

VIEW-BASED ROBOT LOCALIZATION USING ILLUMINATION-INVARIANT SPHERICAL HARMONICS DESCRIPTORS

Holger Friedrich, David Dederscheck, Martin Mutz and Rudolf Mester
*Visual Sensorics and Information Processing Lab, J. W. Goethe University
Robert-Mayer-Str 10, Frankfurt, Germany*

Keywords: Robot localization, spherical harmonics, illumination invariance.

Abstract: In this work we present a view-based approach for robot self-localization using a hemispherical camera system. We use view descriptors that are based upon Spherical Harmonics as orthonormal basis functions on the sphere. The resulting compact representation of the image signal enables us to efficiently compare the views taken at different locations. With the view descriptors stored in a database, we compute a similarity map for the current view by means of a suitable distance metric. Advanced statistical models based upon principal component analysis introduced to that metric allows to deal with severe illumination changes, extending our method to real-world applications.

1 INTRODUCTION

For the purpose of robot localization, omnidirectional vision has become popular during the last years. Many approaches rely on compact view descriptors (Pajdla and Hlavac, 1999; Blaer and Allen, 2002; Gonzalez-Barbosa and Lacroix, 2002; Levin and Szeliski, 2004), (Kröse et al., 2001; Jogan and Leonardis, 2003) (using principal component analysis) (Menegatti et al., 2003; Menegatti et al., 2004) (using Fourier descriptors), (Labbani-Igbida et al., 2006) (using Haar integrals) to store and compare views efficiently.

We present a *view-based method* for robot localization in a *known environment* represented by a set of reference views. The contribution shown in this paper is an extension of our previous work (Friedrich et al., 2007) to more realistic image data. Using real images imposes various challenges: First, we have to take care of varying illumination. Second, for practical reasons an interpolation method for reference views had to be developed.

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 setup, omnidirectional views are obtained from usual planar images taken by an upward-facing cam-

era, which are subsequently projected onto a hemisphere. These images are converted into *view descriptors*, i.e. low dimensional vectors (Fig. 1), by an expansion in orthonormal basis functions. The robot localization task is performed by comparing the current view descriptor to those stored in a map of the known environment, i.e. a database of views (Fig. 2).



Figure 1: Computing an omnidirectional image signal from a planar wide angle image. The right image visualizes a low order Spherical Harmonics descriptor that approximates the omnidirectional image signal.

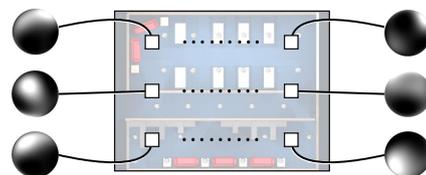


Figure 2: A known environment is represented by a map containing view descriptors. These are obtained from images taken at reference positions.

Given a *suitable distance metric*, for each input view taken at an initially unknown robot position and orientation, a figure of (dis-)similarity to any view stored in the database can be generated directly from the compact vector representation of these views. The resulting likelihood of the robot location/pose allows for more sophisticated sequential self-localization strategies (Thrun et al., 2005), e. g. using particle filters.

2 VIEW DESCRIPTORS

Our view representation is obtained by expanding the spherical image signal $s(\theta, \phi)$ in orthonormal basis functions. The natural choice for spherical basis functions are *Spherical Harmonics* (SH) (Fig. 3). Our approach particularly benefits from using SH since they show the same nice properties concerning rotations which the Fourier basis system provides with respect to translations. Rotations are mapped into a kind of *generalized phase changes*.

Spherical Harmonics (SH). Here, we give a very brief introduction to SH. For group theoretical facts see (Makadia and Daniilidis, 2006) and (Groemer, 1996), for more details on our notation see (Friedrich et al., 2007). Let $N_{\ell m} = \sqrt{\frac{2\ell+1}{2} \frac{(\ell-|m|)!}{(\ell+|m|)!}}$, $\ell \in \mathbb{N}_0$, $m \in \mathbb{Z}$ and $P_{\ell m}(x)$ the Associated Legendre Polynomials (Weisstein, 2007). The SH $Y_{\ell m}(\theta, \phi)$ are defined as

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

with $e^{im\phi}$ being a complex-valued phase term. ℓ ($\ell > 0$) is called *order* and m ($m = -\ell \dots +\ell$) is called *quantum number* for each ℓ . SH have several properties which we exploit in the following sections: Each set of SH of order ℓ forms a *closed orthonormal basis* of dimension $2\ell + 1$; SH of orders $0 \dots \ell$ form a closed orthonormal basis of dimension $(\ell + 1)^2$, hence

$$\int_0^{2\pi} \int_0^{\pi} Y_{\ell m}(\theta, \phi) \cdot Y_{\ell' m'}(\theta, \phi) \cdot \sin \theta \, d\theta \, d\phi = \delta_{\ell \ell'} \cdot \delta_{m m'}, \quad (2)$$

where $\delta_{\ell m}$ is the Kronecker delta function.

3 DoF rotation. Since SH of order ℓ and of order $0 \dots \ell$ form a closed basis, any 3D rotation can be expressed as a *linear transformation* (i. e. multiplication with an unitary matrix U_{ℓ} for each order ℓ) and *does not mix coefficients of different orders*. Hence rotations retain the distribution of spectral energy among different orders (Makadia and Daniilidis, 2006; Kazhdan et al., 2003). This is a unique characteristic of SH which makes them so particularly useful, amongst others for the purpose of robot ego-localization pursued here. Applying a 3D rotation to a spherical func-

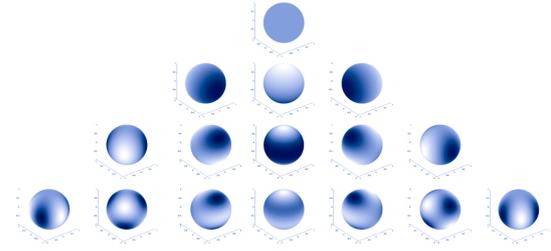


Figure 3: A Spherical Harmonics function is a periodic function on the unit sphere. The rows show SH of orders. $\ell = 0, 1, 2, 3$; columns show $m = 2\ell + 1$ functions for each order ℓ .

tion represented by coefficients a_{jk} yields new coefficients b_{jk} according to

$$\begin{pmatrix} b_{00} \\ b_{10} \\ b_{11} \\ b_{1,-1} \\ \vdots \\ \vdots \\ b_{2,-2} \end{pmatrix} = \underbrace{\begin{pmatrix} U_{\ell=0} & & & & \\ & U_{\ell=1} & & & \\ & & \mathbf{0} & & \\ & & & U_{\ell=2} & \\ & & & & \mathbf{0} \end{pmatrix}}_{U_{tot}} \begin{pmatrix} a_{00} \\ a_{10} \\ a_{11} \\ a_{1,-1} \\ \vdots \\ \vdots \\ a_{2,-2} \end{pmatrix} \quad (3)$$

Rotation estimation has been treated by (Burel and Henoco, 1995), and more recently in (Makadia, 2006; Makadia and Daniilidis, 2006; Makadia and Daniilidis, 2003; Kovacs and Wriggers, 2002).

1 DoF rotation about X_3 -axis. 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. Recalling the definition of the complex-valued SH, the implications of a rotation of ϕ about the X_3 -axis are as follows:

$$Y_{\ell m}(\theta, \phi + \phi) = e^{im\phi} \cdot Y_{\ell m}(\theta, \phi). \quad (4)$$

The rotation matrix changes into a diagonal matrix with elements $e^{-im\phi}$, thus

$$b_{\ell m} = e^{-im\phi} \cdot a_{\ell m}. \quad (5)$$

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

$$s(\theta, \phi) = \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} a_{\ell m} \cdot Y_{\ell m}(\theta, \phi), \quad (6)$$

the coefficients $a_{\ell m}$ are obtained by computing scalar products between the signal and the complex conjugate of each of the basis functions:

$$a_{\ell m} = \int_0^{2\pi} \int_0^{\pi} s(\theta, \phi) \cdot \overline{Y_{\ell m}(\theta, \phi)} \cdot \sin \theta \, d\theta \, d\phi. \quad (7)$$

In practice, this is done using SH of order $\ell = 0$ up to a small number, e. g. $\ell = 4$, resulting in a notable *compression of the input image* to a 25-dimensional vector.

3 LOCALIZATION

3.1 Similarity Measure

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

$$\begin{aligned}
 Q &= \int_0^{2\pi} \int_0^\pi (g(\theta, \phi) - h(\theta, \phi))^2 \cdot \sin \theta \, d\theta \, d\phi \stackrel{Eq.6}{=} \\
 &\int_0^{2\pi} \int_0^\pi \left(\sum_{\ell=0}^\infty \sum_{m=-\ell}^\ell (a_{\ell m} - b_{\ell m}) \cdot \frac{Y_{\ell m}(\theta, \phi)}{\sin \theta} \right)^2 \cdot \sin \theta \, d\theta \, d\phi \\
 &\stackrel{Eq.2}{=} \sum_{\ell=0}^\infty \sum_{m=-\ell}^\ell \sum_{\ell'=0}^\infty \sum_{m'=-\ell'}^{\ell'} (a_{\ell m} - b_{\ell m}) \cdot (a_{\ell' m'} - b_{\ell' m'}) \cdot \delta_{\ell\ell'} \cdot \delta_{mm'} = \|\vec{a} - \vec{b}\|_2^2 \quad (8)
 \end{aligned}$$

This result is of course a consequence of the basis signals forming an orthonormal basis. The measure Q is sensitive to any rotation between the signals. Hence, to find the minimum dissimilarity of two view descriptors we first have to de-rotate them, i. e. compensate for the unknown rotation.

3.2 De-Rotation

Currently our implementation is based on direct non-linear estimation of φ similar to the method described in (Makadia et al., 2004). In this method, the 3D-rotation \mathbf{U}_{tot} for view descriptors \vec{a} and \vec{b} is determined such that $\|\vec{b} - \mathbf{U}_{tot} \vec{a}\|_2^2$ is minimized. The constraint of mere 1-axis rotation, which has been maintained in our experiments so far, 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 de-rotation is possible (Makadia et al., 2004; Makadia and Daniilidis, 2006) for other robot configurations, that is, the SH approach is even more interesting in that case.

3.3 Rotation Invariant Similarity

As coefficients corresponding to different orders of SH are not mixed in rotations, the norms of these subgroups of coefficients are invariant to arbitrary 3D rotations of the signal. Thus L_2 norms, one for each order of SH, 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 (Kazhdan et al., 2003). With a proper metric, spherical signals can be compared to each other even without performing the de-rotation. *Only if* the energy spectrum is identical or similar, the particular spherical signals *can* be

identical. Hence, the energy spectrum can be used as a prefiltering in a matching process.

3.4 Robot Localization Algorithm

We perform the localization task by calculating a distance measure between the current view descriptor and each reference location:

1. First, we use the fast rotation invariant similarity measure to drop all unlikely views.
2. Then, for all reference views which survived this ‘prefiltering’ we estimate the best matching rotation with respect to the current view descriptor and de-rotate it accordingly. Thereafter, we compute the similarity according to the measure introduced in Sec. 3.1.

The resulting similarity map typically shows a distinct extremum at the true location of the robot. As our experiments have shown, there can also be additional extrema of similar likelihood for different poses. This corresponds to the regularities of man-made environments resulting in similar views at more than one position. At each instant, however, we have prior knowledge about the previous course of the robot and its previous pose, which is presumably always sufficient to disambiguate the pose estimation process. Such strategies are well-known in robot navigation and have been, amongst many others, described by Thrun et al. (Thrun et al., 2005) (‘Monte Carlo Localization’), or Menegatti et al. (Menegatti et al., 2003) (using different view descriptors).

4 ILLUMINATION INVARIANCE

Changes in illumination are an inevitable issue to deal with in most vision applications (Mester et al., 2001; Steinbauer and Bischof, 2005). Thus, when performing localization we need to introduce methods that disregard the effects by illumination changes on the measured distance between two view descriptors.

4.1 Multiplicative Illumination Model

The most simple model for illumination changes is a global *multiplicative* change, i. e. the brightness of each pixel in the source image is multiplied by the same number α . This kind of change can of course be easily compensated for by normalization. For typical views at different illumination conditions, we can also expect the factor α to be within certain boundaries, thus limiting illumination invariance to ‘reasonable’ changes.

Eq. (7) dictates that global multiplicative changes of the image signal influence the obtained feature vector in a linear way. Therefore, we may perform the normalization directly in the domain of our feature vectors using the L_2 norm, i.e. each feature vector is normalized to unit length prior to comparison. In the process we can also check if the length of the two compared feature vectors differs so much that it would be unlikely that they refer to the same view.

The normalization is introduced to the rotation sensitive similarity measure Q as follows:

$$\left\| \left(\vec{b} / \|\vec{b}\|_2 \right) - \mathbf{U}_{\text{tot}} \cdot \left(\vec{a} / \|\vec{a}\|_2 \right) \right\|_2^2. \quad (9)$$

4.2 Mahalanobis Distance Using Principal Component Analysis

The similarity measures presented so far only consider global multiplicative variations in illumination. However, typical changes in illumination lead to much more specific and local effects.

Hence, for a fixed position and orientation we no longer deal with a static image, but a multitude of images which can occur under variations of the illumination. This can be interpreted as a distribution on the set of all possible images, which can be much easier described in the space of SH coefficient vectors, i.e. a $(\ell + 1)^2$ -dimensional space (e.g. with $\ell = 4$) instead of a space of $\mathbb{R}^{N \cdot N}$ where N is the image dimension.

The simplest statistical description of this distribution uses the first and second order moments of the distribution, that is $\vec{m}_a = \mathbb{E}[\vec{b}]$ and

$$\mathbf{C}_b = \text{Cov}[\vec{b}] = \mathbb{E}[(\vec{b} - \vec{m}_b)(\vec{b} - \vec{m}_b)^T]. \quad (10)$$

Using these moments does *not* imply a Gaussian assumption on the distribution of \vec{b} , but if we use the Gaussian assumption, we may specify the likelihood of a particular vector \vec{b} to be generated by this distribution:

$$L(\vec{b} | \vec{b} \in \mathcal{N}(\vec{m}_b | \mathbf{C}_b)) = K \cdot e^{-\frac{1}{2}(\vec{b} - \vec{m}_b)^T \mathbf{C}_b^{-1} (\vec{b} - \vec{m}_b)}.$$

This likelihood can be used as a *distance measure* of a given \vec{b} to the mean vector \vec{m}_b of the distribution and thus used as a means to find the correct association of \vec{b} to one of several competing distributions, each one of them representing a particular location and orientation hypothesis.

In the light of our grid of reference frames this means that *each pose* is represented by the *individual mean vector* \vec{m}_b and a *location-specific covariance matrix* \mathbf{C}_b corresponding to \vec{b} .

To compensate for the effects of varying illumination, we introduce such a distance measure that attenuates the effect of those components of the compared

view descriptors, which are affected by varying illuminations. As the covariance matrix \mathbf{C}_b represents this influence, this is accomplished by using the Mahalanobis Distance Δ instead of the normalized L_2 norm.

Let \vec{b} be a feature vector from the reference grid and \vec{a} the currently regarded view descriptor. Consider the covariance matrix \mathbf{C}_b , which has been obtained by sampling a set of typical and different illumination scenarios, thus representing the illumination change model:

$$\Delta_b = \sqrt{(\vec{b} - \vec{a})^T \cdot \mathbf{C}_b^{-1} \cdot (\vec{b} - \vec{a})}. \quad (11)$$

For further investigation on the effects of varying illumination it can be useful to regard the eigenimages obtained by principal component analysis (PCA) (Turk and Pentland, 1991). For a feature vector \vec{b} the transition into the PCA space yields the transformed vector \vec{u} by means of

$$\vec{u} = \mathbf{A}^T (\vec{b} - \vec{m}_b), \quad (12)$$

where \mathbf{A} is a matrix which includes the eigenvectors of the covariance matrix \mathbf{C}_b . This is performed analogously for the currently regarded feature vector \vec{a} , yielding \vec{v} .

In PCA space all components of the transformed feature vector \vec{u} are of course linear independent. Hence, the covariance matrix \mathbf{C}_u in PCA space reduces to a diagonal matrix which only contains the variances σ_i^2 . Obviously, these variances directly correspond to the eigenvalues λ_i of the covariance matrix \mathbf{C}_b .

By using the Mahalanobis Distance in PCA space, the weighting of the components of the feature vectors simply reduces to component-wise multiplying with the inverse of their variances:

$$\Delta_u = \sqrt{\sum_{i=1}^N (u_i - v_i)^2 \lambda_i^{-1}}. \quad (13)$$

As the covariance matrix \mathbf{C}_b has been obtained by training different illumination scenarios, the variances λ_i of the PCA space denote to what extent a component will be affected by illumination changes. Hence, components which are highly affected by illumination changes will be downweighted by using the Mahalanobis Distance.

As recording data for training typical illumination changes is an expensive process, it is not always viable to use an individual training set for *each* reference location of our reference grid. For our experiments we currently only use a global training set recorded at a designated pose (aligned with the grid direction). The PCA transform of reference views and input view descriptors is then obtained using

the corresponding global covariance matrix \mathbf{C}_g and mean vector \vec{m}_g . Different regions \mathcal{R}_i covered by the database of reference views may, however, show completely different behavior under varying illuminations. This would be, for instance, different rooms or parts of rooms in a building. In the light of this, we propose to use different global training sets each recorded at a fixed pose inside the region. This results in the individual $\mathbf{C}_{g,i}$ and $\vec{m}_{g,i}$ to be used for the corresponding region \mathcal{R}_i .

5 INTERPOLATION OF VIEW DESCRIPTORS

The recording of reference views can be a very expensive procedure in real-world applications. Hence, there clearly is a need for interpolating view descriptors in a way so that a sufficiently detailed map can be obtained from a sparse set of actually recorded reference images.

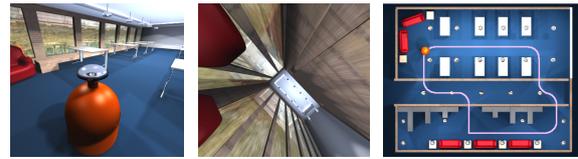
Due to the low-frequency nature of the signal represented in a view-descriptor composed only of lower-order SH coefficients (7), we propose to perform a *bilinear interpolation* between two view descriptors recorded at two sufficiently close positions. Using this simple interpolation method, it should be possible to supplement a rather coarse grid of reference views with additional interpolated views. Consequently, we can obtain a more precise localization of likely robot positions. We can also expect the spatial distribution of our dissimilarity map to be more smooth, giving a benefit to the minima detection over the original set of reference views. A profound investigation on the question to what extent the effects of translations on SH representations can be approximately covered by interpolation has been started.

6 EXPERIMENTS

For our experiments we use both synthetic data obtained by the 3D software (The Blender Foundation, 2007) and real data; we approximate with SH up to order $\ell = 4$.

6.1 Simulated Environment

In our previous work (Friedrich et al., 2007) we have used synthetic image data of an artificial office environment (Fig. 4). An upwards facing wide-angle perspective camera with a field of view (FOV) of approx. 172.5° yields the input images. The resulting



(a) Robot w. camera. (b) View facing upwards. (c) True path of robot.

Figure 4: Views of our simulated office environment.

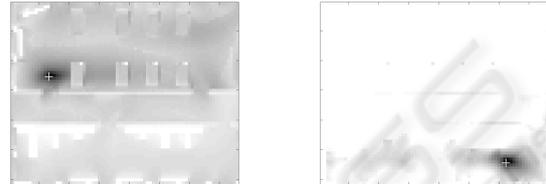


Figure 5: Plots of the dissimilarity between view descriptors obtained at different positions of the simulated environment and all the reference views. The grid consists of 4636 views at a spacing of 0.2m. Dark areas mark likely positions; white crosses mark the true position. The right image additionally uses the rotation invariant prefiltering.

images can be projected onto a hemisphere. To extend this to a full spherical signal we employ suitable reflection at the equator before approximating the signal by a SH expansion. Of course this representation can be obtained directly from the 2D images.

Prior to performing any localization of the robot, we must create a set of reference frames and calculate its corresponding view descriptors. For our localization experiment, we have also rendered a series of frames with the simulated robot moving along a path (Fig. 4). Note that these positions are in general *not* aligned with the grid, neither is the orientation of the robot aligned with the direction of the reference frames.

The images in Fig. 5 are maps of the simulated environment showing a measure corresponding to the likelihood of the robot location, calculated at discrete positions along the motion path.

6.2 Hemispherical Camera System

Hardware. To acquire hemispherical wide-angle images using a real camera, we have designed a low-cost camera system with a fisheye lens based on the ideas presented in (Dietz, 2006).

As shown in Fig. 6, the camera system consists of a cheap door peephole attached to a low-cost b/w CCTV camera with a 12 mm lens, which is used to perform ocular projection.

The optical quality of such a system is of course



Figure 6: Low-cost camera system based on a door peephole lens (a) adapted to a CCTV camera (c). A mounting frame (d) is used to lock the peephole in a calibrated distance to the camera lens (b).

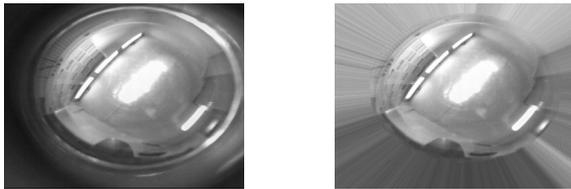


Figure 7: Image data prior to (left) and after preprocessing (right). Note the interpolated area.

not comparable to commercial high-quality fisheye lenses; however, if proper image preprocessing and careful calibration is performed, the results are already quite useful, especially in the light of this low-cost solution. For our application, these are more than acceptable, as we do not make use of the spatial high-frequency components of the images.

Calibration and Preprocessing. The raw images obtained by our hemispherical camera system are of course distorted and like any real camera image require calibration. However, due to the nature of the extreme wide-angle lens, there is a blind area beyond the usable FOV in the images, which is preceded by an unusable region due to reflections of the peephole housing.

For preprocessing, we first perform photometric calibration (multiplicative). To compensate for the effects of decaying brightness in the outer regions of the image, we recorded a set of reference images of a white homogenous area, which were averaged and normalized to the lower boundary of a suitable upper brightness percentile. We use the pixel-wise inverse of the relative brightness as a template, which is then multiplied with future captured images. To discard the garbage induced by reflexes at the rim of the images, we only use a safe area of image data. This leads to an elliptical image area corresponding to a FOV of approx. 160° . As the usable image data is not completely hemispherical, we need to use an appropriate interpolation to extend the signal to the full 180° . We use a radial nearest-neighbor interpolation

from the border of our aforementioned safe area for that purpose (Fig. 7). This makes sure that no additional discontinuities occur at the equator of a consequent spherical projection. To project the input image onto the sphere, we need to calibrate our camera system. We use the INRIA toolbox (Mei, 2006).

6.3 Real Environment

Illumination Invariance. For our experiments, we recorded a grid of 374 reference views in an office while the illumination was kept constant. Thereafter, we recorded several sequences while the robot was driven through that environment. This was performed for two cases – one with the same constant illumination as for the reference grid, whereas in the other case there were substantial changes, such as switching the ceiling lighting on and off (Fig. 8).

Since the CCTV camera we used employs a sub-optimal gain control which cannot be turned off, we had to use the normalization according to the presented multiplicative illumination invariant similarity measure for performing the localization task even under constant illumination. This method fails, however, for the second sequence with more severe changes in illumination, clearly indicating the need for a statistical model.

To model the typical illumination changes in that room, we recorded a global training set at a designated location, which was used as input to a PCA model. Dissimilarity maps using the resulting PCA based distance measure show distinctly better results even where the multiplicative illumination invariant similarity measure could not cope with the input (Fig. 8). The results are very promising in the light that the given illumination changes covered even extreme cases.

Interpolation. To evaluate whether the detour of recording all the reference views of a densely spaced grid is actually necessary to obtain sufficiently distinctive localization results, we computed another reference grid where the view descriptors in every second row and column were obtained through bilinear interpolation.

In Fig. 9, we show that there are only little differences between results using a full resolution reference grid with a spacing of 0.2 m and one obtained by interpolation of a half resolution grid. This applies especially to the location of the dissimilarity minima. This encourages the usage of interpolated grids to obtain smoother localization results with more reliable minima detection.

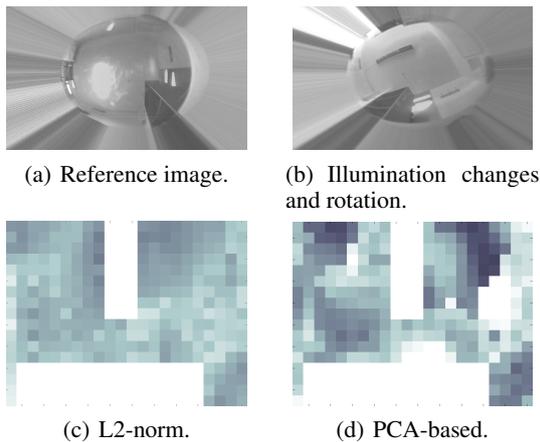


Figure 8: Illumination changes. Dissimilarity maps showing the behavior of different distance measures when applied to views with substantial illumination changes vs. the illumination used to record the reference grid. The true position is at the upper right.

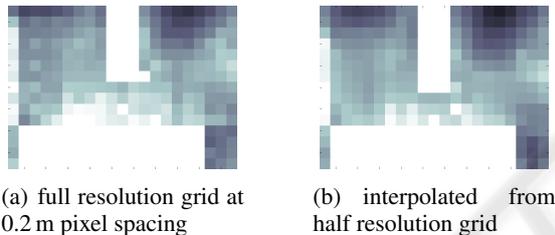


Figure 9: Interpolation. Dissimilarity maps using real data with the same constant illumination as for recording the reference grid. Note that both results show only minor differences.

7 OUTLOOK

So far, only a small fraction of methods to achieve illumination invariant robot localization have been investigated in our work. Learning typical effects of illumination changes for the PCA model at each location induces a high effort during the learning phase. Therefore, other methods should be considered as well: The generalization from a SH representation of image signals to a SH representation of feature images could be useful, e.g. using gradient images (Reisert and Burkhardt, 2006). Furthermore we could investigate the benefits of using color images and photometric invariants (Mileva et al., 2007).

Concerning the interpolation of view descriptors, a better understanding of the effects which translations have on the view descriptors will be crucial. This will be particularly important to further reduce the required density of the reference view grid. In connection to this, a differential analysis will provide

a measure to estimate a viable maximum distance between reference views that still can be interpolated without excessive error. On the other hand, more sophisticated interpolation methods might improve the localization results.

To speed up and robustify the process of robot localization, we propose to use more advanced prefiltering of unlikely positions. Further invariants can be determined from applying point-wise non-linear functions on the original or the spherical image signal before expanding in SH (Schulz et al., 2006). The use of phase information in addition to the energy spectra used so far could make the rotation invariant matching of view descriptors more discriminant, resulting in less views which misleadingly pass the filtering.

An issue that still needs further investigation is the handling of occlusions, and we proceed towards this goal. A statistical model for spherical signals which allows for a correct interpolation of missing data serves as a highly practical means for comparing/matching/correlating *incomplete* omnidirectional data (Mühlich and Mester, 2004). It allows to compare a given 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. The potential and usefulness of a statistically correct procedure for comparing *incomplete* data cannot be overestimated.

REFERENCES

- Blaer, P. and Allen, P. (2002). Topological mobile robot localization using fast vision techniques. In *Proc. ICRA*, volume 1, pages 1031–1036.
- Burel, G. and Henoco, H. (1995). Determination of the orientation of 3D objects using Spherical Harmonics. *Graphical Models and Image Processing*, 57(5):400–408.
- Dietz, H. G. (2006). Fisheye digital imaging for under twenty dollars. Technical report, Univ. of Kentucky. <http://aggregate.org/DIT/peepfish/>.
- Friedrich, H., Dederscheck, D., Krajsek, K., and Mester, R. (2007). View-based robot localization using Spherical Harmonics: Concept and first experimental results. In Hamprecht, F., Schnörr, C., and Jähne, B., editors, *DAGM 07*, number 4713 in LNCS, pages 21–31. Springer.
- Gonzalez-Barbosa, J.-J. and Lacroix, S. (2002). Rover localization in natural environments by indexing panoramic images. In *Proc. ICRA*, pages 1365–1370. IEEE Computer Society.
- Groemer, H. (1996). *Geometric Applications of Fourier Series and Spherical Harmonics*. Encyclopedia of Mathematics and Its Applications. Cambridge University Press.

- Jogan, M. and Leonardis, A. (2003). Robust localization using an omnidirectional appearance-based subspace model of environment. *Robotics and Autonomous Systems*, 45:57–72.
- Kazhdan, M., Funkhouser, T., and Rusinkiewicz, S. (2003). Rotation invariant Spherical Harmonic representation of 3D shape descriptors. In Kobbelt, L., Schröder, P., and Hoppe, H., editors, *Eurographics Symp. on Geometry Proc.*
- Kovacs, J. A. and Wriggers, W. (2002). Fast rotational matching. *Acta Crystallographica Section D*, 58(8):1282–1286.
- Kröse, B., Vlassis, N., Bunschoten, R., and Motomura, Y. (2001). A probabilistic model for appearance-based robot localization. *Image and Vision Computing*, 19(6):381–391.
- Labbani-Igbida, O., Charron, C., and Mouaddib, E. M. (2006). Extraction of Haar integral features on omnidirectional images: Application to local and global localization. In *DAGM 06*, pages 334–343.
- Levin, A. and Szeliski, R. (2004). Visual odometry and map correlation. In *CVPR*, volume I, pages 611–618.
- Makadia, A. (2006). Correspondenceless visual navigation under constrained motion. In Daniilidis, K. and Klette, R., editors, *Imaging beyond the Pinhole Camera*, pages 253–268.
- Makadia, A. and Daniilidis, K. (2003). Direct 3D-rotation estimation from spherical images via a generalized shift theorem. In *CVPR*, volume 2, pages 217–224.
- Makadia, A. and Daniilidis, K. (2006). Rotation recovery from spherical images without correspondences. *PAMI*, 28(7):1170–1175.
- Makadia, A., Sorgi, L., and Daniilidis, K. (2004). Rotation estimation from spherical images. In *ICPR*, volume 3.
- Mei, C. (2006). Omnidirectional calibration toolbox. <http://www.robots.ox.ac.uk/~cmei/Toolbox.html>.
- Menegatti, E., Maeda, T., and Ishiguro, H. (2004). Image-based memory for robot navigation using properties of the omnidirectional images. *Robotics and Autonomous Systems*, 47(4).
- Menegatti, E., Zoccarato, M., Pagello, E., and Ishiguro, H. (2003). Hierarchical image-based localisation for mobile robots with Monte-Carlo localisation. In *Proc. European Conference on Mobile Robots*, pages 13–20.
- Mester, R., Aach, T., and Dübgen, L. (2001). Illumination-invariant change detection using a statistical colinearity criterion. In *Pattern Recognition*, number 2191 in LNCS.
- Mileva, Y., Bruhn, A., and Weickert, J. (2007). Illumination-robust variational optical flow with photometric invariants. In Hamprecht, F., Schnörr, C., and Jähne, B., editors, *Pattern Recognition*, volume 4713 of LNCS, pages 152–162. Springer.
- Mühlich, M. and Mester, R. (2004). A statistical unification of image interpolation, error concealment, and source-adapted filter design. In *Proc. SSIAP*, pages 128–132. IEEE Computer Society.
- Pajdla, T. and Hlavac, V. (1999). Zero phase representation of panoramic images for image based localization. In *Computer Analysis of Images and Patterns*, pages 550–557.
- Reisert, M. and Burkhardt, H. (2006). Invariant features for 3D-data based on group integration using directional information and Spherical Harmonic expansion. In *Proc. ICPR 2006*, LNCS.
- Schulz, J., Schmidt, T., Ronneberger, O., Burkhardt, H., Pasternak, T., Dovzhenko, A., and Palme, K. (2006). Fast scalar and vectorial grayscale based invariant features for 3D cell nuclei localization and classification. In Franke, K. and al., E., editors, *DAGM 2006*, number 4174 in LNCS, pages 182–191. Springer.
- Steinbauer, G. and Bischof, H. (2005). Illumination insensitive robot self-localization using panoramic eigenspaces. In *RoboCup 2004*.
- The Blender Foundation (2007). Blender. <http://www.blender.org>.
- Thrun, S., Burgard, W., and Fox, D. (2005). *Probabilistic Robotics*. The MIT Press.
- Turk, M. and Pentland, A. (1991). Eigenfaces for recognition. *J. of Cognitive Neuroscience*, 3(1).
- Weissstein, E. W. (2007). Legendre polynomial. A Wolfram Web Resource. <http://mathworld.wolfram.com/LegendrePolynomial.html>.