

Optical Rails.

View-based Point-To-Point Navigation using Spherical Harmonics

Holger Friedrich, David Dederscheck, Eduard Rosert, and Rudolf Mester

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

Abstract. We present a view-based method for steering a robot in a network of positions; this includes navigation along a prerecorded path, but also allows for arbitrary movement of the robot between adjacent positions in the network. The approach uses an upward-looking omnidirectional camera; even a very modest quality of the optical system is sufficient, since all views are represented in terms of low-order basis functions (spherical harmonics). Motor control signals for the robot are derived from a differential matching approach; the computation of the involved gradient information is extremely simplified by exploiting the fact that all images are represented in terms of basis functions. The viability of the approach for steering the robot has been shown in extensive simulations using photorealistic views; the validity of these simulations in comparison to a tangible system implementation operating in a real indoor scene has been shown in previous investigations [5].

1 Introduction

Visual perception is the most important capability of humans and animals when it comes to autonomously moving through a complex environment. The question “Where am I?” is in biological systems most often answered by comparing *views*, and it seems less probable that any sort of geometry, e. g. triangulation, plays an essential role in identifying where we are, and in finding our way along a known path through our world. Visual perception based on geometric principles (triangulation, epipolar geometry, etc.) has its particular virtues, so far also gladly exploited by our group, but it seems as if *view-based* approaches were advantageous in other situations.

The present paper shows that guiding a robot or vehicle through a known environment merely based on omnidirectional views and without using any geometric information is possible, and that it is not only a viable, but also an advantageous alternative to using markers, laser scanners, etc. We denote this approach as ‘*Optical Rails*’.

The method shares certain characteristics with landmark-based robot homing approaches [4, 9, 1], view based ones [11], and also with those that use epipolar

geometry [2, 8]. Our approach is particularly efficient by exploiting the representation of the image in spherical harmonics. We obtain a pose derivative necessary for controlling the motion of the robot vehicle from a differential analysis of a matching criterion. However – in contrast to many similar approaches – we acquire differential entities in a simple and computationally inexpensive way using precomputed expressions.

By expanding the image signal in basis functions we avoid the computation of derivatives in image space like in the classical Lucas-Kanade approach [7]. Because we use a truncated expansion, our method works well without a good image resolution or precision optics.

2 Moving Robots on Optical Rails

The concept of *Optical Rails* is based on the omnidirectional views obtained at discrete locations in a network of so-called ‘waypoints’. Instead of using geometric position information, merely the comparison of the omnidirectional views at the current robot location and at a target location allows the robot to control its drive motors to move towards the target waypoint. Furthermore, *paths* can be represented by a sequence of waypoints to visit one after the other (Fig.1). This enables the robot to drive along prerecorded sequences of waypoints, or to arbitrary locations as necessary *ad hoc*. We denote the task of approaching a target location based on the associated view as *differential pose tracking*.

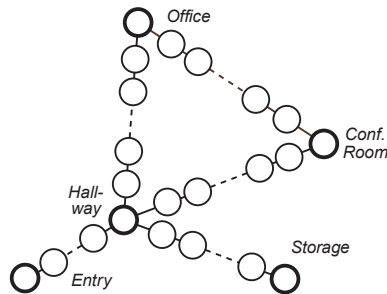


Fig. 1. A navigation graph consisting of a moderate number of reference views and interconnecting ‘Optical Rails’.

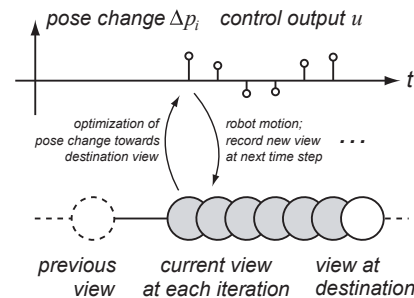


Fig. 2. Differential pose tracking: At each step, the optimal pose change is calculated based on the current view.

In our approach, the omnidirectional views obtained from an upward-looking omnidirectional or ultra-wide angle camera are represented in terms of an expansion into spherical harmonics, analogously to a usual Fourier analysis in the plane. The coefficients of this expansion form *view descriptors*. The similarity (or dissimilarity) between different views can be easily determined from the

difference of the corresponding two coefficient vectors. Considering one view descriptor as a target vector, and the other as the view descriptor of the current robot pose, the differential change of the dissimilarity when performing an infinitesimal pose change of the robot yields a direct information on how the robot should be moved in order to decrease the dissimilarity. This differential process is fully analogous to classical differential matching processes (a.k.a. ‘differential image registration’) such as used by Lucas and Kanade [7]. This way, we can deduce the translation and rotation a robot should move to approach a given target location.

The effect of a pose change (2D translation on the ground plane, rotation about the z axis) on the spherical image signal is in general complex, in particular for the translation part. Furthermore, for an exact prediction of how the spherical signal would look like after the pose change we would need the depth structure of the scene, which is usually not available. Therefore, we reproject the spherical signal on a virtual ‘ceiling plane’ parallel to the ground plane on which the robot is moving, perform the pose change (3 degrees of freedom). The projection onto the ceiling plane is known as ‘gnomonic projection’ in the literature.

It is important to notice that the absolute pose of the robot in the world coordinate frame is not needed for this operation; what we actually deal with are (infinitesimal) *changes* of the pose in terms of the robot coordinate system. For using this to actually move the robot in the ‘right’ direction, only the Euclidean transform between the camera coordinate frame and the robot coordinate frame needs to be known.

3 A Model for Differential Pose Change Estimation

Let \mathbf{p}_c be the current pose, and \mathbf{a} be the associated view descriptor. Let furthermore \mathbf{p}_d be the destination pose, and $\tilde{\mathbf{a}}$ the associated destination view descriptor. (The actual values of \mathbf{p}_c and \mathbf{p}_d will not be needed in the final result). The dissimilarity measure Q between two views is (like in our previous work [5]) the mean squared image signal difference, obtained by integration across the semisphere which is related to the view descriptors in a simple way:

$$Q := \int_{\theta=0}^{\pi/2} \int_{\phi=0}^{2\pi} (s(\theta, \phi) - \tilde{s}(\theta, \phi))^2 \cdot \sin \theta \, d\phi \, d\theta = \|\mathbf{a} - \tilde{\mathbf{a}}\|_2^2. \quad (1)$$

The view descriptor \mathbf{a} is dependent on the current pose of the robot; thus the negative gradient $-\mathbf{g} = -\partial Q / \partial \mathbf{p}_c$ represents the driving direction of the robot to reduce the dissimilarity Q , which corresponds to approaching the destination. Since the relation between the image signal and the pose change is awkward to compute on the sphere, we use the projection approach:

- The spherical image signal corresponding to a particular view descriptor (= a weighted sum of basis functions) is projected onto the ceiling plane,
- differential translations and rotations of the projected signal resemble movements of the camera, which correspond to pose changes of the robot,

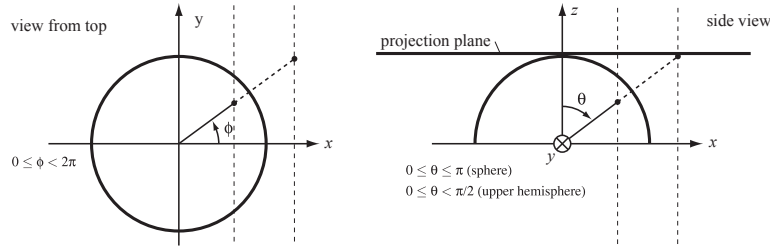


Fig. 3. Gnomonic Projection of a Hemispherical Signal

- the dissimilarity measure Q is computed by integration across the ceiling plane.

Taking the partial derivatives with respect to the components of the current pose vector \mathbf{p}_c yields the gradient \mathbf{g} which represents the (negative) steering direction towards the destination. This approach is of course only an approximation since the effects of the underlying geometry of the environment, i. e. objects at different distances, are disregarded.

Gnomonic Projection of a Hemispherical Signal. If we project an image signal $s(\theta, \phi)$ defined on a semi-sphere from the center of that sphere onto a plane, we obtain the *Gnomonic Projection* (Fig. 3). We obtain spherical coordinates from Cartesian coordinates on the projection plane by

$$\mathbf{G}(x, y) = (\theta(x, y); \phi(x, y))^T \quad \begin{aligned} \theta(x, y) &= \arctan(1, \sqrt{x^2 + y^2}) \\ \phi(x, y) &= \arctan(x, y). \end{aligned} \quad (2)$$

Conversely, we obtain cartesian coordinates on the projection plane from spherical coordinates as follows:

$$\mathbf{G}^{-1}(\theta, \phi) = (x(\theta, \phi); y(\theta, \phi))^T \quad \begin{aligned} x(\theta, \phi) &= \frac{\sin(\theta) \cdot \cos(\phi)}{\cos(\theta)} = \tan(\theta) \cdot \cos(\phi) \\ y(\theta, \phi) &= \frac{\sin(\theta) \cdot \sin(\phi)}{\cos(\theta)} = \tan(\theta) \cdot \sin(\phi). \end{aligned} \quad (3)$$

Integral of a Planar Projection of a Hemispherical Function. The dissimilarity measure Q as defined in (1) shall now be computed by an integration over the projection plane instead of an integration over the sphere. To that purpose, we need to introduce a weight function $w(\mathbf{x})$.

With $\Delta(\theta, \phi) := s(\theta, \phi) - \tilde{s}(\theta, \phi)$, our dissimilarity measure Q is given by

$$Q = \int_{\theta=0}^{\pi/2} \int_{\phi=0}^{2\pi} \Delta^2(\theta, \phi) \cdot \sin \theta \, d\phi \, d\theta. \quad (4)$$

where the factor $\sin \theta$ is introduced by the Jacobian determinant.

If we now perform the gnomonic transform as given above, we obtain

$$Q = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \underbrace{\Delta^2(\theta(x, y), \phi(x, y))}_{=:\mathbf{G}(x, y)} \cdot \underbrace{\det |\mathbf{J}| \cdot \sin(\theta(x, y))}_{=:w(x, y)} \, dx \, dy, \quad (5)$$

where $\det |\mathbf{J}|$ is the determinant of the Jacobian for the gnomonic transform $\mathbf{G}(x, y)$:

$$\mathbf{J} = \begin{pmatrix} \frac{\partial \theta(x, y)}{\partial x} & \frac{\partial \phi(x, y)}{\partial x} \\ \frac{\partial \theta(x, y)}{\partial y} & \frac{\partial \phi(x, y)}{\partial y} \end{pmatrix} = \begin{pmatrix} \frac{x}{\sqrt{x^2 + y^2}(1 + x^2 + y^2)} & -\frac{y}{x^2 + y^2} \\ \frac{y}{\sqrt{x^2 + y^2}(1 + x^2 + y^2)} & \frac{x}{x^2 + y^2} \end{pmatrix} \quad (6)$$

$$\implies \det |\mathbf{J}| = 1 / (\sqrt{x^2 + y^2} \cdot (1 + x^2 + y^2)) \quad (7)$$

For the weight function $w(x, y)$ we finally obtain:

$$w(x, y) = \det |\mathbf{J}| \cdot \sin(\arctan(\sqrt{x^2 + y^2})) = (1 + x^2 + y^2)^{-\frac{3}{2}}. \quad (8)$$

Differential Pose Change Estimation for 3 DoF Motion. Let $b(\mathbf{x})$ and $\tilde{b}(\mathbf{x})$ be the planar image signals taken at the current pose \mathbf{p}_c and destination pose \mathbf{p}_d of the robot, respectively. We define $\mathbf{f}(\mathbf{x}, \mathbf{d})$ a 2D affine transform which represents the pose translation and rotation by the parameter vector \mathbf{d} :

$$\mathbf{f}(\mathbf{x}, \mathbf{d}) = \mathbf{A} \cdot \mathbf{x} \quad \mathbf{A} = \begin{pmatrix} \cos \varphi - \sin \varphi v_1 & \sin \varphi v_1 \\ \sin \varphi & \cos \varphi v_2 \\ 0 & 0 & 1 \end{pmatrix}, \quad \mathbf{d} = \begin{pmatrix} v_1 \\ v_2 \\ \varphi \end{pmatrix}, \quad \mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ 1 \end{pmatrix} \quad (9)$$

We now regard our dissimilarity measure Q for a transformed input image $b(\mathbf{f}(\mathbf{x}, \mathbf{d}))$:

$$Q = \int_A w(\mathbf{x}) \cdot [b(\mathbf{f}(\mathbf{x}, \mathbf{d})) - \tilde{b}(\mathbf{x})]^2 d\mathbf{x}, \quad (10)$$

The partial derivative of Q with respect to the parameter vector \mathbf{d} at $\mathbf{d} = \mathbf{0}$ is:

$$\frac{\partial Q}{\partial d_i} = \int_A 2 \cdot w(\mathbf{x}) \cdot [b(\mathbf{f}(\mathbf{x}, \mathbf{d})) - \tilde{b}(\mathbf{x})] \cdot \left[\frac{\partial b(\mathbf{f}(\mathbf{x}, \mathbf{d}))}{\partial d_i} - \frac{\partial \tilde{b}(\mathbf{x})}{\partial d_i} \right] d\mathbf{x} \Big|_{\mathbf{d}=\mathbf{0}}. \quad (11)$$

As $\tilde{b}(\mathbf{x})$ does not depend on \mathbf{d} , $\frac{\partial \tilde{b}(\mathbf{x})}{\partial d_i}$ is 0. The derivative $\frac{\partial b(\mathbf{f}(\mathbf{x}, \mathbf{d}))}{\partial d_i} - \frac{\partial \tilde{b}(\mathbf{x})}{\partial d_i}$ can be obtained by the generalized chain rule:

$$\frac{\partial b(\mathbf{f}(\mathbf{x}, \mathbf{d}))}{\partial d_i} = \left(\frac{\partial f_1(\mathbf{x}, \mathbf{d})}{\partial d_i}, \frac{\partial f_2(\mathbf{x}, \mathbf{d})}{\partial d_i}, \frac{\partial f_3(\mathbf{x}, \mathbf{d})}{\partial d_i} \right) \left(\frac{\partial b(\mathbf{y})}{\partial y_1}, \frac{\partial b(\mathbf{y})}{\partial y_2}, \frac{\partial b(\mathbf{y})}{\partial y_3} \right)^T \Big|_{\mathbf{y}=\mathbf{f}(\mathbf{x}, \mathbf{d})} \quad (12)$$

where $f_i(\mathbf{x}, \mathbf{d})$ are three individual components of the vector function $\mathbf{f}(\mathbf{x}, \mathbf{d})$, which represents the affine transform. With $\xi_1 := -x_1 \cdot \sin \varphi + x_2 \cos \varphi$ and $\xi_2 := -x_1 \cdot \cos \varphi - x_2 \cdot \sin \varphi$, the individual partial derivatives evaluate as follows:

$$\begin{aligned} \frac{\partial f_1(\mathbf{x}, \mathbf{d})}{\partial p_i} &= \frac{\partial (x_1 \cos \varphi + x_2 \sin \varphi + v_1)}{\partial p_i} & \frac{\partial f_1(\mathbf{x}, \mathbf{d})}{\partial v_1} &= 1 & \frac{\partial f_2(\mathbf{x}, \mathbf{d})}{\partial v_1} &= 0 & \frac{\partial f_3(\mathbf{x}, \mathbf{d})}{\partial v_1} &= 0 \\ \frac{\partial f_2(\mathbf{x}, \mathbf{d})}{\partial p_i} &= \frac{\partial (-x_1 \sin \varphi + x_2 \cos \varphi + v_2)}{\partial p_i} & \frac{\partial f_1(\mathbf{x}, \mathbf{d})}{\partial v_2} &= 0 & \frac{\partial f_2(\mathbf{x}, \mathbf{d})}{\partial v_2} &= 1 & \frac{\partial f_3(\mathbf{x}, \mathbf{d})}{\partial v_2} &= 0 \\ \frac{\partial f_3(\mathbf{x}, \mathbf{d})}{\partial p_i} &= \frac{\partial (0 \cdot x_1 + 0 \cdot x_2 + 1 \cdot 1)}{\partial p_i} & \frac{\partial f_1(\mathbf{x}, \mathbf{d})}{\partial \varphi} &= \xi_1 & \frac{\partial f_2(\mathbf{x}, \mathbf{d})}{\partial \varphi} &= \xi_2 & \frac{\partial f_3(\mathbf{x}, \mathbf{d})}{\partial \varphi} &= 0 \end{aligned}$$

$$\implies \frac{\partial b(\mathbf{f}(\mathbf{x}, \mathbf{d}))}{\partial \mathbf{d}_i} = \left(\frac{\partial b(\mathbf{y})}{\partial y_1}; \frac{\partial b(\mathbf{y})}{\partial y_2}; \xi_1 \frac{\partial b(\mathbf{y})}{\partial y_1} + \xi_2 \frac{\partial b(\mathbf{y})}{\partial y_2} \right)^T \Big|_{\mathbf{y}=\mathbf{f}(\mathbf{x}, \mathbf{d})} \quad (13)$$

By substituting the results from (13) into (11) we then obtain:

$$\frac{\partial Q}{\partial \mathbf{d}} = 2 \int_A w(\mathbf{x}) [b(\mathbf{f}(\mathbf{x}, \mathbf{d})) - \tilde{b}(\mathbf{x})] \left(\frac{\partial b(\mathbf{y})}{\partial y_1}; \frac{\partial b(\mathbf{y})}{\partial y_2}; \xi_1 \frac{\partial b(\mathbf{y})}{\partial y_1} + \xi_2 \frac{\partial b(\mathbf{y})}{\partial y_2} \right)^T \Big|_{\mathbf{y}=\mathbf{f}(\mathbf{x}, \mathbf{d})} \Big|_{\mathbf{d}=\mathbf{0}} d\mathbf{x} \quad (14)$$

As we have now evaluated all derivative terms which depend on the value of the parameter vector, we may now perform the transition to $\mathbf{d} = \mathbf{0}$. As $\mathbf{f}(\mathbf{x}, \mathbf{0})$ is the identity transform, we obtain:

$$\frac{\partial Q}{\partial \mathbf{d}} = \int_A 2 \cdot w(\mathbf{x}) \cdot [b(\mathbf{x}) - \tilde{b}(\mathbf{x})] \cdot \left(\frac{\partial b(\mathbf{x})}{\partial x_1}; \frac{\partial b(\mathbf{x})}{\partial x_2}; x_2 \frac{\partial b(\mathbf{x})}{\partial x_1} - x_1 \frac{\partial b(\mathbf{x})}{\partial x_2} \right)^T d\mathbf{x}, \quad (15)$$

which represents the gradient g for continuous planar image signals. However, by replacing these signals using expansions into basis functions, this can be drastically simplified as will be shown below.

Exploiting the Representation as Linear Combination. We regard now the particular case of image signals being represented by linear combination of basis functions \check{Y}_j defined on the ceiling plane, i. e.

$$b(\mathbf{x}) = \sum_j a_j \cdot \check{Y}_j(\mathbf{x}), \quad \tilde{b}(\mathbf{x}) = \sum_j \tilde{a}_j \cdot \check{Y}_j(\mathbf{x}), \quad (16)$$

where a_j are the coefficients of the linear combination, corresponding to the j -th entry of the view descriptor \mathbf{a} (see Appendix A). Substituting the image signal $\sum_j a_j \cdot \check{Y}_j(\mathbf{x})$ in (15) results in:

$$\begin{aligned} \frac{\partial Q}{\partial \mathbf{d}} &= 2 \int_A w(\mathbf{x}) \left(\begin{array}{c} \sum_j \sum_k (a_j - \tilde{a}_j) \check{Y}_j(\mathbf{x}) a_k \frac{\partial \check{Y}_k(\mathbf{x})}{\partial x_1} \\ \sum_j \sum_k (a_j - \tilde{a}_j) \check{Y}_j(\mathbf{x}) a_k \frac{\partial \check{Y}_k(\mathbf{x})}{\partial x_2} \\ \sum_j \sum_k (a_j - \tilde{a}_j) \check{Y}_j(\mathbf{x}) a_k \left(x_2 \frac{\partial \check{Y}_k(\mathbf{x})}{\partial x_1} - x_1 \frac{\partial \check{Y}_k(\mathbf{x})}{\partial x_2} \right) \end{array} \right) d\mathbf{x} \\ &= 2 \left(\begin{array}{c} \sum_j \sum_k (a_j - \tilde{a}_j) a_k \int_A w(\mathbf{x}) \check{Y}_j(\mathbf{x}) \frac{\partial \check{Y}_k(\mathbf{x})}{\partial x_1} d\mathbf{x} \\ \sum_j \sum_k (a_j - \tilde{a}_j) a_k \int_A w(\mathbf{x}) \check{Y}_j(\mathbf{x}) \frac{\partial \check{Y}_k(\mathbf{x})}{\partial x_2} d\mathbf{x} \\ \sum_j \sum_k (a_j - \tilde{a}_j) a_k \int_A w(\mathbf{x}) \check{Y}_j(\mathbf{x}) \left(x_2 \frac{\partial \check{Y}_k(\mathbf{x})}{\partial x_1} - x_1 \frac{\partial \check{Y}_k(\mathbf{x})}{\partial x_2} \right) d\mathbf{x} \end{array} \right) \quad (17) \end{aligned}$$

Since the integrals in the three sums do not depend on any of the spectral coefficients a_i , we may pre-compute them. We obtain the pre-computed coefficients $u_i(j, k)$:

$$u_1(j, k) = \int_{x_1=-\infty}^{\infty} \int_{x_2=-\infty}^{\infty} w(\mathbf{x}) \cdot \check{Y}_j(\mathbf{x}) \frac{\partial \check{Y}_k(\mathbf{x})}{\partial x_1} dx_2 dx_1 \quad (18)$$

$$u_2(j, k) = \int_{x_1=-\infty}^{\infty} \int_{x_2=-\infty}^{\infty} w(\mathbf{x}) \check{Y}_j(\mathbf{x}) \frac{\partial \check{Y}_k(\mathbf{x})}{\partial x_2} dx_2 dx_1 \quad (19)$$

$$u_3(j, k) = \int_{x_1=-\infty}^{\infty} \int_{x_2=-\infty}^{\infty} w(\mathbf{x}) \check{Y}_j(\mathbf{x}) \left(x_2 \frac{\partial \check{Y}_k(\mathbf{x})}{\partial x_1} - x_1 \frac{\partial \check{Y}_k(\mathbf{x})}{\partial x_2} \right) dx_2 dx_1. \quad (20)$$

With the above precomputed integrals, the computation of the components of the gradient can be regarded as a the following bilinear form:

$$\mathbf{g}(\mathbf{a}, \tilde{\mathbf{a}}) := \frac{\partial Q}{\partial \mathbf{d}} = \begin{pmatrix} \sum_j \sum_k (a_j - \tilde{a}_j) \cdot a_k \cdot u_1(j, k) \\ \sum_j \sum_k (a_j - \tilde{a}_j) \cdot a_k \cdot u_2(j, k) \\ \sum_j \sum_k (a_j - \tilde{a}_j) \cdot a_k \cdot u_3(j, k) \end{pmatrix} \quad (21)$$

Note that the negative gradient $-\mathbf{g}$ represents necessary pose change of the *camera*. This direction $-\mathbf{g}$ has to be transformed to the coordinate system of the vehicle in order to steer the robot in the desired direction.

4 Experiments

We use a complex office environment created using BLENDER [3] in order to provide an experimental area for a simulated robot [5]. The modeled camera yields photo-realistic wide-angle images with a field of view of approx. 172.5° which can be projected onto a hemisphere. This hemispherical signal is extended to a full spherical signal by suitable reflection at the equator and is then being expanded in spherical harmonics (Appendix A). The following experiments are carried out with a low order approximation ($\ell = 4$, i. e. 15 contributing basis functions.).

Differential Translation Estimation. In our first experiment, we demonstrate and verify the behavior of the gradient $-\mathbf{g}$. We have rendered a grid of images around two target locations A and B. Ideally, from each grid point, the vector should point towards the center. The results are shown in Fig. 4. The quality of the gradient estimate depends on the geometric properties of the room.

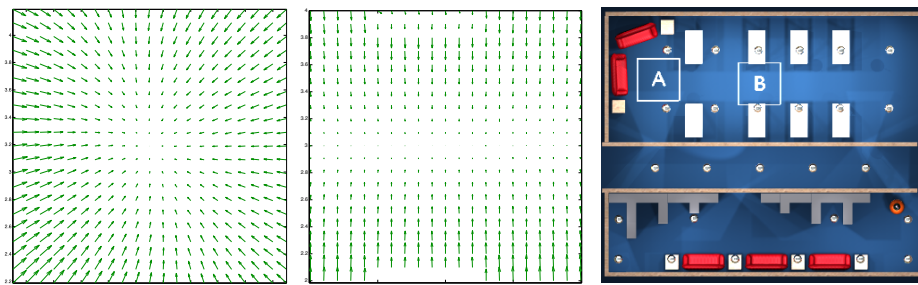


Fig. 4. The vector plots show the estimated direction towards the destination at the center. The grid extends 1 m in each direction at a spacing of 0.1 m. Note that even objects very close to the camera (area A, left plot) do not disturb the estimate, whereas directions of preference may deteriorate the results (area B, right plot).

Differential Pose Tracking. In this experiment we demonstrate the process of differential pose tracking based on iteratively applying the results of our different pose change model to views obtained at successive locations.

We begin with two initial views: The omnidirectional image at the destination pose \mathbf{p}_d and at the initial current pose \mathbf{p}_c are rendered, from those images we obtain the corresponding view descriptors $\tilde{\mathbf{a}}$ and \mathbf{a} . After this, the following steps are iterated:

From the images the gradient $-\mathbf{g}(\mathbf{a}, \tilde{\mathbf{a}})$ is computed and $-\mathbf{g}/\|\mathbf{g}\|_2 \cdot s$ is added to the current pose \mathbf{p}_c , where s denotes a translatory step size. The view at this new pose \mathbf{p}_c is rendered and the view descriptor \mathbf{a} is updated accordingly.

The iteration stops when $\|\mathbf{a} - \tilde{\mathbf{a}}\|_2^2$ falls below a threshold value. Note that for differential pose tracking, also the rotational component of $-\mathbf{g}$ is considered (Fig. 5, left).

A Round-Trip on Optical Rails. In the previous experiments we have shown that the concept of differential pose tracking can serve as a reliable method of approaching a single destination from within a local vicinity. The implementation of Optical Rails is possible by extending the destination to a *sequence of views* approached one after the other. To that end, a criterion is required to decide whether we have reached a destination so that we may proceed to the next destination (*waypoint handover*). Different measures have been investigated, including thresholding on similarity, a quotient of dissimilarities of the current view to the destination vs. next destination, and finally a criterion that meets

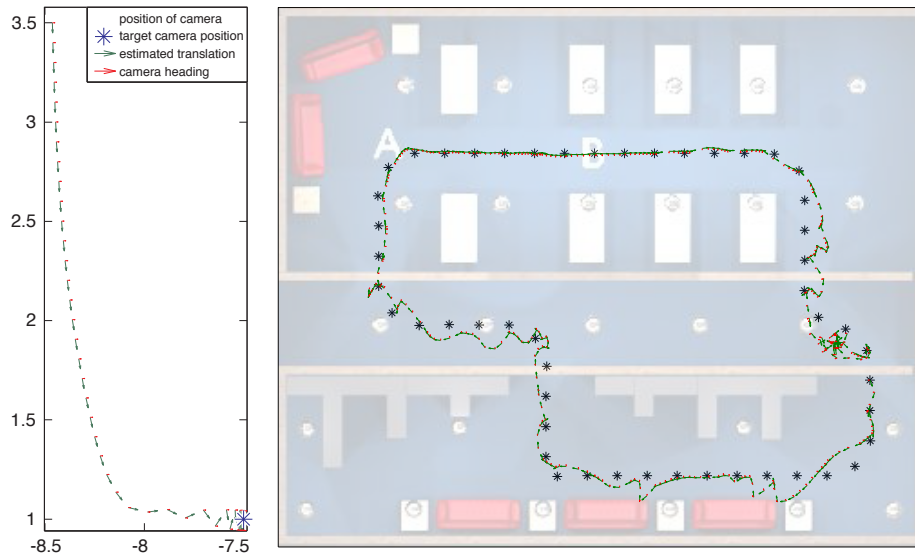


Fig. 5. Left: Visual pose tracking from initial pose (top) to target position (lower right, started from approx. 4 m distance, detail plot). Right: Round-trip using Optical Rails.

out needs best. When steering from a current view \mathbf{a} towards a destination view $\tilde{\mathbf{a}}$ with the predecessor destination view $\tilde{\mathbf{a}}'$, our criterion for advancing towards the next destination view $\tilde{\mathbf{a}}''$ is

$$\kappa > \frac{\|\mathbf{a} - \tilde{\mathbf{a}}\|_2}{\|\tilde{\mathbf{a}} - \tilde{\mathbf{a}}'\|_2 \cdot \|\mathbf{g}(\tilde{\mathbf{a}}', \tilde{\mathbf{a}})\|_2}, \quad (22)$$

where we call κ a threshold of *successive relative similarity*. This similarity measure has proven to be sufficiently robust for handover in most situations.

In Fig. 5 we show a round-trip through our virtual office environment using a set of equidistant waypoints. Due to the spatial characteristics in the different areas of the environment, waypoint handover is difficult in some cases (note the lower part of Fig. 5). Premature waypoint handover can cause the robot to leave the Optical Rail. However, convergence towards the proper destination is usually retained and should improve further if we make use of a vehicle model and an adaptive waypoint selection scheme [1]. In case of occlusions – e. g. caused by doors – a systematic waypoint placement would yield in a local augmentation of the Optical Rail with additional reference views providing greater stability.

5 Outlook

We have shown that Optical Rails is a very advantageous concept for robot view-based navigation without geometry, and verified this claim in a realistic simulated environment. The expansion of the image signal into basis functions enables us to incorporate precomputed differential expressions of the basis functions. This yields a computationally inexpensive way of obtaining steering information towards a destination.

The next obvious step to perform is to port this approach to a real robot. The hardware requirements for Optical Rails are quite modest: Due to the low-frequency character of the used omnidirectional image representation, no high-resolution camera or high-quality lens system are required. It has already been verified in our previous work that even a low-cost camera system is sufficient for omnidirectional vision using spherical harmonics. As there are no markers or scanning devices involved, the Optical Rails approach imposes only very few requirements on the structure of the environment. Other issues of practical importance have been considered in our investigations, but are not discussed here: efficient ways of achieving illumination invariance have been investigated thoroughly [6], and the robustness of waypoint handover, and handling of occlusions are subjects currently under investigation. All experimental results obtained so far indicate that Optical Rails is truly a versatile and powerful navigation concept for robots in a known environment.

A Expansion in Basis Functions, Spherical Harmonics

A very compact representation of views can be obtained by expanding the spherical image signal $s(\theta, \phi)$ in orthonormal basis functions. The natural choice for

spherical basis functions are *spherical harmonics*. The complex-valued spherical harmonics $Y_{\ell m}(\theta, \phi)$ are defined as

$$Y_{\ell m}(\theta, \phi) = \frac{1}{\sqrt{2\pi}} \cdot N_{\ell m} \cdot P_{\ell m}(\cos \theta) \cdot e^{im\phi} \quad N_{\ell m} = \sqrt{\frac{2\ell+1}{2} \frac{(\ell-|m|)!}{(\ell+|m|)!}}, \ell \in \mathbb{N}_0, m \in \mathbb{Z} \quad (23)$$

with $P_{\ell m}(x)$ the Associated Legendre Polynomials [12] and $e^{im\phi}$ being a complex-valued phase term. ℓ ($\ell > 0$) is called *order* and m ($m = -\ell.. +\ell$) is called *quantum number* for each ℓ . To approximate a signal $s(\theta, \phi)$, i. e.

$$s(\theta, \phi) = \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} a_{\ell m} \cdot Y_{\ell m}(\theta, \phi), \quad (24)$$

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

$$a_{\ell m} = \int_0^{2\pi} \int_0^{\pi} s(\theta, \phi) \cdot \overline{Y_{\ell m}(\theta, \phi)} \cdot \sin \theta \, d\theta \, d\phi. \quad (25)$$

As the image signals are real-valued, it is sufficient to use real-valued spherical harmonics which can be obtained from the complex-valued ones by combination of complex conjugate functions. However, for real-valued signals the coefficient vectors for both representations can be easily converted into each other [10].

References

1. A. Argyros, K. Bekris, S. Orphanoudakis, and Lydia E. Kavraki. Robot homing by exploiting panoramic vision. *Autonomous Robots*, 19(1):7–25, 2005.
2. Ronen Basri, Ehud Rivlin, and Ilan Shimshoni. Visual homing: Surfing on the epipoles. *International Journal of Computer Vision*, 33(2):117–137, 1999.
3. The Blender Foundation. Blender. Website www.blender.org, 2007.
4. Matthias O. Franz, Bernhard Schölkopf, H. A. Mallot, and Heinrich H. Bülthoff. Where did I take that snapshot? Scene-based homing by image matching. *Biological Cybernetics*, 79:191–202, 1998.
5. Holger Friedrich, David Dederscheck, Kai Krajsek, and Rudolf Mester. View-based robot localization using spherical harmonics: Concept and first experimental results. In *DAGM 2007*. Springer Verlag, 2007.
6. Holger Friedrich, David Dederscheck, Martin Mutz, and Rudolf Mester. View-based robot localization using illumination-invariant Spherical Harmonics descriptors. In *VISAPP 2008*, 2008.
7. B.D. Lucas and Takeo Kanade. An iterative image registration technique with an application to stereo vision. In *IJCAI81*, pages 674–679, 1981.
8. Ameesh Makadia, Christopher Geyer, and Kostas Daniilidis. Radon-based structure from motion without correspondences. In *Proc. CVPR*, 2005.
9. Ralf Möller, Andrew Vardy, Sven Kreft, and Sebastian Ruwisch. Visual homing in environments with anisotropic landmark distribution. *Autonomous Robots*, 23(3):231–245, 2007.
10. Peter Nillius. *Image Analysis using the Physics of Light Scattering*. PhD thesis, Royal Institute of Technology (KTH), Stockholm, Sweden, 2004.
11. W. Stürzl and H. A. Mallot. Efficient visual homing based on Fourier transformed panoramic images. *Robotics and Autonomous Systems*, 54:300–313, 2006.
12. Eric W. Weisstein. Legendre polynomial. A Wolfram Web Resource. <http://mathworld.wolfram.com/LegendrePolynomial.html>, 2007.