

Slicing the View: Occlusion-Aware View-Based Robot Navigation

David Dederscheck, Martin Zahn, Holger Friedrich, and Rudolf Mester

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

Abstract. Optical Rails [1] is a purely view-based method for autonomous track following with a mobile robot, based upon compact omnidirectional view descriptors using basis functions on the sphere. We address the most prominent points of criticism towards holistic methods for robot navigation: Dealing with occlusions and varying illumination. This is accomplished by slicing the omnidirectional view into segments, enabling dynamic visual fields capable of masking out occlusions while preserving proven, efficient paradigms for holistic view comparison and steering.

1 Introduction: Beyond The View-Based Approach

In the field of visual robot navigation, view-based, or ‘holistic’ approaches have recently become increasingly popular: They replace the tedious task of matching prominent points or performing accurate geometric measurements by a comparison of entire images in the spectral domain, utilizing the possibilities of omnidirectional sensors, which provide a large rotation-invariant visual field. The vulnerability of holistic methods to occlusions of the view, or local changes in illumination, is commonly considered a major obstacle for their successful deployment. Dederscheck et. al. [1] have shown that illumination-invariant view-based navigation is possible by using vector-valued image representations. Dynamic occlusions pose a different challenge: They require *adaptive* visual fields.

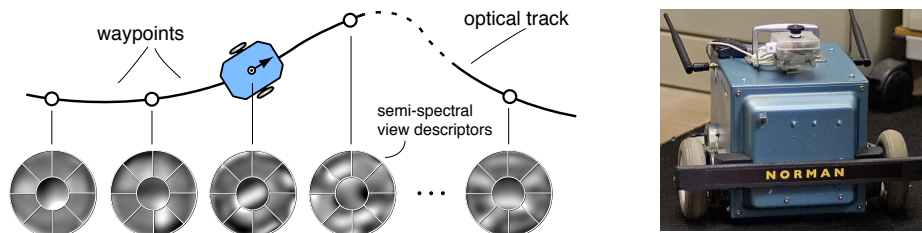


Fig. 1. *Left:* Concept of view-based robot navigation using segmented, semi-spectral views. *Right:* Mobile robot vehicle with upward-facing low-cost omnidirectional camera.

In the present paper we propose a simple solution for this task: We slice an omnidirectional view into multiple segments; while the spatial arrangement of the slices is known, each segment is represented in the spectral domain, i. e. by the coefficients of an expansion into low-frequency basis functions on the sphere. Hence, we denote the proposed hybrid concept as a *semi-spectral* view representation, extending the track following approach of Dederscheck et. al. (*Optical Rails*) towards occlusion-awareness: When comparing views stored in a database to newly seen images, only those view segments free of occlusions are regarded. At the same time, handling of illumination changes is facilitated, enabling segment-wise compensation of lighting conditions.

Semi-spectral views can efficiently be recombined (Sec. 2.1) yielding the coefficients of expansion of a single hemispherical view. Thus, our new approach easily connects to existing methods harnessing the traditional virtues of spectral representations, such as efficient rotation estimation.

The Navigation Concept. A path to be followed is represented as a sparse sequence of omnidirectional *reference views* serving as waypoints during track following. Instead of memorizing the shape of the trajectory, steering is purely based on gradient descent in view dissimilarities, which is performed efficiently without any operations in image space. A similar navigation approach has been proposed by Matsumoto et. al. [2] (omnidirectional view sequence). Both approaches can be regarded as topological navigation methods, related to biologically inspired approaches (e. g. Möller and Vardy [3]) based on the *snapshot theory* proposed by Cartwright and Collett [4]. In contrast to the traditional feature-based or landmark-based approaches (e. g. Shah and Aggarwal [5]) we present a *purely view-based* approach not relying on geometry.

2 The Semi-Spectral View Representation

Expansion of Spherical Signals and Visual Fields. As in Dederscheck et. al. [1], the representation of omnidirectional views is based upon expansion of image signals on the sphere S^2 into real-valued spherical harmonics, which form an orthonormal basis on the sphere.

We define spherical coordinates $\boldsymbol{\eta} = (\theta, \phi)^T \in S^2$, where $\theta \in [0, \pi]$ denotes the angle of *colatitude* and $\phi \in [0, 2\pi)$ denotes the angle of *longitude*. Image signals on the sphere are denoted as $s(\boldsymbol{\eta})$, and the utilized basis functions as Y_j , $j \in \{1, \dots, M\}$. A truncated expansion of a spherical signal $s(\boldsymbol{\eta})$ is given by

$$s(\boldsymbol{\eta}) = \sum_{j=0}^M a_j \cdot Y_j(\boldsymbol{\eta}). \quad (1)$$

The weights a_j of the expansion form a *coefficient vector* $\mathbf{a} := (a_1, \dots, a_M)^T$.

The *visual field* of an omnidirectional sensor is defined as the subset G of the sphere S^2 where image data can be acquired; in our case of a webcam with a low-cost fisheye lens, an almost hemispherical field of view limited by a maximum angle θ_{\max} is given by $G := \{(\theta, \phi)^T \in S^2 \mid \theta < \theta_{\max}\}$. The visual field G is then

partitioned into segments H_i . Let I be an index set enumerating the segments; we define the partition P of the visual field G by

$$P := \{H_i \subseteq G \mid \bigcup_{i \in I} H_i = G, \forall i \neq j : H_i \cap H_j = \emptyset\}. \quad (2)$$

Each of the view segments H_i will be represented by a coefficient vector, forming the *semi-spectral* view representation for Optical Rails. The utilized segmentation scheme (Fig. 1) consists of a central area surrounded by a set of radial slices, corresponding to the characteristics of typical occlusions: each outer slice can consume a person standing close to the robot.

Mask functions. For any visual field $K \subseteq S^2$, we define a mask function $m_K : S^2 \mapsto \{0, 1\}$ which is one within K , zero otherwise. For m_G and m_{H_i} we obtain

$$\sum_{i \in I} m_{H_i}(\theta, \phi) = m_G(\theta, \phi). \quad (3)$$

The defined mask functions will be used throughout the following sections to describe areas of interest when comparing visual fields.

Semi-spectral view descriptors. In our approach each segment is treated as an individual signal $s(\boldsymbol{\eta})$ only valid within its respective visual field H_i . Correspondingly, a spherical harmonics expansion is provided for each segment H_i , the resulting coefficient vector is denoted as $\mathbf{a}^{(H_i)}$.

A stacked representation of the individual $\mathbf{a}^{(H_i)}$ forms the *semi-spectral view descriptor* $\mathbf{a}^{\otimes} = (\mathbf{a}^{(H_1)}; \dots; \mathbf{a}^{(H_9)})^T$ consisting of all segments: The expansion of a full visual field G into a semi-spectral view descriptor is simply defined as a weighted sum of the expansions of the individual segments

$$\hat{s}(\boldsymbol{\eta}) := \sum_{i \in I} m_{H_i}(\boldsymbol{\eta}) \cdot \sum_{j=1}^M a_j^{(H_i)} \cdot Y_j(\boldsymbol{\eta}), \quad (4)$$

i.e. weighting the individual expansion with the respective mask functions m_{H_i} ensures that each expansion only occupies its respective visual field.

Since spherical harmonics are orthonormal, the coefficients of expansion for a full spherical view are obtained as the inner product of the signal and the basis functions: $a_i = \langle s(\boldsymbol{\eta}), Y_i(\boldsymbol{\eta}) \rangle$. However, for signals defined on only part of the sphere this orthonormality is violated and the above result no longer applies.

In the next section, we derive the optimal expansion of arbitrary visual fields into the same set of basis functions. Finally, the expansion will be performed efficiently, directly using pixel vectors of input images without prior unwrapping.

2.1 Optimized Expansion of Individual Visual Fields

Let $K \subseteq S^2$ and let $\mathbf{a}^{(K)}$ be a coefficient vector of the expansion of an image signal $s(\boldsymbol{\eta})$ in the visual field denoted by K . For such $\mathbf{a}^{(K)}$ we require that it correctly represents the original signal within K , whereas outside the expansion may assume arbitrary values. The desired view descriptor can be determined using least-squares minimization of the error of approximation within area K .

This approach allows optimal expansion of arbitrary visual fields, where the orthogonality of the basis is violated, such as small segments of the sphere.

Let $s(\boldsymbol{\eta})$ be an image signal to be approximated in region K ; the expanded image signal represented by the resulting view descriptor $\mathbf{a}^{(K)}$ is defined as

$$\hat{s}(\boldsymbol{\eta}) = \sum_{i=1}^M a_i^{(K)} \cdot Y_i(\boldsymbol{\eta}). \quad (5)$$

We define the squared residual error of the expansion in region K by

$$R = \int_{\boldsymbol{\eta} \in S^2} m_K(\boldsymbol{\eta}) \cdot (\hat{s}(\boldsymbol{\eta}) - s(\boldsymbol{\eta}))^2 d\boldsymbol{\eta} \quad (6)$$

where $d\boldsymbol{\eta}$ corresponds to $\sin \theta d\theta d\phi$. Since m_K is zero outside region K , the residual approximation error is only determined within region K .

Let us now regard the image signals in the cartesian coordinate frame of a planar camera sensor. We define the mapping \mathbf{h} between spherical coordinates and the image plane of the sensor by

$$\mathbf{h} : S^2 \mapsto \mathbb{R}^2, \quad \mathbf{x} = \mathbf{h}(\boldsymbol{\eta}) \quad \mathbf{h}^{-1} : \mathbb{R}^2 \mapsto S^2, \quad \boldsymbol{\eta} = \mathbf{h}^{-1}(\mathbf{x}). \quad (7)$$

Let the Jacobian of the projection $(\theta, \phi)^T = \mathbf{h}^{-1}(\mathbf{x})$ be denoted as $\mathbf{J}_{h^{-1}}(\mathbf{x})$. We define the weight function $w(\mathbf{x}) = \sin(\mathbf{h}^{-1}(\mathbf{x})^T \cdot (1, 0)^T) \cdot \det(\mathbf{J}_{h^{-1}}(\mathbf{x}))$. Now we can directly perform the transition to pixel coordinates of the input image of the omnidirectional camera. Let \mathbf{x}_k be the coordinate of the k -th pixel in a continuous planar image signal, where the pixels are numbered in an arbitrary order with indices $k \in \{1, \dots, N\}$. We define the vectors and matrices

$$\begin{aligned} \dot{\mathbf{y}}_i &= (-Y_i(\mathbf{h}^{-1}(\mathbf{x}_k)) -)^T, & \mathbf{w} &= (- (w(\mathbf{x}_k) \cdot A_p) -)^T, \\ \mathbf{b} &= (-s(\mathbf{h}^{-1}(\mathbf{x}_k)) -)^T, & \dot{\mathbf{m}}_K &= (-m_K(\mathbf{h}^{-1}(\mathbf{x}_k)) -)^T, \end{aligned} \quad (8)$$

$$\dot{\mathbf{Y}} = (\dot{\mathbf{y}}_1, \dots, \dot{\mathbf{y}}_M), \quad \mathbf{W}_K = \text{diag}\{\mathbf{w}\} \cdot \text{diag}\{\dot{\mathbf{m}}_K\}, \quad (9)$$

where A_p denotes the pixel area and \mathbf{W}_K is a diagonal weight matrix including $\dot{\mathbf{m}}_K$. Using the ‘pixel vectors’, R simply results in a quadratic form:

$$R = (\dot{\mathbf{Y}} \cdot \mathbf{a}^{(K)} - \mathbf{b})^T \cdot \mathbf{W}_K \cdot (\dot{\mathbf{Y}} \cdot \mathbf{a}^{(K)} - \mathbf{b}) \quad (10)$$

We minimize R by setting the derivative of R with respect to the expansion coefficients to be zero: $\mathbf{0} \stackrel{!}{=} \partial R / \partial \mathbf{a}^{(K)}$. This leads to the normal equation

$$(\mathbf{W}_K \cdot \dot{\mathbf{Y}})^T \cdot \mathbf{b} = \mathbf{G}_K \cdot \mathbf{a}^{(K)}, \quad \text{where} \quad \mathbf{G}_K := (\dot{\mathbf{Y}}^T \cdot \mathbf{W}_K \cdot \dot{\mathbf{Y}}) \quad (11)$$

is a Gramian matrix. The resulting coefficient vector $\mathbf{a}^{(K)}$ is denoted as the *optimal expansion* of an input image (pixel vector) \mathbf{b} in K :

$$\mathbf{a}^{(K)} = \mathbf{G}_K^{-1} \cdot (\mathbf{W}_K \cdot \dot{\mathbf{Y}})^T \cdot \mathbf{b}. \quad (12)$$

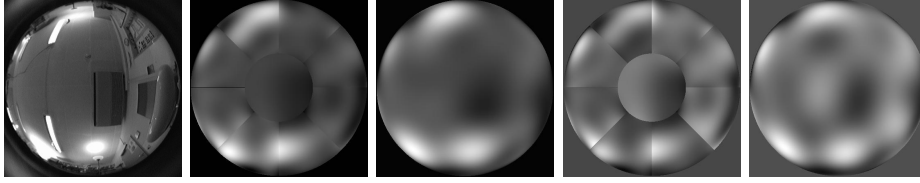


Fig. 2. *Left to right:* Input image, segmented image expansion, recombined view, segmented expansion with normalized illumination (cf. Sec. 3.3), and recombined view. Expansion into spherical harmonics up to order 9 (9×100 resp. 1×100 coefficients).

Recombining Slice View Descriptors. The semi-spectral representation consisting of spherical harmonics view expansions of multiple visual fields is not simply a composite form of representing multiple visual fields: The same representation as for a single spherical harmonics expansion of the entire visual field is easily obtained directly from the coefficient vectors.

Replacing K by region H_i in (11) and summarizing both sides leads to

$$\sum_{i \in I} (\mathbf{W}_{H_i} \cdot \hat{\mathbf{Y}})^T \cdot \mathbf{b} = \sum_{i \in I} (\hat{\mathbf{Y}}^T \cdot \mathbf{W}_{H_i} \cdot \hat{\mathbf{Y}}) \cdot \mathbf{a}^{(H_i)} \quad (13)$$

$$((\sum_{i \in I} \mathbf{W}_{H_i}) \cdot \hat{\mathbf{Y}})^T \cdot \mathbf{b} = \sum_{i \in I} \mathbf{G}_{H_i} \cdot \mathbf{a}^{(H_i)}; \quad (14)$$

by equation (3) and $\mathbf{W}_G = \text{diag}\{\mathbf{w}\} \cdot \text{diag}\{\hat{\mathbf{m}}_G\} = \text{diag}\{\mathbf{w}\} \cdot (\sum_{i \in I} \text{diag}\{\hat{\mathbf{m}}_{H_i}\}) = \sum_{i \in I} \text{diag}\{\mathbf{w}\} \cdot \text{diag}\{\hat{\mathbf{m}}_{H_i}\} = \sum_{i \in I} \mathbf{W}_{H_i}$ we obtain

$$(\mathbf{W}_G \cdot \hat{\mathbf{Y}})^T \cdot \mathbf{b} = \sum_{i \in I} \mathbf{G}_{H_i} \cdot \mathbf{a}^{(H_i)} \quad (15)$$

$$\mathbf{G}_G^{-1} \cdot (\mathbf{W}_G \cdot \hat{\mathbf{Y}})^T \cdot \mathbf{b} = \mathbf{G}_G^{-1} \cdot \sum_{i \in I} \mathbf{G}_{H_i} \cdot \mathbf{a}^{(H_i)}, \quad (16)$$

which according to (12) yields the view descriptor for the full visual field

$$\mathbf{a}^{(H)} = \mathbf{G}_H^{-1} \cdot \sum_{i \in I} \mathbf{G}_{H_i} \cdot \mathbf{a}^{(H_i)}. \quad (17)$$

In general, arbitrary visual fields composed of segments H_i can be recombined if only some of the available \mathbf{W}_{H_i} are summed up on both sides of (13).

3 Occlusion-Aware View Comparison and Steering

In Optical Rails, robot navigation for visual track following is based exclusively on the comparison of views, i. e. the currently observed view and a destination view chosen from a sequence of waypoints. First, a dissimilarity metric Q is defined, which then leads to a gradient of view dissimilarity with respect to robot motions, providing the steering direction towards a waypoint.

For both of these tasks, we introduce the extension to segmented visual fields, yielding figures of dissimilarity and steering vectors for each of the individual ‘slices’. Within the comparison result for the full view, each slice can be enabled or disabled individually, considering (both in the current and in the destination view) only those parts of the visual field which are free of occlusions.

We define a *clearance vector* $\mathbf{c} = (-c_i -)^T$ for all i in index set I , where $c_i = 1$ if the visual field H_i is free of occlusions, otherwise $c_i = 0$.

3.1 View Dissimilarities

We compare views using the total squared difference of image signals in the visual field G of the camera. Let $s(\boldsymbol{\eta})$ be the currently observed view, and let $\tilde{s}(\boldsymbol{\eta})$ be a destination view. We define a dissimilarity metric Q_K for an arbitrary visual field K (and analogously Q_G for the full visual field G):

$$Q_K(s(\boldsymbol{\eta}), \tilde{s}(\boldsymbol{\eta})) := \int_{\boldsymbol{\eta} \in S^2} m_K(\boldsymbol{\eta}) \cdot (s(\boldsymbol{\eta}) - \tilde{s}(\boldsymbol{\eta}))^2 d\boldsymbol{\eta}. \quad (18)$$

We now subdivide Q_G into the individual view segments according to (3) and introduce c_i as a weight to each summand to include the i -th segment only if it is free of occlusions. We define the occlusion-aware dissimilarity metric

$$Q = \int_{\boldsymbol{\eta} \in S^2} \sum_{i \in I} c_i \cdot m_{H_i}(\boldsymbol{\eta}) \cdot (s(\boldsymbol{\eta}) - \tilde{s}(\boldsymbol{\eta}))^2 d\boldsymbol{\eta} = \sum_{i \in I} c_i \cdot Q_{H_i}(s(\boldsymbol{\eta}), \tilde{s}(\boldsymbol{\eta})). \quad (19)$$

To determine the individual Q_{H_i} , we substitute the two signals with their respective expansions into basis functions according to (5):

$$\begin{aligned} Q_{H_i} &= \int_{\boldsymbol{\eta} \in S^2} m_{H_i}(\boldsymbol{\eta}) \cdot \left[\left(\sum_{k=1}^M a_k^{(H_i)} \cdot Y_k(\boldsymbol{\eta}) \right) - \left(\sum_{k=1}^M \tilde{a}_k^{(H_i)} \cdot Y_k(\boldsymbol{\eta}) \right) \right]^2 d\boldsymbol{\eta} \\ &= \sum_{k=1}^M \sum_{j=1}^M (a_k^{(H_i)} - \tilde{a}_k^{(H_i)}) \cdot (a_j^{(H_i)} - \tilde{a}_j^{(H_i)}) \cdot G_{kj}^{(H_i)}, \end{aligned} \quad (20)$$

where the integrals $G^{(H_i)} = \int_{\boldsymbol{\eta} \in S^2} m_{H_i}(\boldsymbol{\eta}) \cdot Y_k(\boldsymbol{\eta}) \cdot Y_j(\boldsymbol{\eta}) d\boldsymbol{\eta}$ correspond to the entries of a Gramian matrix \mathbf{G}_{H_i} as defined in Sec. 2.1.

Finally, the dissimilarity Q for segmented semi-spectral views is obtained as

$$Q(\mathbf{a}^{\otimes}, \tilde{\mathbf{a}}^{\otimes}) = \sum_{i \in I} c_i \cdot (\mathbf{a}^{(H_i)} - \tilde{\mathbf{a}}^{(H_i)})^T \cdot \mathbf{G}_{H_i} \cdot (\mathbf{a}^{(H_i)} - \tilde{\mathbf{a}}^{(H_i)}). \quad (21)$$

3.2 The Pose Change Gradient

To determine the steering direction towards a destination waypoint in Optical Rails, gradient descent in dissimilarities Q of two omnidirectional views is performed, with respect to the pose change of the robot. Let $s(\boldsymbol{\eta})$ be the currently acquired view and let $\tilde{s}(\boldsymbol{\eta})$ be the destination view. Let $\mathbf{p} = (x_1, x_2, p_\varphi)^T$ be the local pose (translation, rotation) of the robot, $\mathbf{p}_0 = \mathbf{0}$ denoting the current pose.

We define the gradient $\mathbf{g} = \partial Q / \partial \mathbf{p} \big|_{\mathbf{p}=\mathbf{p}_0}$ which points into the direction of robot pose change leading to the steepest change of the view dissimilarity Q ; hence, we denote \mathbf{g} as the *pose change gradient*. If the original pose \mathbf{p}_0 is situated in the vicinity of the destination waypoint, motions in the direction $-\mathbf{g}$ lead towards that waypoint.

By predicting the changes of the current view $s(\boldsymbol{\eta})$ upon small motions of the robot, we can determine \mathbf{g} analytically. For that purpose the environment is modeled as a hollow semisphere with the robot situated in the center (Fig. 3). This model is closely related to assumptions common also to many landmark-based navigation approaches and view-based visual homing (Möller et al. [3]).

We define a displacement transform \mathbf{f} by $\boldsymbol{\eta}' = \mathbf{f}(\boldsymbol{\eta}, \mathbf{p})$, such that if $s(\boldsymbol{\eta})$ denotes the current view at pose \mathbf{p}_0 , the view predicted after performing a pose

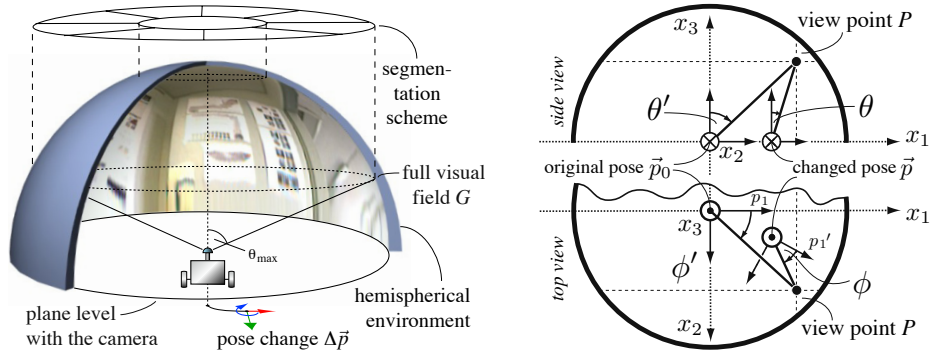


Fig. 3. Predicting view changes due to local robot motions: The hemispherical environment model provides a reasonable depth assumption for a variety of indoor scenarios.

change from \mathbf{p}_0 to \mathbf{p} is obtained as $s(\mathbf{f}(\boldsymbol{\eta}, \mathbf{p}))$. The transform \mathbf{f} is obtained by projecting a view point P (3D coordinates of the environment) seen at coordinates $\boldsymbol{\eta} = (\theta, \phi)^T$ from the changed pose \mathbf{p} to the camera center at the original pose \mathbf{p}_0 . The resulting view angles yield the spherical coordinates $\boldsymbol{\eta}' = (\theta', \phi')^T$ at which P is seen from \mathbf{p}_0 (Fig. 3).

Occlusion-aware pose change gradient for semi-spectral view. Corresponding to the definition of Q in (19), the occlusion-aware pose change gradient \mathbf{g} is obtained as the sum of pose change gradients \mathbf{g}_{H_i} for the individual view segments:

$$\mathbf{g} = \frac{\partial Q}{\partial \mathbf{p}} \Big|_{\mathbf{p}=\mathbf{p}_0} = \frac{\partial}{\partial \mathbf{p}} \sum_{i \in I} c_i \cdot Q_{H_i} \Big|_{\mathbf{p}=\mathbf{p}_0} = \sum_{i \in I} c_i \cdot \mathbf{g}_{H_i}, \quad \mathbf{g}_{H_i} := \frac{\partial Q_{H_i}}{\partial \mathbf{p}} \Big|_{\mathbf{p}=\mathbf{p}_0} \quad (22)$$

Computation of the individual \mathbf{g}_{H_i} is performed using the generalized chain rule:

$$\begin{aligned} \mathbf{g}_{H_i} &= \frac{\partial Q_{H_i}}{\partial \mathbf{p}} \Big|_{\mathbf{p}=\mathbf{p}_0} = \int_{\boldsymbol{\eta} \in S^2} 2 \cdot m_{H_i}(\boldsymbol{\eta}) \cdot [s(\mathbf{f}(\boldsymbol{\eta}, \mathbf{p})) - \tilde{s}(\boldsymbol{\eta})] \cdot \left[\frac{\partial s(\mathbf{f}(\boldsymbol{\eta}, \mathbf{p}))}{\partial \mathbf{p}} - \frac{\partial \tilde{s}(\boldsymbol{\eta})}{\partial \mathbf{p}} \right] \Big|_{\mathbf{p}=\mathbf{p}_0} d\boldsymbol{\eta} \\ &= \int_{\boldsymbol{\eta} \in S^2} 2 \cdot m_{H_i}(\boldsymbol{\eta}) \cdot [s(\boldsymbol{\eta}) - \tilde{s}(\boldsymbol{\eta})] \cdot \frac{\partial \mathbf{f}(\boldsymbol{\eta}, \mathbf{p})^T}{\partial \mathbf{p}} \cdot \frac{\partial s(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}} \Big|_{\mathbf{p}=\mathbf{p}_0} d\boldsymbol{\eta}. \end{aligned} \quad (23)$$

Since views to be compared are represented by view descriptors consisting of spectral coefficients of the original signal for each segment H_i , we replace signals $s(\boldsymbol{\eta})$, $\tilde{s}(\boldsymbol{\eta})$ by their respective expansions according to (4).

Due to the linearity of integration we can determine a factor which is independent of the signal observed in the respective visual field:

$$\mathbf{u}_p^{H_i}(j, k) := 2 \cdot \int_{S^2} m_{H_i}(\boldsymbol{\eta}) \cdot Y_j(\boldsymbol{\eta}) \cdot \frac{\partial \mathbf{f}(\boldsymbol{\eta}, \mathbf{p})^T}{\partial \mathbf{p}} \cdot \frac{\partial Y_k(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}} \Big|_{\mathbf{p}=\mathbf{p}_0} d\boldsymbol{\eta}. \quad (24)$$

The steering gradient \mathbf{g}_{H_i} of the comparison in segment H_i is thus yielded as

$$\mathbf{g}_{H_i}(\mathbf{a}^{(H_i)}, \tilde{\mathbf{a}}^{(H_i)}) = \sum_{j=1}^M \sum_{k=1}^M [a_j^{(H_i)} - \tilde{a}_j^{(H_i)}] \cdot a_k^{(H_i)} \cdot \mathbf{u}_p^{H_i}(j, k), \quad (25)$$

where the entries $\mathbf{u}_p(j, k)$ are precomputed expressions common to each visual field H_i . For the occlusion-aware pose change gradient \mathbf{g} we finally obtain

$$\mathbf{g}(\mathbf{a}^{\otimes}, \tilde{\mathbf{a}}^{\otimes}) = \sum_{i \in I} c_i \cdot \sum_{j=1}^M \sum_{k=1}^M [a_j^{(H_i)} - \tilde{a}_j^{(H_i)}] \cdot a_k^{(H_i)} \cdot \mathbf{u}_p^{H_i}(j, k). \quad (26)$$

3.3 Illumination Invariance

With the semi-spectral signal representation in Optical Rails, illumination invariance of views can be achieved simply by a segment-wise preprocessing step: Bias removal and subsequent normalization of the image signal results in

$$\mathbf{b}_{H_i}^{\text{norm}} = \frac{\mathbf{b}_0}{\sqrt{\mathbf{b}_0^T \cdot \text{diag}\{\hat{\mathbf{m}}_{H_i}\} \cdot \mathbf{b}_0}} \quad \text{with bias removed signal } \mathbf{b}_0 := \mathbf{b} - \frac{\mathbf{b}^T \cdot \hat{\mathbf{m}}_{H_i}}{\hat{\mathbf{m}}_{H_i}^T \cdot \hat{\mathbf{m}}_{H_i}} \quad (27)$$

for each segment $\mathbf{a}^{(H_i)}$, and $\mathbf{b}_{H_i}^{\text{norm}}$ replaces \mathbf{b} in the subsequent expansion. Thus, the comparison of the resulting views is resilient also to inhomogeneous changes in illumination, since the brightness is normalized independently in the view segments. Hence, occlusions in single segments also do not affect illumination compensation in the remaining view segments.

Compared to global illumination compensation with a linear model, our experiments have shown the normalization of individual segments to be favorable, since the magnitude of dissimilarities and the pose change gradient depends less on the distribution of illumination patterns and the large-scale appearance of views: A certain degree of scale invariance of these measures is achieved.

4 Experiments

In our experiments, we show that robust robot navigation with tolerance to occlusions is feasible by selectively masking out the affected visual fields. First, we compare the behavior of the pose change gradient when steering towards an occluded destination view; finally, a comparison of a challenging track following scenario both using full and partially masked out visual fields is shown.

Pose Change Gradient: Effects of Occlusions and Mitigation. Our first experiment provides a systematic comparison of the pose change gradient \mathbf{g} used for steering the robot (i) without occlusions, (ii) with occluded views and changed illumination, and (iii) with the affected view segments masked out in the comparison. Fig. 4 shows steering vectors for these three cases, which have been obtained by comparing views from a regularly spaced grid to a destination view.

Results show severe deviations of the steering direction for the occluded case, leading away from the actual destination. Excluding the affected views mitigates this effect, leading to a less pronounced, yet essentially correct pose gradient.

Track Following Experiments. Visual track following in Optical Rails consists of a recording phase, where the robot is manually steered along the desired route while semi-spectral views are acquired in rapid succession, and a driving phase, where the robot autonomously follows the route based upon this view sequence. Both phases have been implemented in real time (≥ 15 frames/sec.) on different mobile robot vehicles running MATLAB for all image processing tasks using the precomputed entities defined in Sec. 2.1 and 3.2.

Prior to track following, a set of distinctive views are automatically selected by filtering the teach-in sequence with a threshold criterion; the visual dissimilarity between adjacent waypoints is required to monotonically decrease in the

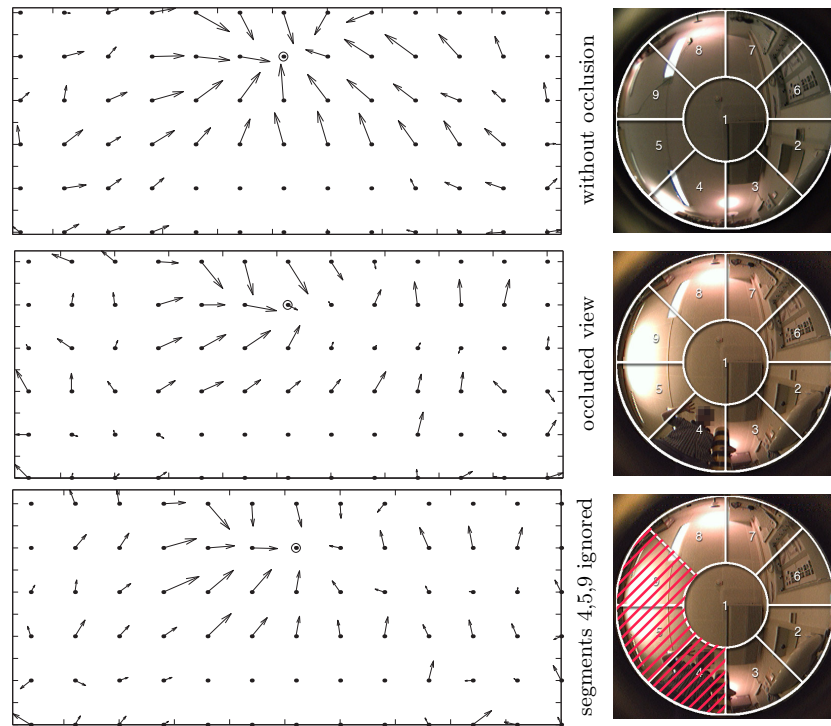


Fig. 4. Results of the pose change gradient using a view grid recorded in our lab, demonstrating the effects of occlusions and severe illumination changes and their mitigation by masking out the affected segments. Arrows show the translation components of the gradient \mathbf{g} , which ideally should point towards the true destination (circle mark).

driving direction, yielding only waypoints within a suitable radius of convergence. While approaching a waypoint by successively computing the pose change gradient, arrival detection (waypoint handover) is performed by comparing the current view to a set of adjacent waypoints from the filtered view sequence.

In Fig. 5, we show a particularly challenging scenario of visual track following in our laboratory environment: An arbitrary path originates in a small closet with a narrow door and ends below a desk, where the complete view of the ceiling has been covered. Despite the resulting dramatic variations in brightness and appearance, the original trajectory could always be reproduced accurately. This specifically includes the part of the route while ‘parking’ under the desk, where successful navigation was possible only by the introduction of segment-wise normalized illumination.

Finally, to demonstrate that in case of occlusions sufficiently accurate steering would still be possible, two adjacent view segments at the front were masked out, simulating a typical occlusion by a person walking in front: Only slight deterioration of the accuracy of the resulting trajectory occurred.

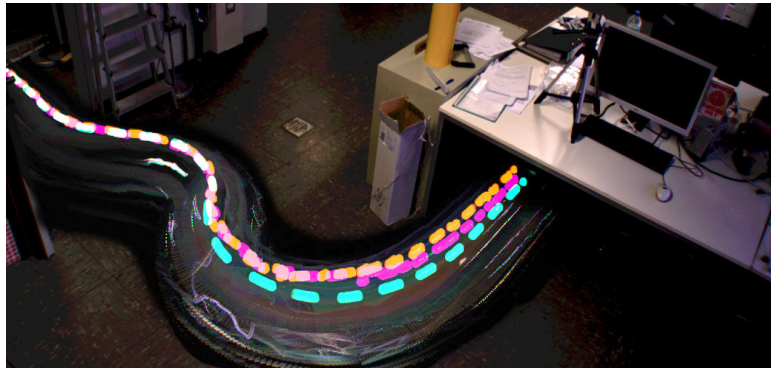


Fig. 5. Results of track following, documented by beacon (blinking at 1 Hz): *Cyan line*: teach in track, *orange line*: reverse track following, full view, *magenta line*: reverse track following with two consecutive segments (at front) masked out, simulating an occlusion.

5 Conclusion

In the present paper we have successfully shown the extension of the traditional view-based navigation paradigm to a segmented hybrid approach with adaptive visual fields. This enables selective masking of occluded areas and largely simplifies the task of attaining illumination invariance.

Obviously, the detection of occlusions and obstacles is a formidable problem in itself; it could be solved by tentatively assuming each segment to be affected by an occlusion, and performing the comparison only with the remaining segments as described in Sec. 3.1, leading to a threshold decision. The presented recombination of views Sec. 2.1 enables application of the method to purely holistic comparisons such as rotation estimation and rotation-invariant matching, which is required for more complex navigation tasks involving self-localization.

In the light of the obtained results and the highly versatile yet simple nature of the approach, view-based navigation appears to us as an attractive field for future challenges.

References

1. Dederscheck, D., Friedrich, H., Lenhart, C., Zahn, M., Mester, R.: ‘Featuring’ Optical Rails: View-based robot guidance using orientation features on the sphere. In: OMNIVIS workshop, ICCV, Kyoto, Japan, IEEE (2009)
2. Matsumoto, Y., Inaba, M., Inoue, H.: View-based navigation using an omniview sequence in a corridor environment. *Machine Vision and Applications* **14** (2003)
3. Möller, R., Vardy, A.: Local visual homing by matched-filter descent in image distances. *Biological Cybernetics* **95** (2006) 413–430 ISSN 0340-1200.
4. Cartwright, B., Collett, T.: Landmark maps for honeybees. *Biological Cybernetics* **57** (1987) 85–93
5. Shah, S., Aggarwal, J.K.: Mobile robot navigation and scene modeling using stereo fish-eye lens system. *Machine Vision and Applications* **10** (1997) 159–173