Advances in Intelligent Systems and Computing

Volume 253

Series Editor

Janusz Kacprzyk, Polish Academy of Sciences, Warsaw, Poland
e-mail: kacprzyk@ibspan.waw.pl

For further volumes:
http://www.springer.com/series/11156

About this Series

The series "Advances in Intelligent Systems and Computing" contains publications on theory, applications, and design methods of Intelligent Systems and Intelligent Computing. Virtually all disciplines such as engineering, natural sciences, computer and information science, ICT, economics, business, e-commerce, environment, healthcare, life science are covered. The list of topics spans all the areas of modern intelligent systems and computing.

The publications within "Advances in Intelligent Systems and Computing" are primarily textbooks and proceedings of important conferences, symposia and congresses. They cover significant recent developments in the field, both of a foundational and applicable character. An important characteristic feature of the series is the short publication time and world-wide distribution. This permits a rapid and broad dissemination of research results.

Manuel A. Armada · Alberto Sanfeliu
Manuel Ferre
Editors

ROBOT2013: First Iberian Robotics Conference

Advances in Robotics, Volume 2

Springer

Editors
Manuel A. Armada
State Agency National Research Council (CSIC)
Madrid
Spain

Manuel Ferre
University Polytechnic of Madrid (UPM)
Madrid
Spain

Alberto Sanfeliu
University Technical of Catalonia (UPC)
Barcelona
Spain

ISSN 2194-5357 ISSN 2194-5365 (electronic)
ISBN 978-3-319-03652-6 ISBN 978-3-319-03653-3 (eBook)
DOI 10.1007/978-3-319-03653-3
Springer Cham Heidelberg New York Dordrecht London

Library of Congress Control Number: 2013952937

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Contents

Part XVII: Grasping and Dexterous Manipulation

Topological Height Estimation Using Global Appearance of Images

Francisco Amorós, Luis Payá,
Oscar Reinoso, Luis Miguel Jiménez, and Miguel Juliá

Departamento de Ingeniería de Sistemas Industriales, Miguel Hernández University,
Avda. de la Universidad s/n, 03202, Elche (Alicante), Spain
{famoros,lpaya,o.reinoso,luis.jimenez,mjulia}@umh.es
http://arvc.umh.es

Abstract. In this work we present a comparison of different methods for topological height estimation applicable in UAVs navigation tasks using omnidirectional images. We take profit of the camera calibration information in oder to obtain different projections of the visual information from the omnidirectional images. The descriptors used to collect the visual information are based on the global appearance of the scenes. We test the algorithms using a real and dealing database.

Keywords: UAV, global appearance descriptors, zooming, omnidirectional image, topological navigation.

1 Introduction

Visual systems are commonly used in robotics navigation tasks. The richness of the information that a camera provides and the multiple possibilities of configurations and applications make them a popular sensing mechanism. We focus our work in omnidirectional vision and global appearance descriptors. In the literature, we can find numerous examples where omnidirectional visual systems are employed in navigation tasks, such as [19] and [9].

Classical research into mobile robots provided with vision systems has focused on local features descriptors, extracting natural or artificial landmarks from the image. Recent approaches propose processing the image as a whole, without local feature extraction. These global appearance techniques have demonstrate a good accuracy on the floor plane navigation in both location and orientation estimation [5], [3].

Nowadays, Unmanned Aerial Vehicles (UAVs) are becoming very popular as a platform in the field of robotic navigation research. In this sense, we can find in [12], [6], [18] different works that study the motion and attitude of UAVs using visual systems. Specifically, they are based on image feature extraction or image segmentation in order to extract valuable information of the scenes.

The aim of this paper is to extend the use of the global appearance descriptors to experiments where the height of the mobile robot changes. For that purpose,

M.A. Armada et al. (eds.), *ROBOT2013: First Iberian Robotics Conference*,
Advances in Intelligent Systems and Computing 253,
DOI: 10.1007/978-3-319-03653-3_7,

we suppose that the UAV is stabilized and the visual sensor has the same attitude, which corresponds with the perpendicular regarding the floor plane. In particular, we study the ability of height estimation using global appearance descriptors.

The experimental database is composed of omnidirectional images acquired using a catadioptric system composed of an hyperbolic mirror and a camera.

The remainder of the paper is structured as follows: Section 2 includes the global appearance descriptors we use in order to compress the visual information. Section 3 discusses the different methods used with the purpose of finding the relative height between images acquired in a same point in the floor plane. In the next section, the database used in the experiments is presented. Section 5 gathers the experimental results, and finally, the main conclusions are included in section 6.

2 Global Appearance Descriptors

In this section we summarize some techniques to extract the most relevant information from images. In particular, we describe descriptors based on the global appearance of scenes. These descriptors are computed working with the image as a whole, avoiding segmentation or landmarks extraction, trying to keep the amount of memory to a minimum.

2.1 Fourier Signature

In [11] the Fourier Signature is defined. It is possible to represent an image using the Discrete Fourier Transform of each row. So, we can expand each row of an image image $\{a_n\} = \{a_0, a_1, \ldots, a_{N-1}\}$ into the sequence of complex numbers $\{A_n\} = \{A_0, A_1, \ldots, A_{N-1}\}$:

$$\{A_n\} = \mathcal{F}[\{a_n\}] = \sum_{n=0}^{N-1} a_n e^{-j\frac{2\pi}{N}kn}, \quad k = 0, \ldots, N-1. \tag{1}$$

Taking profit of the Fourier Transform properties, we just keep the first coefficients to represent each row since the most relevant information concentrates in the low frequency components of the sequence. Moreover, when working with omnidirectional images, the modulus of the Fourier Transform of the image's rows is invariant against rotations in the perpendicular plane of the image. Representing each row of the original image as $\mathcal{F}[\{a_n\}]$ and $\mathcal{F}[\{a_{n-q}\}]$ the same row shifted q pixels, being q proportional to the relative rotation between images, the rotational invariance can be expressed with the shift theorem as:

$$\mathcal{F}[\{a_{n-q}\}] = A_k e^{-j\frac{2\pi qk}{N}}, \quad k = 0, \ldots N-1, \tag{2}$$

where $\mathcal{F}[\{a_{n-q}\}]$ is the Fourier Transform of the shifted sequence, and A_k are the components of the Fourier Transform of the non-shifted sequence.

2.2 2D Fourier Transform

When we have an image f(x,y) with Ny rows and Nx columns, the 2D discrete Fourier Transform is defined through:

$$\mathcal{F}[f(x,y)] = F(u,v) = \frac{1}{N_y} \sum_{x=0}^{N_x-1} \sum_{y=0}^{N_y-1} f(x,y) e^{-2\pi j \left(\frac{ux}{N_x} + \frac{vy}{N_y} \right)}, \\ u = 0, \ldots, N_x - 1, v = 0, \ldots, N_y - 1. \tag{3}$$

The components of the transformed image are complex numbers so it can be split in two matrices, one with the modules (power spectrum) and other with the angles. The most relevant information in the Fourier domain concentrates in the low frequency components. Furthermore, removing high frequency information can lead to an improvement in localization because these components are more affected by noise. Another interesting property when we work with panoramic images is the rotational invariance, which is reflected in the shift theorem:

$$\mathcal{F}[f(x-x_0, y-y_0)] = F(u,v) \cdot e^{-2\pi j \left(\frac{ux_0}{N_x} + \frac{vy_0}{N_y} \right)}, \\ u = 0, \ldots, N_x - 1, v = 0, \ldots, N_y - 1. \tag{4}$$

According to this property, the power spectrum of the rotated image remains the same of the original image and only a change in the phase of the components of the transformed image is produced, whose value depends on the shift on the x-axis (x_0) and the y-axis (y_0). Taking into account Eq. (4), the first row of the bidimiensional Fourier Transform, which corresponds with $v = 0$, is only affected by shifts on the x-axis, whereas the first column of the transform, which corresponds with $u = 0$, is only affected by shifts on the y-axis.

2.3 Spherical Fourier Transform

Omnidirectional images can be projected onto the unit sphere when the intrinsic parameters of the vision system are known. Being $\theta \in [0, \pi]$ the colatitude angle, and $\phi \in [0, 2\pi)$ the azimuth angle, the projection of the omnidirectional image in the 2D sphere can be expressed as $f(\theta, \phi)$. In [4], it is shown that the spherical harmonic functions Y_{lm} form a complete orthonormal basis over the unit sphere. Any square integrable function defined on the sphere $f \in L^2(s^2)$ can be represented by its spherical harmonic expansion as:

$$f(\theta, \phi) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} \widehat{f}_{lm} Y_{lm}(\theta, \phi), \tag{5}$$

with $l \in \mathbb{N}$ and $m \in \mathbb{Z}$, $|m| \leq l$. $\widehat{f}_{lm} \in \mathbb{C}$ denotes the spherical harmonic coefficients, and Y_{lm} the spherical harmonic function of degree l and order m defined in Eq. 6.

$$Y_{lm}(\theta, \phi) = \sqrt{\frac{2l+1}{4\pi} \frac{(l-m)!}{(l+m)!}} P_l^m(\cos\theta) e^{im\theta}, \tag{6}$$

where $P_l^m(x)$ are the associated Legendre functions.

It is possible to obtain a rotationally invariant representation from the Spherical Fourier Transform. Considering B the band limit of f, the coefficients of $e = (e_1, ..., e_B)$ are not affected by 3D rotations of the signal, where

$$e_l = \sqrt{\sum_{|m| \leq l} |\widehat{f}_{lm}|^2}. \tag{7}$$

We can find more information and examples of applications of the Spherical Fourier Transform in navigation tasks in other works such as [8], [10], [15], [7] and [16]

3 Height Estimation Methods

In this section we explain the different techniques used to obtain a measurement of the relative height of images captured in a same point. We make use of functions included in the Matlab Toolbox *OCamCalib* [14] to calibrate the camera and obtain different views of the visual information from the omnidirectional image.

3.1 Zooming of the Orthographic View

In [1], a method to obtain relative distance between images using zooming is presented. We propose to make use of the zooming concept with the purpose of measuring the vertical shift of a UAV.

However, we can not extract valuable information zooming the omnidirectional image directly. We need a representation of the image perpendicular to the robot movement. For that reason, we use the orthographic view [13] of the scene. In [9], [2], we find examples where orthographic view is used in robot navigation tasks.

We vary the distance of the plane where the omnidirectional image is projected to obtain different zooms of the bird-eye view by changing the focal distance.

The indicator of the vertical distance between two images using this method is the focal difference between both images.

After obtaining the orthographic view, we need to describe the scene. The descriptors we use are the Fourier Signature and the 2D Fourier Transform.

3.2 Camera Coordinate Reference System Movement

As shown in [17], given an image, it is possible to simulate the movement of the coordinate reference system (CRS) of the camera using the epipolar geometry, modifying the projection of the original image. The reprojected image, that uses the new CRS, reflects the movement of the camera.

Fist of all, we estimate the coordinates of the image in the real world in pixels. $m = [m_{x_{pix}}, m_{y_{pix}}]$ are the pixel coordinates regarding the omnidirectional image center. The camera calibration allows us to obtain the coordinates in the real world of the image. The image will be represented in the unit sphere $M \in \mathbb{R}^3$.

Then, we apply a change in the camera reference system:

$$M' = M + \rho \cdot T, \tag{8}$$

being T the unitary displacement vector in the z-axis, ($T = [0, 0, 1]^T$), and ρ a scale factor proportional to the displacement of the CRS.

Once we have the new coordinates of the image M', we can obtain the new pixel coordinates m'. Doing the association of the pixels of m with the new coordinates m', we obtain the new omnidirectional image that includes the camera CRS movement.

We have to take into account that, when we match the correspondences between m and m', some pixel coordinates of the new image might lay outside the image frame, and some other pixels might not have associated any value. We interpolate the values of the pixels that have not any association.

After obtaining the new coordinates of the image, we need to gather the visual information using a descriptor. Note that from M', we can obtain different representations of the visual information. Specifically, we use three different representations of the scene: the orthographic view of the omnidirectional image, the panoramic image, and the unit sphere. In Fig. 3, an example of each projection is shown.

We use the Fourier Signature and the 2D Fourier Transform to describe the orthographic and the panoramic views, whereas the Spherical Fourier Transform describes the unit sphere projection.

To obtain the height difference of two scenes captured in the same (x,y) position, we simulate different vertical CRS movements of the reference image, and compare them with the test image. Then, we look for the best image association, using the minimum Euclidean Distance of the image's descriptors.

The height difference using this technique is represented by the displacement scale factor ρ of the reference image.

4 Experimental Database

In order to carry out the experiments, we have acquired our own database of omnidirectional images in outdoor locations. We use a catadioptric system composed of a hyperbolic mirror and a camera with a resolution of 1280x960 pixels. The camera has been coupled to a tripod that allow us to have a range of 165cm in height.

The image acquisition has been done in 10 different locations. From every position, we capture 12 images in different heights. The minimum height is 125cm (h=1), and the maximum is 290cm (h=12), with a step of 15cm between consecutive images. Therefore, the database is composed of 120 images captured in real conditions. We do not vary the orientation of the images captured in a same location, although small rotations and short displacements have been unavoidable.

In the database, we include images near and far from buildings, garden areas and a parking. We also vary the time when the images are captured to change the illumination conditions and to have a more dealing database.

Fig. 1 and Fig. 2 include some examples of database images.

(a) (b) (c)

Fig. 1. Example of images captured in three different locations varying the relative position with the nearest building and the illumination conditions

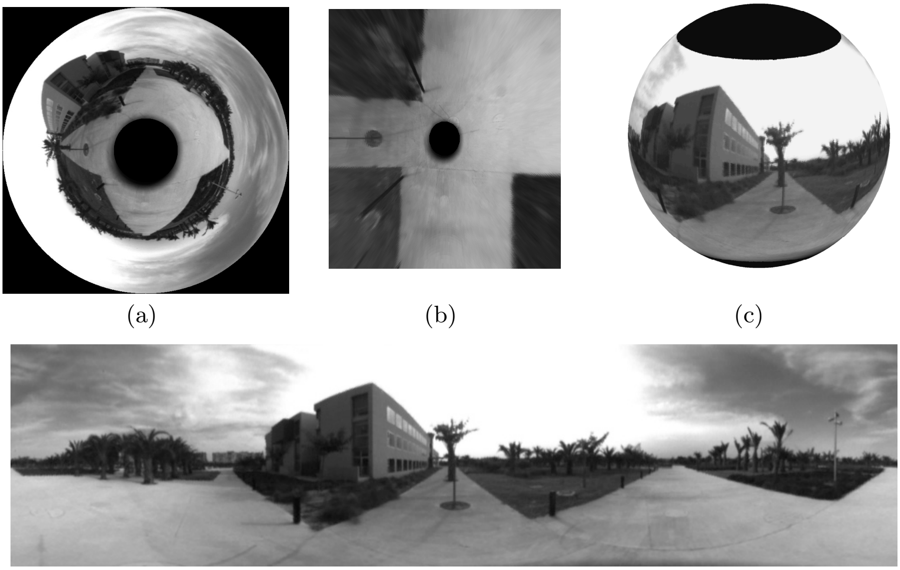
(a) (b) (c)

Fig. 2. Example of images captured at three different heights in the same location. (a) is at a height of 125 cm, (b) is at a height of 200 cm and (c) is at a height of 290 cm.

In the experiments, we use different representations of the original visual information. Specifically, we compute the panoramic image, the orthographic view (or bird-eye view) and the projection onto the unit sphere. Fig. 3 includes an example of each representation.

5 Experiments and Results

Our goal is to check whether the different techniques provide a topological measurement in the image space proportional to the real change in height of the scenes. The topological indicator will depend on the height estimation method. For that purpose, we test all the methods included in Section 3 using three different experiments.



(a) (b) (c)

(d)

Fig. 3. Different projections of the same image. (a) Omnidirectional image, (b) Orthographic view, (c) Unit Sphere projection and (d) Panoramic view.

In the first experiment, we estimate the height of the images regarding lowest image in height ($h = 1$) for each location. We simulate several vertical shifts (with focal distance change or else the constant ρ variation depending on the technique). We compute the descriptor of the resulting images and create a comparison base with them. After that, we compare the other images captured in each localization, which are in different heights, with the base.

The match criteria is the minimum Euclidean Distance of the descriptors. In Fig. 4 we include the mean value and standard deviation of the height estimation the 10 different locations.

The second experiment is analogous to the first one, but we vary the reference image that forms the base. In this case, we choose the image corresponding to $h = 5$ (185 cm) as reference, having test images both below and above the comparison image. Fig. 5 includes the mean value and standard deviation of the results for the 10 different locations.

In the third experiment, we focus the analysis in the gradient of heights. For each location, we carry out as many comparisons as possible given a difference of heights, taking the reference images at different heights. Specifically, for $\Delta h = 2$ (i.e., 30 cm), we have 100 experiments; for $\Delta h = 4$, 80 experiments; for $\Delta h = 6$, 60; and for $\Delta h = 8$, 40 different comparisons. The results are included in Fig. 6.

Taking into account all the experimental results, we can confirm that all the methods present a monotonically increasing tendency as we increment the height lag between the compared images. Moreover, considering the results of the second

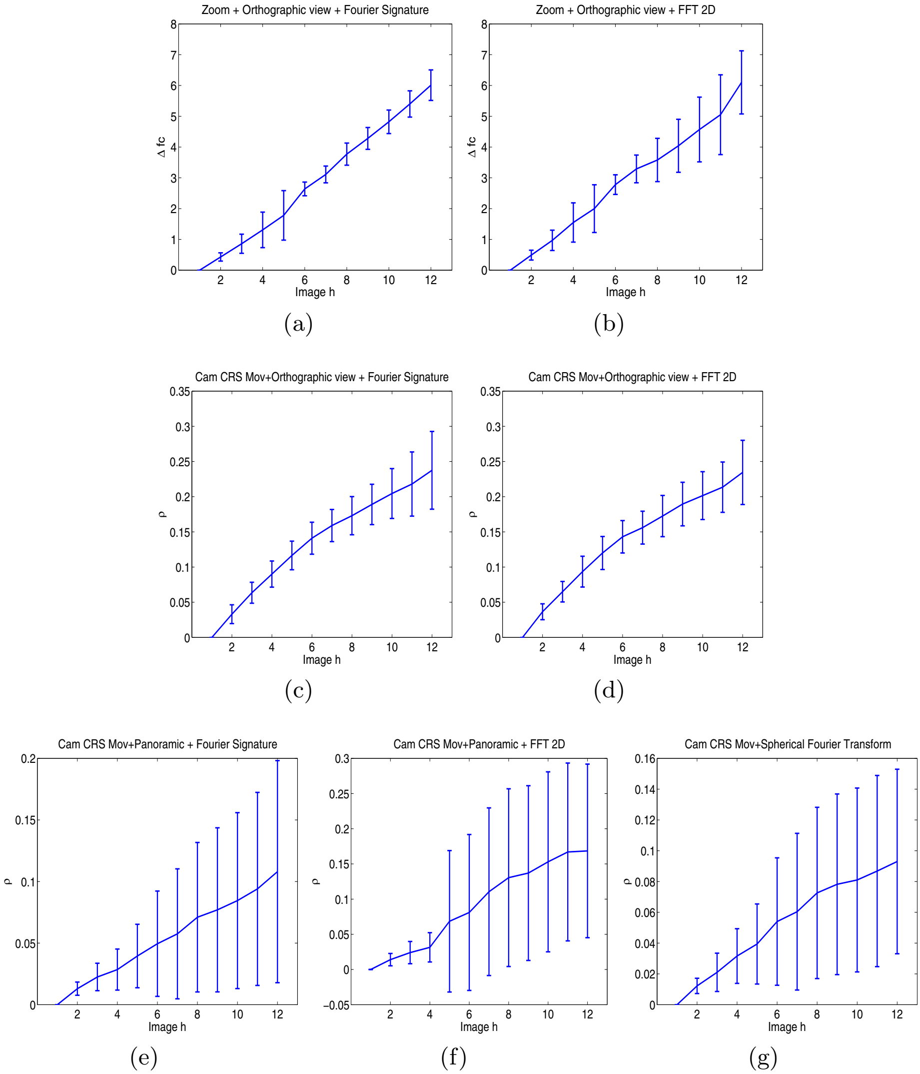


Fig. 4. Experimental results estimating the height regarding the image with h=0. Mean and standard deviation of all the different locations using the different methods: (a) Zooming over the the Orthographic view using the Fourier Signature descriptor, (b) Zooming over the the Orthographic view using the 2D Fourier Transform descriptor, (c) Camera CRS Movement with Orthographic view using the Fourier Signature descriptor, (d) Camera CRS Movement with Orthographic view using the 2D Fourier Transform descriptor, (e) Camera CRS Movement with Panoramic view using the Fourier Signature descriptor, (f) Camera CRS Movement with Panoramic view using the 2D Fourier Transform descriptor, and (g) Camera CRS Movement with Unit Sphere projection using Spherical Fourier Transform descriptor.

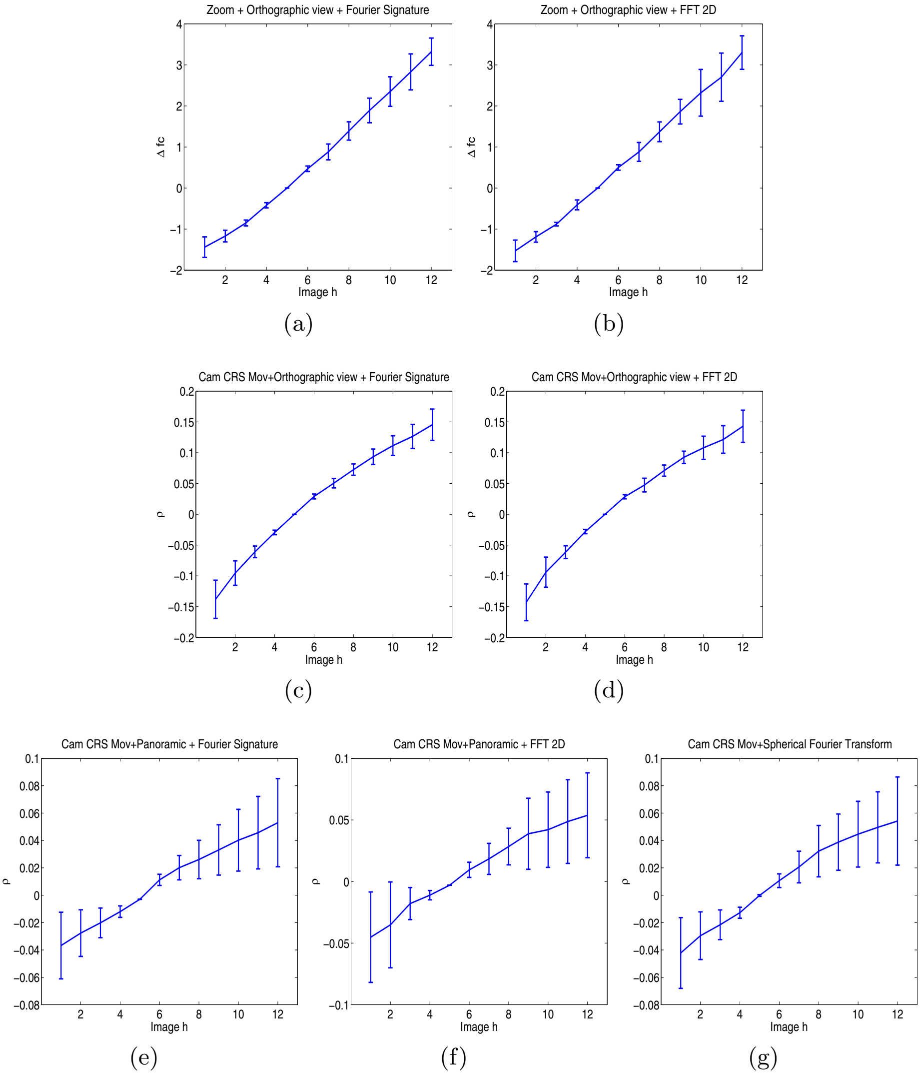


Fig. 5. Experimental results estimating the height regarding the image with h=5. Mean and standard deviation of all the different locations using the different methods: (a) Zooming over the the Orthographic view using the Fourier Signature descriptor, (b) Zooming over the the Orthographic view using the 2D Fourier Transform descriptor, (c) Camera CRS Movement with Orthographic view using the Fourier Signature descriptor, (d) Camera CRS Movement with Orthographic view using the 2D Fourier Transform descriptor, (e) Camera CRS Movement with Panoramic view using the Fourier Signature descriptor, (f) Camera CRS Movement with Panoramic view using the 2D Fourier Transform descriptor, and (g) Camera CRS Movement with Unit Sphere projection using Spherical Fourier Transform descriptor.

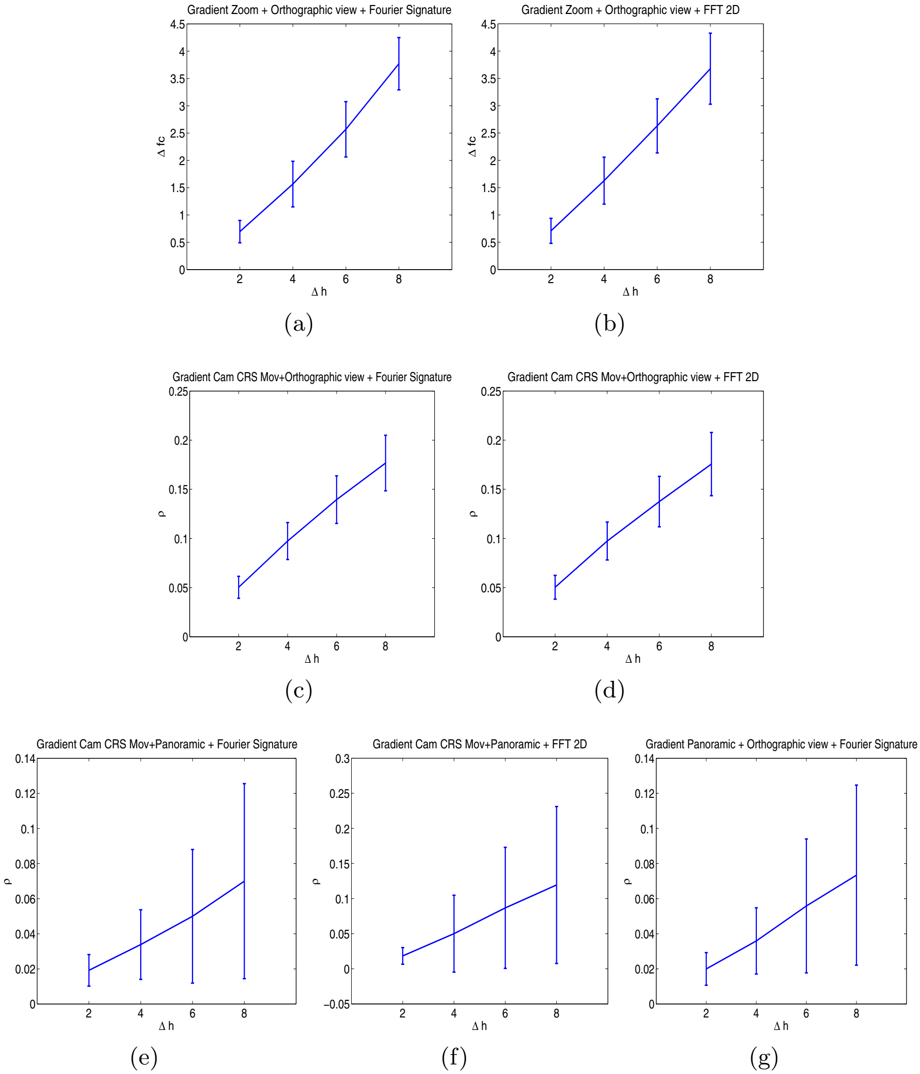


Fig. 6. Experimental results estimating four different gradients. Mean and standard deviation of all the possible experiments using the different methods: (a) Zooming over the the Orthographic view using the Fourier Signature descriptor, (b) Zooming over the the Orthographic view using the 2D Fourier Transform descriptor, (c) Camera CRS Movement with Orthographic view using the Fourier Signature descriptor, (d) Camera CRS Movement with Orthographic view using the 2D Fourier Transform descriptor, (e) Camera CRS Movement with Panoramic view using the Fourier Signature descriptor, (f) Camera CRS Movement with Panoramic view using the 2D Fourier Transform descriptor, and (g) Camera CRS Movement with Unit Sphere projection using Spherical Fourier Transform descriptor

experiment included in Fig. 5, when the test images are below the reference, the height indicator has negative sign. This allow us to determine the direction of the height difference.

The methods based on the orthographic view present better results than the techniques based on other image projections. As a rule, when we increase the height difference, the standard deviation increases. This is specially remarkable in the method based on the camera CRS movement that uses the panoramic view and the unit sphere projection.

When we simulate the CRS movement described in Eq.(8), we are applying the same displacement in all the pixels of the image, independently of the distance of the object depicted in the scene. However, when we change the height of the camera in the real world, the objects vary their position in the image depending on their relative position with the vision system. As an instance, the projection of objects that are far away from the camera changes less than the projection of closer objects when we vary the sensor location.

This is particularly notable when we work with the panoramic view or the unit sphere projection, as we use almost the whole image, that usually includes information of objects placed in different distances from the camera system. On the contrary, the orthographic view usually include elements that are at a similar distance (near the floor plane).

Despite this fact, the performance of all the algorithms are acceptable until a height lag of 45cm ($\Delta h = 3$).

Regarding to the descriptor used to represent the image, the Fourier Signature presents better accuracy than Fourier 2D, although there is no important difference in their performance.

In the experiments, we can also realize that the Spherical Fourier Transform over the unit sphere outperforms the Fourier Signature and the FFT 2D over the panoramic image. However, as stated above, the handicaps derived of the camera CRS movement technique affect the results.

6 Conclusions and Future Work

In this work we have presented a comparison of different topological height estimation techniques applicable in UAVs navigation tasks using omnidirectional images. The approaches we include in this work describe the visual information using global appearance descriptors. The experiments have been carried out using our own database captured in a real environment under challenging conditions.

The experimental results demonstrate that all methods proposed are able to estimate the relative height between two scenes captured in the same location for small height lags. However, the techniques based on the orthographic view of the scene present a better accuracy. Moreover, the Fourier Signature outperforms as a descriptor of the scenes.

The algorithm can deal with small rotations and short displacements, although it is not designed to work under bigger camera rotations or displacements. However, it would be to include the height estimation algorithm in a localization system in order to locate the mobile and estimate the phase lag between the reference map and the current image. That way, we would be able to correct the phase lag between scenes and to use the height estimation algorithms proposed in this work.

The future work we should include the height estimation algorithm in a localization system in order to locate the mobile and estimate the phase lag between the map and the current image. It also should extend this research to include topological distance estimation taking into account 6D movements and topological mapping.

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

1. Amorós, F., Payá, L., Reinoso, Ó., Mayol-Cuevas, W., Calway, A.: Topological map building and path estimation using global-appearance image descriptors. In: ICINCO (2) (2013)
2. Bonev, B., Cazorla, M., Escolano, F.: Robot navigation behaviors based on omnidirectional vision and information theory. Journal of Physical Agents 1(1), 27–36 (2007)
3. Chang, C.K., Siagian, C., Itti, L.: Mobile robot vision navigation amp; localization using gist and saliency. In: 2010 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 4147–4154 (2010)
4. Driscoll, J., Healy, D.: Computing fourier transforms and convolutions on the 2-sphere. Advances in Applied Mathematics 15(2), 202–250 (1994), http://www.sciencedirect.com/science/article/pii/S0196885884710086
5. Fernández, L., Payá, L., Valiente, D., Gil, A., Reinoso, Ó.: Monte carlo localization using the global appearance of omnidirectional images - algorithm optimization to large indoor environments. In: ICINCO (2), 439–442 (2012)
6. Han, K., Aeschliman, C., Park, J., Kak, A., Kwon, H., Pack, D.: Uav vision: Feature based accurate ground target localization through propagated initializations and interframe homographies. In: 2012 IEEE International Conference on Robotics and Automation (ICRA), pp. 944–950 (2012)
7. Huhle, B., Schairer, T., Schilling, A., Strasser, W.: Learning to localize with gaussian process regression on omnidirectional image data. In: 2010 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 5208–5213 (2010)
8. Makadia, A., Sorgi, L., Daniilidis, K.: Rotation estimation from spherical images. In: Proceedings of the 17th International Conference on Pattern Recognition, ICPR 2004, vol. 3, pp. 590–593 (2004)
9. Maohai, L., Han, W., Lining, S., Zesu, C.: Robust omnidirectional mobile robot topological navigation system using omnidirectional vision. Engineering Applications of Artificial Intelligence (2013), http://www.sciencedirect.com/science/article/pii/S0952197613000973

10. McEwen, J., Wiaux, Y.: A novel sampling theorem on the sphere. IEEE Transactions on Signal Processing 59(12), 5876–5887 (2011)
11. Menegatti, E., Maeda, T., Ishiguro, H.: Image-based memory for robot navigation using properties of omnidirectional images. Robotics and Autonomous Systems 47(4), 251–267 (2004)
12. Mondragón, I.F., Olivares-Méndez, M., Campoy, P., Martínez, C., Mejias, L.: Unmanned aerial vehicles uavs attitude, height, motion estimation and control using visual systems. Autonomous Robots 29(1), 17–34 (2010)
13. Roebert, S., Schmits, T., Visser, A.: Creating a bird-eye view map using an omnidirectional camera, pp. 233–240 (2008)
14. Scaramuzza, D., Martinelli, A., Siegwart, R.: A flexible technique for accurate omnidirectional camera calibration and structure from motion. In: IEEE International Conference on Computer Vision Systems, ICVS 2006, p. 45 (January 2006)
15. Schairer, T., Huhle, B., Strasser, W.: Increased accuracy orientation estimation from omnidirectional images using the spherical fourier transform. In: 3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video, pp. 1–4 (2009)
16. Schairer, T., Huhle, B., Vorst, P., Schilling, A., Strasser, W.: Visual mapping with uncertainty for correspondence-free localization using gaussian process regression. In: 2011 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 4229–4235 (2011)
17. Valiente, D., Gil, A., Fernández, L., Reinoso, Ó.: View-based slam using omnidirectional images. In: ICINCO (2), pp. 48–57 (2012)
18. Wang, C., Wang, T., Liang, J., Chen, Y., Zhang, Y., Wang, C.: Monocular visual slam for small uavs in gps-denied environments. In: 2012 IEEE International Conference on Robotics and Biomimetics (ROBIO), pp. 896–901 (2012)
19. Winters, N., Gaspar, J., Lacey, G., Santos-Victor, J.: Omni-directional vision for robot navigation. In: Proceedings of the IEEE Workshop on Omnidirectional Vision, pp. 21–28 (2000)