter can be used in compression and segmentation of digital images (Manjunath and Ma, 1996), (Kruizinga and Petkov, 1999).

MAP BUILDING AND LOCALIZATION USING GLOBAL-APPEARANCE DESCRIPTORS APPLIED TO PANORAMIC IMAGES

Mathematically, the Gabor filter is expressed as a complex function in the space domain as:

$$g(x, y) = s(x, y) \omega_r(x, y) \tag{14}$$

being $s(x, y)$ a complex sinusoid and $\omega_r(x, y)$ a Gaussian bidimensional function, known as envelope. The complex sinusoid can be written as:

$$s(x, y) = \exp(j2\pi(u_0x + v_0y))$$

where (u_0, v_0) define the spatial frequency. Eq. 15 can be also expressed in polar coordinates:

$$s(x, y) = \exp(j2\pi F_0(x \cos w_0 + y \sin w_0)) \tag{16}$$

being $F_0 = \sqrt{u_0^2 + v_0^2}$ the magnitude of the sinusoid, and $w_0 = \tan^{-1}\left(\frac{v_0}{u_0}\right)$ its direction.

In eq. 17 is expressed the Gaussian envelope.

$$\omega_r(x, y) = K \exp(-\pi(a^2(x - x_0)_r^2 + b^2(y - y_0)_r^2))$$

with K the magnitude scale, (a, b) the scale of the axis, and (x_0, y_0) the position of the maximum of the function.

Figure 5. (a) Complex sinusoid, (b) Gaussian envelope and (c) Gabor filter resulting of the convolution of both functions.

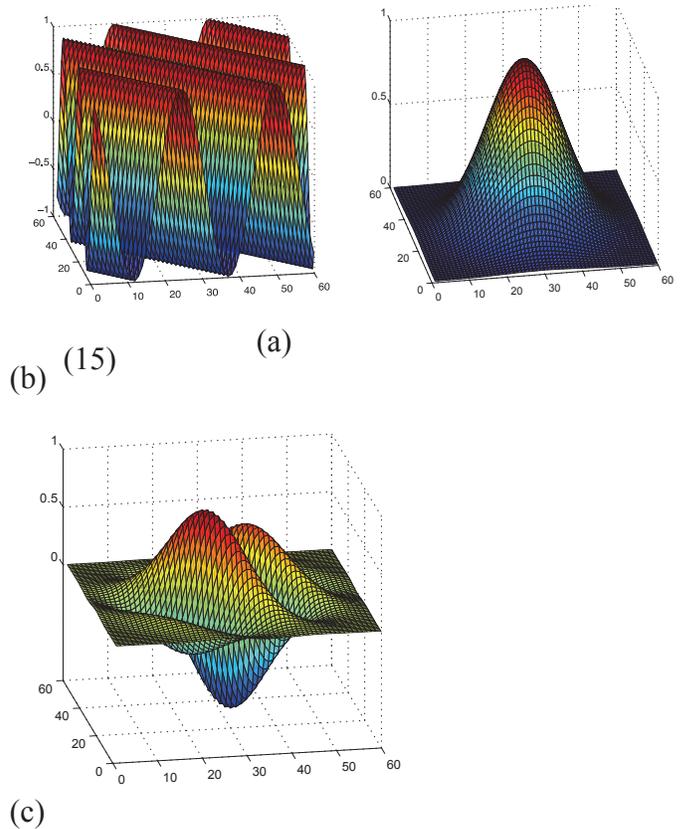


Figure 6. Gabor bank with 2 spatial scales ad different orientations in frequency domain. (a) Limits of the masks, and (b) 3D representation.

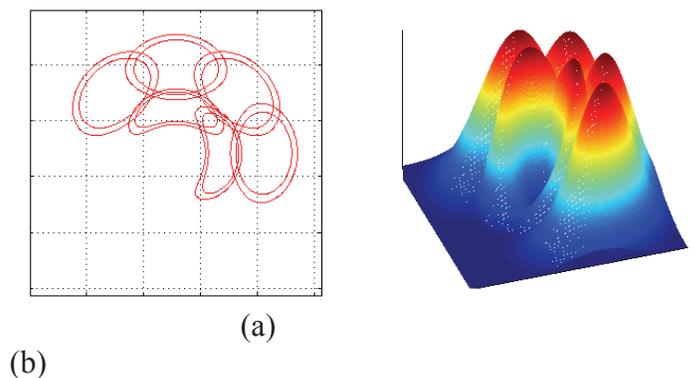
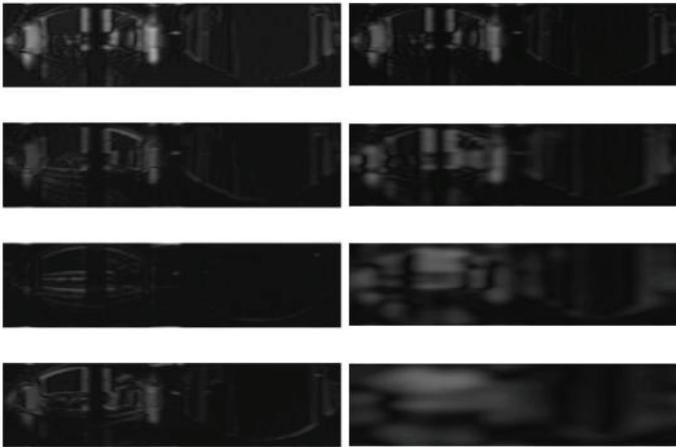


Figure 7. Example of an image filtered with (a) different orientations (0°, 45°, 90° and 135°) and (b)

MAP BUILDING AND LOCALIZATION USING GLOBAL-APPEARANCE DESCRIPTORS APPLIED TO PANORAMIC IMAGES

different scales.



(a)

(b)

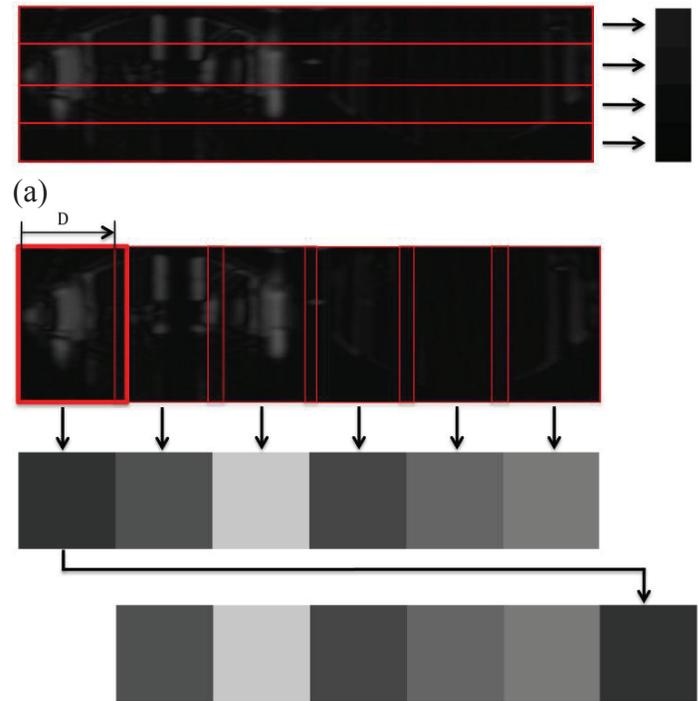
The first step to create the descriptor consists in creating a bank of the Gabor masks with different resolutions and orientations. After that, we filter the image with the set of masks. The orientation of the filter depends on the number of masks of each level because they are equally distributed between 0 and 180 degrees (fig. 6). The filtering is done in the frequency domain, so we compute the bidimensional Fourier Transform of the scene, and multiply each element of the image with its corresponding element of the mask. The results encode different structural information depending on the mask used (fig. 7).

An omnidirectional image contains the same pixels in a row although the image is rotated. So, to create the descriptor, we calculate the average pixel's value within cells with the same width as the image. We repeat this operation for every filtered image, obtaining an array of rotational invariant characteristics. To know the relative orientation between two rotated images, we use vertical. These

vertical windows have the same height of the window. The distance between vertical windows and their width are variables (Fig. 8).

The location is estimated by calculating the minimum distance between the horizontal cells' descriptor of the database and the current image.

Figure 8. Extraction of the Descriptor Values from a filtered image for (a) location and (b) phase estimation.



(b)

The orientation is obtained by comparison and rotation of vertical cells' vector of the current image and the nearby image in the map. The angle accuracy we are able to detect between two shifted images is proportional to the distance between consecutive vertical cells, D :

$$Min. Angle(^{\circ}) = \frac{D * 360}{With\ of\ the\ image} \tag{18}$$

MAP BUILDING AND LOCALIZATION USING GLOBAL-APPEARANCE DESCRIPTORS APPLIED TO PANORAMIC IMAGES

MAP BUILDING

In this section, we compare the performance of the different descriptors in the task of creating a map from an image database collected in several living spaces under realistic illumination conditions. The database belongs to the Technique Faculty of

Area	$n_x \times n_y$	Grid A	Grid B	Grid C	Grid D
		10cm	20cm	30cm	40cm
L.R.	22x11	242	66	32	18
K.	12x9	108	30	12	9
C.A.	36x11	396	118	48	27
TOT		746	204	92	54

Bielefeld University (Moeller et al., 2007). The images are structured in a 10x10 cm rectangular grid. Specifically, there are examples from a living room (L.R.), a kitchen (K.) and a living room and kitchen area combined (C.A.). In Fig. 9 we can see an example of an image corresponding to each area. They were captured with an omnidirectional camera, and later converted into the panoramic format with a resolution of 41x256 pixels.

In order to evaluate the behaviour of the descriptors in different situations, we vary the distance between elements of the database in the map building, with a total of four different grids. Table 1 shows the number of images in every area (n_x and n_y), the distance between elements depending on the grid and the number of images included in the map for each configuration.

Once the images of the database are selected, we apply the compression method to obtain a dataset, which represents the map of the environment that the robot will use to localize itself. All the functions and simulations and simulations have been obtained using Matlab R2009b under Mac OS X.

Figure 9. Image of (a) living room, (b) kitchen and (c) the combined area.



Table 1. Size of the database and number of images selected depending on the grid.

Fig. 10 shows the time and memory requirements of the different algorithms in order to build the map with regard to its main variables.

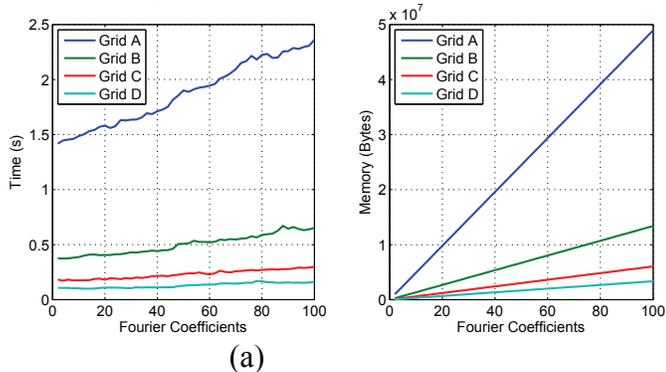
For Fourier Signature, the main variable is the number of coefficient per row we keep to represent the image. In fig. 10(a) and 10(b) we can see the elapsed time and memory to compress and store the map with the Fourier Signature. It is composed of a matrix with the module values of the Fourier Signature to carry out the location, and another matrix that contains the phase of each coefficient of the Fourier Transform to find out the orientation of the robot. There is a proportional increase of the memory and time necessities as we get more coefficients of each row since we are storing and computing more information. The changes between different grids are proportional to number of images the map includes.

The PCA map is made up of the matrix $P \square C^{k \times p}$, being k the main eigenvectors retrieved and p the number of different locations of the map. P contains the projections in the new subspace of all the images of

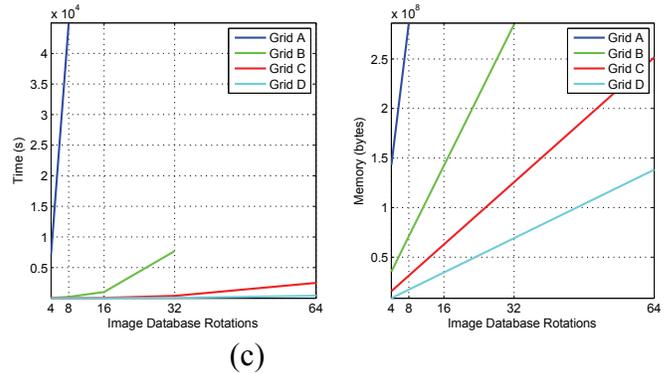
MAP BUILDING AND LOCALIZATION USING GLOBAL-APPEARANCE DESCRIPTORS APPLIED TO PANORAMIC IMAGES

the map regarding one orientation. We also need the change of basis matrix $V \square C^{k \times M}$, being M the elements of an image, to project the input images in the new subspace. Because P represents only one orientation of each scene, we also need to include the vector of the phase lag between components of consecutive rotations, $\vec{p}_{ph} \square \square^{k \times 1}$. In the PCA technique, the main variable is the number of rotations per image we use to create the dataset that we compress by means of PCA. Fig. 10(c) and fig. 10(d) show the time and memory necessities to build the location map using Rotational PCA. The amount of memory is linearly dependant on the number of images and rotations we include in the database. We have considered all the eigenvectors in the measurements.

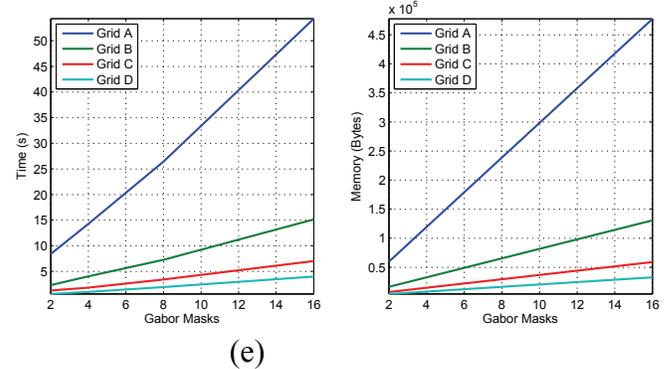
Figure 10. (a) Elapsed time and (b) memory necessities to build the map using Fourier-based algorithm. (c) Elapsed time and (d) memory necessities to build the map using PCA-based algorithm. (e) Elapsed time and (f) memory necessities to build the map using Gist-Gabor varying location parameters. (g) Elapsed time and (h) memory necessities to build the map using Gist-Gabor varying orientation parameters.



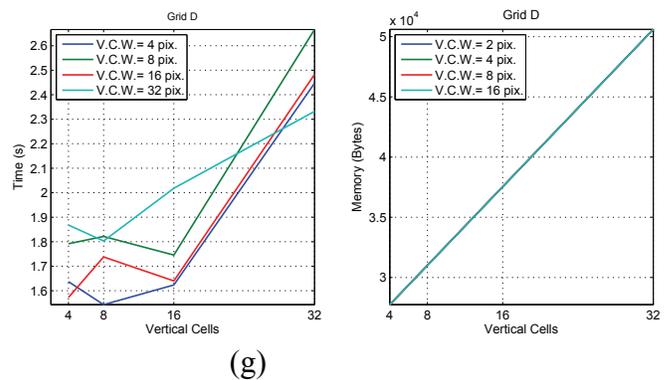
(b)



(d)



(f)



(h)

The memory will vary proportionally to the selected eigenvectors. However, the elapsed time is not linear because the eigenspace calculation spends considerably more time as the data matrix increases.

MAP BUILDING AND LOCALIZATION USING GLOBAL-APPEARANCE DESCRIPTORS APPLIED TO PANORAMIC IMAGES

We have neither estimated the map with more than 8 rotations in Grid A nor with more than 32 rotations in Grid B because of the computational requirements make it unfeasible in real experiments.

The last charts of fig. 10 study the compression of the database with Gist-Gabor. Since the parameters of location and orientation recovery are independent, its measurements are separated.

In fig. 10(e) and fig. 10(f) main location parameter, the number of masks we use in order to filter the image, is varied. They show a linear dependency between the number of Gabor masks and the memory and time requirements. Phase variables are the distance between vertical cells, and its width (Vertical Cells Width or V.C.W.). We set the location parameters with Grid D and a fixed number of four Gabor Masks. Fig. 10(g) shows that the memory size of the map is not affected by the width of the cells since we store the mean value of its pixels whichever its size is, although the elapsed time differs (fig. 10(h)). In both cases results are more dependent on the number of cells than on its width.

Comparatively, PCA compression is the computationally heaviest method since the SVD decomposition, despite of using the properties of the circulant matrices, is a time expensive process. Fourier signature the fastest algorithm, and Gist-Gabor the most compact descriptor, because we concentrate the scene information in a vector with the mean value of cells of the image.

LOCALIZATION AND ORIENTATION RECOVERY

Once we have compared the three methods in terms of the database creation, in this section we measure the goodness of each algorithm by assessing the results of calculating the pose of the robot using an input image to compare with the map previously created, and the time requirements. The test images we have used are all the available images in the database, independently

on the grid selected to build the map, and 15 rotations of each one (every 22.5°). The total number of test images is 11,936 images. Because the pose includes the position and the orientation of the robot, both are studied separately.

In fig. 12, fig. 13 and fig. 14 we study the position retrieval of each descriptor for the different map configurations. The results are presented as binary results considering if we obtain the best match as possible or not, and the information is showed with recall and precision measurement (Gil et al., 2009). In eq. (19) and (20) we can find the definition of recall and precision concepts:

$$recall = \frac{\#correct \ matches \ retrieved}{\#total \ correct \ matches}$$

(19)

$$precision = \frac{\#correct \ matches \ retrieved}{\#correct \ matches}$$

(20)

Precision can be considered as a fidelity measurement, and recall is an indicator of how exhaustive the different methods are.

Each chart provide information about if a correct location is the Nearest Neighbour (N.N.), i.e., if it is the first result selected, or it is between Second or Third Nearest Neighbours (S.N.N or T.N.N).

Regarding the rotation, we represent the results accuracy in bar graphs depending on how much they differ from the correct ones. The results are in percentage over correct locations. Fig. 15 presents the best phase lag accuracy results obtained for each method and grid. The parameters for the position retrieval in the phase study are the same as used in fig. 11, fig. 12 and fig. 13 respectively for each method. The phase lag is just studied in the experiments that

the position has been correct in order to represent the accuracy of the phase estimation process. The elapsed time in both localization and pose estimation is studied in fig.11. In the first column is represented the time for only the position retrieval, whereas in the second column appears the elapsed time for the position and phase estimation.

Figure 11. Elapsed time for (a) location and (b) pose estimation using Fourier-based algorithm. Elapsed time for (c) location and (d) pose estimation using PCA-based algorithm. Elapsed time for (e) location and (f) pose estimation using Gist-Gabor.

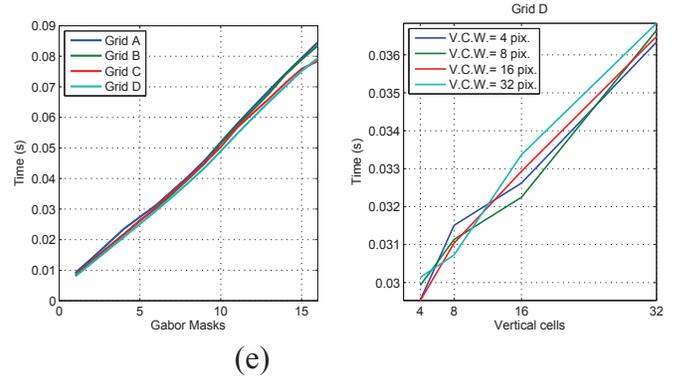
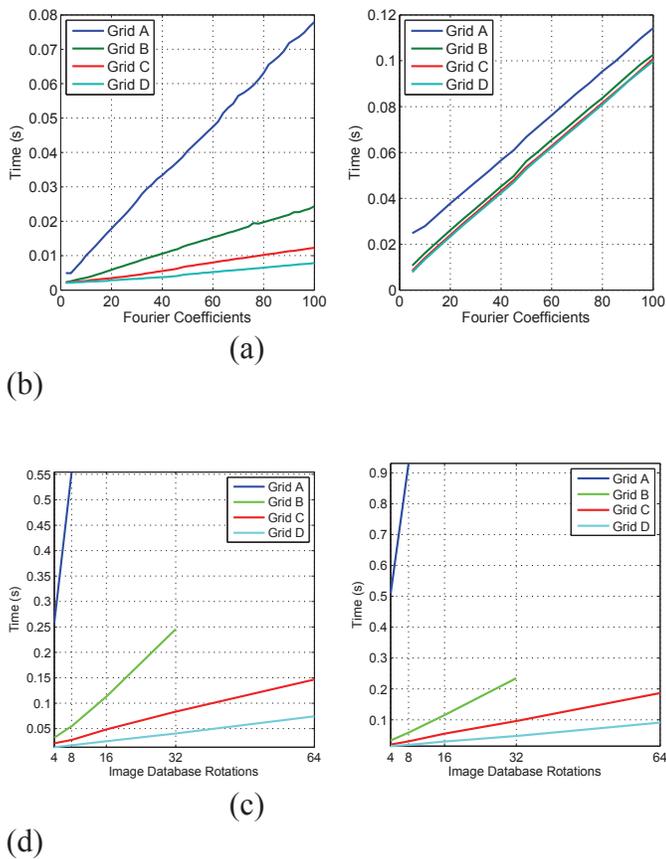
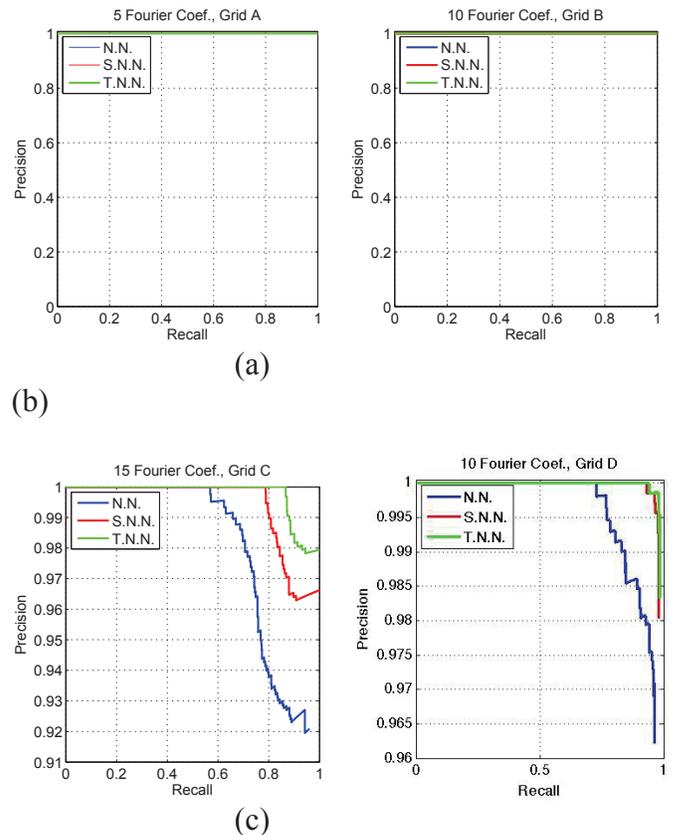


Figure 12. Recall-Precision charts for Fourier-based algorithm using (a) Grid A, (b) Grid B, (c) Grid C and (d) Grid D.



MAP BUILDING AND LOCALIZATION USING GLOBAL-APPEARANCE DESCRIPTORS APPLIED TO PANORAMIC IMAGES

(d)
Figure 13. Recall-Precision charts for PCA-based algorithm using (a) Grid A, (b) Grid B, (c) Grid C and (d) Grid D.

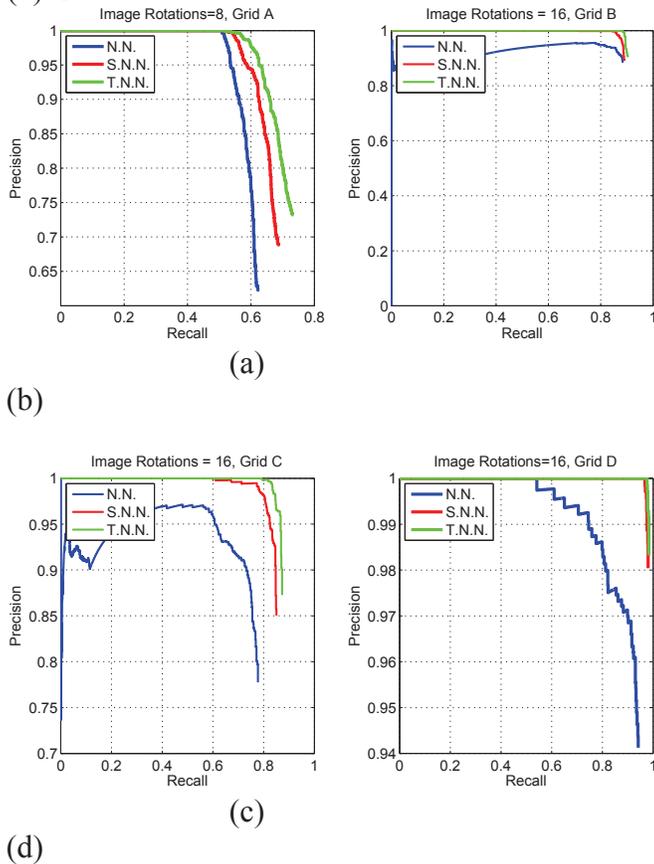
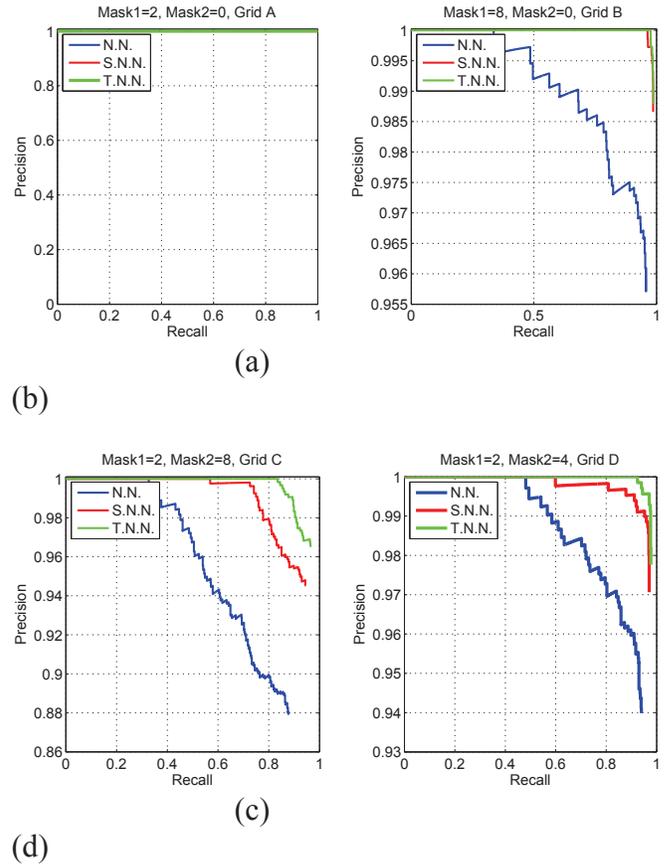


Figure 14. Recall-Precision charts for Gist-Gabor using (a) Grid A, (b) Grid B, (c) Grid C and (d) Grid D.



Fourier Signature technique

We use two matrices in order to represent the map obtained from the Fourier Signature compression that includes the module and the phase of the selected Fourier Coefficients respectively. When a new image arrives, the first step is to compute its Fourier Signature. Then, the location is estimated by calculating the Euclidean distance of the power spectrum between the image and the spectra of the map. We obtain the best match calculating the minimum distance. Once the position of the robot is known, the phase's vector associated with the most similar image retrieved is used to compute the orientation of the robot regarding the map created previously using eq. 2.

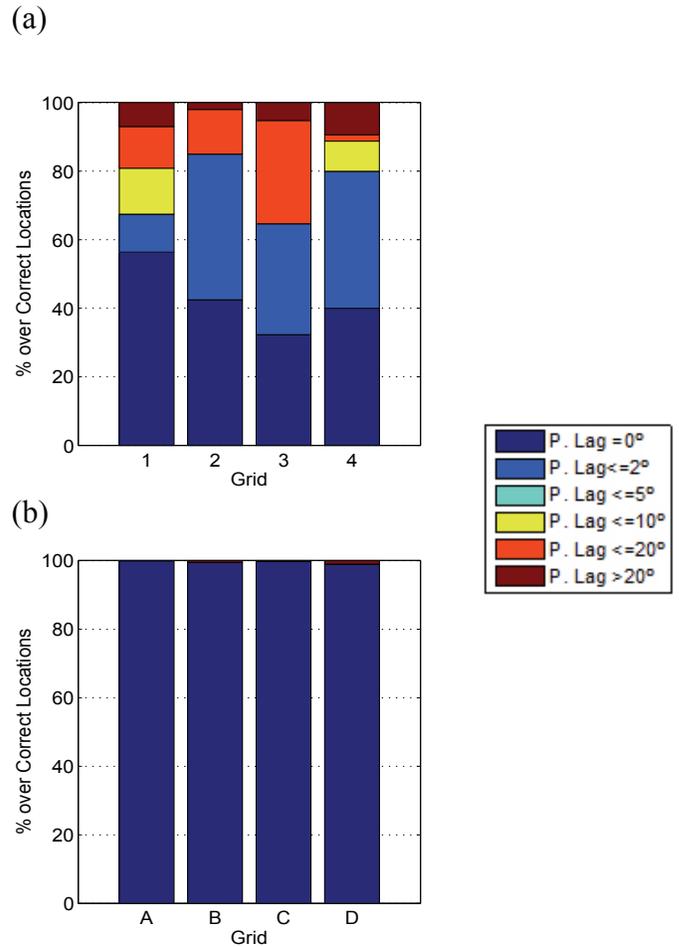
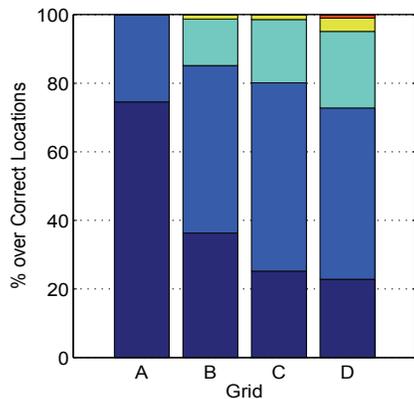
In fig.10 (a) and (b) we can see the time the algorithm

MAP BUILDING AND LOCALIZATION USING GLOBAL-APPEARANCE DESCRIPTORS APPLIED TO PANORAMIC IMAGES

spends to estimate the position and the pose respectively according to the number of Fourier coefficients it uses. To find the position, the elapsed time rises in accordance with the number of images the map stores, i.e. the grid, and the number of Fourier components. But the pose depends almost only on the number of coefficients per row. The reason is that the phase lag estimation is the heaviest computational part of the algorithm, and it depends only on the number of components we use. Regarding the position recovering, fig. 11 shows that in the experiments with the different grids the algorithm is able to find the best match using a relatively low number of Fourier components (no more than 15 coefficients per row), keeping an accuracy above the 92 percent.

The phase lag appears in fig. 15(a). In the grid A, with 5 Fourier coefficients the error remains less than or equal to two degrees in all the experiments. As we increase the distance between images in the map, the accuracy decreases. In any case, the algorithm is able to recover the orientation using a maximum of 5 components in the grid D with an error less than or equal to 5 degrees in the 92 percent of the correct locations.

Figure 15. Error in the phase estimation over correct locations using Fourier-based algorithm in (a) Grid A and (b) Grid D. Error in the phase estimation over correct locations using PCA-based algorithm in (c) Grid A and (d) Grid D. Error in the phase estimation over correct locations using Gist-Gabor in (e) Grid A and (f) Grid D.



(c)

PCA-Based Technique

The first step to estimate the pose using rotational PCA is to project the input image $\vec{I}^j \square \square^{M \times 1}$ onto the new eigenspace $\vec{p} = V^T \times \vec{I} \square \square^{M \times 1}$. The location is estimated by computing the module of \vec{p} and comparing with the modules of the projections of the map. The criterion is the minimum Euclidean distance. Once the position is known, we use the phase vector \vec{p}_{ph} to simulate the projections of the rotated siblings of the image to determine the orientation. Fig. 10(c) and (d) show the time spent on location and pose estimation. Comparing both charts we can see that, except in Grid A, the measurements

are similar, demonstrating that the phase recovering is computationally fast. Even so, this algorithm is the slowest in the majority of the experiments. In fig.12 we can see that with 8 rotations and the Grid A the database we obtain presents the worst position recovering results although we use the majority of the eigenvectors. We would need a map with more rotation per position to obtain acceptable results, but the computational requirements for computing this map make it unfeasible. However, fig. 12(b), fig. 12(c) and fig. 12(d) shows that with 16 rotations and 100 eigenvectors the results notably improve.

Regarding the phase estimation, with 8 rotations per image the error in grid A remains equal than or less than one degree in the 70 percent of cases, as we can see in fig. 15(b). The problem is that the error in the rest of the experiments is equal to or greater than 10 degrees. A similar situation happens in the other map configurations. Using 16 rotation per image for building the base, the error keeps less than or equal to two degrees in the 82%, 64% and 79% of the correct locations in grid B, C and D respectively, but for the other experiments it is greater than or equal to 10 degrees. If we want to increase the precision in the phase estimation, we have to increase the number of rotations per image, with the consequent computational cost.

Gist-Based Technique

We filter the input images with the same Gabor masks used to build the map. The maximum number of Gabor spatial scales used to extract the scenes information is two. After that, we compute the descriptor of the horizontal cells to estimate the position, and the vertical cells to the orientation. The elapsed time in the position recovering (fig. 10(e)) depends on the number of Gabor masks we use in order to filter the image. Fig. 10(f) shows the relationship between the elapsed time in pose estimation and the orientation parameters. As it

happens in the base creation, the number of vertical cells determines the results more than its size. The position estimation presents good results in both grids with few masks, as fig. 13 shows. In all the grids configurations, the precision of the descriptors is superior to 88 percent for the Nearest Neighbour, and around the 98 percent if we consider the three first results of the experiments.

The phase retrieval results appear in fig. 15(c). In all the grids, the descriptor is able to estimate the orientation of almost all the experiments without error using 16 vertical cells. They are binary results, as they present only experiments with error equal 0 or failures, i.e., error greater than 20° (it can be seen in the results of grid D). This is because the angle is discretized (eq. 18), and the rotations of the test images agreed with the distance between windows when using 16 vertical cells. However, if the artificial rotations of the images had a gap lower, we would have to use a larger number of vertical windows in order to increase the angle accuracy

CONCLUSIONS

This article presents the study and simulation results of the application of three different appearance-based algorithms to application of descriptors based on the global-appearance of panoramic images for robotic navigation tasks. We have studied the computational requirements of each method for the creation of dense maps of a real environment and the results of the pose estimation of a robot regarding the created map.

All of them have demonstrated to be perfectly valid to carry out the map building and pose estimation of a robot within the map. However, the PCA-based descriptor needs to include a minimum number of image rotations to be able to recover the location and orientation of the robot with an acceptable accuracy. As a consequence, when the number of images

MAP BUILDING AND LOCALIZATION USING GLOBAL-APPEARANCE DESCRIPTORS APPLIED TO PANORAMIC IMAGES

included into the map grows, the computational cost of the method can make it application unfeasible. Moreover, due to the fact that is a non-incremental method, we need to have the entire database before starting the robot navigation. By this reason, this method may be not advisable for certain tasks, such as SLAM. Gist-Gabor presents the most compact representation. Fourier Signature needs more memory to store the coefficients of the map images, but we need few components per row. Hence the memory requirements of both methods are similar. PCA technique outperforms the memory needs.

Regarding the elapsed time to create the map, rotational PCA clearly exceeds the other methods. Gist-Gabor lasts longer than Fourier Signature, and it is more dependant on the quantity of information it stores, i.e. the number of masks we use to filter the image.

In the position retrieval, the three algorithms present a high rate of correct location, except in the case of PCA using the densest map, since the time and memory requirement does not permit building a map with more information. Fourier Signature can be stood out, as it presents the greater accuracy in all the maps configurations, followed by Gist technique.

In the orientation estimation task, the results worse for the maps with a larger grid distance, i.e., with larger distance between images. PCA technique has the lowest accuracy. Although Gist-Gabor presents better results in the experimental part than Fourier Signature, it is important the fact that Gist-Gabor angle's estimation is sampled with regard to the number of cells we use, and it could increase time and memory consumptions as we need higher accuracy.

The results achieved in this article show how appearance-based methods can extract the information of panoramic images, and its feasible application to map building and localization. The promising results obtained in these experiments invite us to continue with the study of new descriptors and its performance using databases with illumination changes, noise or occlusions.

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

Dalal, N. and Triggs, B. (2005). "Histograms of oriented gradients for human detection". In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, San Diego, USA. Vol. II, pp. 886-893.*

Friedman, A. (1979). "Framing pictures: The role of knowledge in automatized encoding and memory for gist". In *Journal of Experimental Psychology: General, 108:316-355.*

Gabor, D. (1946). "Theory of communication". In *Electrical Engineers. Part III: Radio and Communication Engineering, Journal of the Institution of, vol.93, no.26, pp.429-457.*

Gil, A., Martinez, O., Ballesta, M., and Reinoso, O. (2009). "A comparative evaluation of interest point detectors and local descriptors for visual slam." *SPRINGER Machine Vision and Applications.*

Jogan, M. and Leonardis, A. (2000). "Robust localization using eigenspace of spinning-images". In *Proc. IEEE Workshop on Omnidirectional Vision, Hilton Head Island, USA, pp. 37-44. IEEE.*

Jogan, M. and Leonardis, A. (2003). "Robust localization using an omnidirectionalappearance-

**MAP BUILDING AND LOCALIZATION USING
GLOBAL-APPEARANCE DESCRIPTORS APPLIED TO PANORAMIC IMAGES**

based subspace model of environment.” In *Robotics and Autonomous Systems*, vol. 45, no. 1, pp. 51-72.

Kirby, M. (2000). “Geometric data analysis. an empirical approach to dimensionality reduction and the study of patterns”. *Wiley-Interscience, New York, USA*.

Krose, B., Bunschoten, R., Hagen, S., Terwijn, B., and Vlassis, N. (2004). “Household robots look and learn: environment modeling and localization from an omnidirectional vision system”. *Robotics Automation Magazine, IEEE*, 11(4):45 – 52.

Kruizinga, P. and Petkov, N. (1999) , "Nonlinear operator for oriented texture," *Image Processing, IEEE Transactions on* , vol.8, no.10, pp.1395-1407,

Leonardis, A. and Jogan, M. (2000). Robust localization using eigenspace of spinning-images. In *IEEE Workshop Omnidirectional Vision*. Ed. INTICC PRESS ISBN: 978-989-674-000-9 - pp. 250-255.

Manjunath, B. and Ma, W. (1996). “Texture features for browsing and retrieval of image data”. *Pattern Analysis and Machine intelligence, IEEE Transactions on*, 18(8):837 –842.

Menegatti, E., Maeda, T., and Ishiguro, H. (2004). Image based memory for robot navigation using properties of omnidirectional images. In *Robotics and Autonomous Systems*. Vol. 47, No. 4, pp. 251-276.

Moeller, R., Vardy, A., Kreft, S., and Ruwisch, S. (2007). “Visual homing in environments with anisotropic landmark distribution”. In *Autonomous Robots*, 23(3), 2007, pp. 231-245.

Oliva, A. and Schyns, P. (1997). “Coarse blobs or fine edges? evidence that information diagnosticity changes the perception of complex visual stimuli”. In *Cogn. Psychol.* 34, 72-107.

Oliva, A. and Torralba, A. (2001). “Modeling the shape of the scene: a holistic representation of the spatial envelope”. In *International Journal of Computer Vision*, Vol. 42(3): 145-175.

Payá, L., Fenandez, L., Reinoso, O., Gil, A., and Ubeda, D. (2009). ”Appearance-based dense maps creation: Comparison of compression techniques with panoramic images”. In *6th Int Conf on Informatics in Control, Automation and Robotics ICINCO 2009*. Ed. INTICC PRESS ISBN: 978-989-674-000-9 - pp. 250-255.

Thrun, S. (2003). Robotic mapping: A survey, in exploring artificial intelligence. In *The New Milenium*, pp. 1-35. Morgan Kaufmann Publishers, San Francisco, USA.

Torralba, A. (2003). Contextual priming for object detection. In *International Journal of Computer Vision*, Vol. 53(2), 169-191.

Turk, M. and Pentland, A. (1991), “Eigenfaces for Recognition”, *Journal of Cognitive Neuroscience*, Vol. 3, No. 1, 1991, pp. 71-86.

Uenohara, M. and Kanade, T. (1998). "Optimal approximation of uniformly rotated images: relationship between Karhunen-Loeve expansion and discrete cosine transform”, *Image Processing, IEEE Transactions on* , vol.7, no.1, pp.116-119.