View-Based Robot Localization Using Spherical Harmonics: Concept and First Experimental Results

Holger Friedrich, David Dederscheck, Kai Krajsek, and Rudolf Mester

Visual Sensorics and Information Processing Lab
J.W. Goethe University, Frankfurt, Germany
{holgerf,davidded,krajsek,mester}@vsi.cs.uni-frankfurt.de
http://www.vsi.cs.uni-frankfurt.de

Abstract. Robot self-localization using a hemispherical camera system can be done without correspondences. We present a view-based approach using view descriptors, which enables us to efficiently compare the image signal taken at different locations. A compact representation of the image signal can be computed using Spherical Harmonics as orthonormal basis functions defined on the sphere. This is particularly useful because rotations between two representations can be found easily. Compact view descriptors stored in a database enable us to compute a likelihood for the current view corresponding to a particular position and orientation in the map.

1 Introduction

Omnidirectional vision has become increasingly popular for the purpose of robot localization during the last years. Many approaches rely on compact image descriptors [21,1,5], [11,8] (using principal component analysis) [19,18] (using Fourier descriptors), [13] (using Haar integrals) to store and compare views efficiently. There are also approaches combining both compact descriptors and local features, e.g. [14].

We present a view-based method for robot localization in a known environment. A mobile robot equipped with an omnidirectional camera system provides a spherical image signal $s(\theta, \phi)$, i. e. an image signal defined on a sphere. In our experiments performed so far, the omnidirectional images were obtained from a simulated ultra-wide angle lens camera mounted face up on the robot, yielding rectangular images which can be mapped on the semi-sphere in a straightforward manner. These images were converted into *view descriptors*, i. e. *low dimensional vectors* (Fig. 1). The robot localization task is performed by comparing the current view descriptor to those stored in a database of views. Given a suitable distance metric, this yields a likelihood of the robot location. The image descriptors used here are not rotation invariant; due to their particular structure it is possible to estimate the orientation (rotation compared to a reference pose) of the current view. Our view representation is obtained by performing a linear

F.A. Hamprecht, C. Schnörr, and B. Jähne (Eds.): DAGM 2007, LNCS 4713, pp. 21–31, 2007.

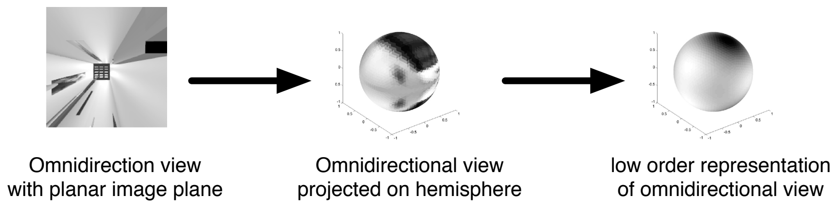


Fig. 1. Computing an omnidirectional image signal from a planar wide angle image. The resulting hemispherical image signal is reflected at the equator to obtain an spherical image signal. The right image is a visualization of a low order Spherical Harmonic descriptor that approximates the omnidirectional image signal.

spectral transform, that is by expanding the spherical image signal $s(\theta, \phi)$ in orthonormal basis functions $b_i(\theta, \phi)$ according to

$$s(\theta, \phi) = \sum_i a_i \cdot b_i(\theta, \phi). \tag{1}$$

This is possible for any square integrable signal $s(\theta, \phi)$ defined on the sphere. Let $\overline{b}$ denote the complex conjugate of b. We obtain the coefficients a_i by

$$a_i = \int_0^{2\pi} \int_0^{\pi} s(\theta, \phi) \cdot \overline{b_i(\theta, \phi)} \cdot \sin\theta \; \mathrm{d}\theta \, \mathrm{d}\phi. \tag{2}$$

Our approach benefits from using *Spherical Harmonics* (Fig. 3) as basis functions b_i since they show the same nice properties concerning rotations which the Fourier basis system has with respect to translations. Rotations are mapped into a kind of *generalized phase changes*.

All views in the database are labeled with their corresponding location in the map (see Fig. 2); thus finding a match in the database – in principle – solves the localization task. We briefly discuss some obvious problems such as variations in illumination and impact of occlusions in Sec. 6. For each given view at an initially unknown robot position and orientation, a figure of (dis-)similarity to any other view in the database can be generated directly from the compact

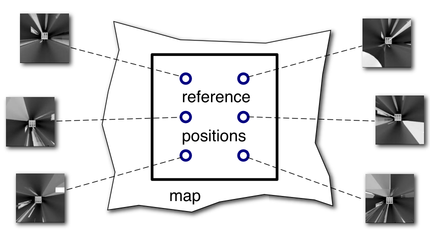


Fig. 2. A known environment is represented by a map containing view descriptors. These are obtained from images taken at reference positions.

vector representation of these views; this allows for more sophisticated temporal self-localization strategies [22], e. g. using particle filters.

In order to compare a given descriptor to those stored in a database, an operation has to be performed on the descriptor that corresponds to (virtually) *de*-rotating the corresponding view. Hence, for finding a measure of similarity between two views, de-rotation has to be performed as an integral part of the comparison. The estimation of 3D rotation between spherical signals has been investigated in different contexts in the last years [16,17,15,10]. For our application, a fast solution is of particular importance since we have to compare many pairs of descriptors. It is therefore useful to exploit group theoretical properties of Spherical Harmonics in order to eliminate those descriptors which cannot correspond to the same image signal (=pruning a search tree). This can be done by comparing the 'amplitude spectrum' while disregarding the 'phase'.

The following sections deal with some essential mathematical characteristics of Spherical Harmonics, de-rotation and similarity measurement. The paper is concluded with a description of our experimental setup and the experimental results that we have obtained so far.

2 Spherical Harmonics

Here we emphasize some facts about Spherical Harmonics (Fig. 3) which are of particular interest for the matching and self localization task. For further group theoretical facts see [16] and [6]. Let

$$N_{\ell m} = \sqrt{\tfrac{2\ell+1}{2}\tfrac{(\ell-|m|)!}{(\ell+|m|)!}},\ \ell \in \mathbb{N}_0,\, m \in \mathbb{Z} \tag{3}$$

and $P_{\ell m}(x)$ the Associated Legendre Polynomials [23].

The Spherical Harmonics $Y_{\ell m}(\theta,\phi)$ are defined as

$$Y_{\ell m}(\theta,\phi) = \tfrac{1}{\sqrt{2\pi}} \cdot N_{\ell m} \cdot P_{\ell m}(\cos\theta) \cdot e^{im\phi} \tag{4}$$

with $e^{im\phi}$ being a complex-valued phase term. ℓ ($\ell > 0$) is called *order* and m ($m = -\ell..+\ell$) is called *quantum number* for each ℓ. Note that slightly different notations of this definition exist. Some authors disregard the so called Condon-Shortley phase $(-1)^m$ in the definition of the associated Legendre polynomials. We do not omit this factor and conform to the notation of [23,12].

Spherical Harmonics have several properties that we would like to exploit in the following sections: Each set of Spherical Harmonics of order ℓ forms an orthonormal basis of dimension $2\ell+1$; Spherical Harmonics of orders $0\ldots\ell$ form an orthonormal basis of dimension $(\ell+1)^2$, i. e.

$$\int_0^{2\pi}\!\!\int_0^{\pi} \overline{Y_{\ell m}(\theta,\phi)} \cdot Y_{\ell' m'}(\theta,\phi) \cdot \sin\theta \,\mathrm{d}\theta\,\mathrm{d}\phi \;=\; \delta_{\ell\ell'} \cdot \delta_{mm'} \tag{5}$$

where $\delta_{\ell m}$ is the Kronecker delta function. The complex conjugate of a Spherical Harmonic function is simple to obtain:

$$Y_{\ell,-m}(\theta,\phi) = (-1)^m \cdot \overline{Y_{\ell m}(\theta,\phi)}. \tag{6}$$

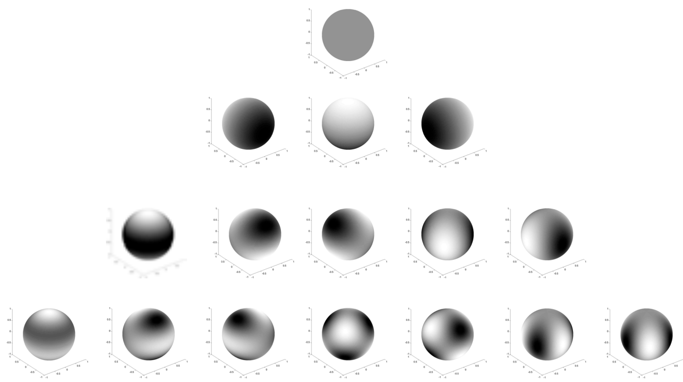


Fig. 3. A Spherical Harmonic function is a periodic function on the unit sphere which has ℓ maxima. The rows show Spherical Harmonics of *orders* $\ell = 0, 1, 2, 3$; columns show $m = 2\ell + 1$ functions for each order ℓ.

To approximate a signal $s(\theta, \phi)$, i. e.

$$s(\theta, \phi) = \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} a_{\ell m} \cdot Y_{\ell m}(\theta, \phi) \tag{7}$$

we need to compute the coefficients $a_{\ell m}$ using Eq. 2

$$a_{\ell m} = \int_0^{2\pi} \int_0^{\pi} s(\theta, \phi) \cdot \overline{Y_{\ell m}(\theta, \phi)} \cdot \sin\theta \, \mathrm{d}\theta \, \mathrm{d}\phi. \tag{8}$$

In practice, this is done using Spherical Harmonics of order $\ell = 0$ up to a small number, e. g. $\ell = 4$. It may also be useful to use real-valued Spherical Harmonics as defined in [12]. Eq. 6 implies that the number of coefficients stays the same for real-valued or complex-valued Spherical Harmonics. Hints on implementation can be found in [3,7].

3 Rotation Estimation

For general robot self-localization, we have to determine the 3D rotation between two Spherical Harmonic representations of image signals. This problem has already been investigated [2], and more recently in [15,16,10]. As an initial test case, we have chosen a mobile robot moving on a plane. For this particular application we only need to deal with 1D rotation estimation.

3.1 Rotations

The 3D case. Recall that Spherical Harmonics of order ℓ form a basis. Any 3D rotation can be expressed as a *linear transformation* (i. e. multiplication with an unitary matrix U_ℓ) and *does not mix coefficients of different order* ℓ. Hence rotations retain the distribution of spectral energy among different orders [16]. This is a unique characteristic of Spherical Harmonics which makes them so particularly useful, amongst others for the purpose of robot ego-localization pursued here. Applying a 3D rotation to a spherical function represented by coefficients a_{jk} yields new coefficients b_{jk} according to

$$\begin{pmatrix} b_{00}\\ b_{10}\\ b_{11}\\ b_{1,-1}\\ \vdots\\ \vdots\\ b_{2,-2}\end{pmatrix} = \underbrace{\begin{pmatrix} \boxed{\mathbf{U}_{\ell=0}} & & \mathbf{0}\\ & \boxed{\mathbf{U}_{\ell=1}} & \\ \mathbf{0} & & \boxed{\mathbf{U}_{\ell=2}}\end{pmatrix}}_{\mathbf{\Lambda}_R} \begin{pmatrix} a_{00}\\ a_{10}\\ a_{11}\\ a_{1,-1}\\ \vdots\\ \vdots\\ a_{2,-2}\end{pmatrix} \tag{9}$$

The 1D case: rotation about Z-axis. Since the robot moves on a plane in our current configuration, the problem of de-rotation is simplified somewhat. Recalling the definition of the complex-valued Spherical Harmonics, the implications of a rotation of φ about the Z-axis are as follows:

$$Y_{\ell m}(\theta, \phi + \varphi) = \tfrac{1}{\sqrt{2\pi}} \cdot N_{\ell m} \cdot P_{\ell m}(\cos\theta) \cdot e^{i\,m\,(\phi+\varphi)} \;=\; e^{i\,m\,\varphi} \cdot Y_{\ell m}(\theta, \phi)\,.$$

The rotation matrix becomes much simpler because it changes into a diagonal matrix with elements $e^{-im\varphi}$:

$$\begin{pmatrix} b_{00}\\ b_{10}\\ b_{11}\\ b_{1,-1}\\ \vdots\\ \vdots\\ b_{2,-2}\end{pmatrix} = \begin{pmatrix} \boxed{1} & & & & & & & & \\ & 1 & & & & & \mathbf{0} & & \\ & & e^{-i\varphi} & & & & & & \\ & & & e^{i\varphi} & & & & & \\ & & & & 1 & & & & \\ & & & & & e^{-i\varphi} & & & \\ & \mathbf{0} & & & & & e^{i\varphi} & & \\ & & & & & & & e^{-2i\varphi} & \\ & & & & & & & & e^{2i\varphi} \end{pmatrix} \begin{pmatrix} a_{00}\\ a_{10}\\ a_{11}\\ a_{1,-1}\\ \vdots\\ \vdots\\ a_{2,-2}\end{pmatrix} \tag{10}$$

3.2 De-rotation

Currently our implementation is based on direct non-linear estimation of φ similar to the method described in [17]. In this method, the 3D-rotation $\mathbf{\Lambda}_R$ for view descriptors $\boldsymbol{a}$ and $\boldsymbol{b}$ is determined such that $||\boldsymbol{b} - \mathbf{\Lambda}_R\,\boldsymbol{a}||_2^2$ is minimized. This corresponds to the mean square signal difference between both signal approximations integrated over the sphere, as it will be discussed in more detail later in Sec. 4.1. The constraint of mere 1-axis rotation which has been maintained in our experiments so far, leaving full 3D, 6 DoF pose estimation to future investigations, leads to simplifications: we have to determine the angle φ that minimizes $\sum_\ell \sum_{m=-\ell}^{\ell} (b_{\ell m} - e^{-im\varphi}\,a_{\ell m})^2$. We emphasize that full 3D derotation is possible [17,16] for other robot configurations, that is, the spherical harmonic approach is even more interesting and attractive in that case.

4 Localization

4.1 Similarity Measure

Similarity between two image descriptors, $\boldsymbol{a}$ for signal $g(\theta,\phi)$ and $\boldsymbol{b}$ for signal $h(\theta,\phi)$, can be defined in a natural way. We define the *dissimilarity* Q as the squared difference of the two regarded image signals in the Spherical Harmonic domain up to order ℓ:

$$
\begin{aligned}
Q &= \int_0^{2\pi}\int_0^{\pi} (g(\theta,\phi) - h(\theta,\phi))^2 \cdot \sin\theta \, \mathrm{d}\theta \, \mathrm{d}\phi \\
&\overset{Eq.\,7}{=} \int_0^{2\pi}\int_0^{\pi} \left(\sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} (a_{\ell m} - b_{\ell m}) \cdot Y_{\ell m}(\theta,\phi) \right)^2 \cdot \sin\theta \, \mathrm{d}\theta \, \mathrm{d}\phi \\
&\overset{Eq.\,5}{=} \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} \sum_{\ell'=0}^{\infty} \sum_{m'=-\ell'}^{\ell'} (a_{\ell m} - b_{\ell m}) \cdot (a_{\ell' m'} - b_{\ell' m'}) \cdot \delta_{\ell\ell'} \cdot \delta_{mm'} = ||\boldsymbol{a} - \boldsymbol{b}||_2^2
\end{aligned}
$$

This result is of course not very astonishing, taking into account the fact that the regarded basis signals form an orthonormal basis. The measure Q is of course sensitive to any rotation between the signals. Hence, to find the minimum dissimilarity of two view descriptors we must de-rotate them first.

4.2 A Concept for a Rotation Invariant Similarity Measure

As we mentioned in Sec. 3.1, the norms of the subgroups of coefficients belonging to Spherical Harmonics of the same order are invariant to arbitrary 3D rotations of the signal. Thus L_2 norms, one for each order of Spherical Harmonics, can be considered as a kind of *energy spectrum* of the omnidirectional signal.

This energy spectrum is an efficient means for comparing pairs of spherical signals [9]. With a proper metric which should be derived from statistical models of the signal and the expected noise, spherical signals can be compared to each other even without performing the 'de-rotation'. If the energy spectrum is identical or similar, the particular spherical signals *can* be identical but they need not to be so. However, if their energy spectra are significantly different, both signals cannot be identical.

4.3 Robot Localization Algorithm

Robot localization can be done the following way:

For each reference location

1. use the fast rotation invariant similarity measure to drop unlikely views,
2. try to find the best matching rotation for the current image descriptor and de-rotate the current descriptor,
3. compute the similarity according to Sec. 4.1.

This yields a similarity map, which in all our experiments performed so far, has a distinct extremum at the true location of the robot. It can, however, also contain other extrema, i. e. different poses which have a similar likelihood. Considering the fact that man-made environments have certain regularities, which may result in similar views at several distinct positions, this is not too astonishing, forming a general problem of view-based navigation. At each instant, however, we have prior knowledge about the robots previous course and its previous pose(s), which is presumably always sufficient to disambiguate the current pose estimation process. Such strategies are well-known in robot navigation, and have been, amongst many others, described by Thrun et al. [22] ('Monte Carlo Localization'), or Menegatti et al. [19] (using other image descriptors).

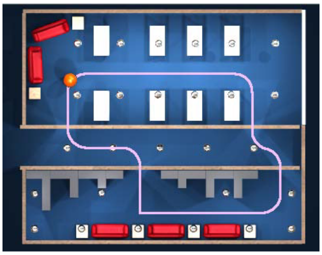
5 Experimental Results

For our experiments, we currently use simulated image data rendered by ray tracing software. Using the 3D modeling software Blender [4], we have created an artificial environment resembling an office area, which provides an experimental area for a simulated robot (see Fig. 4). An upwards facing wide-angle perspective camera with a field of view of approx. 172.5° yields the simulated input images of the robot. The resulting images can be projected onto a hemisphere. This hemispherical signal is extended to a full spherical signal by suitable reflection at the equator. Subsequently, the spherical signal can be approximated by Spherical Harmonics. Of course, a direct expansion of the 2D wide-angle images into Spherical Harmonics is possible without the detour of projecting the perspective signal onto the sphere. We use Spherical Harmonics up to order $\ell = 4$. The reflection across the equator introduces an additional symmetry to the spherical image signal. Hence, additional constraints exist on the coefficients of the Spherical Harmonics.

Prior to performing a localization of the robot, we must create a set of reference frames and calculate its corresponding view descriptors. The view-attributed map needed for performing the robot localization must be computed beforehand; in a real application, the robot and a precision localization device will be driven through the envisaged environment while the views and the corresponding poses are recorded.

For our localization experiment, we have rendered a series of frames with the robot moving along a fixed path (Fig. 4(a)). To obtain realistic sequences of images, each taken at a definite position, a sequence of poses is recorded by a control script while the robot moves along a given path. The resulting list is then used to place the camera for the rendering process.

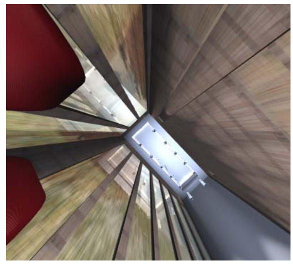
The images in Fig. 5 are maps of the simulation environment showing a measure corresponding to the likelihood of the robot location, calculated at discrete positions along the motion path. Note that these positions are in general *not* aligned with the grid and the heading direction of the robot is not aligned with the direction the grid was built with.



(a) Environment from bird's eye perspective. The path of the robot is marked.

(b) Simulated view of a normal camera facing forwards.

(c) Robot with camera facing upwards.



(d) Wide angle view facing upwards.

Fig. 4. Views of our simulated office environment

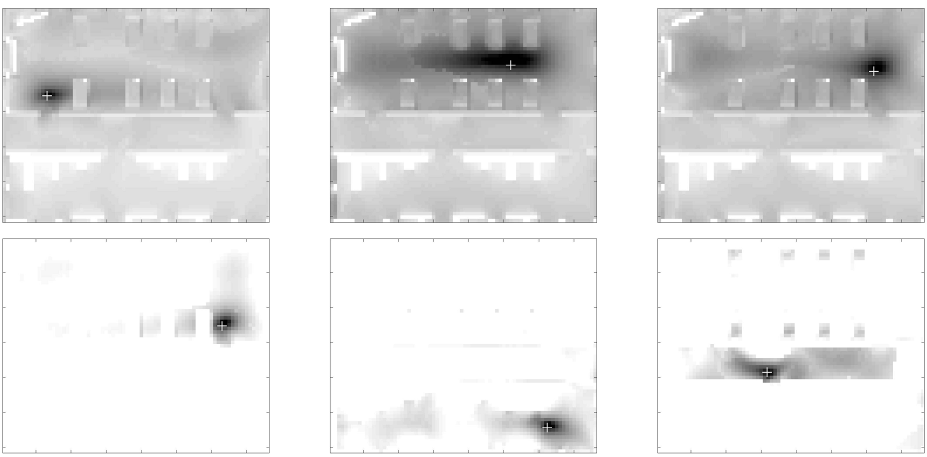


Fig. 5. These plots show the dissimilarity between current view descriptors obtained at six different positions of the path of the robot and the reference views from the database. The shown results have been obtained using a position grid with a spacing of 0.2 m and a total of 4636 view descriptors. Dark areas mark likely positions; white crosses mark the true position. The lower row additionally uses the rotation invariant measure to drop view descriptors beforehand if their energy spectra deviate excessively.

6 Next Steps Towards Realistic View Representations: An Outlook

So far, the view-based representation of the environment in which self-localization shall be performed has been described as a static one, where only one spherical view for each reference location has to be regarded. However, every realistic environment in which a robot will move is subject to various changes, particularly in illumination. Therefore, the content of the database has to consider this appropriately. We propose to represent the set of possible views (under varying illumination etc.) by a stochastic model of low order, as extracted from a larger training set of images. We denote such a model here as a *'dynamic spherical view model' (DSVM)*. In the case of 'normal' rectangular image areas, such models have already been successful for (non-spherical) background modeling [20]. Such a stochastic model inherently induces a suitable and statistically correct metric for the matching process, i. e. the Mahalanobis distance induced by the covariance matrix of the dynamic spherical view model.

In conventional 1D or n-D signal processing a second order description of the signal statistics in terms of covariance functions is sufficient to derive a canonical representation of the signal. This canonical representation is basically the result of a linear transform into a new coordinate system such that the covariance between any pair of different spectral coefficients is zero (*principal component analysis [PCA]* or *Karhunen-Loeve transform [KLT]*). The transfer of this approach to the spherical domain leads to a very practical statistical model for signal processes defined on the spherical domain. Using such a model, a PCA representation of spherical stochastic processes (here: spherical stochastic models for typical omnidirectional signals) can be developed.

The spherical PCA model for omnidirectional signals is a highly practical means for performing any kind of signal processing for *incomplete* spherical data. For example, it allows to compare a given spherical signal with other signals stored in a database even if the input signal contains areas where the signal value is not known or very largely destroyed (occlusions, . . .). The potential and usefulness of a statistically correct procedure for comparing *incomplete* data cannot be overestimated.

These statistical extensions of the self-localization approach using Spherical Harmonics still remain to be performed in investigations planned for the near future. We hope that by the paper presented here the feasibility of the baseline approach and the attractiveness of using Spherical Harmonics for omnidirectional vision and recognition could be conveyed.

References

1. Blaer, P., Allen, P.: Topological mobile robot localization using fast vision techniques. In: International Conference on Robotics and Automation, vol. 1, May 2002, pp. 1031–1036 (2002)
2. Burel, G., Henoco, H.: Determination of the orientation of 3D objects using Spherical Harmonics. Graph. Models Image Process. 57(5), 400–408 (1995)

3. Driscoll, J.R., Healy Jr., D.M.: Computing Fourier transforms and convolutions on the 2-sphere. Adv. Appl. Math. 15(2), 202–250 (1994)
4. The Blender Foundation. Blender (2007), http://www.blender.org
5. Gonzalez-Barbosa, J.-J., Lacroix, S.: Rover localization in natural environments by indexing panoramic images. In: Proceedings of the ICRA '02 IEEE International Conference on Robotics and Automation, pp. 1365–1370. IEEE, Los Alamitos (2002)
6. Groemer, H.: Geometric Applications of Fourier Series and Spherical Harmonics. In: Encyclopedia of Mathematics and Its Applications, Cambridge Univ. Press (1996)
7. Healy Jr., D.M., Rockmore, D.N., Kostelec, P.J., Moore, S.: FFTs for the 2-sphere – improvements and variations. Journal of Fourier Analysis and Applications 9(4), 341–385 (2003)
8. Jogan, M., Leonardis, A.: Robust localization using an omnidirectional appearance-based subspace model of environment. Robotics and Autonomous Systems 45, 57–72 (2003)
9. Kazhdan, M., Funkhouser, T., Rusinkiewicz, S.: Rotation invariant Spherical Harmonic representation of 3D shape descriptors. In: Kobbelt, L., Schröder, P., Hoppe, H. (eds.) Eurographics Symposium on Geometry Processing (June 2003)
10. Kovacs, J.A., Wriggers, W.: Fast rotational matching. Acta Crystallographica Section D 58(8), 1282–1286 (2002)
11. Kröse, B., Vlassis, N., Bunschoten, R., Motomura, Y.: A probabilistic model for appearance-based robot localization. Image and Vision Computing 19(6), 381–391 (2001)
12. Kudlicki, A., Rowicka, M., Gilski, M., Otwinowski, Z.: An efficient routine for computing symmetric real Spherical Harmonics for high orders of expansion. Journal of Applied Crystallography 39, 501–504 (2005)
13. Labbani-Igbida, O., Charron, C., Mouaddib, E.M.: Extraction of Haar integral features on omnidirectional images: Application to local and global localization. In: DAGM-Symposium, pp. 334–343 (2006)
14. Levin, A., Szeliski, R.: Visual odometry and map correlation. In: IEEE Conf. on Comp. Vision and Pattern Recognition, June 2004, vol. I, pp. 611–618. IEEE Computer Society Press, Los Alamitos (2004)
15. Makadia, A., Daniilidis, K.: Direct 3D-rotation estimation from spherical images via a generalized shift theorem. In: IEEE Comp. Society Conference on Computer Vision and Pattern Recognition (CVPR '03), vol. 2, pp. 217–224 (2003)
16. Makadia, A., Daniilidis, K.: Rotation recovery from spherical images without correspondences. IEEE Transactions on Pattern Analysis and Machine Intelligence 28(7), 1170–1175 (2006)
17. Makadia, A., Sorgi, L., Daniilidis, K.: Rotation estimation from spherical images. In: Proceedings ICPR'04, vol. 3 (2004)
18. Menegatti, E., Maeda, T., Ishiguro, H.: Image-based memory for robot navigation using properties of the omnidirectional images (2004)
19. Menegatti, E., Zoccarato, M., Pagello, E., Ishiguro, H.: Hierarchical image-based localisation for mobile robots with monte-carlo localisation. In: Proceedings ECMR'03, Warsaw, Poland, September 2003, pp. 13–20 (2003)
20. Oliver, N., Rosario, B., Pentland, A.: A bayesian computer vision system for modeling human interaction. In: Proceedings ICVS'99, pp. 255–272 (1999)

21. Pajdla, T., Hlavac, V.: Zero phase representation of panoramic images for image based localization. In: Comp. Analysis of Images and Patterns, pp. 550–557 (1999)
22. Thrun, S., Burgard, W., Fox, D.: Probabilistic Robotics. The MIT Press, Cambridge (2005)
23. Weisstein, E.W.: Legendre Polynomial. A Wolfram Web Resource (2007), http://mathworld.wolfram.com/LegendrePolynomial.html