Toward topological localization with spherical Fourier transform and uncalibrated camera

[1,3]Francesco Rossi, [2]Ananth Ranganathan,
[2]Frank Dellaert, and [1]Emanuele Menegatti

[1] Intelligent and autonomous system lab,
Padova University, Italy
emg@dei.unipd.it
rossifrances@gmail.com
[2] College of Computing,
Georgia Institute of Technology
{ananth,dellaert}@cc.gatech.edu
[3] Ital TBS s.p.a.,
Trieste, Italy

Abstract. The paper presents a new metric for computing the image similarity using the Spherical Fourier Transform (SFT) of omnidirectional images projected on a sphere. This metric is designed for image-based localization of mobile robot. Given a set of omnidirectional images, generated from an uncalibrated catadioptric sensor composed of an hyperbolic mirror and a prospective camera, we map them on the sphere to calculate the Spherical Fourier Transform. The transform coefficients are used to calculate the similarity between two images, and the rotation around optical axis of the camera is calculated using the $SO(3)$ transform (SOFT).

Key words: spherical Fourier transform, catadioptric camera

1 Introduction

Localization is a fundamental problem in robotics. Mobile robots need to determine their current robot position to solve most of their tasks. In this paper we investigate a vision-based solution to the localization problem.

In the localization process, a robot can use either a metric or a topological map of the environment. The metric map represents geometrically the environment, the topological one is a subdivision in zones of an environment connected among them, differentiated on the basis of physical characteristics (appareance in the space) and/or for the presence of special features (landmarks). The partition obtained is transposed on a graph whose nodes are zones and edges are geographical connections among them.

Among methods using metrics maps there are [1] and [2] where a robot, moving in an environment, compares images acquired in two consecutive instants and calculates the rototranslation, integrating the result on the map. In [5] a

Workshop Proceedings of SIMPAR 2008
Intl. Conf. on SIMULATION, MODELING and PROGRAMMING for AUTONOMOUS ROBOTS
Venice(Italy) 2008 November,3-4
ISBN 978-88-95872-01-8
pp. 319-330

perspective camera is used to grab images from the environment. For every image a list of descriptor is created (using SIFT, see [6]). The list will be used to decide if two images are equal. Other approaches that use SIFT can be found in [7] and [8].

Among methods using topological maps we can cite: the work of Dudek in [4] uses Principal Component Analysis to classify images comparing images in a space of dimension lower than the number of pixels that form. This is done using covariance matrix Q composed from a training set images. Input images are projected on the basis formed by images in the training set. All the images are transformed in polar coordinates then, to obtain rotational invariance, a bidimensional Fourier Transform is applied. Another approach of topological localization is presented in [9]. In this work, image comparison is based on color histogram. Omnidirectional camera are used, so a rotation does not change the histogram. This approach allows to reduce the amount of information to store since color can be discretized, but, although fast and reliable, it looses geometrical information.
Certainly two of the main works on topological mapping are [27] and [26], both use bayesian inference for building up the map, but the first method uses global feature recognition, while the second one resolves the same problem with local feature recognition.

1.1 Spherical Fourier Transform on omnidirectional images

In a previous work, we used the magnitude coefficients of the circular Fourier transform of panoramic images to compare two images [10], following the idea of measuring similarity up to a rotation around a single axis by using the Fourier transform on a cylinder elaborated in [25]. An omnidirectional image f is mapped on a panoramic cylinder to calculate the Fourier transform of each row. If the initial image is rotated, obtaining f', rows of projected image are function of initial rows

$$f(x) = f'(x + a),$$

Now the transform of f' is related to that of f as

$$F'(\theta) = e^{i2\pi a} F(\theta)$$

thus noting that the magnitude does not change. That work showed that not all the coefficients of the Fourier tranform are needed for a correct localization, allowing to speed-up the compare and recognition tasks.

Recently, in [11], [12], [13], [14], [15], [16], [17], [18] Spherical Fourier Transform is applied on two omnidirectional images with single view point (SVP) mapped on the sphere, which differs for a rotation. Using the $SO3$ (the group of rotation matrix) Fourier transform (SOFT) for calculating correlation in $SO(3)$ to obtain a first estimation of the rotation existing between two images. The importance of Makadia's research consists in having found a global evaluation method to compare images, in contrast with methods that use local feature to

Workshop Proceedings of SIMPAR 2008
Intl. Conf. on SIMULATION, MODELING and PROGRAMMING for AUTONOMOUS ROBOTS
Venice(Italy) 2008 November,3-4
ISBN 978-88-95872-01-8
pp. 319-330

estimate rotation (epipolar geometry).
For a tutorial on SOFT and SFT refer to [21], [20], [22], [23].

2 Proposed approach

In this paper it is proven that rotational invariance of Spherical Fourier transform (SFT) can be used as the basis for a new similarity measure between images differing for a rototranslation.

In particular we will show that it is possible:

- to map an omnidirectional image on the sphere without knowing the parameters of the sensor generating it, so without applying the unified theory of catadioptric projection of Geyer ([19]), and still having a representation usefull for localization;
- to apply SFT to image projected on the sphere and to use the energy of block coefficients could be a similarity measure for topological localization;
- to apply the SOFT to calculate the rotation existing between two images differing by a rototranslation.

Exploiting precedent facts will result in the proposal of a framework for robot navigation in unknown environment based exclusively on spherical harmonics theory: ego-motion information can be extracted using [17], thus a metric map of the environment can be built, then topological information are extracted from global comparing of images based on spherical invariant. The latest assertion has been tested with uncalibrated camera, while all works done till now on ego-motion is based on calibrated camera, so it remains to experiment which results can be achieved in the non calibrated case.
Another particularity of the framework is that is based on global feature comparing: in [28] and [29] navigation of insects are studied and conclusions stated that basically they rely on global matched filters for doing that.

2.1 Previous work on localization using spherical harmonics

Recently in [24] spherical harmonics signature of a panoramic image has been choosen as similarity measures. Having calculated Q as the squared difference of two image signals in the spherical harmonics domain up to order l

$$Q = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} \sum_{l'=0}^{\infty} \sum_{m'=-l'}^{l'} (f_k^l - h_k^l)(f_{k'}^{l'} - h_{k'}^{l'})\lambda_{l'l}\lambda_{m'm} \tag{1}$$

for the localization task two operations are performed: a fast rotation invariant similarity measure to drop all unlikely views, then, for all reference views which

Workshop Proceedings of SIMPAR 2008
Intl. Conf. on SIMULATION, MODELING and PROGRAMMING for AUTONOMOUS ROBOTS
Venice(Italy) 2008 November,3-4
ISBN 978-88-95872-01-8
pp. 319-330

survived prefiltering, an estimation of the best matching rotation with respect to the current view descriptor is computed and de-rotate it accordingly. The dissimilarity is computed according to 1.
The approach proposed in [24] required up to $O(l^2)$ coefficients for a similarity estimation and a two step algorithm, while the similarity measure of this paper needs to store only l coefficients (energy block) for each reference image.

In the next section the procedure for obtaining the new similarity measure is described.

3 Similarity measure

Suppose we are given two images, where f is the image taken in a unknown location and h is a reference image. The Eq. 4 defines a similarity measure to compare the images f and h, where $E_l(f)$ is the sum of coefficients energy of degree l; f_k^l is the SFT coefficient of degree l and order k of f.

$$E_l(f) = \sum_{k=0}^{k \leq l} f_k^l \overline{f_k^l} \tag{2}$$

$$Dissim(f,h) = \sum_{k=0}^{k<B} \frac{|E_k(f) - E_k(h)|}{E_k(f)}. \tag{3}$$

$$Sim(f,h) = 1000 - 1000 \frac{Dissim(f,h) - minDissim}{maxDissim - minDissim} \tag{4}$$

$$minDissim = min \{Dissim(h,p) : h \in S_{input},\ p \in S_{ref}\}$$
$$maxDissim = max \{Dissim(h,p) : h \in S_{input},\ p \in S_{ref}\}$$

In Eq. 3 and Eq. 4, B is the bandwidth of the SFT. S_{ref} is the images training set and S_{input} is the images input set.

Several parameters can influence the of the image similarity. Here, we present result for the following list:

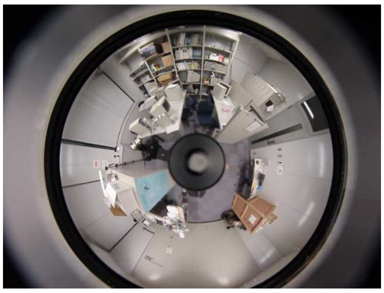
- maximun latitude of mapping of omnidirectional image on the sphere, so how much the sphere is covered by the image;
- the available visual information (image pixels) and the bandwidth of SFT;
- the number of energy block used for similarity measure.

3.1 Latitude of projection

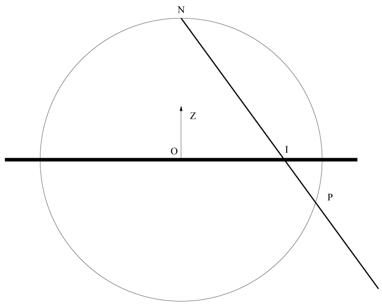
The omnidirectional images are projected on the sphere using the stereographic projection depicted in Fig. 1(b). The image plane is on the equatorial plane, a point I on this plane is projected from the north pole N to the point P on the sphere. The result is plotted in Fig. 1(c). If the image is scaled the sphere can be

Workshop Proceedings of SIMPAR 2008
Intl. Conf. on SIMULATION, MODELING and PROGRAMMING for AUTONOMOUS ROBOTS
Venice(Italy) 2008 November,3-4
ISBN 978-88-95872-01-8
pp. 319-330

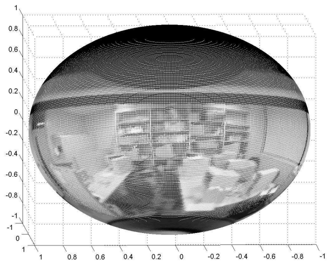
covered more or less. Our omnidirectional images are taken from an uncalibrated catadioptric sensor (see Fig. 1(a)), whose resolution is 640×480 pixels, but the image which reflects the mirror is composed by 400×400 pixels. The result of our tests is that scaling the image does not affect the localization performances.



(a) Omnidirectional image (Wakayama University, Japan)



(b) Stereographic projection



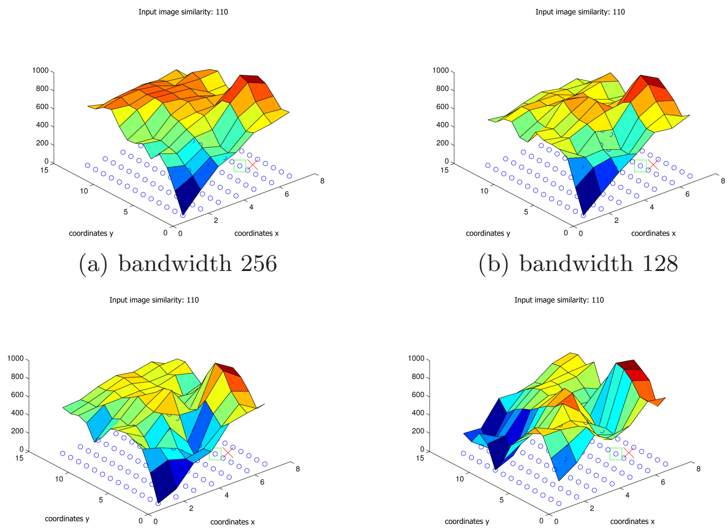
(c) An image projected on the sphere

Fig. 1. The stereographic project of an omnidirectional image on the sphere.

3.2 Resolution and bandwidth

We investigated also the the minimal resolution needed to localize correctly an image. Images were undersampled to resolution of $2B \times 2B$, with $B \in$ 8, 16, 32, 64, 128, 256, then the SFT at bandwidth B is calculated (for having a ratio equal to one between number of pixels and number of point on the sphere). Tests show that with bandwidth less than 32 is not possible to localize. In Fig. 2(c), Fig. 2(b), and Fig. 2(a) the values of similarity between an image of input and a set of reference images plotted over the map (which is a grid of point). The cross in the image identifies the point where the input image is taken, while the square identifies the reference image most similar to input image.

Workshop Proceedings of SIMPAR 2008
Intl. Conf. on SIMULATION, MODELING and PROGRAMMING for AUTONOMOUS ROBOTS
Venice(Italy) 2008 November,3-4
ISBN 978-88-95872-01-8
pp. 319-330



(a) bandwidth 256 (b) bandwidth 128

(c) bandwidth 64 (d) bandwidth 32

Fig. 2. localization varying bandwidth

3.3 Saturation of energy block

Our test shows that not all the energy blocks are needed for similarity measurement, so it is possible to truncate the sum defined in Eq. 4 to a certain index k lower than the maximum bandwidth B. In Fig. 3 are reported the numbers of correct localizations varying the number of energy block used. After the twentieth block there is no improvement in localization.

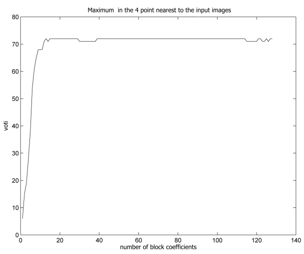
3.4 Calculating rotation

The test images differ only by a translation, so they were manually rotated around the actual image center (i.e. the optical axis of camera) to test the calculation of a rotation of an input image with respect to the reference image. Since the rotation is around optical axis, using the Eulero representation for rotation, the rotation calculated with SOFT has the component of β (rotation around axis y, lying on the image plane) near to zero. Thus we can assume that the β is null and sum the components α and γ (rotations around z axis, the optical one). The result of this assumption is reported in Fig. 4(a), 4(b), where for different bandwidth the rotation estimation (as sum of α and γ) among all the pairs of input image and the most similar reference image (input images are rotated of 30 degrees) is reported.

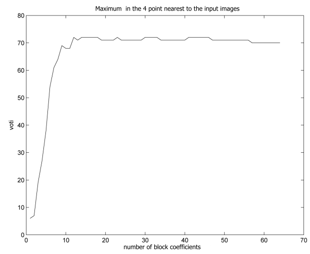
3.5 Consideration

Omnidirectional images keep all their potential on localization, presenting with SFT another invariance property usefull for similarity comparison. Images align-

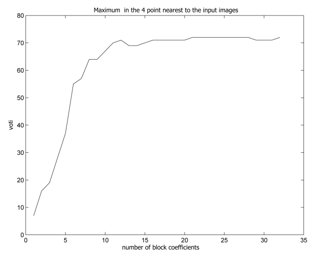
Workshop Proceedings of SIMPAR 2008
Intl. Conf. on SIMULATION, MODELING and PROGRAMMING for AUTONOMOUS ROBOTS
Venice(Italy) 2008 November,3-4
ISBN 978-88-95872-01-8
pp. 319-330



(a) bandwidth 128

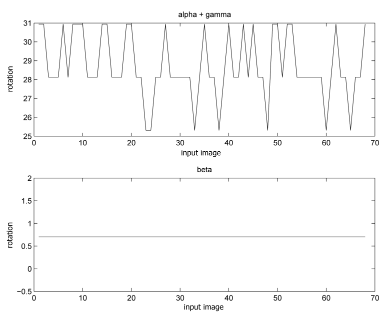


(b) bandwidth 64

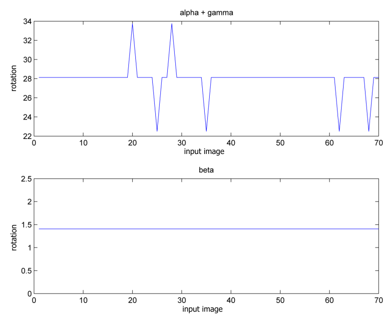


(c) bandwidth 32

Fig. 3. Correct localization varying the number of block coefficients.



(a) bandwidth 64



(b) bandwidth 32

Fig. 4. Rotation estimation

Workshop Proceedings of SIMPAR 2008
Intl. Conf. on SIMULATION, MODELING and PROGRAMMING for AUTONOMOUS ROBOTS
Venice(Italy) 2008 November,3-4
ISBN 978-88-95872-01-8
pp. 319-330

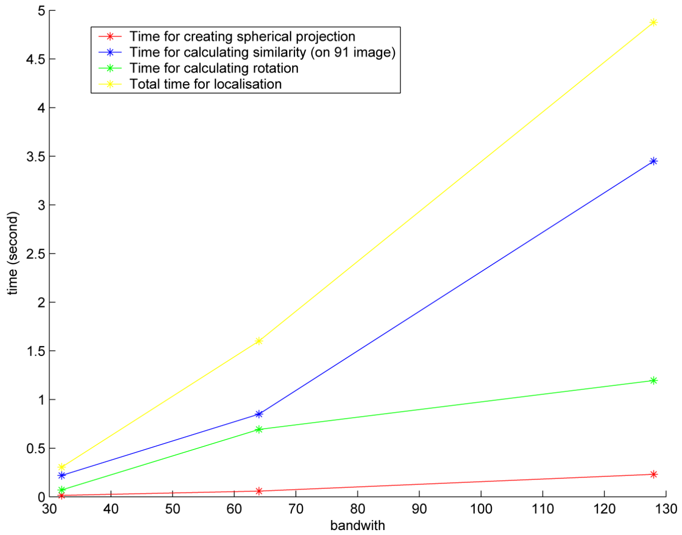


Fig. 5. Time for executing algorithm

ment is done using another interesting property of $SOFT$ (correlation via inverse $SO(3)$ fourier transform): this is a method that refers only to global appareance of images and does not use any image pre-elaboration (like $SIFT$ methods do).

Once again, analyzing images in the spectral domain, richness of information allows to retain frequencies in function of the precision of the localization. That leads to present hierarchical localization, using more spectral information for landmarks of major interest.

4 An algorithm for topological localization

Results explained before are exploited in the following algorithm for topological localization:

1. given an omnidirectional image, obtain its representation on the sphere by a stereographic projection;
2. calculate the coefficients of SFT of spherical image;
3. given the SFT coefficients of a reference image, calculate the similarity with the input image as defined in Eq. 4. The reference image with maximum similarity is candidated to be the most similar to input image and the nearest in the space;
4. calculate the rotation between input image and most similar reference image with SOFT.

Some tests which use this algorithm are shown: a robot makes a path on a grid (the blue dotted line in Fig. 6(c), Fig. 6(b), and Fig. 6(a)), the red solid line

Workshop Proceedings of SIMPAR 2008
Intl. Conf. on SIMULATION, MODELING and PROGRAMMING for AUTONOMOUS ROBOTS
Venice(Italy) 2008 November,3-4
ISBN 978-88-95872-01-8
pp. 319-330

path is the estimation of this path on the topological grid. The red segments are the estimated orientation (the real orientation, plotted in blue, is partially overlapped by red orientation segment). In the case with bandwidth 128 the orientation is estimated with bandwidth 64 (because of lack of computer RAM). Even with an undersampling of the transform the algorithm works.

Execution times are reported in Fig. 5.

5 Conclusion

SFT (and SOFT) can be used for topological localization. SFT has been applied for images comparison, noting that SFT energy block of higher degree can be omitted from the similarity formula. Then SOFT gives an estimation of the rotation. Results of this phase are then integrated to establish the path followed. In summary, the original contribution of this paper is that four properties of SFT of omnidirectional images are highlighted:

- the magnitude of the Fourier components are related to the position of the robot;
- the coefficients of the Fourier components are related to the heading of the robot;
- by using the SFT signatures a high data compression can be achieved;
- a hierarchical localization is embedded in this approach;
- the similarity function we defined is effective in the proposed method to self-organise the visual memory.

5.1 Future work

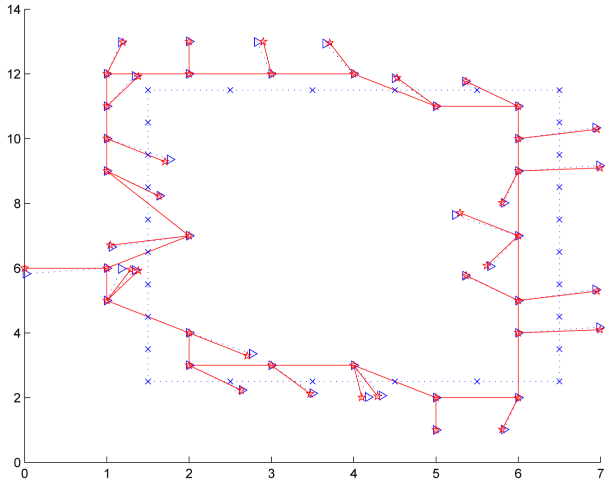
Extension to work are:

1. implement an ibrid localization algorithm (metric and topological) based on SFT and SOFT;
2. test it on a robot, both in outdoor and indoor environments;
3. try to parallelize the computation of the SFT and SOFT;
4. test topological localization on images which differ by an arbitrary rotation.

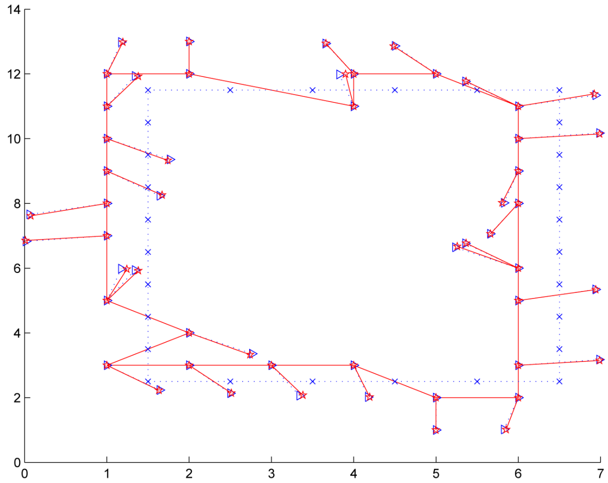
References

1. M. Montemerlo and S. Thrun, Tracts in Advanced Robotics, The FastSLAM Algortihm for Simultaneous localization and Mapping, Springer, 2007
2. A. J. Davidson, Real-time simultaneous localization and mapping with a single camera, International conference on computer vision, pp. 14031410, 2003
3. A. J. Davidson, Y. G. Cid e N. Kita. Real-time 3d slam with wide-angle vision, 5th IFAC/EURON Symposium on Intelligent Autonomous system, 2004
4. G. Dudek and D. Jugessur, Robust place recognition using local appearance based methods, International Conference on Robotics and Automation,San Francisco, april, 2000

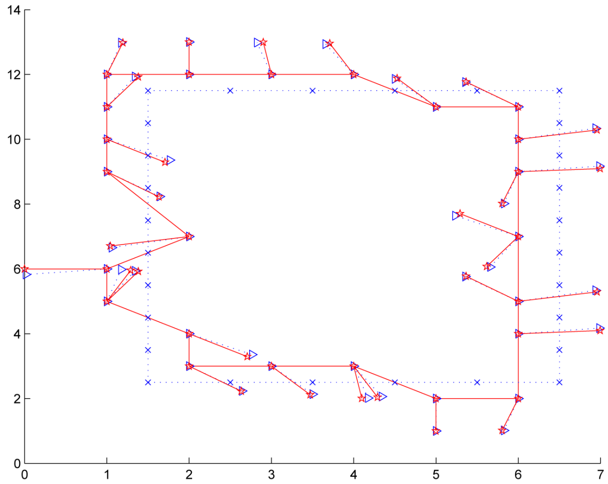
Workshop Proceedings of SIMPAR 2008
Intl. Conf. on SIMULATION, MODELING and PROGRAMMING for AUTONOMOUS ROBOTS
Venice(Italy) 2008 November,3-4
ISBN 978-88-95872-01-8
pp. 319-330



(a) bandwidth 128



(b) bandwidth 64



(c) bandwidth 32

Fig. 6. Reconstruction of the path on a topological map with SFT

Workshop Proceedings of SIMPAR 2008
Intl. Conf. on SIMULATION, MODELING and PROGRAMMING for AUTONOMOUS ROBOTS
Venice(Italy) 2008 November,3-4
ISBN 978-88-95872-01-8
pp. 319-330

5. P. Newman, D. Cole e K. Ho, Outdoor slam using visual appearance and laser ranging, International Conference on Robotics and Automation, 2006
6. D. G.Lowe, Object recognition from local scale-invariant features, Proc. of the 7th Int. Conf. on Computer Vision, Kerkyra, pp. 11501157, 1999
7. C. Silpa-Anan and R. Hartley, Localization using an image-map, International Conference on Robotics and Automation, 2005
8. J. Sivic and A. Zisserman. Video google: A text retrieval approach to object matching in videos, Proc. of the 9th Int. Conf. on Computer Vision, 2003
9. I. Ulrich and I. Nourbakhsh, Appearance-based place recognition for topological localization, International Conference on Robotics and Automation,San Francisco, april, 2000, Best vision paper award.
10. E. Menegatti, M. Zoccarato, E. Pagello and H. Ishiguro, Robotics and Autonomous Systems, cap. Image-based monte-carlo localization with omnidirectional images, Elsevier, 2003
11. A. Makadia and K. Daniilidis, Direct 3d rotation estimation from spherical images via a generalized shift theorem, IEEE Conf. Computer Vision and Pattern Recognition, pp. 217224, 2003
12. A. Makadia, Imaging Beyond the Pinhole Camera, cap. Correspondenceless Visual Navigation under Constrained Motion, Kluwer Press, 2006
13. A. Makadia and K. Daniilidis, Correspondenceless ego-motion estimation using an imu, IEEE International Conference on Robotics and Automation, 2005
14. A. Makadia and K. Daniilidis, Rotation recovery from spherical images without correspondences, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 28(7), 2006
15. A. Makadia and K. Daniilidis, Correspondenceless structure from motion, International Journal of Computer Vision, 2007
16. A. Makadia, C. Geyer, S. Sastry and K. Daniilidis, Radon-based structure from motion without correspondences, IEEE Conference on Computer Vision and Pattern Recognition, 2005
17. A. Makadia, D. Gupta and K. Daniilidis, Planar ego-motion without correspondences, IEEE Workshop on Motion and Video Computing, 2005
18. A. Makadia, I. Sorgi and K. Daniilidis. Rotation estimation from spherical images, Proc. Int. Conf. on Pattern Recognition, 2004
19. C. M. Geyer, Catadioptric projective geometry, theory and applications. phd thesis, University of Pennsylvania, 2003
20. Martin J. Mohlenkamp, A fast transform for spherical harmonics, J. Fourier Anal. Appl.,1999, volume 5,pp. 159-184
21. Peter J. Kostelec and Daniel N. Rockmore, SOFT: $SO(3)$ Fourier Transform, Department of Mathematics, Dartmouth College, Hanover, NH 03755,2003
22. Peter J. Kostelec and Daniel N. Rockmore, S2kit: A Lite Version of SpharmonicKit, Department of Mathematics, Dartmouth College, Hanover, NH 03755,2004
23. J. R. Driscoll and D. Healy. Computing Fourier transforms and convolutions on the 2-sphere, 34th IEEE FOCS, pp. 344349, 1989
24. H. Friedrich, D. Dederscheck, K. Krajsek, and R. Mester, View-based robot localization using illumination-invariant spherical harmonics descriptors, In A. Ranchordas and H. Araujo, editors, Proceedings of the international joint conference on computer vision and computer graphics theory and applications, volume 2, pages 543-550, INSTICC and University of Madeira, 2008
25. T. Pajdla, V. Hlavac, Zero phase representation of panoramic images for image based localization, CAIP 1999, Springer LNCS 1689, pp. 550-557, September 1999

Workshop Proceedings of SIMPAR 2008
Intl. Conf. on SIMULATION, MODELING and PROGRAMMING for AUTONOMOUS ROBOTS
Venice(Italy) 2008 November,3-4
ISBN 978-88-95872-01-8
pp. 319-330

26. T. Goedem, M. Nuttim, T. Tuytelaars and L. Van Gool, Omnidirectional vision based topological navigation, IJCV, 74(3):219-236, September 2007
27. A. Ranganathan, E. Menegatti, F. Dellaert, Bayesian inference in the space of topological maps, IEEE Transactions on Robotics, 2005
28. M.O. Franz and H.G. Krapp, Wide-field, motion- sensitive neurons and matched filters for optic flow fields. Biological Cybernetics, 83:185197, 2000
29. M.V. Srinivasan. Ants match as they march. Nature, 392:660661, 2001

Workshop Proceedings of SIMPAR 2008
Intl. Conf. on SIMULATION, MODELING and PROGRAMMING for AUTONOMOUS ROBOTS
Venice(Italy) 2008 November,3-4
ISBN 978-88-95872-01-8
pp. 319-330