

Unsupervised Learning of a Hierarchy of Topological Maps Using Omnidirectional Images

Aleš Štívec, Matjaž Jogan, Aleš Leonardis
*Visual Cognitive Systems Laboratory,
Faculty of Computer and Information Science,
University of Ljubljana,
Ljubljana, SI-1001, Slovenia,
{ales.stivec, matjaz.jogan, ales.leonardis}@fri.uni-lj.si*

This paper presents a novel appearance-based method for path-based map learning by a mobile robot equipped with an omnidirectional camera. In particular we focus on an unsupervised construction of topological maps, which provide an abstraction of the environment in terms of visual aspects. An unsupervised clustering algorithm is used to represent the images in multiple subspaces, forming thus a sensory grounded representation of the environment's appearance. By introducing transitional fields between clusters we are able to obtain a partitioning of the image set into distinctive *visual aspects*. By abstracting the low-level sensory data we are able to efficiently reconstruct the overall topological layout of the covered path. After the high level topology is estimated, we repeat the procedure on the level of visual aspects to obtain local topological maps. We demonstrate how the resulting representation can be used for modeling indoor and outdoor environments, how it successfully detects previously visited locations and how it can be used for the estimation of the current visual aspect and the retrieval of the relative position within the current visual aspect.

Keywords: Mobile robot; robot localization; topological mapping; hierarchical methods; unsupervised learning; visual learning; appearance-based recognition; omnidirectional vision; clustering.

1. Introduction

In this paper we present a hierarchical mapping method based on the appearance of a path covered by a mobile robot. By using *omnidirectional images*, which represent locations on the robot's path, and the orientations of these images as input, the method learns (in an unsupervised manner) a hierarchical topological model, which reflects the structure of the covered path at different levels of detail. The nodes at the top level of the topological map represent *visually distinctive places* of the environment, which we call *visual aspects*, while the nodes at the bottom level of the topological map represent specific locations where omnidirectional images were captured. At both levels the model represents the connectivity, the spatial arrangement and the information on heading directions between nodes. Because of its hierarchical nature, the model can be built efficiently even for large scale environments. Since the method is capable of loop closing, useful models can be built

from data obtained on looped circuits. Due to a multiple subspace representation of visual aspects, the representation is optimized in terms of reconstructive power and can be efficiently used in conjunction with appearance based recognition methods for visual localization.

Methods for building a representation of the robot’s environment can be roughly classified into three groups: metric, topological and hybrid methods. Metric maps, such as statistical grid maps [30], maps built by alignment of range scans [10, 24], and 3D feature maps, are viable for positioning, however they do not intrinsically support *reasoning* about places, topology, and related actions. In addition, metric representations are often not feasible for large unstructured environments due to their spatial and time complexities [42] and the inaccuracy of odometry and sensory readings in real-world scenarios. In contrast, topological maps [20, 34] model places and transitions between places in a loose layout and provide an efficient *cognitive* representation that can immediately be used for planning. In mobile robotics, the idea of a topological map as an abstraction of an environment is perhaps most concisely stated by Kuipers and Byun in [20], where a continuous environment is modeled as a set of distinctive states which represent perceptually or semantically significant aspects of the environment. Recently, hybrid approaches [8, 34, 37, 41] have been used to combine advantages of both metric and topological approaches.

The hierarchical structure of the representation, which we introduce in this paper, has been inspired by many authors [14, 19, 25, 32, 36], who agree that the spatial representations in humans and animals are hierarchical in nature, providing different levels of abstraction for planning and reasoning. As it has been shown, hierarchical representations support efficient hierarchical path planning in robotics [4, 15] and may facilitate the human-robot interaction, provided that the nodes at the higher level of hierarchy correspond to spatial concepts understandable by humans. A recent example of application of hierarchical maps is the work of Zivkovic *et. al.* ([44, 45]), where a graph cut algorithm is used to construct higher level maps from a connected graph, which represents relations between images based on landmark features in images.

We ground the formation of the hierarchical representation on the visual sensory experience acquired on a path covered by the robot. Path-based visual mapping has been introduced by Zheng and Tsuji [43] and some recent examples can be found in the literature [33, 35], however in contrast to our approach, they all create a nonhierarchical representation.

The method presented in this paper builds a hierarchy of topological maps from a sequence of *omnidirectional images*. We estimate relevant visual aspects of the path using an unsupervised algorithm, *multiple eigenspace analysis* [23], that automatically determines the number of clusters from the underlying data and creates an efficient representation (in the *MDL* sense) by representing the whole set of images using compact linear subspaces. Depending on the structure of the environment, the obtained image clusters may not have well defined boundaries (some images may be equally well encoded by two or more linear subspaces). We therefore define

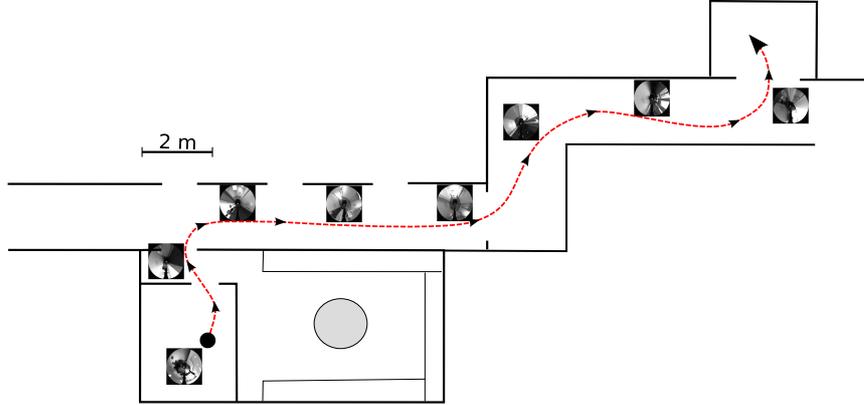


Fig. 1. An example of a path taken by the robot during the exploration of an office environment and some examples of captured omnidirectional images.

a set of complementary *transitional fields*, which describe transition areas between clusters, obtaining thus a partitioning of the captured set of images. Each partition represents a *visual aspect*, a visually distinctive part of the path, and functions as an abstraction of underlying sensory readings (in our case omnidirectional images).

The obtained partitioning enables the creation of a two-level hierarchy of topological maps. At the *top level* the method estimates the topological layout of visual aspects, while at the *bottom level* the method estimates the topological layout of locations where images belonging to the same visual aspect were captured. At each level, topology is estimated using a *physical force model* [13] by promoting the information about image orientation calculated using the *visual compass method* [3]. Because of the complexity of the optimization problem we first estimate the top level and then recursively the bottom level topology.

As we show, the hierarchical structure of the representation not only provides different levels of abstraction of the environment, but also reduces the computational complexity of topological map creation at both levels. The use of multiple eigenspaces method for creating efficient subspace representations of individual visual aspects also results in the increase of the reconstructive power of the hierarchical representation compared to a single subspace representation with the same storage requirements.

The paper is structured as follows. In Section 2 we describe the method in detail, starting with the multiple eigenspace analysis step, followed by the introduction of transitional fields and topology estimation using the physical force model. Examples of building a hierarchy of topological maps in a set of challenging environments and of localization using the resulting representation are presented in Section 3. We conclude with a summary and an outline of the work in progress in Section 4.

2. Hierarchical Mapping

The first step towards building a meaningful interpretation of the path covered during exploration is clustering and modeling of captured images in linear subspaces. The obtained clusters serve as an abstract description of the environment capable of emphasizing its topological characteristics [26].

2.1. Multiple eigenspaces analysis

One of the major issues in clustering of images is the estimation of the optimal number of clusters that would result in a meaningful interpretation of the environment. While Kuipers and Beeson [21] use k -means and estimate the optimal number of clusters using an iterative calculation process, we use an unsupervised *multiple eigenspaces method* [23]. An important aspect of this method is that no *a priori* knowledge of the number of clusters and consequentially no *a priori* knowledge of the location nor spatial extent of clusters is required. This method creates an optimal^a low-dimensional model of each cluster, thus reducing the storage requirements for the hierarchical representation.

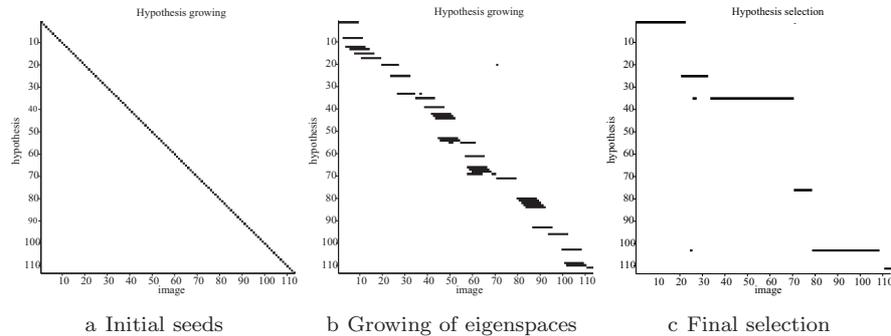


Fig. 2. The partitioning of images for the path shown in Fig. 1. The X axis denotes the images while the Y axis denotes the growing clusters.

Here we briefly review the multiple eigenspace analysis as presented by Leonardis *et. al* [23]. Let $\mathcal{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n | \mathbf{x}_i \in \mathbb{R}^N\}$ be a set of image vectors which correspond to omnidirectional views obtained at locations i , $i = 1 \dots n$ as in Fig. 1 (images are aligned in a reference direction using the visual compass method described in Appendix A). The set of omnidirectional images \mathcal{X} is represented in terms of a set of low-dimensional eigenspaces, where each image is represented as a linear combination of eigenimages spanning the eigenspace \mathcal{U}_j associated with that

^ain an MDL sense

image:

$$\hat{\mathbf{x}}_{ij} = \sum_{k=1}^{d_j} c_{jk}^{(i)} \mathbf{u}_{jk} , \quad (1)$$

where d_j denotes the dimension of the j -th eigenspace and c_{jk} and \mathbf{u}_{jk} are the corresponding coefficients and eigenimages, respectively. Often, only p_j eigenimages ($p_j < d_j$; in our examples the average p_j was between 2 and 3) are needed to represent an image to a sufficient degree of accuracy, yielding an image approximation $\hat{\mathbf{x}}_i$.

The clustering is found using a *recover-and-select* method, by *growing* multiple low dimensional eigenspaces and selecting the optimal subset according to the *minimum description length* (MDL) criterion. The system is initialized with each omnidirectional image as a seed hypothesis (Fig. 2a). These hypotheses iteratively grow by systematically adding images and encoding them in multiple eigenspaces. The growth process is regulated by a pair of error measures, which relate to the reconstruction error of each of the eigenspace and to the distance of the newly added images to that eigenspace (Fig. 2b). After completing the growth of all hypotheses the MDL selection is run, which takes into account the number of images encoded by each eigenspace, dimensionalities of eigenspace and the residual errors of images encoded by each eigenspace.

The set of images is therefore clustered into m clusters \mathcal{G}_j , ($\mathcal{G}_j \subset \mathcal{X}$, $|\mathcal{G}_j| = k_j$), each cluster containing images that lie on a linear subspace of low dimensionality (Fig. 2c). As can be seen in Fig. 3b, the final number of clusters reflects the changes in appearance along the path. The obtained clustering is used as a structural basis for further computations.

2.2. Transitional fields

As the robot moves along the path it continuously captures a sequence of omnidirectional images. Similarity of consecutive images depends on the appearance of the environment and the rate at which images were captured. Due to the nature of the clustering algorithm, the boundaries between obtained clusters might not be clearly defined since some images may be encoded equally well by two or more clusters with respect to the reconstruction error. A similar effect can be noticed in human perception of spatial semantic categories where the boundaries may be graded and subject to dispute [38]. It is therefore crucial to define a set of *transitional fields*, which appropriately model the transition areas between clusters.

If we observe the clustering obtained by multiple eigenspace analysis (Fig. 2c) we notice that some images belong to more than one subset \mathcal{G}_j ($\sum_j \mathbf{1}_{ji} > 1$). To resolve this and obtain clear boundaries between clusters we calculate the reconstruction error $e_{ij} = \|\mathbf{x}_i - \hat{\mathbf{x}}_{ij}\|_2$, where $\hat{\mathbf{x}}_{ij}$ denotes the reconstruction of image \mathbf{x}_i in eigenspace \mathcal{U}_j , for each image \mathbf{x}_i and all eigenspaces \mathcal{U}_j obtained by the multiple eigenspace algorithm. If $\mathbf{r}_i = \frac{e_{ik}}{e_{il}} > \eta$, where e_{ik} is the minimum reconstruction

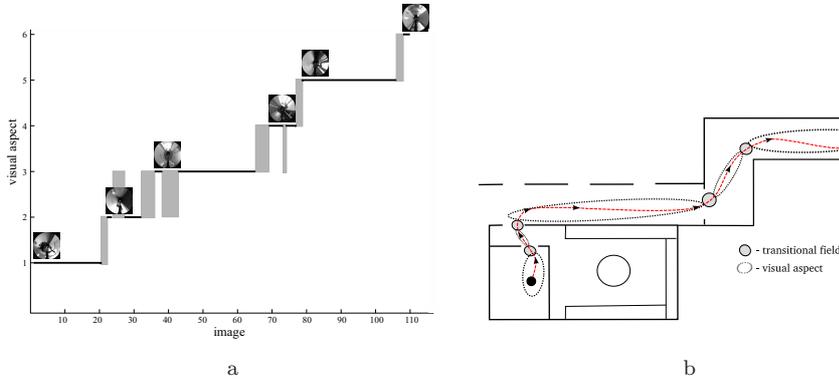


Fig. 3. Figure 3a shows membership of images in subsets \mathcal{G}_i after the calculation of transitional fields. The x -axis denotes the sequence of images and the y -axis denotes the index of the visual aspect to which the image belongs. If $y = 0$ the image falls in a transitional field. Figure 3b shows the physical extent of visual aspects and transitional fields on the path taken by the robot.

error for some eigenspace \mathcal{U}_k , e_{il} is the second-smallest reconstruction error for some eigenspace \mathcal{U}_l ($l \neq k$) and η denotes a threshold, we place this image into a transitional field. In our case η was set to $\bar{\mathbf{r}} + 1.5 \cdot \text{iqr}(\mathbf{r})$, where \mathbf{r} denotes the vector of quotients \mathbf{r}_i for all images \mathbf{x}_i and iqr denotes the interquartile range. The x -axis of Fig. 3a shows the sequence of images as they were captured on the path shown in Fig. 1, while the y -axis shows the index i of the subset \mathcal{G}_i to which each image belongs. Gray areas in the figure show images that were placed into transitional fields according to the above-mentioned criterion.

In this way we obtain a partitioning of a set of images taken during exploration, where each partition represents a visually distinctive place, a *visual aspect* of the path. Figure 3b shows the approximate physical extent of the visual aspects on the path taken by the robot, while the images belonging to the first three visual aspects and first two transitional fields are shown in Fig. 4. Locations where images belonging to transitional fields were captured are not used in map building due to high uncertainty of cluster membership, which might prevent the creation of a stable topological layout.

2.3. Topological mapping

At this stage our method represents sensory data (omnidirectional images) at two different levels, raw data and visual aspects. To estimate the topological maps we employ a physical force model that promotes local information on orientation of images (the heading direction ρ_{jk} of the robot between consecutive locations j and k) to an overall topological arrangement of the map. First, we create the topological layout of visual aspects, the top level topological map, which can be used for high-level path planning, then we recursively create topological layouts of locations belonging to each visual aspect, the bottom level topological maps, which

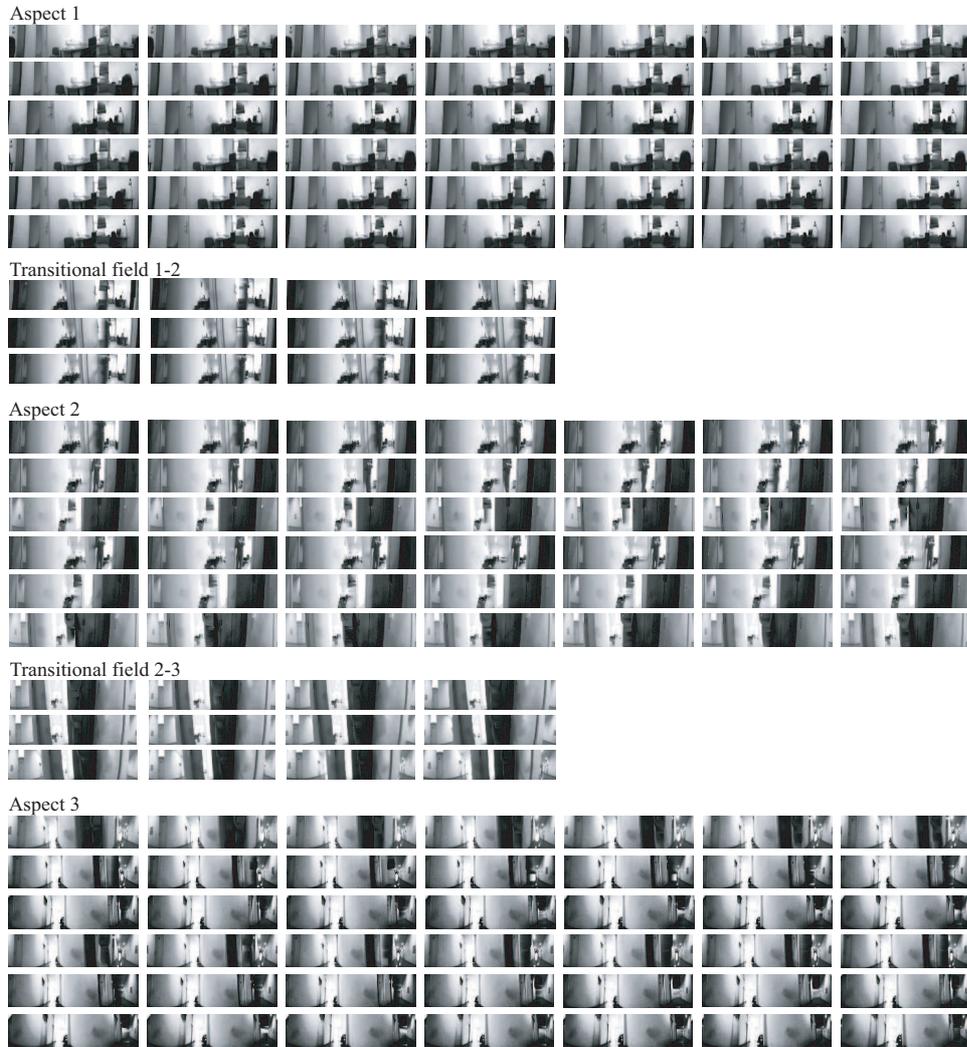


Fig. 4. Figure shows the images belonging to the first three visual aspects and first two transitional fields as determined by our method for the image set captured on the path shown in Fig. 1.

can be used for navigation within a visual aspect. This *divide-and-conquer* approach significantly reduces the computational complexity in comparison to creating a nonhierarchical topological layout (see the discussion in Section 2.3).

The orientation of an image ρ_{jk} is obtained using the improved visual compass method (detailed description in Appendix A), which is also used to align panoramic images and to fill in parts of images corrupted by self-occlusion. The visual compass method works by incrementally building an eigenspace of optimally [2, 3] aligned images. At each step the method projects all possible rotations of an image into

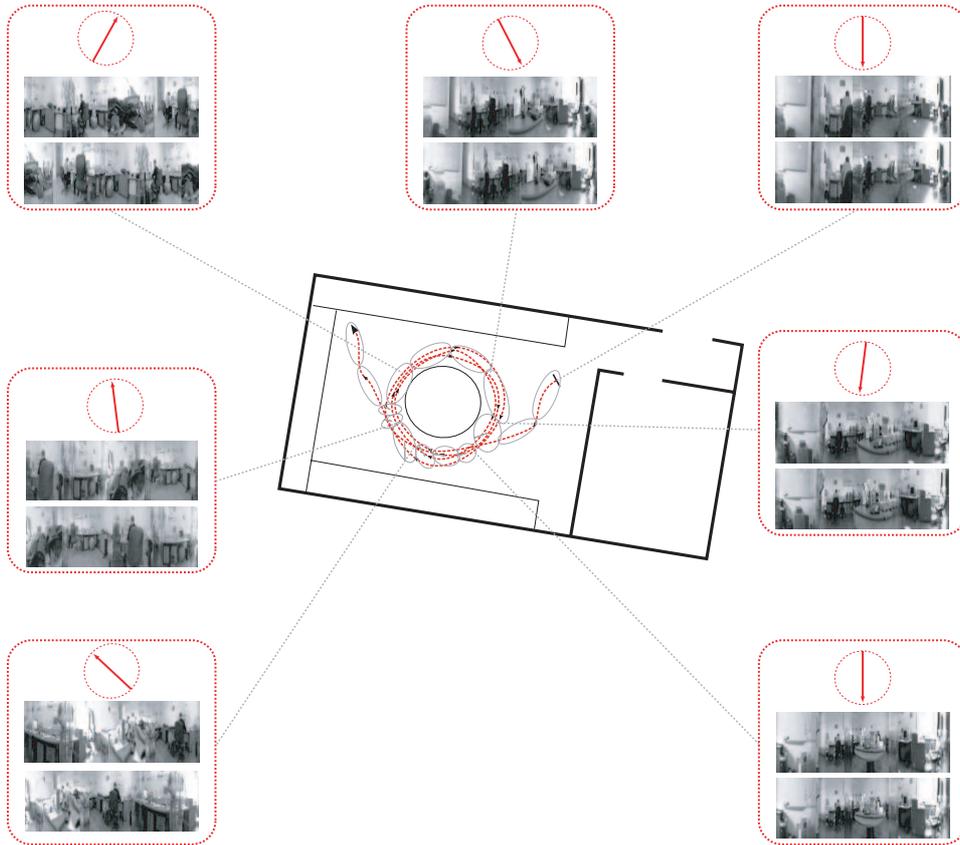


Fig. 5. Calculated orientation information and the alignment of images using the modified visual compass method.

the existing eigenspace, finds the rotation with the smallest reconstruction error and incrementally adds the rotated image into the eigenspace. In contrast to other appearance-based image alignment methods (e.g. [27,31]) the visual compass method is robust to occlusions in images since it uses robust projection [16], and is comparable to some feature-based image alignment methods, e.g. [28,40], which are also robust to occlusions. Figure 5 shows an example of aligning images for a circular path around a table (the full experiment is depicted in Fig. 10a). The top arrow in each of the boxes denotes the calculated direction, the top image is the original panoramic image, followed by the aligned image.

For the top level topological map, the orientation between neighboring visual aspects is estimated using the orientation information of images that were captured during robot's transitions between visual aspects.

Physical force model

In this section we briefly review the basic physical force model as presented by Hafner [13] and propose its extension that enables the calculation of topology at both levels of the hierarchy. A similar method was also presented by Golfarelli et al. [12], who used *elastic correction* to compensate for the dead-reckoning error made by a robot during exploration.

The topological map is modeled as a graph $G = (\mathcal{V}, \mathcal{E})$, where \mathcal{V} is the set of map nodes and \mathcal{E} is the set of connecting edges between consecutive nodes. At the local topological level each node represents a location where an image was taken, while at the global topological level each node represents a visual aspect. The model assigns to each node a repulsive charge and models each connecting edge as a spring. Let $\nu(v, w) \in \mathbb{R}^2$ be the vector connecting nodes v and w and $d(v, w) = \|\nu(v, w)\|$ its Euclidean length. Between each pair of connected nodes the model assumes a force

$$F(v, w) = F_{rep}(v, w) + F_{attr}(v, w) \quad , \quad (2)$$

the sum of the repulsive force, caused by charges at each node, and the attractive force caused by springs between nodes.

Let κ be a constant denoting the predefined length of edges (note that we have no information about distances between nodes, therefore we use $\kappa = 1$). The attractive and the repulsive force between nodes v and w are calculated as

$$F_{attr}(v, w) = \frac{d(v, w) \nu(v, w)}{\kappa} \quad , \quad F_{rep}(v, w) = \frac{-\kappa \nu(v, w)}{d(v, w)} \quad . \quad (3)$$

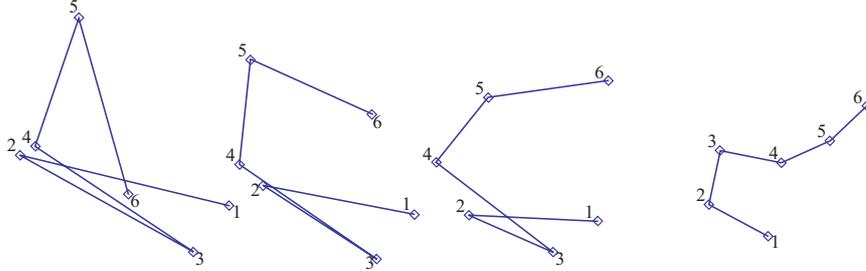


Fig. 6. Figure shows excerpts from the simulated annealing process for the example in Fig. 1. The initial layout of the map is in the leftmost image and the stable solution is in the rightmost image.

The aim is to find a topological configuration of nodes which minimizes the sum of forces between all nodes. To find the optimal solution we use the algorithm described by Fruchterman and Reingold [11]. Their algorithm, however, does not solve for edge orientation. Therefore, a rotational force $F_{rot}(v, w)$ is introduced, which takes into account the angle information between connected map nodes

$$F_{rot}(v, w) = \nu_{\perp}(v, w) d(v, w) \Delta\tau(v, w) \quad . \quad (4)$$

$\Delta\tau(v, w)$ denotes the difference between the current and the preferred edge orientation and $\nu_{\perp}(v, w)$ denotes the unit vector normal to $\nu(v, w)$ whose cross product with the preferred edge direction vector does not contain any negative components. Since $\Delta\tau(v, w) = \pi - \Delta\tau(w, v)$ it follows that $F_{rot}(v, w) = -F_{rot}(w, v)$. We therefore minimize for

$$\sum_{v, w \in V, v \neq w} (F_{rep}(v, w) + F_{attr}(v, w) + F_{rot}(v, w)) . \quad (5)$$

An example of finding a stable solution using *simulated annealing* [18] is shown in Fig. 6. The resulting two levels (Fig. 7) of the hierarchy reflect the topological relations on the bottom level (Level 0) as well as on the top level of the representation (Level 1).

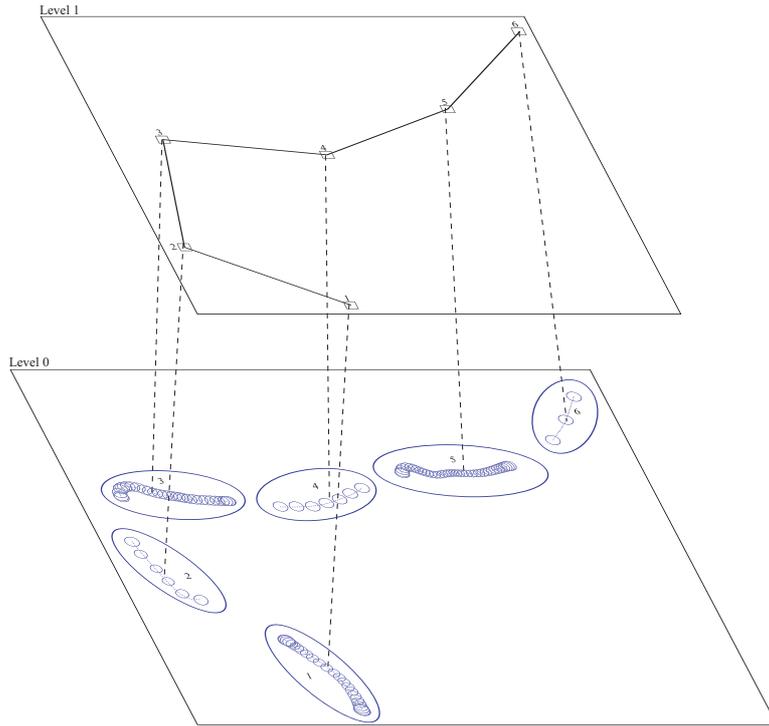


Fig. 7. The top level topological map (Level 1) of the path shown on Fig. 1 is calculated first, then the bottom level topological maps (Level 0) belonging to the 6 visual aspects are calculated recursively.

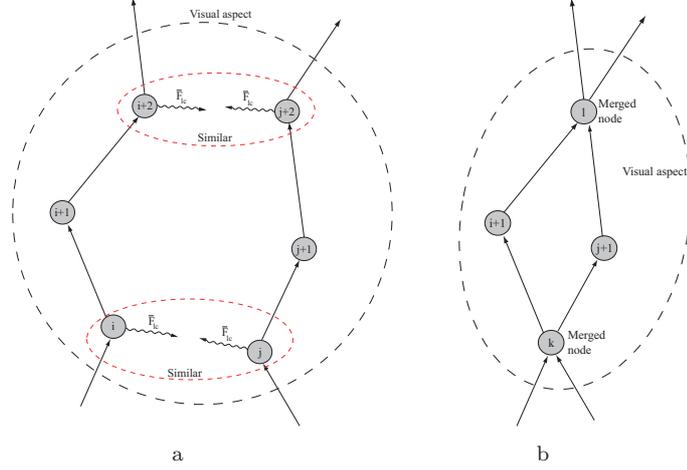


Fig. 8. Figure 8a shows the application of the loop closing force F_{lc} (on the bottom level of the hierarchy) between nodes representing similar images within a visual aspect. Figure 8b shows the final stable solution with both pairs of nodes merged.

Loop closing

In order to create a consistent representation of the path the robot must be able to recognize a previously visited location. On the level of visual aspects, loop detection and closing is handled by the multiple eigenspace analysis, which efficiently clusters similar images into the same visual aspect. For the method to be able to detect loop closings at the bottom level, where each node of the map represents a location at which an image was captured, we need to introduce an additional loop closing force

$$F_{lc}(v, w) = \gamma \frac{d(v, w) \nu(v, w)}{\kappa}, \quad (6)$$

which acts as an attractive force between nodes representing images that are significantly more similar in appearance than any two consecutive images in the sequence (γ is a constant, $0 < \gamma < 1$). In the stable solution of the topological map such nodes are merged if they are significantly closer to each other than any other two consecutive nodes in the map. We therefore optimize for

$$\sum_{v, w \in V, v \neq w} (F_{rep}(v, w) + F_{attr}(v, w) + F_{rot}(v, w) + F_{lc}(v, w)) . \quad (7)$$

Figure 8 shows an example of the application of the loop closing force.

Hierarchical map building

The hierarchical approach to map building allows to build maps of large environments, where a nonhierarchical approach would typically fail or would result in a prohibitive complexity. The initial partitioning enables a divide-and-conquer calculation, which speeds up the minimization of the model significantly. A comparative

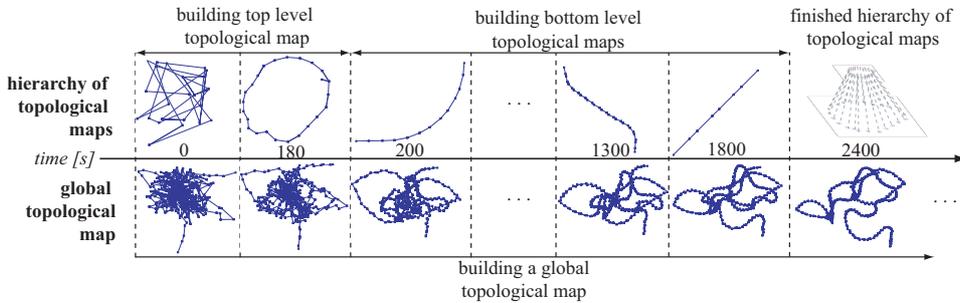


Fig. 9. Figure shows the comparison between creating a topological layout of the path using the proposed hierarchical approach opposed to a nonhierarchical approach for the path shown in Fig. 13.

study of the execution time of both methods can be found in Fig. 9, which shows the stages of map building for both approaches on a time scale. While it is not possible to provide an exact estimate of computational complexity due to the heuristic optimization step, it is clear that the divide-and-conquer approach enables the distribution of the computational load to a number of smaller subtasks. In fact, in all of the cases tested the nonhierarchical approach did not converge to a stable solution in time proportional to 100 times the time needed for a hierarchical solution. Note that for practical reasons we did not measure the final time of execution. However, as one can see from Fig. 9, the hierarchical approach outperforms the nonhierarchical approach, which was not capable to calculate a stable topology in a reasonable time.

3. Experimental Results

This section describes further experiments on map building, followed by an experiment on visual localization using the resulting model. First we present an indoor experiment which demonstrates the loop closing ability of our method on the local and the global level. Next, we show results of two large-scale outdoor experiments. Finally, we study a practical example of using the hierarchical map for the task of identifying the current visual aspect and for retrieving the relative position of the robot.

3.1. Building a hierarchy of topological maps

For the first experiment, the robot made three loops around a table located in the center of an office (Fig. 10a). The results show that the multiple eigenspaces method clusters images taken on nearby locations into same visual aspects closing the loop successfully on the level of a global map in Fig. 10b. On the level of a local map we can see that the application of the loop closing force F_{lc} causes several nodes to be merged effectively closing the loop (Fig. 10c). Great variation in images caused by

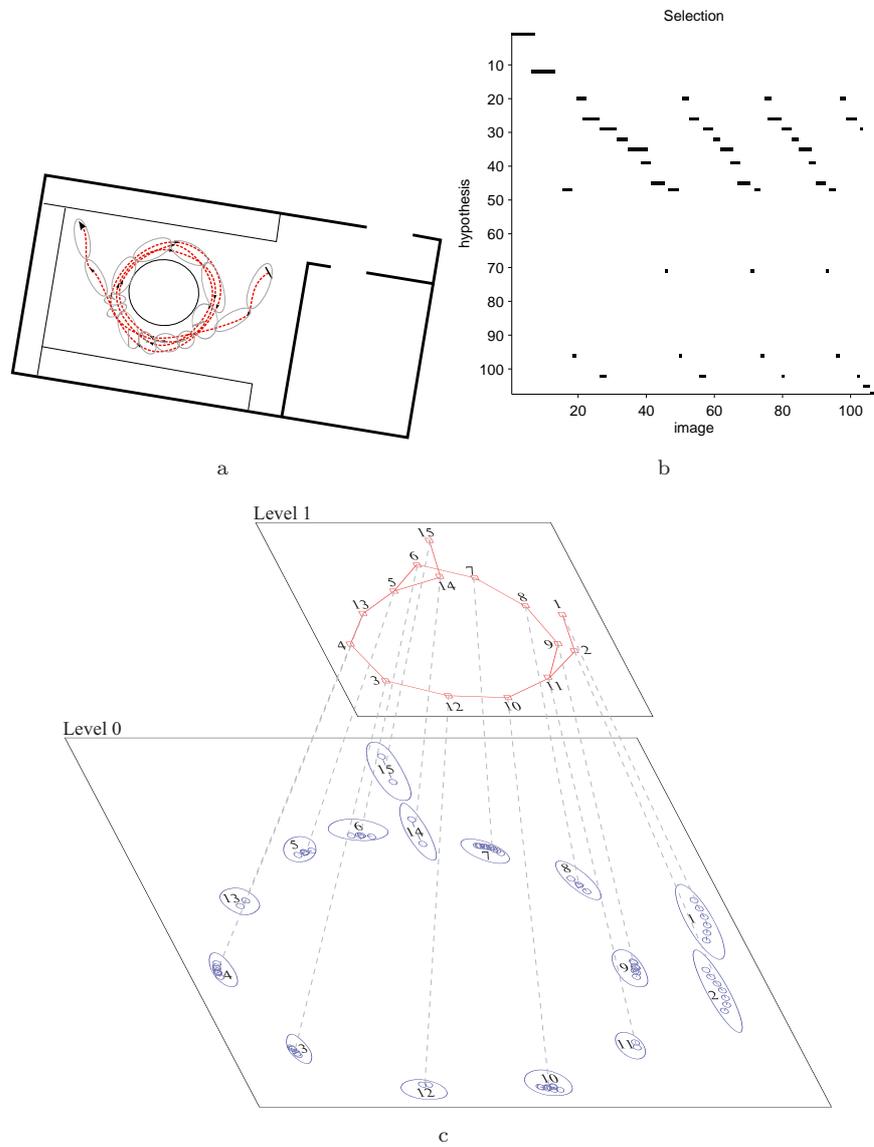


Fig. 10. Results of the second experiment demonstrate that our method successfully closes loops in an office environment on the level of global topology as well as on the level of local topology. Figure 10c also shows an error made by the image alignment algorithm in direction calculation between visual aspects 5 and 14.

the vicinity of large objects (the table) results in a relatively high number of visual aspects.

The second and third experiment were designed to show the feasibility of mapping paths in large unstructured outdoor environments. The length of the path the

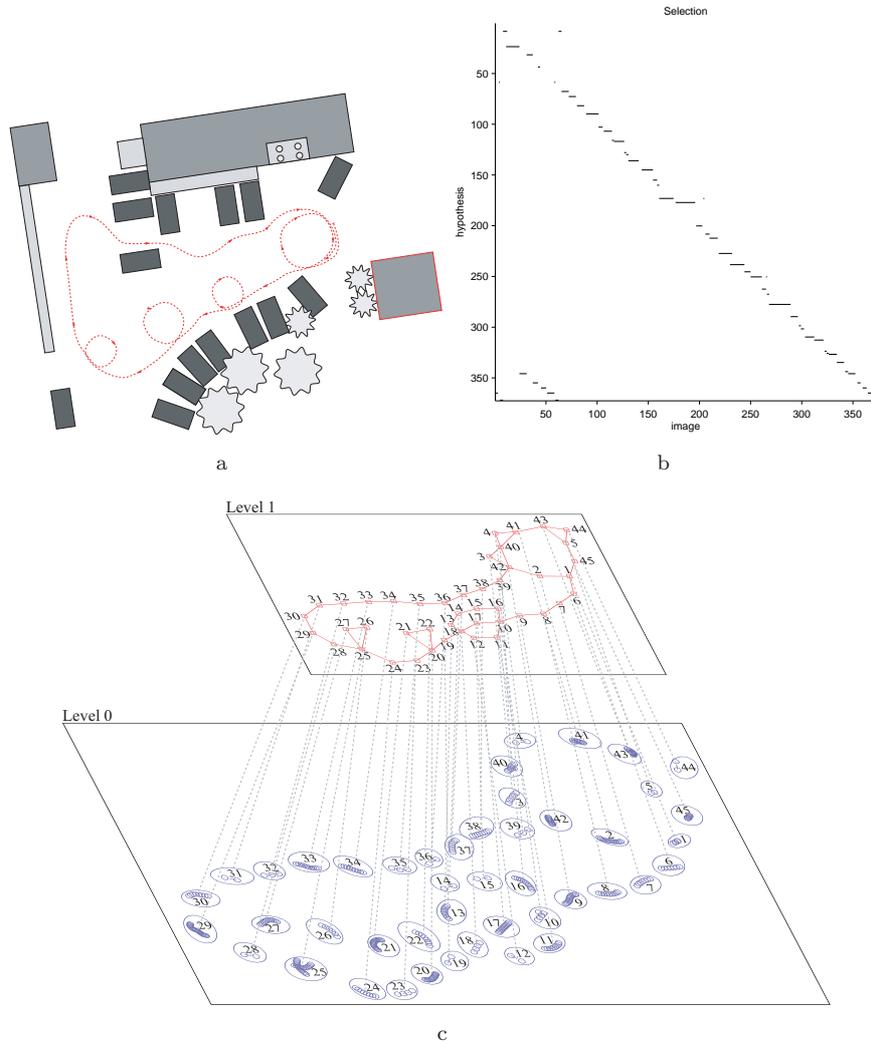


Fig. 11. Results of the third experiment in an outdoor environment demonstrates the feasibility of building a map hierarchy of the path in an unstructured outdoor environment and successful loop closing.

robot took during the second experiment was approximately 70 meters (Fig. 11a) and included multiple loops of varying sizes. The results of the multiple eigenspace analysis (Fig. 11b) show that the method successfully detects and closes loops on the global topological level. The application of the loop closing force on the local topological level also results in successful loop closing by merging nodes representing similar omnidirectional images. The resulting hierarchy of topological maps is shown on Fig. 11c.

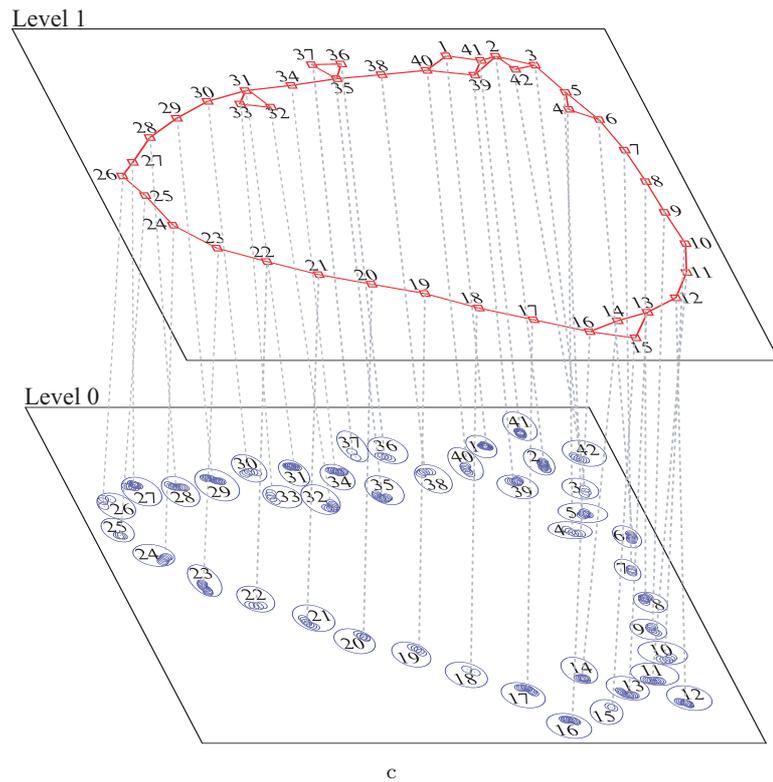
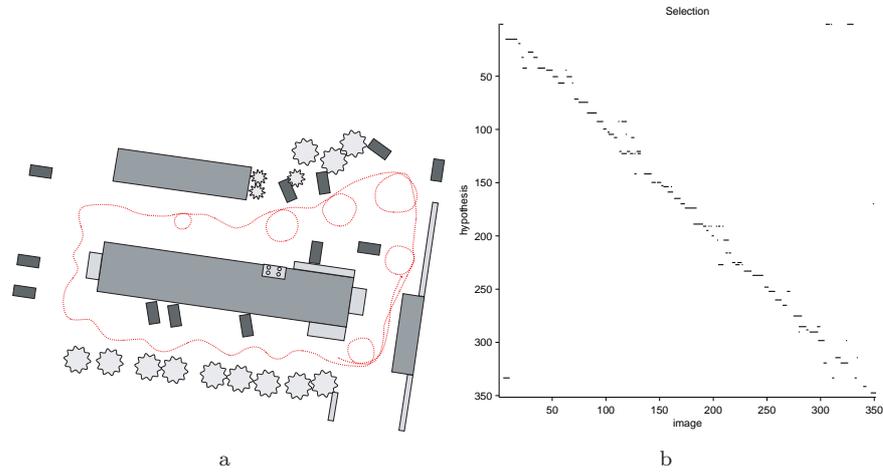


Fig. 12. Results of the fourth experiment which demonstrates the feasibility of building a map hierarchy of path in a large outdoor environment as well as successful detection of loops of various sizes.

During the third experiment, the robot was driven around the faculty building on a path in Fig. 12a, which was approximately 170 meters in length and included multiple loops of varying sizes. The results of the multiple eigenspace analysis (Fig. 12b) show that the method successfully detects and closes loops on the level of global topology. Figure 12c shows the resulting hierarchy of topological maps from which we can see that the application of the loop closing force results in successful loop closing on the level of local topological maps.

3.2. Localization

Once the hierarchical map is built, it can be used for estimation of the current visual aspect, therefore retrieving a high level information on the robot's current position in the topological layout. This information can be used for orientation, planning and human-robot communication. For navigation within a visual aspect a relative position can be retrieved from the low level map modeling positions within that aspect. Positioning is performed by subspace based localization [16], where the position is estimated by projecting the momentary view into a subspace modelling the appearance of the environment within the visual aspect.

In order to demonstrate the efficiency of localization using a hierarchical representation, we test it for robustness against occlusion and compare its performance with the performance of the standard localization method using a single global eigenspace. We compared the performance of solving the kidnapped robot problem [9] using the hierarchy of topological maps and using a single global eigenspace representation of the training set of images (using both non-robust and robust method) at different levels of occlusions in the test images.

The standard localization method is performed by projecting the momentary view \mathbf{y} into the eigenspace $\mathbf{w}_y = \mathbf{V}^T(\mathbf{y} - \bar{\mathbf{x}})$, where $\bar{\mathbf{x}}$ denotes the mean training image, and retrieving the nearest position from the learned model. This calculation, however, is sensitive to occlusions due to objects moving in the environment and self-occlusions. Therefore we used the robust approach for projection as described by Leonardis and Bischof [22]. Instead of using whole image vectors, the method generates and evaluates a set of hypotheses \mathbf{r}^l as subsets of image points $\mathbf{r}^l = (\mathbf{r}_1^l, \mathbf{r}_2^l, \dots, \mathbf{r}_k^l)$, where $k \geq p$. The method tries to find the solution vector $\mathbf{w}(\mathbf{x})$ that minimizes

$$E(\mathbf{r}) = \sum_{i=1}^k \left(\mathbf{x}_{\mathbf{r}_i^l} - \sum_{j=1}^p \mathbf{w}_j(\mathbf{x}) \mathbf{v}_{\mathbf{j}, \mathbf{r}_i^l} \right)^2$$

The initial set of points is selected randomly. After finding a robust solution of the set of equations, an α -trimming step selects points with an arbitrary small error, which contribute to further computation. These hypotheses are then subject to a selection procedure based on the MDL principle as described in [22].

Localization in the top level topological map is performed by calculating robust projections $\mathbf{w}^{\mathcal{U}_i}(\mathbf{y})$ of the momentary view \mathbf{y} into all eigenspaces \mathcal{U}_i that model

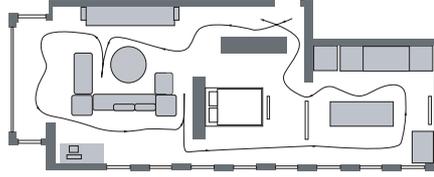


Fig. 13. Figure shows the path on which the robot was taken through the environment.

visual aspects. The location in the top level map is determined as

$$\mathcal{U} = \operatorname{minarg}_{\mathcal{U}_i} (\mathbf{y} - E_{\mathcal{U}_i} \mathbf{w}^{\mathcal{U}_i}(\mathbf{y})) ,$$

where $E_{\mathcal{U}_i}$ denotes the eigenvector matrix of eigenspace \mathcal{U}_i .

Localization at the bottom level topological maps is performed by comparing the projection $\mathbf{w}^{\mathcal{U}}$ to projections of training images belonging to the corresponding visual aspect.

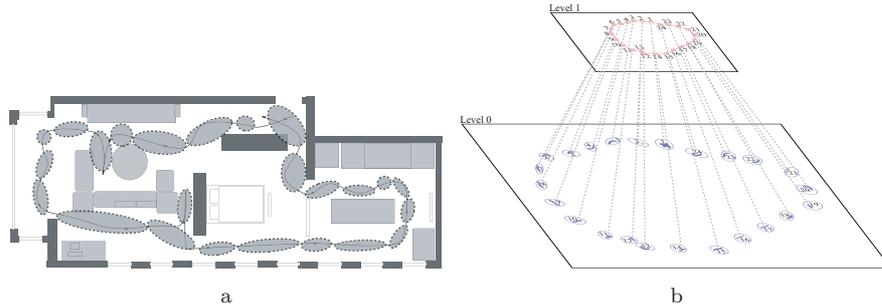


Fig. 14. Figure 14a shows the extent of visual aspects determined by the presented method from omnidirectional images captured on the path show in Fig. 13. Figure 14b shows the created hierarchy of topological maps.

To demonstrate localization we used the FS2HSC - IEEE/RSJ IROS 2006 Workshop dataset [1] which includes accurate odometry. To create the hierarchy of topological maps we used 409 omnidirectional images extracted from the occlusion-free video sequence included in the dataset. Map of the path through the environment is shown in Fig. 13. Our method automatically retrieved 24 visual aspects and modeled 24 eigenspaces of an average dimension of 1.6 that efficiently encoded the training images. Figure 14a shows the spatial extents of visual aspects, the calculation of the topological layout using the physical force model at both levels is shown in Fig. 9. The hierarchy of topological maps is shown in Fig. 14b. Results show that the method successfully detected and closed loops in the robots path. For the purpose of testing we used 100 images from the same video sequence that were not included in the training set and artificially added different levels of occlusions. Replacement noise consisted of randomly generated rectangular areas which were

filled with randomly generated pixels values. For the purpose of comparison the dimension of the single global eigenspace was set to be equal to 39, the sum of dimensions of eigenspaces representing visual aspect.

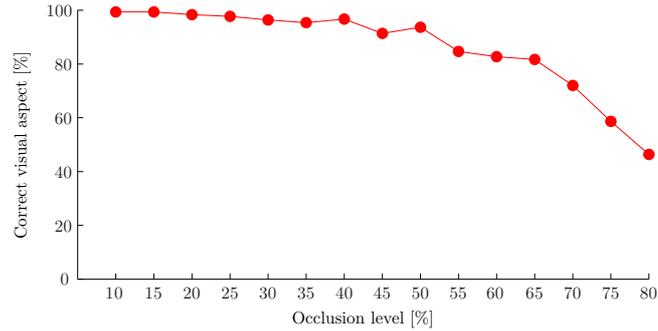


Fig. 15. Figure shows the probability of correct localization in the top level topological at different levels of occlusions in the input image.

First we tested the robustness of top level localization. Figure 15 shows the probability of determining the correct visual aspect to which the test image belongs at different levels of occlusion. The ground truth was determined from the odometry. Results show that the localization at the top hierarchical level achieves a high level of accuracy with reasonable levels of occlusion, while it breaks down at about 60% of occlusion, which corresponds to results regarding the robustness of the robust method reported in [22].

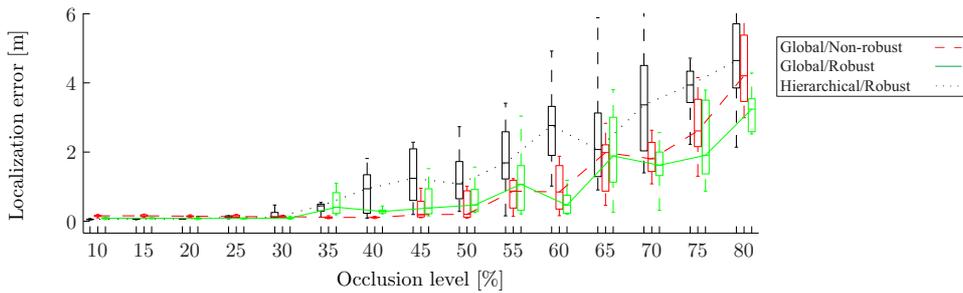


Fig. 16. Figure shows results of localization tests for different levels of occlusions in the image for three methods: global eigenspace/non-robust projection - dotted line, global eigenspace/robust projection - full line, hierarchy of topological maps/robust projection - dashed line.

Results of localization in the bottom level topological maps can be seen in Fig. 16, which shows the localization error (in meters) for different level of occlusions and all three methods of localization: global eigenspace/non-robust, global

eigenspace/robust and hierarchy of topological maps/robust. Results show that the *breaking point* of localization using a single global eigenspace and non-robust projection is at around 40% of occlusion in the image, while localization using a single eigenspace and robust projection breaks down around 60% of occlusions. Localization using the presented hierarchy also breaks down at 60% of occlusions. Note that the accuracy of localization in this case is limited by the ability to estimate the correct visual aspect, as the relative position is searched for only within that aspect.

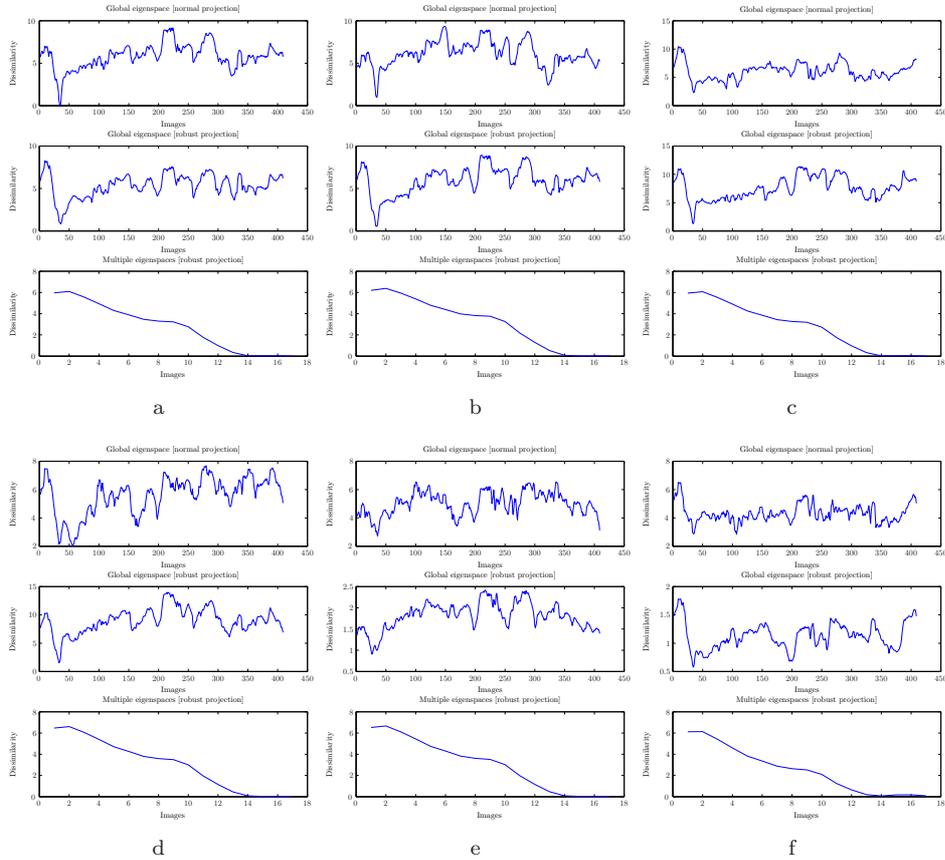


Fig. 17. Figure shows Euclidean distances between the projection of the test image and projections of training images for all three methods at 0% (17a), 15% (17b), 30% (17c), 45% (17d), 60% (17e) and 75% (17f) of occlusion.

The results in Fig. 16 show that localization using the hierarchy of topological maps performs slightly worse than localization using a global eigenspace with robust projection of test images at low levels of occlusions. To explain this phenomena, we compared the Euclidean distances between projections of the test image

and training images into a global eigenspace and projections into the eigenspace modelling the correct visual aspect (according to the odometry). At low levels of occlusion (*e.g.* Fig. 17a), a distinctive minimum for the global eigenspace (non-robust and robust projection) can be observed. On the other hand, the minimum for our representation is less obvious, since all images modeled by the local subspace are very similar in appearance, and are therefore all similar to the test image. However, for high levels of occlusion (*e.g.* Fig. 17e), we can observe that the minimum becomes less obvious for the global eigenspace, while the minimum in our method remains stable. This corresponds to results shown in Fig. 16, since at low levels of occlusions it is more difficult to find the correct location using our method (hence a slightly higher localization error), while at higher levels of occlusions our method outperforms the localization using a global eigenspace.

4. Conclusion

We introduced an unsupervised computational model which maps the topology of a path at different levels of abstraction. As we show, unsupervised clustering using the multiple eigenspaces algorithm gives a set of clusters that, represented as low dimensional subspaces, efficiently encode the visual aspects of the path. We also show that the topology can be efficiently estimated using a physical force model. As the results show, the presented model gives an intuitive representation, which is feasible for navigation in micro-locations, as well as for reasoning about the environment as a collection of interconnected visual aspects at a higher level.

The model has already been extended to be robust to occlusions using the approach described by Jogan and Leonardis [22]. To achieve robustness to illumination variations gradient filtering of eigenspaces can be applied as described by Jogan *et al.* [17].

The constructed topological map hierarchy can be used for robot navigation in various ways. We could use hierarchical decomposition of motion planning described by Miura and Shirai [29] and decompose the planning task into *global route selection* and *local path planning*. Another approach to hierarchical path planning is described by Bakker *et al.* [4] where they formalize robot maps as Markov Decision Processes, such that the planning task becomes a dynamic programming problem [5]. All these approaches show the feasibility and promise of the hierarchical mapping and planning.

An issue that needs to be addressed is the spatial extent of the obtained visual aspects. The results from the previous section show that visual aspects tend to have a smaller spatial extent when the robot is moving close to objects in the environment (*e.g.* cars, walls, chairs, *etc.*). This can be explained by the fact that the appearance of the environment changes faster when the robot is moving close to objects. On the other hand open space (*e.g.* middle of a room/corridor) tends to get represented with a low number of visual aspects, since the appearance does not change significantly as the robot moves along its path.

In future we plan to investigate how to add an arbitrary number of levels to the hierarchy including multiple modalities (*e.g.* range sensors) or levels based on the tutor’s representation of spatial semantic categories, which would support semi supervised learning and human-robot communication. We also plan to devise an incremental method that would allow to augment existing maps with novel information. Such an approach would need to overcome the intrinsic limitation of the multiple eigenspace method, *i.e.* that the representation is built from a batch of captured images, which, once represented in a low dimensional subspace, cannot be completely reconstructed from their subspace coefficients. Another interesting research issue is how to promote novel information through the hierarchy.

Appendix A. Improved visual compass

Since we use an appearance-based method we need to align the omnidirectional images in some reference direction. This could be achieved by using odometry or external compass devices, but odometry has proven to be quite inaccurate [6, 7] and the performance of compass devices suffers greatly in indoor environments due to disturbances from electromagnetic sources (*e.g.* power lines) or large ferro-magnetic structures (*e.g.* metal bookshelves) [39].

A.1. Aligning images with incremental eigenspace approach

In our method we use a modified visual compass approach described by Artač [3]. As originally described the method makes the assumption that we have a set of aligned images which are used to build the initial eigenspace \mathcal{U} . Let $\Phi = \{\phi_1, \dots, \phi_N\}$ be the set of N possible rotations of each omnidirectional image. For each new image \mathbf{x}_i , we create a set of all possible rotations of the i -th image $\mathcal{X}_i^\Phi = \{\mathbf{x}_i^{\phi_1}, \mathbf{x}_i^{\phi_2}, \dots, \mathbf{x}_i^{\phi_N}\}$, where $\mathbf{x}_i^{\phi_j}$ denotes the j -th possible rotation of the image. We then proceed by projecting rotated images into eigenspace \mathcal{U} and calculating residual errors

$$\Psi(\mathbf{x}_i^{\phi_j}) = \mathbf{x}_i^{\phi_j} - \sum_{k=1}^d (c_k^{\phi_j} \mathbf{u}_k) - \mathbf{x}_m, \quad (\text{A.1})$$

where \mathbf{x}_m denotes the mean image and $c_k^{\phi_j}$ and \mathbf{u}_k are the corresponding coefficients and eigenimages, respectively. The rotation of the image \mathbf{x}_i with the smallest residual error $\Phi(\mathbf{x}_i^{\phi_j})$ is taken as the true orientation and the rotated image $\mathbf{x}_i^{\phi_j}$ is incrementally added to the eigenspace \mathcal{U} as described by Artač [2].

A.2. Aligning the initial set of images

In our method, however, we do not assume that we have an initial set of aligned images since we do not want to constrain the movement of the robot in the initial steps of the exploration. To create an initial eigenspace we take the first M images of the image sequence and try to find a set rotation values $\Phi_0 = \{\phi_1, \phi_2, \dots, \phi_M\}$

for these images that minimizes the sum of residual errors when projecting these images into a common eigenspace. To achieve this we use an iterative procedure that repeats until the set of rotations Φ_0 converges. The pseudocode for the procedure is shown in Procedure 1.

Procedure 1 Align the initial set of images

```

%  $\Phi_0$  - the output set of rotations
 $\Phi_0 = \{\phi_1, \phi_2, \dots, \phi_M | \phi_i \in \{1, 2, \dots, N\}\}$ 
%  $\mathcal{X}_0^\Phi$  - set of rotated images  $\mathbf{x}_i^{\phi_i}$ 
repeat
  for  $i = 0$  to  $M$  do
    % EV - eigenvectors
    %  $x_m$  - mean image
     $[EV, x_m] = \text{buildEigenspace}(\mathcal{X}_0^\Phi)$ ;
    for  $j = 1$  to  $N$  do
      % calculate the residual error
       $\xi_j = \text{residualError}(\mathbf{x}_i^j, EV, x_m)$ ;
    end for
    % find the rotation with the smallest
    % residual error
     $\phi_i = \arg \min_j (\xi_j)$ ;
  end for
until convergence of  $\Phi_0$ 

```

The output of this procedure is a set of rotations Φ_0 that give us the initial set of aligned images \mathcal{X}^{Φ_0} . These aligned images are used to create the initial eigenspace \mathcal{U} and we can continue in the same way as described in section A.1. We do not have a theoretical proof of convergence for this procedure, but experimental results show that for small values of M (for experiments described in section 3 we used $M = 15$) the procedure converges in ten to twenty iterations.

Acknowledgments

This research has been supported in part by the following funds: Research program Computer Vision P2-0214 (RS), the EU FP6-004250-IP project CoSy, the ECVision, EU (IST-2002-35454) research network, CONEX project, and SI-A project.

References

1. FS2HSC - IEEE/RSJ IROS 2006 Workshop dataset. <http://www2.science.uva.nl/sites/cogniron>, October 2006.
2. M. Artač, M. Jogan, and A. Leonardis. Mobile robot localization using an incremental eigenspace model. In *Proceedings 2002 IEEE International Conference on Robotics and Automation*, pages 1025–1030, Washington, D.C., May 11-15 2002.

3. M. Artač. Localisation and basic navigation for a mobile robot using visual learning and recognition. Master's thesis, Faculty of computer and information science, University of Ljubljana, March 2003.
4. B. Bakker, Z. Zivkovic, and B. Kröse. Hierarchical dynamic programming for robot path planning. In *Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 3720–3625, 2005.
5. D. P. Bertsekas. *Dynamic Programming and Optimal Control*. Athena Scientific, 1995.
6. J. Borenstein and L. Feng. Measurement and correction of systematic odometry errors in mobile robots. *IEEE Journal of Robotics and Automation*, 12(6):869–880, December 1996.
7. K. Chong and L. Kleeman. Accurate odometry and error modelling for a mobile robot. In *Proceedings of the 1997 IEEE International Conference on Robotics and Automation*, pages 2783–2788, 1997.
8. G. Dudek, M. Jenkin, E. Milios, and D. Wilkes. On building and navigating with a globally topological but locally metric map. In *Proc. 3rd ECPD Int. Conf. on Advanced Robotics, Intelligent Automation and Active Systems*, pages 132–144, Bremen, Germany, 1997.
9. S. P. Engelson. *Passive map learning and visual place recognition*. PhD thesis, New Haven, CT, USA, 1994.
10. U. Frese, P. Larsson, and T. Duckett. A multilevel relaxation algorithm for simultaneous localisation and mapping. *IEEE Transactions on Robotics*, 21(2):196–207, 2005.
11. T. M. J. Fruchteman and E. M. Reingold. Graph-drawing by force directed placement. *Software - Practise and Experience*, (21):1129–1164, 1991.
12. M. Golfarelli, D. Maio, and S. Rizzi. Correction of dead-reckoning errors in map building for mobile robots. *IEEE Transactions on Robotics and Automation*, 17(1):37–47, 2001.
13. V. V. Hafner. Cognitive maps for navigation in open environments. In *Proceedings of the 6th International Conference on Intelligent Autonomous Systems (IAS-6)*, pages 801–808, Venice, 2000. IOS Press.
14. S. C. Hirtle and J. Jonides. Evidence of hierachies in cognitive maps. *Memory and cognition*, 13:208–217, 1985.
15. W. K. Hyun and I. H. Suh. A hierarchical collision-free path planning algorithm for robotics. *IROS*, 02:2488, 1995.
16. M. Jogan and A. Leonardis. Robust localization using an omnidirectional apperance-based subspace model of environment. *Robotics and Autonomous Systems*, 45(1):51–72, October 2003.
17. M. Jogan, A. Leonardis, H. Wildenauer, and H. Bishof. Mobile robot localization under varying illumination. In *Proceedings Intl. Conference on Pattern Recognition ICPR02*, pages 741–744, Quebec, Canada, 2002.
18. S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi. Optimization by simulated annealing. *Science*, Number 4598, 13 May 1983, 220, 4598:671–680, 1983.
19. B. Kuipers. Modeling spatial knowledge. In S. Chen, editor, *Advances in Spatial Reasoning (Volume 2)*, pages 171–198. Ablex, Norwood, NJ, 1990.
20. B. Kuipers and Y. T. Byun. A robot exploration and mapping strategy based on a semantic hierarchy of spatial representations. *Toward learning robots*, pages 47–63, 1993.
21. B. J. Kuipers and P. Beeson. Bootstrap learning for place recognition. In *Eighteenth national conference on Artificial intelligence*, pages 174–180, Menlo Park, CA, USA, 2002.

22. A. Leonardis and H. Bischof. Robust recognition using eigenimages. *Computer Vision and Image Understanding*, 78(1):99–118, 2000.
23. A. Leonardis, H. Bischof, and J. Maver. Multiple eigenspaces. *Pattern Recognition*, 35(11):2613–2627, 2002.
24. B. Linketkai, L. Liao, and D. Fox. Relational object maps for mobile robots. In *Proc. of the International Joint Conference on Artificial Intelligence (IJCAI-05)*, pages 1471–1476, 2005.
25. K. Lynch. *The image of the city*. M.I.T. Press, Cambridge, 1971.
26. D. Maio and S. Rizzi. Map learning and clustering in autonomous systems. *IEEE Trans. Pattern Anal. Mach. Intell.*, 15(12):1286–1297, 1993.
27. E. Menegatti, M. Zoccarato, E. Pagello, and H. Ishiguro. Image-based monte carlo localisation with omnidirectional images. *Robotics and Autonomous Systems*, 48(1):17–30, 2004.
28. K. Mikolajczyk and C. Schmid. A performance evaluation of local descriptors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27:1615–1630, 2005.
29. J. Miura and Y. Shirai. Hierarchical vision-motion planning with uncertainty: Local path planning and global route selection. In *Proc. 1992 IEEE/RSJ Int. Conf. on Intelligent Robots and System*, Raleigh, NC, USA, July 1992.
30. H. P. Moravec. The Stanford cart and the CMU rover. In *Proceedings of the IEEE*, number 71, pages 872–884, 1983.
31. T. Pajdla and V. Hlaváč. Zero phase representation of panoramic images for image based localization. In *8-th International Conference on Computer Analysis of Images and Patterns*, number 1689 in Lecture Notes in Computer Science, pages 550–557. Springer Verlag, September 1999.
32. J. Piaget and B. Inhelder. *The child's conception of space*. Norton, New York, 1967.
33. J. M. Porta and B. J. A. Kröse. Appearance-based concurrent map building and localization. *Robotics and Autonomous Systems*, 54(2):159–164, 2006.
34. E. Remolina and B. Kuipers. Towards a general theory of topological maps. *Artif. Intell.*, 152(1):47–104, 2004.
35. P. E. Rybski, S. I. Roumeliotis, M. Gini, and N. Papanikolopoulos. Appearance-based minimalistic metric SLAM. In *Proceedings of IROS'03, the IEEE International Conference on Intelligent Robots and Systems*, pages 194–199, Las Vegas, Nevada, October 2003.
36. A. W. Siegel and S. H. White. The development of spatial representations of large-scale environments. In H.W. Reese, editor, *Advances in Child Development and Behavior*, number 10, pages 10–55. Academic Press, 1975.
37. S. Simhon and G. Dudek. A global topological map formed by local metric maps. In *IROS*, volume 3, pages 1708–1714, Victoria, Canada, October 1998.
38. B. Smith and D. Mark. Ontology with human subjects testing: an empirical investigation of geographic categories. *The American Journal of Economics and Sociology*, 58(2):245–272, April 1999.
39. S. Suksakulchai, S. Thongchai, D. M. Wilkes, and K. Kawamura. Mobile robot localization using an electronic compass for corridor environment. In *Proceedings 2000 IEEE International Conference on Systems, Man and Cybernetics*, 2000.
40. R. Szeliski. Image alignment and stitching. In N. Paragios *et al.*, editor, *Handbook of Mathematical Models in Computer Vision*, pages 273–292. Springer, 2005.
41. S. Thrun. Learning metric-topological maps for indoor mobile robot. In *Artificial Intelligence 99*, pages 21–71, 1999.
42. S. Thrun and A. Buecken. Learning maps for indoor mobile robot navigation. Technical Report CMU-CS-96-121, Computer Science Department, Carnegie Mellon Uni-

- versity, Pittsburgh, PA, 1996.
43. J. Y. Zheng and S. Tsuji. Panoramic representation for route recognition by a mobile robot. *Int. J. Comput. Vision*, 9(1):55–76, 1992.
 44. Z. Zivkovic, B. Bakker, and B. Kröse. Hierarchical map building using visual landmarks and geometric constraints. In *In Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 7–12, 2005.
 45. Z. Zivkovic, B. Bakker, and B. Kröse. Hierarchical map building and planning based on graph partitioning. In *IEEE International Conference on Robotics and Automation*, pages 803–809, 2006.
-



Aleš Štivec received the B.Sc. degree in computer science from the University of Ljubljana, Slovenia, in 2004. He is a junior researcher at the Visual Cognitive Systems Laboratory at the Faculty of Computer and Information Science, University of Ljubljana. His research is primarily in cognitive computer vision with focus on cognitive spatial representations, hierarchical and unsupervised methods.



Aleš Leonardis received the B.Sc. and M.Sc. degrees in electrical engineering, in 1985 and 1988 respectively, and the Ph.D. in computer and information science in 1993, all from University of Ljubljana, Slovenia. Currently, he is a full professor and the head of the Visual Cognitive Systems Laboratory with the Faculty of Computer and Information Science, University of Ljubljana. He is also an adjunct professor at the Faculty of Computer Science, Graz University of Technology.



Matjaž Jogan received the B.Sc. and M.Sc. degrees in computer science from the University of Ljubljana, Slovenia, in 1999 and 2002, respectively. As a member of the Visual Cognitive Systems Laboratory and the Computer Vision Laboratory at the Faculty of Computer and Information Science, University of Ljubljana, he is active as a Researcher in the field of cognitive computer vision. Within this area, he is particularly interested in hierarchical methods for visual learning, cognitive spatial representations, and omnidirectional vision.