Localization by Spatial Context Analysis\textsuperscript{1)}

Lucas Paletta\textsuperscript{†}, Simone Frintrop\textsuperscript{‡†}, and Joachim Hertzberg\textsuperscript{‡†}

\textsuperscript{†}Joanneum Research
Institute of Digital Image Processing
A-8010 Graz, Austria
lucas.paletta@joanneum.ac.at

\textsuperscript{‡†}GMD-NIL. Research Center for Information Technology
Institute for Autonomous intelligent Systems
D-53754 Sankt Augustin, Germany
\{simone.frintrop,joachim.hertzberg\}@gmd.de

Abstract:
This work presents the concept to recover and utilize the visual context in panoramic images. Omnidirectional imaging has become recently an efficient basis for robot navigation. The proposed Bayesian reasoning over local image appearances enables to reject false hypotheses which do not fit the structural constraints in corresponding feature trajectories. The methodology is proved with real image data from an office robot to dramatically increase the localization performance in the presence of severe occlusion effects, particularly in noisy environments, and to recover rotational information on the fly.

1 Introduction

Navigation is a fundamental task for autonomous mobile robots [9]. Particularly, it relies on robust localization methods, e.g. in the presence of occlusion and under noisy and uncertain environment conditions. The interpretation of a single sensor pattern always depends on the ambiguity of a specific view, the imprecision of the environment model, and the uncertainty in the image formation [4, 20]. Instead of generating arbitrarily complex classifiers to solve an a priori ill-posed problem, one may analyze the context within the sensor pattern which provides much more efficient cues for localization.

Localization methods are discriminated as geometric or topological. Geometric approaches require a comparison between the local geometry model extracted from the sensor information with the predetermined model of the entire environment. Typical geometry models are occupancy grids for sonar or laser based localization [15, 23], and models based on local features for visually guided navigation [2]. These models either lack complexity of information or are difficult to be built. In a topological approach, maps are represented by adjacency

\textsuperscript{1)} This work is funded by the European Commission’s TMR project VIRGO under grant number ERBFMRXCT960049, and the Austrian Competence Center for Vision Technologies ACV. This support and cooperation is gratefully acknowledged.
Figure 1: (a) Robot kurt-2 equipped with omnidirectional camera on top. (b) Topological map of the office environment (left) and grid for position measurements (right) in the corresponding navigation modules env1-env3.

graphs, with nodes denoting locations and arcs representing transitions between them [24, 25]. Memory-based methods [6, 1, 7, 12, 13] for localization are biologically inspired and do not need an explicit model of the world, but rather match the sensory input with the stored information in memory. Concurrently, progress on the mosaicing of manifolds [21] suggests even to represent entire environments of video shots in single static images for further analysis.

Omnidirectional imaging has become recently an efficient basis for robot navigation [1, 7, 24, 25, 8]. The sensor characteristics enable to provide information not only about a restricted field of view ahead but rather to integrate multi-directional appearances where nearby positions are strongly correlated [12, 8]. Recent advances in robust omnidirectional navigation have been achieved towards illumination invariance [25] and reasonable occlusion tolerance [8], respectively.

The original contribution of this work is to take advantage of the local context in omnidirectional sensing, by reasoning on the structure of ordered ‘uni-directional’ views in the panoramic image. One may understand panoramic images at certain environment positions in analogy to multi-view captures in object recognition. The proposed method for robot localization thus corresponds to the interpretation of probabilistic dependencies between local structures in 2D [22] and multiple views of 3D [20] objects, respectively. The probabilistic framework enables to take advantage of even partial evidence about a position and to reject false hypotheses which do not fit the structural constraints. These characteristics render it fairly insensitive to severe occlusion and noise in the visual input. Furthermore, as a consequence of the multi-view representation, the robot’s orientation is implicitly recovered by localization.
Each panoramic image is first partitioned into a fixed number of overlapping uni-directional camera views, i.e., appearance sectors. Distributions of sector images in eigenspace [16, 1] are then represented by mixture of Gaussians [3] to provide a posterior distribution over potential locations. The ambiguity in a local sector interpretation as well as occlusion defects are then resolved by Bayesian reasoning over the spatial context of the current position. Probabilistic navigation modules for localization in small environments are being considered integrated within a global topological navigation concept.

2 Localization by panoramic eigenimages

Complex environments are favourably modelled by topological representations, with nodes representing locations and arcs denoting adjacency of locations. Fig. 1(b) illustrates a topological map of a small test environment used in the experiments. Within each map module, locations are directly learned from images captured at robot runs. Image based navigation derives a sensor based model of the environment [6, 12], without the load of additional pre-processing and interpretation. Eigenspace analysis of the reference image memory [16, 11] enables significant reduction of the storage requirements and to focus on statistically relevant features. Panoramic eigenimages [1] result then from principal component analysis on omnidirectional sensor images.

The eigenspace of an image training set \( \mathbf{X} = \{ \mathbf{x}_1, \ldots, \mathbf{x}_N \} \) of size \( N \) is a low-dimensional representation of the original image space. It captures the maximum variations in the presented data set whereas distances are a measure of image correlation [16]. Recognition is supported by the property that close points in subspace correspond to similar appearances. The eigenspace is spanned by a selection of the \( \epsilon \) most prominent eigenvectors \( \mathbf{e}_i \) of the covariance matrix on \( \mathbf{XX}^T \). The projection of image \( \mathbf{x} \) into the eigenspace of dimension \( \epsilon \) is \( \mathbf{y}(\mathbf{x}) = (\mathbf{e}_1, \ldots, \mathbf{e}_\epsilon)^T \mathbf{x} \). For localization, samples \( \mathbf{x} \) are mapped into eigenspace to \( \mathbf{y}(\mathbf{x}) \), and assigned the label of the closest position representative \( \mathbf{y}_r \).

3 Extraction of local appearance sectors

Recent progress on the robustness of navigation has been made to enable the dealing with illumination variance [25] and occlusion patterns [8]. However, rotational information concerning the orientation of the observer with respect to the environment had either to be incorporated in the representation or to be recovered in the localization process. Matching of on-journey captured panoramic images to reference images of the environment model is not rotationally invariant since the orientation of the sensor determines the shift between the two images, therefore, matching requires a search which is linear in the number of the images and proportional to their horizontal resolution [12]. Alternatively, an approach called Zero Phase
Figure 2: The presented local window approach continually splits the panoramic image (bottom) into overlapping appearance sectors (top).

Figure 3: Feature trajectory of sector projections from images $x_1$, $x_2$ (Fig. 2) of nearby positions $\phi_1$ and $\phi_2$, illustrating the ambiguity in their appearance and feature trajectory in eigenspace, respectively.

Representation (ZPR) [18] finds a transformation that projects differently oriented panoramic images into one representative image, though with the lack of robustness to partial occlusions.

In contrast, local context in the omnidirectional image enables to recover in parallel the rotational and the positional information, being capable of handling both severe occlusions and changes in the robot’s orientation. Therefore, the original contribution of this work is to exploit the spatial relation between uni-directional views in panoramic images. Due to its wide visual field, occlusion of the entire panoramic view becomes very unlikely. Finding evidence for a position from subwindows of the original image, then reasoning over the local decisions, is the method followed in the sequel. Finding evidence for a position from subwindows of the original image, then reasoning over the local decisions, is the method followed in the sequel.
The panoramic image is first partitioned into a fixed number of overlapping uni-directional camera views (Fig. 2), i.e., local appearance sectors $s$ at $\phi(x_i) = \{s_{i,1}, \ldots, s_{i,\Sigma}\}$, for a partition into $j = 1..\Sigma$ sector parametrizations $\kappa_j$, starting at field angle $360^\circ/\Sigma$ and with predefined width. Eigenspace analysis is then performed on the set of sectors from images captured at all $N$ individual camera positions, i.e., on $X = \{s_{1,1}, \ldots, s_{N,\Sigma}\}$. To each position $\phi_i$, $i = 1..\Phi$, corresponds a parametric set of points in eigenspace which can be interpolated to a closed curve. Curves of neighboring positions in the environment may overlap due to the similar appearance of their respective sector projections (Fig. 3). Since appearance sectors convey less information than the global image itself, one expects a higher degree of ambiguity in this representation, but a higher robustness to occlusion, too.

4 Probabilistic localization under Bayesian context

In uncertain and noisy environments, localization on the basis of a crisp mapping from observations $y$ to positions $\phi_i$ becomes unreliable, and soft computing methods are required to quantify the ambiguity by means of beliefs for multiple position hypotheses [7, 23, 10].


4.1 Bayesian interpretation

Given the measurement about position $\phi_i$ and appearance sector $\kappa_j$, the likelihood of obtaining feature vector $y$ is denoted by $p(y|\phi_i, \kappa_j)$. The likelihood is estimated from a set of sample images with fixed $\phi_i$ and $\kappa_j$, capturing the inaccuracies such as moderate light or positioning variations. From the learned likelihoods one obtains then via Bayesian inversion

$$P(\phi_i, \kappa_j|y) = p(y|\phi_i, \kappa_j)P(\kappa_j|\phi_i)P(\phi_i)/p(y),$$

and a posterior estimate with respect to the position hypotheses $\phi_i$ is given by $P(\phi_i|y) = \Sigma_j P(\phi_i, \kappa_j|y)$.

In practice, the likelihoods are estimated only at selected values of $\kappa$, e.g. at most informative or periodic settings $\kappa_j \equiv k \times \Delta \kappa$ [4, 19]. The posterior estimates for intermediate values $\kappa$ are determined using a Gaussian mixture model [3], $P(\phi_i, \kappa_j|y) \approx \Sigma_g P(g) \lambda_{\mu_g, \sigma_g}(y)$, where $P(g)$ are the mixing coefficients and $\lambda_{\mu_g, \sigma_g}(y)$ are Gaussians described by parameters $\mu_g, \sigma_g$ which can be adjusted in a training session.
4.2 Bayesian decision fusion process

Multiple measurements enable to resolve ambiguities in the initial interpretation. In a most general framework on belief integration, the discrimination status is iteratively updated with newly upcoming evidence, fusing the posterior beliefs in the position hypotheses \( P(\phi_i|y) \) obtained from each single sector measurement \( s \) with new evidence [4, 19, 20].

**Spatial context** in the combination of sector evidences corresponds in general to the specific impact of shift actions \( a \equiv \Delta \kappa \) (Fig. 7) on the sector feature trajectory in eigenspace (Fig. 3). Introducing the representation of actions into Bayesian decision fusion [20] leads to

\[
P(\phi_i, \kappa_j | y_1, a, y_2) = \alpha P(\phi_i, \kappa_j | y_1, a) p(y_2 | \phi_i, \kappa_j, y_1, a),
\]

where \( \alpha = 1/p(y_2|y_1, a) \) is the normalizing factor. Spatial context is now exploited using the conditional term \( P(\phi_i, \kappa_j | y_1, a) \): The probability for observing appearance sector \( (\phi_i, \kappa_j) \) as a consequence of deterministic action \( a = \Delta \kappa \) must be identical to the probability of having measured at the action’s starting point before, i.e. at appearance sector view \( (\phi_i, \kappa_j - \Delta \kappa) \), thus

\[
P(\phi_i, \kappa_j | y_1, a) \equiv P(\phi_i, \kappa_j - \Delta \kappa | y_1).
\]

Furthermore, the probability density of \( y_2 \), given the knowledge of view \( (\phi_i, \kappa_j) \), is conditionally independent of previous observations and actions, and therefore \( p(y_2 | \phi_i, \kappa_j, y_1, a) = p(y_2 | \phi_i, \kappa_j) \). The recursive update rule for \( M \) conditionally dependent observations accordingly becomes

\[
P(\phi_i, \kappa_j | y_1, a_1, \ldots, a_{M-1}, y_M) = \alpha p(y_M | \phi_i, \kappa_j) P(\phi_i, \kappa_j - \Delta \kappa_{M-1} | y_1, a_1, \ldots, y_{M-1}),
\]

and the posterior distribution on position hypotheses, using \( Y_T^a \equiv \{y_1, a_1, \ldots, a_{M-1}, y_M\} \), is then given by

\[
P(\phi_i | Y_T^a) = \sum_j P(\phi_i, \kappa_j | Y_M^a).
\]

In practice, occlusions may induce ambiguity in single appearance sectors, but they barely represent sector trajectories comparable with the training structures. Analogously, localization at intergrid points and direction measurements are considered being resolvable by extracting structural constraints from fusion steps.

Once the appearance sector \( P(\phi_i, \kappa_j) \) is recovered from the positional measurement, the rotation \( \Delta \kappa \) between a stored training image \( x_s \) and a test sample \( x_t \) is easily determined by comparing the location of sector \( \kappa_{j,s} \) within \( x_s \) with the corresponding sector \( \kappa_{j,t} \) within \( x_t \).

The computational complexity for a single updating step depends crucially on the sector representation of the panoramic image. For \( \Sigma \) sectors, \( \Sigma (\Sigma - 1)/2 \) sector combinations are computed (each with \( \Sigma \Phi \) probability updates) to result in \( O(\Phi \Sigma^3) \), which enabled real-time processing in the experiments described in Section 5.
5 Experiments

The experiments were conducted in the hallway of the GMD-AIS department (Fig. 1(b)) with a Neuronics omnidirectional camera installed on top of the robot platform kurt-2 (Fig. 1(a)). Images were captured from a grid of sensing points each 50 cm apart, and assigned to the corresponding topological structure depicted in Fig. 1(b). Note that the robot was aligned to only one direction, since rotational information is implicitly encoded within the sequence of sectors.

The images of size $360 \times 200$ were split into 36 sectors of size $20 \times 200$, each covering $20^\circ$ of the panoramic field of view, and each within a translational shift of $10^\circ$ (Fig. 2). The images were normalized to $||x|| = 1$, to account for illumination variations - according to [16]. Principal component analysis was then applied to images of a corresponding topological module (env1-3) to determine the most prominent eigenvectors (Section 2), and each sector image was projected to the 10-dimensional eigenspace.

To test the navigation system, the training images (Fig 4) were occluded by images of an artificial image database (Fig. 5). The impact of occlusion effects was gradually controlled by the percentage of image content covered by artificial images (e.g., Fig. 7).

The likelihood $p(y|\phi, \kappa)$ (Eq. 4) of a test sector sample $y$, given field view $(\phi, \kappa)$, was modelled assuming Gaussian distributed error, with mean $\mu_{\phi, \kappa}$ and variance $\sigma_{\phi, \kappa}^2$ indicating the sector specific sample distribution. This distribution can also be estimated from illumination or positioning variances actually experienced at view $(\phi, \kappa)$ [19, 20].

Occlusion tolerance A first experiment compares the navigation performance of the traditional 'global window' approach - classifying the panoramic image - with the one obtained using 'local window' data - classifying the sector image. Both representations were evaluated by a nearest neighbor (NN) classifier, i.e., test samples $y_t$ were assigned the label of the training representative $y_s$ which is closest in eigenspace. Fig. 6(a) depicts the results w.r.t. recognition accuracy and various degrees of occlusion. While the performance of the 'global window' approach rigorously decreases with slight occlusion effects, the recognition of single sectors remains optimal until $70\%$ occlusion. Thereafter, performance decreases (with some perturbation due to the test set configuration) since every sector is affected by severe occlusion.

Robustness against noise The second experiment evaluates the robustness of the interpretation against Gaussian noise in the visual input. A probabilistic interpretation of single sectors (section 4.1) is compared with the localization using Bayesian context (section 4.2)
between 2 sectors (Fig. 6(b)). With increasing ratio $\sigma_n^2/\sigma_i^2$, where $\sigma_n^2$ denotes the noise and $\sigma_i^2$ the learned variance respectively, the advantage of the structural evaluation within the context based approach ('2-sector classification') strictly outperforms a single sector evaluation. '1-sector classification' substantially represents a NN classifier which fails with increasing noise.

**Implicit rotational information**  With the decision for a most likely sector in the training image set, one is able to recover the rotation of the robot w.r.t. the training image. Fig. 7 depicts the test image of 70% occlusion (top) and the recovered training image (bottom). The Bayesian context classifier determined sectors $\kappa_1$ and $\kappa_2$ as most discriminating sectors, which enables - under knowledge of the corresponding sectors ($\kappa_1',\kappa_2'$) in the train image - to recover the rotation $\Delta \kappa$ between these images.

The local sector navigation system enables real-time localization. The Bayesian update procedure for $\Sigma = 36$ sectors and $\Phi = 41$ positions required $\Phi \Sigma^2 (\Sigma - 1)/2 \approx 9.3 \times 10^8$ computations, which represented a tractable load for a 300 Mhz PC.
Figure 6: (a) Comparison of occlusion sensitivity between the global and the local window approach. (b) Comparison of noise tolerance between a 1-sector probabilistic interpretation and a 2-sector approach utilizing Bayesian context.

Figure 7: Test image at position $\phi_2$ with 70% occlusion (top) and recovered train image (bottom) plus corresponding rotation component $\Delta K$.

6 Conclusions

Spatial context is presented as an important cue for robot localization which is naturally derived from a sequence of appearance sectors within the panoramic image. The Bayesian framework for the decision fusion process enables to quantify the uncertainty in the discrimination of each position.

The results from navigation experiments using the office robot demonstrate that the Bayesian reasoning allows highly occlusion and noise tolerant localization, vigorously improving the navigation performances in comparison to previous approaches.

The localization system is considered to enable accurate visual navigation of autonomous
robots even at crowded places such as offices, factories, and urban environments. The detection of the occluding image regions may then provide a starting point to apply object recognition for further interpretation.

References


