COMPARISON OF GLOBAL-APPEARANCE TECHNIQUES APPLIED TO VISUAL MAP BUILDING AND LOCALIZATION

Extracting the Most Relevant Information from Panoramic Images

Francisco Amorós, Luis Payá, Oscar Reinoso and Luis M. Jiménez
Departamento de Ingeniería de Sistemas y Automática,
Miguel Hernández University, Avda. de la Universidad s/n, 03202, Elche (Alicante), Spain

Keywords: Robot Mapping, Appearance-based Methods, Omnidirectional Vision, Spatial Localization.

Abstract: Techniques based on the global appearance of visual information have proved to be a robust alternative in the field of robotic mapping and localization. However, they present some critical issues that must be studied when trying to build an application that works in real time. In this paper, we review and compare several methods to build a global descriptor of panoramic scenes and we study the critical parameters that make their applicable or not in real mapping and localizations tasks, such as invariance against rotations, computational costs and accuracy in robot localization. All the experiments have been carried out with omnidirectional images captured in a real environment under realistic lighting conditions.

1 INTRODUCTION

When a robot or a team of robots has to carry out a task that implies autonomous navigation through an environment, an internal representation of this environment is needed. This representation has to allow the robot to estimate its position and orientation using the information provided by the sensors it is equipped with. Omnidirectional visual systems are commonly used with this goal due to the richness of the information they provide and the relatively low cost they have. Classical researches into mobile robots provided with vision systems have focused on local features descriptors, extracting natural or artificial landmarks from the image to build the map and carry out the localization of the robot (Thrun, 2003).

Recent approaches propose processing the image as a whole without local feature extraction. These appearance-based techniques are interesting when dealing with unstructured environments where it may be hard to find patterns to recognize the scene. But we have to work with a large amount of information, having a high computational cost. That is the reason why we need to study compression techniques.

The localization task requires techniques that present rotational invariance in order to recognise the most similar image regardless of the robot's orientation in the ground plane. But some orientation information to estimate the pose of the robot is also necessary. Incremental methods are also advisable, since some navigation tasks require to add or modify elements of the map as the robot moves through the environment.

Several approaches to compress the visual information can be found in the literature. For example, PCA (Principal Components Analysis) has demonstrated being robust applied to image processing, as (Krose et al., 2007) shows. Other authors use the Fourier Transform to extract the most relevant information of an image (Menegatti et al., 2004).

(Paya et al., 2009) present a comparative study about appearance-based techniques. We complement this study and take into account three methods: Fourier Signature, Rotational PCA and Gist-Gabor. The last technique has proved previous promising results, although we have no notice it has been previously used in localization and mapping tasks.

2 REVIEW OF COMPRESSION TECHNIQUES

In this section we summarize some techniques to extract the most relevant information from a database made up of panoramic images.

Amorós F., Payá L., Reinoso O. and M. Jiménez L..
COMPARISON OF GLOBAL-APPEARANCE TECHNIQUES APPLIED TO VISUAL MAP BUILDING AND LOCALIZATION.
DOI: 10.5220/0003864703950398
In *Proceedings of the International Conference on Computer Vision Theory and Applications* (VISAPP-2012), pages 395-398
ISBN: 978-989-8565-04-4

2.1 Fourier-based Techniques

As shown in (Menegatti et al., 2004) it is possible to represent an image using the Discrete Fourier Transform of each row. Taking profit of the Fourier Transform properties, since the most relevant information concentrates in the low frequency components of the sequence, we keep the first coefficients to represent each row. Moreover, as we work with omnidirectional images, when the Fourier Transform of each row is computed, another interesting property appears: rotational invariance. Comparing the transform of a row and the transform of the same sequence rotated, the modulus are the same and just the phase changes. The modulus let the position estimation, and with the phase coefficients we can find out the relative rotation.

2.2 PCA-based Techniques

PCA-based techniques have proved to be a very useful compressing method (Krose et al., 2007). They make possible that, having a set of N images with M pixels each, $\vec{I}^j \in \Re^{Mx1}, j = 1 \dots N$, we could transform each image in a feature vector (also named projection of the image) $\vec{p}^j \in \Re^{kx1}, j = 1 \dots N$, being k the PCA features containing the most relevant information of the image, $k \leq N$. However, if we apply PCA directly over the matrix that contains the images, we obtain a database with information of just one orientation of each scene. To solve this problem, in (Leonardis and Jogan, 2000) the use of the *Eigenspace of Spining-Images* is proposed. This technique creates a set of spinning images from every image included in the map. After that, the database is compressed by means of PCA analysis. The robustness in localization and angular resolution of the map depends on the number of rotated siblings of each image we include.

2.3 Gist-based Techniques

Gist is another concept that can be used to compress visual information as (Friedman, 1979) details. It can be defined as an abstract representation that activates the memory of scenes' categories. They try to obtain the essential information of the image simulating the human perception system, i.e., identifying a scene through its colour or remarkable structures, avoiding the representation of specific objects. In (Oliva and Torralba, 2001) this idea is developed under the name of *holistic representation of the spatial envelope* to create a descriptor. In (Torralba, 2003) this model is 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. The descriptor we propose is named Gist-Gabor since it uses Gabor filtering in order to obtain frequency and orientation information using the global image.

The first step consists in creating a bank of the Gabor masks with different resolutions and orientations. Then, the image is filtered with the set of filters. The results encode different structural information. To create the descriptor, we calculate the average pixel's value within cells with the same width as the omnidirectional image, obtaining an array of rotational invariant characteristics. To know the relative orientation between two rotated images, vertical windows with the image's height are used, making up a vector with the mean value the pixels they contains. By rotating the order of its components and comparing with the database we estimate the orientation.

3 LOCALIZATION AND ORIENTATION RECOVERING

In this section we assess each algorithm by calculating the pose of the robot within a map created previously, and the time they spent. The image database we have used to carry out the experiments belongs to the Technique Faculty of Bielefeld University ((Moeller et al., 2007). It has been collected in three different living spaces under realistic illumination conditions. The images are structured in a 10x10 cm rectangular grid. In the experiments, we have varied the distance between images of the database when building the map to simulate different conditions. Table 1 shows the different grids information.

Table 1: Grid's size and number of images selected.

	GridA	GridB	GridC	GridD
Distance	10cm	20cm	30cm	40cm
Images	746	204	92	54

The test set is made up of all the available images in the database and 15 artificial rotations of each one (every 22.5°), but the images included in the map. The simulations have been obtained using Matlab R2009b under Mac OS X. The position retrieval accuracy is studied 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), with information about if a correct location is in the Nearest Neighbour (N.N.), i. e., if it is the first result selected, or between Second or Third Nearest Neighbours (S.N.N or T.N.N).

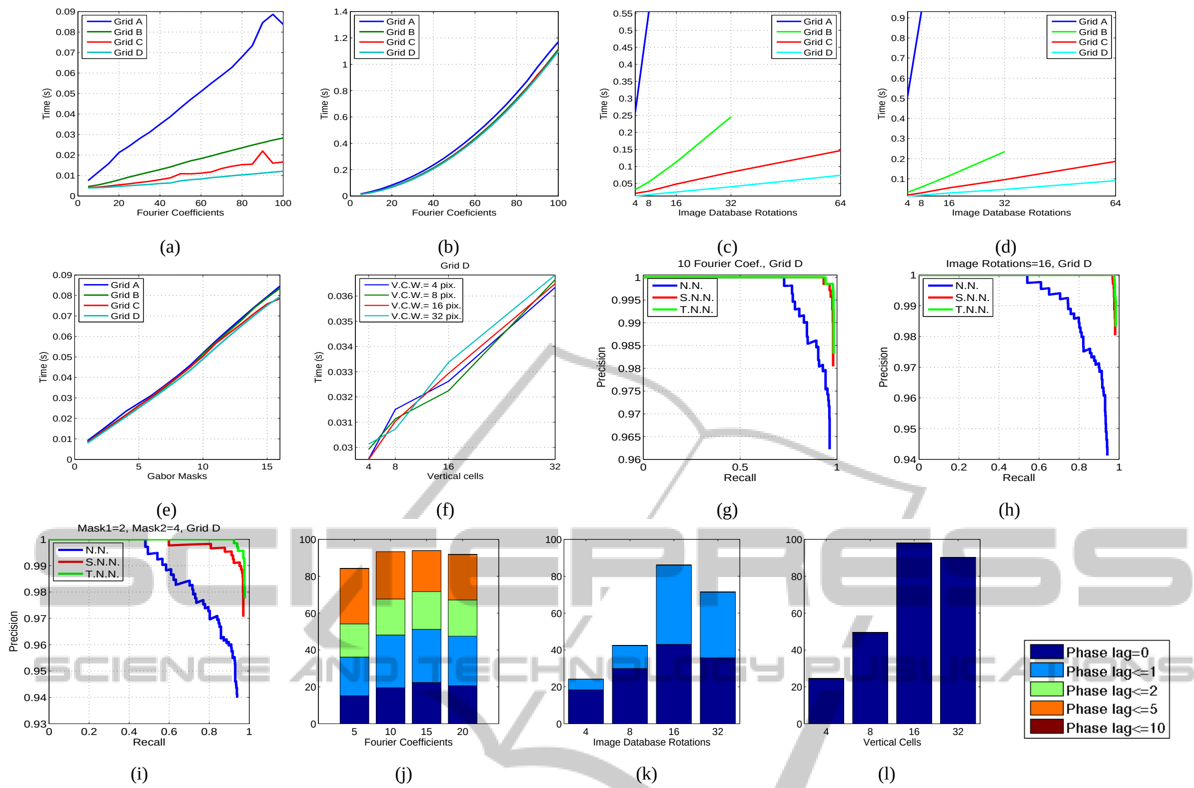

Figure 1: Elapsed time for (a) location and (b) pose estimation using Fourier-based algorithm. Elapsed time for (c) location and (d) pose estimation using PCA-based algorithm. Elapsed time for (e) location and (f) pose estimation using Gist-Gabor. Recall-Precision charts using grid D for (g) Fourier-based algorithm, (h) PCA-based algorithm and (i) Gist-Gabor. Phase error over correct locations using Grid D for (j) Fourier-based algorithm, (k) PCA-based algorithm and (l) Gist-Gabor.

Regarding the rotation, we represent the results accuracy in bar graphs depending on how much they differ from the correct ones, in percentage over correct locations. In order to avoid redundant information, we include only the pose estimation experiments in the most critical case, i.e., using the Grid D.

3.1 Fourier Signature Technique

The map obtained with the Fourier Signature is represented with two matrices: the module and the phase of the selected Fourier Coefficients. The location is estimated by calculating the minimum Euclidean distance of the power spectrum between the image and the spectra of the map. The phases' vector associated with the most similar image retrieved in the map is used to compute the orientation. In fig.1(a) we can see that, 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 (fig.1(b)). This is due to the orientation estimation, since it is the computationally heaviest part of the algorithm and it depends only on the number of components we use. Regarding the position recovering, fig.1(g) show that in both experiments the algorithm is able to find the best match using a relatively low number of Fourier components. The phase lag appears in fig.1 (j). The algorithm is able to recover the orientation using 10 components with an error less than or equal 5 degrees in the 92 per cent of correct locations.

3.2 PCA-based Techniques

When a new image $\vec{I} \in \Re^{1xM}$ arrives, it is projected onto the eigenspace $\vec{p} = V^T \cdot \vec{I} \in \Re^{kx1}$. 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 phases vector $\vec{p}_{ph}$ to simulate the projections of the rotated siblings of the image to determine the orientation. Fig. 1(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 quite fast. Even so, this algorithm is the slowest in the majority of the experiments. Fig.1(h) shows that with 16 rotations and 100 eigenvectors the position estimation presents good accuracy. Fig.1(k)) shows that, with 16 rotations, the percentage of experiments equal than or under one degree is 86% although in the rest of the experiments is greater than 10 degrees.

3.3 Gist-based Techniques

To extract the information of a test image, we filter the image with the same Gabor masks used to built the map. The maximum number of spatial escales used is two. After that, we compute the descriptor using the same horizontal and vertical cells as in the map. The elapsed time in the position recovering (fig. 1(e)) depends on the number of Gabor masks we use in order to filter the image. Fig. 1(f) shows the relationship between the elapsed time in pose estimation and the orientation parameters. The number of vertical cells determines the results over its size. The position estimation presents good accuracy with few masks (fig. 1(i)). The phase retrieval results appear in fig.1(l). The descriptor is able to estimate the orientation of almost all the experiments without error using 16 vertical cells. But they are binary results, since the angle is discretized depending on the number of vertical cells we apply to the image.

4 CONCLUSIONS

In this paper we have presented the comparison of different appearance-based algorithms applied to the creation of a descriptor using panoramic images. We have studied the elapsed and the accuracy in the pose estimation regarding a previously created map.

All of them have demonstrated to be perfectly valid to carry out the estimation of the pose of a robot within the map. However, when the number of images included in the map grows, the computational cost of PCA descriptor can make it application unfeasible. Moreover, it is a non-incremental method.

Regarding the elapsed time, rotational PCA 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. The three algorithms present a high rate of retrieved positions, being Fourier Signature remarkable.

In the orientation estimation task, PCA technique has the lowest accuracy. Although Gist-Gabor outperforms Fourier Signature, 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.

To finish, this paper proves again the possibilities that appearance-based techniques offer.The results achieved encourage us to continue studying new possibilities and deepening in its development, looking for new available techniques and improving its robustness to illumination change, noise or occlusions.

ACKNOWLEDGEMENTS

This work has been supported by the Spanish government through the project DPI2010-15308.

REFERENCES

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*.

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.

Krose, B., Bunschoten, R., Hagen, S., Terwijn, B., and Vlassis, N. (2007). Visual homing in enviroments with anisotropic landmark distrubution. In *Autonomous Robots, 23(3), 2007, pp. 231-245*.

Leonardis, A. and Jogan, M. (2000). Robust localization using eigenspace of spinning-images. In *IEEE Workshop Omnidirectional Vision*. Proceedings of the IEEE Workshop on Omnidirectional Vision, IEEE Computer Society,pp. 37-44.

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 enviroments with anisotropic landmark distrubution. In *Autonomous Robots, 23(3), 2007, pp. 231-245*.

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*.

Paya, 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*. Ed. INSTICC 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*.