Noname manuscript No.
(will be inserted by the editor)

Combining Invariant Features and the ALV Homing Method for Autonomous Robot Navigation Based on Panoramas

Alex Goldhoorn · Arnau Ramisa · David Aldavert · Ricardo Toledo · Ramon Lopez de Mantaras

Received: date / Accepted: date

Abstract Biologically inspired homing methods, such as the Average Landmark Vector, are an interesting solution for local navigation due to its simplicity. However, usually they require a modification of the environment by placing artificial landmarks in order to work reliably. In this paper we combine the Average Landmark Vector with invariant feature points automatically detected in panoramic images to overcome this limitation. The proposed approach has been evaluated first in simulation and, as promising results are found, also in two data sets of panoramas from real world environments.

Keywords visual homing · biologically inspired methods · local features · robot navigation

1 Introduction

There is significant research in robotic navigation using methods based on animal navigation techniques. For instance, Carwright and Collet (1983) studied how the honeybees learned and used landmarks to navigate. From this research they created the *snapshot model*. The idea of this model is to calculate the home vector, which is the

This work was partially supported by the FI grant from the Generalitat de Catalunya, the European Social Fund, the MID-CBR project grant TIN2006-15140- C03-01 and FEDER funds, the grant 2005-SGR-00093, the MIPRCV Consolider Imagennio 2010 and the Marco Polo fund from the University of Groningen.

A. Goldhoorn · A. Ramisa · R. L. de Mantaras
IIIA, Campus UAB, 08193, Bellaterra, Spain
Tel.: +34 93 580 9570
Fax: +34 93 580 9661
E-mail: aramisa@iiia.csic.es, mantaras@iiia.csic.es

D. Aldavert · R. Toledo
CVC, Campus UAB, 08193, Bellaterra, Spain
Tel.: +34 93 581 1828
Fax: +34 93 581 1670
E-mail: aldavert@cvc.uab.cat, ricard@cvc.uab.cat

vector pointing to the home position. A panoramic image of the target location is created and stored by the animal. Then, when the insect wants to go back to the stored position it uses a matching mechanism to compare the current retinal image to the stored panorama. Another example is the work on robotic navigation of Lambrinos et al (1998, 2000), that took inspiration from the different navigation techniques of the ant species *Cataglyphis* described by Wehner (1987). These techniques have the advantage of being computationally cheap.

Lambrinos et al (1998, 2000) suggest the Average Landmark Vector (ALV) as a way to model these navigation techniques. This model assumes that the animal stores an average landmark vector instead of a snapshot. Landmarks can be (simple) features like edges. The direction to the destination is the difference of the ALV at the destination and the ALV at the current location. The advantages of this model are that it is simple to calculate and that only the orientation and the ALV at the home location have to be stored instead of a whole image. A third advantage is that no matching of the landmarks has to be done.

In robot homing research, artificial landmarks are often used. This is a strong limitation as it requires setting up the environment beforehand. Instead, in our work the goal is to create a simple homing method that can be used without having to rely on artificial landmarks. For this we propose the combination of the ALV homing technique with visual invariant feature detectors, like the ones described by Mikolajczyk et al (2005), in panoramic images.

More concretely, we propose to use the invariant feature based ALV in conjunction with a global localization method proposed by one of the authors (Ramisa 2006). The ALV will serve as a costless navigation technique to make the robot go from a starting position to a destination position navigating through a path of place signatures. Another advantage of ALV is that the same visual information as the global localization method of Ramisa (2006) is used, so no additional image processing time is required.

Local homing could also improve the method proposed by Ramisa (2006) in case of ambiguity in the global localization by trying to go to the position in the room where the panorama was made. The most likely hypothesis panorama can be used as a reference for the homing method.

Another contribution of this work is an evaluation of the ALV homing method using feature points extracted from panoramic images of unmodified indoor environments. This is a necessary step towards combining it with the global localization technique of Ramisa (2006), that would complement the localization method by giving it a computationally efficient method to travel from one node to another of the global map.

Experiments with the ALV homing method were first done in simulation (Goldhoorn et al 2007a,b) and because the results were promising, experiments were also done with real robots (Goldhoorn 2008) in an office environment. Additionally, experiments with artificial landmarks were also done for comparison purposes.

This paper is divided in the following sections: first, some background of the methods used is provided, this includes the panorama acquisition techniques, the local features, the global localization method of Ramisa (2006) and the Average Landmark Vector homing method; next, the proposed method is explained; then the experiments performed to evaluate the method are presented followed by a discussion of the results and finally the conclusions and future work are described.

2 Background

In this section we describe the different concepts and techniques that have been used in this work.

2.1 Panorama

One way to acquire panoramas consists in stitching images taken with a camera rotating around a fixed point of view until the full 360° have been covered. The next step is to stitch those images together to complete the panorama. The images should be projected onto a smooth surface such as a cylinder to avoid discontinuities or inhomogeneous sampling. A cylindrical representation offers some advantages. In the first place it can be created relatively easily, and also, in opposition to other plenoptic representations such as a sphere, can be unrolled and stored in an efficient way as a conventional rectangular image (McMillan and Bishop 1995). It is important to have a fixed optical center, otherwise motion parallax would be introduced. However the small translations can be tolerated when the objects are far enough from the camera. In this work we used the method from Ramisa (2006) to construct the panoramas. Follows a brief explanation.

First the coordinates have to be transformed from the Cartesian system of the images to the cylindrical coordinate system:

$$\theta = \tan^{-1}\left(\frac{x}{f}\right), \quad v = \frac{y}{\sqrt{x^2 + f^2}} \tag{1}$$

Where (x,y) is the pixel position in the image, f the focal distance (in pixels), θ the angular position and v the height on the cylinder. The radius of the cylinder is equal to the focal length of the camera to optimise the aspect ratio of the image (Shum and Szeliski 1997). The next step is to stitch the images, but for this the displacement

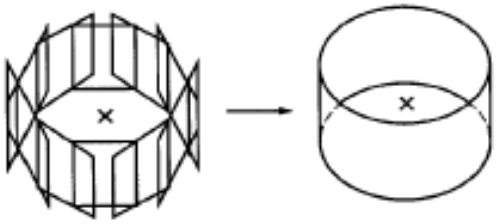


Fig. 1 The projection from the image sequence to a cylinder.

vectors $\Delta t = (t_x, t_y)$ have to be calculated for each succeeding image pair. In theory t_x can be deduced from the panning angle and $t_y = 0$, however in reality this is not true due to camera twist and not perfect panning.

Local features (see section 2.2) can be used to estimate the translation between two images. The advantage of using local features instead of the more conventional iterative maximization of the normalized correlation (McMillan and Bishop 1995; Szeliski and

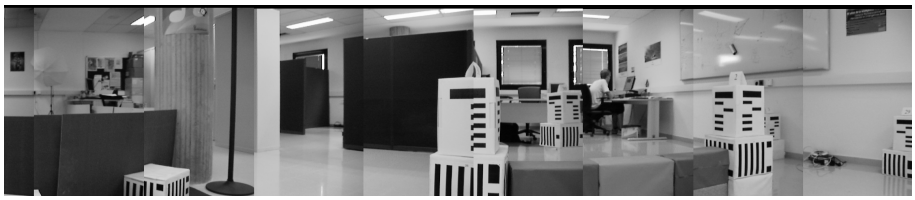


Fig. 2 Part of a panorama image created by stitching several images together. The image is made in the robot laboratory.

Shum 1997) is its lower computational complexity (provided that the local features will have to be computed anyway) and higher robustness to several image transformations such as illumination changes, noise and zoom. However, in the case of few texture in the image it is not possible to use the feature-based approach and the iterative method is used. When the translations have been calculated, the images can be stitched to produce the whole panorama.

An example of such a panorama created by stitching is shown in Figure 2. As can be seen there are still small distortions due to not perfect shifting of the images. The difference in intensity is because of the automatic camera gain. To avoid artifacts created in the stitching process, the features from the original images are used. Features from overlapping regions are only added to the constellation once. Another way to acquire panoramas is by using an omnidirectional camera. There are two approaches to do this, first by using a fish-eye lens; and secondly by using a conventional camera pointed to a hyperbolic mirror above it. These methods have some clear advantages such as the speed of creation and that no images have to be stitched, and therefore no artifacts will be introduced. A disadvantage is the lower resolution.

Finally, another alternative to acquire panoramas is using a camera ring. This is a ring of synchronized cameras and offers a high speed of acquisition without sacrificing the high resolution. A disadvantage of this method is the high price of the whole system.

2.2 Local Visual Features

Local visual features can be points or regions of an image which correspond to a local extrema function over it. The main interest of these features is that are detectable under several transformations and illumination changes. This robustness makes these features very suitable for the purpose of matching and recognition. Moreover, representations made with such local features are robust to partial occlusions and background clutter. Extracting features from an image reduces the dimensionality of the information and adds robustness against noise, aliasing and acquisition conditions.

The feature region detectors Maximally Stable Extremal Regions (MSER) and Differences of Gaussians (DoG) are used in this work to test the homing method because they are fast to compute and yet robust. These detectors, among others, were also used in (Ramisa 2006) and we briefly describe them below.

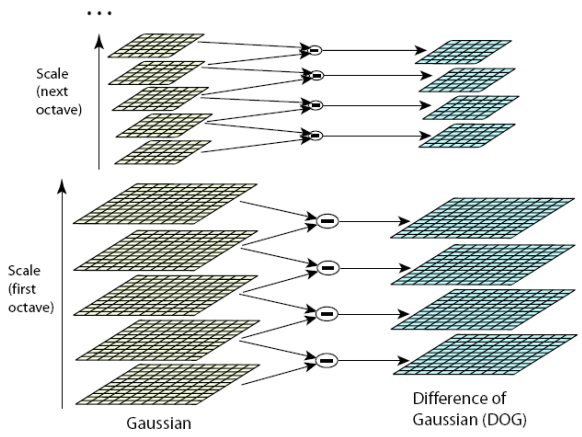


Fig. 3 At the left the initial image is incrementally convolved with Gaussians. The adjacent image scales are subtracted to produce the DoGs, which are shown at the right. After each octave, the Gaussian image is down-sampled by a factor of 2, and the process repeated. (Taken from Lowe (2004)).

2.2.1 Differences of Gaussians

The Scale-Invariant Feature Transform (SIFT) algorithm proposed in (Lowe 1999, 2004) is based on a biologically inspired model of complex neurons in the primary visual cortex proposed by Edelman et al (1997). These neurons are activated by a gradient in a particular orientation if it appears within a small range of positions in the retina.

Although the SIFT algorithm includes both an interest region detector and a descriptor, we are only interested in the detector part for this work: the extrema of the differences of Gaussians. The standalone version of this detector is known in the computer vision literature as Differences of Gaussians or simply *DoG*.

Differences of Gaussians D are produced subtracting every two neighbour levels of the scale-space of the image, separated by a factor k:

$$D(x, y, \sigma) = (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y), \tag{2}$$

where $G(x, y, \sigma)$ is a Gaussian kernel with sigma σ and $I(x, y)$ is the input image. Figure 3 shows an efficient approach to construct D. To avoid detecting multiple times the same feature at different scales, they are only detected at their characteristic scale (Lindeberg and Gårding 1997). Local extrema of D are detected by comparing each sample point to its eight neighbours in the current image and the nine neighbours in the above and below scales of the DoG. The point is selected only if it is the maximum or minimum in its neighbourhood. Finally unstable feature points are rejected. These correspond, for example, to feature points localised along an edge or feature points with low contrast.

2.2.2 MSER

The Maximally Stable Extremal Regions (MSER) proposed by Matas et al (2002) can be defined informally as image regions in which the pixels have an intensity value

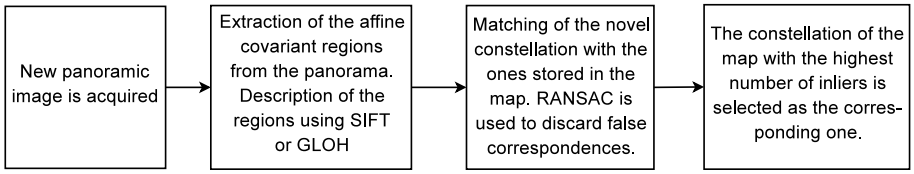


Fig. 4 Overview of the Ramisa (2006) global localization algorithm.

much higher or lower than neighboring pixels. Although apparently very simple, MSER feature points are very stable to change of viewpoint (they are perspective-invariant) and to affine illumination changes.

Furthermore, the algorithm proposed by Matas et al (2002) to compute the MSER feature points has a near linear complexity. The algorithm works as follows: First the pixels are sorted by intensity, then the pixels are placed in the image (in decreasing or increasing order) and the list of connected components and their areas are maintained using an efficient union find algorithm. Each connected component is stored as a function of intensity. By doing intensity thresholds we find the parts of the function where no changes in the area of connected components occur, i.e. they are not merged with others. These parts are the maximally stable extremal regions. Murphy-Chutorian and Trivedi (2006) propose an even more efficient version of the algorithm to compute MSER feature points using N-Tree Disjoint-Set Forests structure.

The MSER detector was tested by Mikolajczyk et al (2005) and found to be one of the best in their repeatability experiments. A notable advantage of this method over DoG is that the regions found are much more robust and faster to compute. On the negative side, MSER feature points are usually sparse, which makes this type of feature points unsuitable (at least when used without complementary features) for certain applications such as object localization and recognition (Vinyals et al 2007).

2.3 Global Localization Method

The ALV homing technique is used in this work to complement the global localization method described in (Ramisa 2006) and (Ramisa et al 2008). This previous work is similar to the one of Tapus and Siegwart (2005), where the authors defined a fingerprint of a room as a character sequence where each character describes a certain type of feature in the relative position of the panorama and a 2D range laser scan. However, Ramisa (2006) and Ramisa et al (2008) used only vision sensors, and the fingerprint consisted of a constellation of feature regions extracted from a panoramic image of the room. Such representation is convenient for many reasons. In the first place, using local feature points yields a signature robust to a certain degree of change in the scene, that can handle occlusions, change in the point of view or variations in the objects present in the room. To find the current location of the robot in a map, first a panoramic image must be acquired, and local feature regions have to be extracted from the image and described with a local region descriptor such as SIFT (Lowe 2004). Once the new signature has been computed, it is compared to each signature of the map by matching feature region descriptors and then computing the essential matrix with the eight point algorithm and RANSAC to filter the outliers. If the robot is moving in

a flat environment, that is usually the case in indoor environments, the four points algorithm with flat-world assumption can be used to obtain a more robust estimate of the essential matrix as done by Valgren and Lilienthal (2008). An overview of the method can be seen in Figure 4.

2.4 Average Landmark Vector (ALV)

In this section we describe the biologically inspired homing technique Average Landmark Vector by Lambrinos et al (1998, 2000). The ALV is defined as the average of the landmark (or feature) position vectors:

$$ALV(F, \overrightarrow{x}) = \frac{1}{n} \sum_{i=0}^{n} \overrightarrow{f_i} \tag{3}$$

Where $F = \{\overrightarrow{f_1}, \overrightarrow{f_2}, \ldots, \overrightarrow{f_n}\}$ is the collection of features that define the signature taken at the current position $\overrightarrow{x}$ and f_i are the coordinates of the i^{th} landmark position vector. In this equation F contains the global feature positions to explain and proof the homing technique. This is the robot centred version, but it is made world centred by subtracting the current position $\overrightarrow{x}$ to easily proof that the homing technique works :

$$ALV(F, \overrightarrow{x}) = \frac{1}{n} \sum_{i=0}^{i=n} \overrightarrow{f_i} - \overrightarrow{x} \tag{4}$$

To differentiate between the world coordinate system and the (self centred) coordinate system of the robot, the home vector is defined as follows:

$$homing(F, \overrightarrow{x}, \overrightarrow{d}) = ALV(F, \overrightarrow{x}) - ALV(F, \overrightarrow{d}) \tag{5}$$

Where $\overrightarrow{x}$ is the current location of the robot and $\overrightarrow{d}$ the destination. When the ALV functions are substituted by Eqn. 4 then $\overrightarrow{d} - \overrightarrow{x}$ remains, which is exactly the home vector. Figure 5 shows an example of the calculation of the home vector. To simplify the image only the average landmark (the gray square) is shown. In this example it is also assumed that the depth of the landmarks is known. The ALVs are calculated for the current (C) and the $Home$ position, these are A_1 and A_2 respectively. The home vector (H) is calculated by subtracting the ALV at the destination position (A_2) from the ALV at the current position (A_1). This results in the home vector H which points to the destination location.

One important prerequisite of the ALV is that it is necessary to have the panoramic images aligned to an external compass reference before computing the homing direction. The Sahara ant *Cataglyphis*, for example, use the polarization patterns of the blue sky to obtain the compass direction (Wehner 1994).

ALV homing does not work when the ALV at the current location and at the goal location are the same (after correction for orientation differences), because this results in a zero vector. An exceptional theoretical case in which this could happen is when the ALV point, the current location and the goal location are aligned, in practice however this is very unlikely. To let the robot move anyway in such situations a random vector could be used to move the robot a small distance, and then continue the homing procedure.

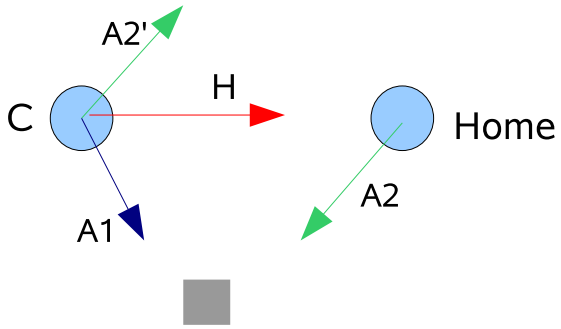


Fig. 5 The calculation of the home vector. Both ALVs (A_1 and A_2) point to the average feature position, which is drawn as a gray block. The home vector (H) is calculated by subtracting the ALV at the destination location (A_2) from the ALV at the current location (A_1). This subtraction is shown, by the addition of the reverse vector, A'_2, to A_1. The robots are aligned in this example.

In this work we propose to use the ALV method with natural feature points automatically extracted form images acquired with the mobile robot camera, without the need of artificial landmarks in the environment.

The feature points evaluated are the Differences of Gaussians from Lowe (2004) and the Maximally Stable Extremal Regions from Matas et al (2002). Only the x and z coordinates of the feature points are used to compute the ALV because of the flat world assumption. These local feature points possess qualities which make them interesting for the ALV. In the first place they are fast to compute (and even faster hardware-based approaches are being built), the second is that many higher-level processes are based on information from these interesting regions. Examples could be global localization (Ramisa et al 2008; Valgren and Lilienthal 2008) or object recognition (Lowe 2004; Csurka et al 2004). Therefore there is no overhead in reusing them for the ALV. As a way to solve the constant orientation prerequisite, in our work all test panoramic images have been acquired with the robot facing a constant direction as is common practice in similar works (Möller et al 2001; Hafner and Moller 2001). In order to apply the ALV method in a navigation experiment a magnetic compass, or another system to acquire the global orientation, is required to align the panoramas.

3 Related Work

To the best of our knowledge no other work has addressed the combination of the ALV homing method with invariant feature points such as the MSER or the DoG.

So far, in most works that studied the ALV homing method, artificial landmarks have been used. For example Lambrinos et al (2000) used as landmarks four black vertical cylinders, and in (Möller 2000) experiments were done inside of a white box with several wide black vertical stripes on the walls. Möller et al (2001) did extensive experiments in a desert type outdoor scenario with four black cylinders as landmarks. In this same work an experiment was attempted in an indoor scenario. Natural landmarks where found by vertically averaging a certain area of the image and finding edges (i.e. intensity jumps) in the unidimensional graylevel profile.

Hafner and Moller (2001) investigated if a Multi-Layer Perceptron with backpropagation and a Perceptron with Delta Rule were able to learn a homing strategy both in simulation and in real world experiments. For the real-world experiments panoramic images acquired by the robot camera were reduced to a single line by vertically averaging (similarly to what Möller et al (2001) did), thus the input of the neural networks is a unidimensional image. Both neural networks successfully learned a homing strategy with the same characteristics as ALV.

Usher et al (2003) used a version of ALV augmented with depth information to guide a car-like vehicle in an outdoor experiment. Landmarks were salient color blobs and the depth information was acquired directly from the distance of the landmark to the center of the omnidirectional image (no unwrapping is performed) using a flat-world assumption. The authors performed real-world experiments using red traffic cones (witch hat model) as landmarks.

Vardy (2005) did an extensive study for a variety of biologically plausible visual homing methods in his PhD thesis, both for local and associative methods, in a real office environment. Among the methods evaluated in his work, there is the one proposed in Hafner and Moller (2001), referred to as *Center of Mass ALV*. In the experiments it performed similarly to other local homing methods, although it was found that an extra learning phase was necessary to determine which area of the panoramic image should be used to generate the unidimensional image in certain environments.

4 Experiments Performed and Results Obtained

4.1 Simulation

To evaluate how well the ALV homing method works with our type of visual features, a series of simulation experiments were performed first. Here we report the most important findings of these experiments. A more detailed explanation and discussion of the simulation experiments can be found in (Goldhoorn et al 2007b; Goldhoorn 2008).

The experiments were done in a simulated environment (see Fig. 6) with different distributions of feature points. The environment is a room composed of a flat floor, in which the robot moves, and four walls. We have done experiments in this room changing the number of visible walls. The simulated robot was said to be successful if it found the destination point within the following three limitations:

1. The robot is not allowed to use more than 2000 steps (iterations)
2. The projection of the world should not be empty more than five times in a row (in that case either the previous home vector or a vector with random orientation and length was used)
3. The robot should travel at most a distance ten times the Euclidean distance between the start and destination position.

Although the feature points used are robust to most image variations, there are almost always changes due to noise in the localization or occlusions.

Adding Gaussian noise to the positions of the feature points with a standard deviation of 0.001 m or less resulted in a 90% successful runs. However a standard deviation of 0.05 m or more resulted in only 5% or less of successful runs.

Occlusions were simulated by removing randomly chosen feature points before every projection. Removing 50% of the feature points resulted in a mean success rate of 85%.

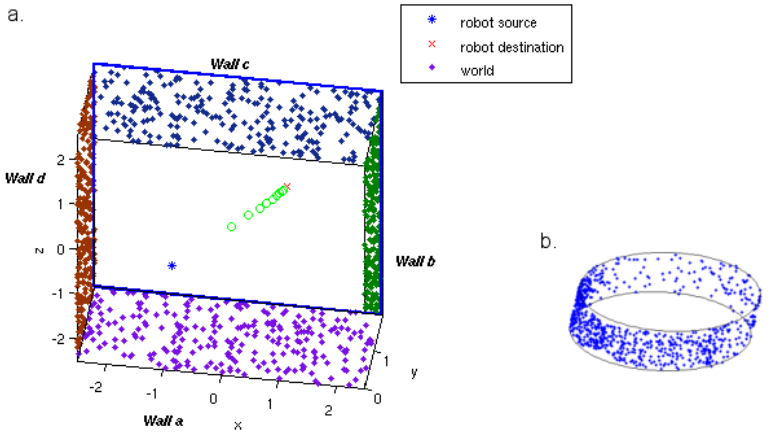


Fig. 6 a) The simulated environment with uniformly randomly spread feature points. b) Panoramic projection of the world used as input for the robot homing system.

The method also was robust to adding randomly placed feature points, which can be thought of as reappearing previously occluded objects.

Having more reliable feature points present in the world increases the performance of the robot (higher success rate, less iterations and a smaller difference with the ideal distance). For the simulation the range for the number of feature points is between 500 and 1000 for a success rate of 100%. Although having only 20 feature points in the world still resulted in 50% to 80% successful runs. However it has to be taken into account that these runs were without any noise and without any other disturbances.

Because no depth is used, the ALV method implies an equal distance assumption of the landmarks. Franz et al (1998) also mentions the isotropic feature distribution, which can explain why results in a world with only one wall were worse than in the other configurations. The robot used more iterations when more feature points were removed, but this was expected since the ALV every time has a different error.

From these experiments can be concluded that visual feature points are suitable for the ALV for visual homing. Therefore the next step was to try this method on a real robot.

4.2 IIIA Panoramas Database

This section first explains the experimental setup and robot used, then the results are presented and discussed.

4.2.1 Experimental Setup

In these experiments several panorama were acquired at a grid of known points in the rooms. The orientation of the robot was kept constant for each panorama so no alignment step is necessary between the panoramas.

Three types of landmarks/feature points were used: 1) DoG feature points; 2) MSER feature points; and, only in the robot laboratory, 3) artificial landmarks. The

experiments were done in three rooms of different sizes: the *robot laboratory* , the *square room* and the *corridor*. A scaled map of the rooms can be seen in Figures 8, 10 and 11.

The locations where the panoramas were created are marked as circles with its identifying number and a line starting at the center of the circle and pointing to the direction of the estimated home vector. The home location is shown as a red circle without line and is also indicated in the figures captions. The biggest objects in the rooms, such as desks, are also shown in the maps to give a rough idea of the environment. Finally, the squares in Figure 8 show the landmarks positions and its ID number.

Like in the simulation, only the direction of a feature is known and not its distance, therefore the home vector will not contain distance information either. The home angle calculated by the homing method is compared to the ground truth home angle which is calculated by geometry.

$$\theta_{\text{diff}}(h_{\text{h}}, h_{\text{c}}) = \min\left(|h_{\text{c}} - h_{\text{h}}|; 360 - |h_{\text{c}} - h_{\text{h}}|\right) \tag{6}$$

All angles are in degrees and counter-clockwise; h_{c} is the correct homing direction calculated by using the positions (geometry), and h_{h} is computed by the homing method. To find out how well the method works for each room and each type of feature, all the panorama positions per data set are used. For each data set (the *square room*, the *robot laboratory* and the *corridor*) all the locations where a panorama was created are used to calculate the home vector to each of the other locations. From the error calculated with Eqn. 6 for each possible panorama pairings in one room, the mean, median, standard deviation and a score are calculated. The score is calculated by using the proportion of the maximum error and ranges between 0 and 1 where 1 is best. Namely:

$$s = 1 - \frac{\sum_{i=1}^{n} \sum_{j=1; i \neq j}^{n} \theta_{\text{diff}}(h_{\text{h}}(P_i, P_j), h_{\text{c}}(P_i, P_j))}{180n(n-1)} \tag{7}$$

where n is the number of panoramas in the set and P the set of panoramas. The numerator is the sum of the difference of the home angle calculated by the ALV homing method and by geometry. This difference, i.e. error, is calculated for each panorama pair, which in total are $n(n-1)$ pairs. The sum of errors is divided by that factor to get an average and, to normalise the score between 0 and 1, it is also divided by the maximum possible error, which is 180°.

4.2.2 Robot

A Pioneer 2AT robot (Figure 7.a) is used with a pan tilt unit (Directed Perception PTU 46-70) mounted on it and on top of this PTU a camera (Sony DFW-VL500; with a resolution of 640 × 480 pixels). The robot is controlled from a laptop (Acer Travelmate C110 with an Intel Pentium M 1000MHz, 799 MHz, 760 MB RAM) placed on top of the robot. The programs, written in C++, are run under Microsoft Windows XP.

The experiments were done in three mentioned different areas in the IIIA-CSIC research center. The room in which most experiments were done is the robotics laboratory. The panorama in Fig. 2 shows this room as seen from the robot and in Figure 8 a map can be seen. As can be observed in the figure, artificial landmarks are present in the room. These landmarks were used in a set of experiments for comparison purposes with the local feature based approach.

(a) (b)

Fig. 7 (a) The Pioneer 2AT robot as used in the experiments. A pan tilt unit is mounted on the robot with a camera on top. (b) An example of a landmark in the *robotics laboratory*.

4.2.3 Landmarks

In order to compare our proposed approach to an artificial landmark based one, extra experiments were done using six artificial landmarks in the *robotics laboratory* (see Figure 7.b) available from previous experiments (Busquets 2003; Busquets et al 2003).

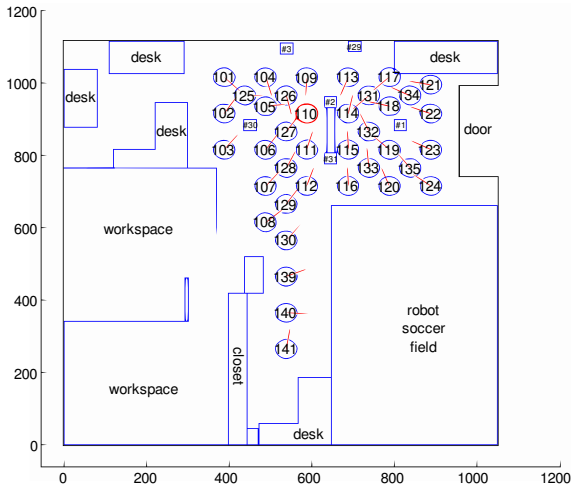
The landmarks contain a bar code from which an ID number can be extracted. Since the size of the bars is known, the distance to the landmark can be calculated. In order to make the artificial landmark approach comparable to the feature based one, neither the landmark number (for matching) nor the distance information was used in our experiments.

4.2.4 Results

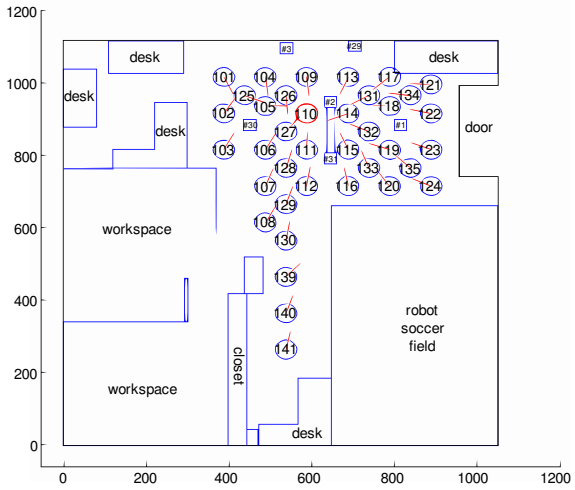
When calculating the home vector between two points, for example a and b, the home vector from a to b will obviously always point in opposite direction of the home vector from b to a. This means that these are dependent values and therefore only one of them was used in the analysis. Next we discuss the results for the three different areas.

Robotics laboratory: Most panoramas, 38 in total, were acquired in the *robotics laboratory*, a room of 10.5 m $\times$ 11.2 m. Only the half of the room is really used for this experiment because the other part is filled with working places and the robot soccer field as can be seen in Figure 8.

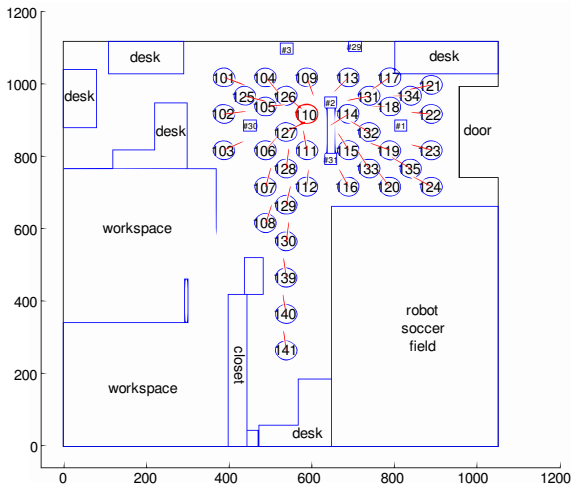
The home vectors have an error equal to or lower than 90° in 89.3% of the cases when the DoG detector was used, 92.6% for the MSER detector and 99.6% when the landmarks were used. An error of 10° or less was obtained in 22.6% of the cases for the DoG detector, 32.7% for the MSER detector and 64.3% for landmarks. Table 1 shows the results for each type of detector used. The homing errors for the three methods are all significantly different ($p < 0.001$) according to the rank sum test, and the t-test after bootstrapping ($n = 1000$). From this can be concluded that the homing method worked best with the artificial landmarks, as expected, and worst with the DoG detector.



(a)



(b)



(c)

Fig. 8 Homing to panorama 110 in the *robotics laboratory* using DoG feature points (a), MSER feature points (b) and the landmarks (c). All measures are in cm.

	DoG	MSER	Landmarks
Mean error	35.60°	27.84°	14.88°
Median error	22.85°	16.03°	10.17°
Standard deviation	36.67°	35.51°	14.86°
Score	0.8022	0.8454	0.9173
Best home	117	117	110

Table 1 The homing error using the panoramas from the *robotics laboratory*. The *best home* field shows the number of the panorama (see Figure 8 for the numbers in the *robotics laboratory*), which when chosen as home, resulted in the lowest average error.

Fig. 9 Part of the panorama 137 from the *square room*.

	DoG	MSER
Mean error	13.78°	9.65°
Median error	12.00°	12.03°
Standard deviation	11.31°	7.84°
Score	0.9234	0.9464
Best home	138	138

Table 2 The error of the homing method using the panoramas which were made in the *square room*.

Square room: The square room is 4.0 m × 3.4 m big. Figure 9 shows a panorama made in this room. Figure 10 shows the map of the room and the home vectors to panorama 137. Table 2 shows the statistics of the homing method using both feature types. MSER feature points achieved lower error rates than DoG feature points, but this is not significant (confirmed by the rank sum test and the t-test) and it must be noted that only three panoramas were created in this room.

Corridor: Although the simulation showed that the ALV homing method works better in square rooms, we wanted to find out what the impact of a very long and very narrow room in a real environment would have on the method. A corridor was chosen for that reason as last experiment room. The part of the corridor in which the robot moved is 2.2 m wide and about 22.5 m long. In Figure 11 the map of the corridor can be seen. Additionally, Figure 12 shows the panoramas acquired in the corridor.

In Figure 11 the home vectors to panorama 203 are shown. An error of 90° or less was obtained in 73.3% of the cases for both feature types, an error of 10° or less was only obtained in one case (6.7%). Table 3 shows the average error of this data set; the differences between the results with DoG and MSER are not statistically significant.

From the results at the different rooms, it can be seen that the ALV homing method worked better in both the *square room* and the *robotics laboratory* than in the *corridor*. This difference might be explained by the previously found conclusion, in the simulated experiment (Section 4.1), that the method works better in approximately square rooms.

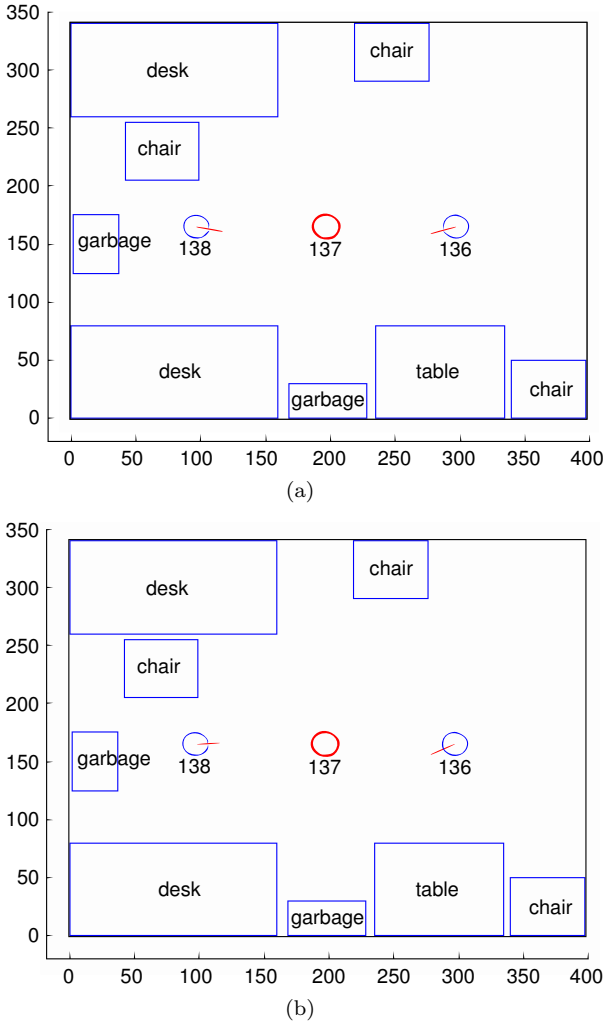


Fig. 10 Homing to panorama 137 in the *square room* (a) using DoG points and (b) MSER points. All measures are in cm.

	DoG	MSER
Mean error	56.26°	52.67°
Median error	44.58°	35.71°
Standard deviation	43.64°	44.90°
Score	0.6874	0.7074
Best home	203	200

Table 3 The average error of the homing method in the *corridor* for the different feature types.

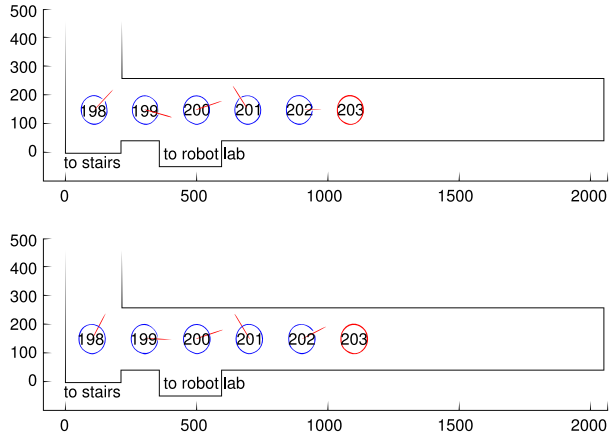


Fig. 11 Homing to panorama 203 in the *corridor* using (a) DoG feature points and (b) MSER feature points. All measures are in cm.

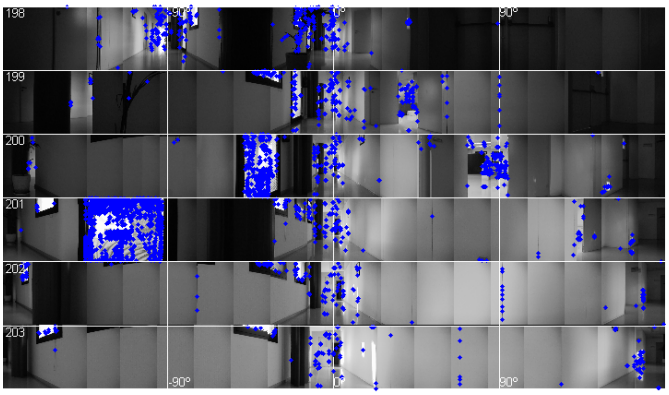


Fig. 12 All the panoramas made in the corridor. The dots are MSER feature points.

This is due to the equal distance assumption. In what follows we provide additional details regarding the analysis of the results.

Corridor results: The panoramas acquired in the *corridor* (Figure 12) show that there are several *disturbing* factors on which numerous MSER feature points were found. Panorama 198 is the only panorama taken at a corridor intersection, and therefore the MSER detector finds considerable more feature points than in the other panoramas. In panoramas 200 and 201 a door with blinds is visible, and the MSER detector also found a great amount of feature points on these blinds; in panoramas 199 and 200 the robotics laboratory is visible through an open door which again has many feature points. Figures 11.a and 11.b confirm this, because here the home direction from panoramas 199, 200

and 202 to 203 were good, but from panoramas 198 and 201 really bad. The reason for this is that in panoramas 198 and 201 the most MSER feature points are located on one side only (as commented earlier in this paragraph), while in the other panoramas (199, 200, 202 and the home panorama 203) the feature points are more or less equally distributed. Although only the MSER detector was mentioned here, the DoG detector generated even more feature points, but with a more or less similar distribution.

In Table 4 (Annex), can be seen that the best *corridor* of the IIIA data sets is at rank 25, but this is below the best of the data sets *robot lab* and *square room*.

Upper and lower part: In an attempt to improve the results, the view of the image was limited to only the lower half of the panorama. This part contains objects which are closer to the robot and therefore decrease the size of the visible world, for this reason a room may look more square.

In the *robotics laboratory* using only the lower half of the panorama resulted in a lower error than using all feature points of the panorama ($p < 0.001$ with the t-test and the rank sum test for both DoG and MSER). For the other rooms there was no significant difference in performance. Also here the best results were when the MSER detector was used ($p < 0.005$ for the *robotics laboratory* and *corridor*) except for the *square room* where DoG was the best detector ($p < 0.001$, rank sum test).

Also the use of only the upper half part of the panorama was tested, but these results were significantly worse than using the whole panorama for the *robotics laboratory* ($p < 0.001$, t-test and rank sum test). There was again no significant difference in the *square room* and *corridor*.

Depth: When only the position of the feature points on the panorama are used then only the direction of the home vector can be found. However, when the distance to the feature points is available, a more precise estimation of the distance to home can be calculated.

4.3 Vardy's Panorama Database

As an additional test, we used the image database of Vardy (2005)[1] which he discussed and used to test several homing techniques in his thesis.

Vardy's image database consists of panoramic images acquired over a grid of equally separated feature points from the hall and the robotics laboratory of Bielefeld University. He created six data sets of the laboratory and two of the hall, all under slightly different conditions, such as the amount of light and added objects. In the robotic laboratory the data set consisted of a 10×17 image grid with 30 cm separation between each image (horizontally and vertically); in the hall 10×21 images in a grid were created per data set with 50 cm separation between images. In contrast to the *IIIA database*, Vardy's database was acquired with an *ImagingSource DFK 4303* camera pointing towards an hyperbolic mirror. This system directly acquires omnidirectional images, and therefore spares the panorama creation step. However it suffers from a much lower resolution. Figure 13 shows a panorama from the *hall1* data set. In our experiments, first all the feature points are extracted from the images. As can be seen

[1] Vardy's *Panoramic Image Database* is available at `http://www.ti.uni-bielefeld.de/html/research/avardy/index.html`.

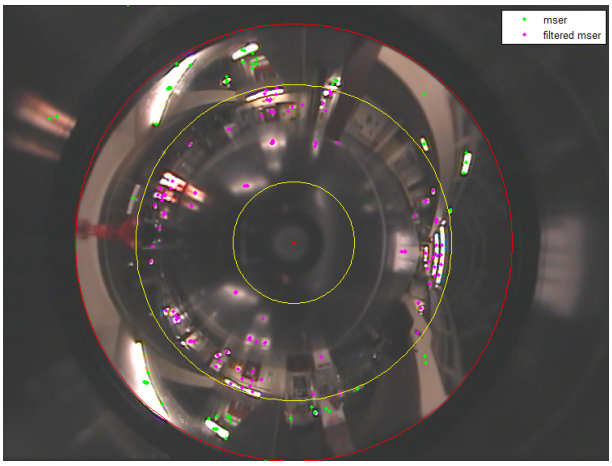


Fig. 13 A panorama from Vardy's image database. The outer red circle shows the border of the parabolic mirror, the two inner yellow circles show the 20° line above and below the horizon. The points show the location of the MSER feature points; the filtered feature points are the ones between the yellow circles (best viewed in color).

in Figure 13, the image also contains non relevant parts which lay outside the mirror. To focus on the informative area of the image, the field of view is reduced to a limited number of degrees above and below the horizon, which is the line between the centre of the spherical mirror and the outer circle of the mirror. Only feature points which fall in this area are used for the homing method.

The vector of a feature has its origin in the image centre (shown as the red dot in Figure 13) and points to the feature point. These vectors have to be normalised to 1 before calculating the ALV, because the length of the vectors only shows the distance in pixels on the image. After this, the ALVs and the home vector can be calculated as described in section 2.4. Table 4 in the Annex shows the results with all data sets sorted by score.

4.3.1 Results

As can be seen from Table 4 (Annex), the scores for Vardy data set vary from 0.85 to 0.3 and the home angle error from $28.2^\circ \pm 27.6$ to $126.0^\circ \pm 43.3$. The results are worse than the results with the previously discussed data sets, but it must be noticed that Vardy's data sets contain more samples.

It can be seen that a wider vertical view angle gives better results. When MSER feature points were used, a view angle of 15° (above and below the horizon) worked significantly better than a lower angle ($p < 0.001$, t-test and rank sum test). For all data sets except for *doorlit* and *hall1* the best view angle was 20°. This is also the case when DoG feature points were used, except for the data sets *day*, *hall2* and *screen*. In the data set *day* the difference was not significant enough; using a view angle of 5° had the best results in the sets *hall2* ($p < 0.001$, rank sum test) and *screen* ($p < 0.05$, rank sum test) when DoG feature points were used.

It is also clear from Table 4 (Annex) that the performance is better when using the MSER detector than the DoG detector. This difference is significant for all data sets

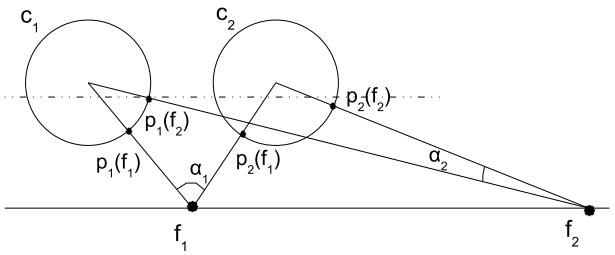


Fig. 14 The position where points are projected in different panoramas varies less (and therefore are less informative) if the points are far away. This is a problem for narrow and long corridors with most texture at the extremes.

with a view of more than 5° above and below the horizon (using the *t*-test and rank sum test; $p < 0.001$). It also can be seen from the table that the best of the IIIA sets are all above the data sets of Vardy, however this is only significant for the *robotics laboratory*. Comparing the results of Vardy's data set with the results of the IIIA data set is difficult because of several reasons. First of all there are two big differences between them: the environments and the way the panoramas are made. The rooms could be assumed to be quite similar since they both are flat "office like" with several desks, chairs and computers, but the landmarks present in the *robotics laboratory* are not present in Vardy's rooms for example.

4.4 Overall Discussion of Experimental Results

As has been seen, with the real robot experiments the ALV homing method gave very positive results. The best results were obtained with the panoramas from the *square room*. The results from the *corridor* were worst, as expected. In the simulation already was found that the performance of the homing method is better in square rooms than in rooms with big differences in width and length. The problem with long rooms such as a corridor is that the projections of the feature points onto a panorama are closer to each other the further away they are from the robot (see Figure 14).

Looking at the difference in performance using DoG and MSER feature points it can be concluded that the use of MSER feature points significantly outperforms the use of DoG feature points. The artificial landmarks in the *robotics laboratory* were used to find out how well the method worked in comparison with invariant feature points. The results with the artificial landmarks were significantly better than using invariant feature points, the error was about 7° less than using MSER feature points (with only the lower half of the panorama).

Normally one should expect the homing method to work worse when the distance between the current location and the home is lower, but this relation could not be found. This might be because the room is too small or because objects occlude a big part of the field. Further work would be needed to find out if there is any relation between the distance and error.

An attempt to improve the results was done by trying to make the rooms, such as the corridor, more square by only using the lower half of the panorama, because then the closer objects are more prominent. This however had no significant improvement

in the *corridor*, and neither in the *square room*. Only in the *robotics laboratory* there was a significant lower error ($p < 0.001$).

The images of Vardy (2005) data sets were also used to test the ALV homing method. Although the different panorama acquisition system, in practice the performance of these sets was not much worse than the results of the IIIA ones. From these images also SIFT and MSER feature points were extracted and used to calculate the ALV. It was found that using almost the whole image (20° above and below the horizon) resulted in the best performance.

The scores (with 1 being best and 0 begin worst) of the IIIA data sets varied from 0.67 to 0.96, whereas the results of Vardy's data sets varied from 0.30 to 0.85 (see Table 4 in the Annex). Looking at the best parameters however, such as using the lower half of the panorama for the IIIA data sets and using a view angle of 20° above and below the horizon of Vardy's data, the scores of the IIIA data sets vary from 0.73 to 0.96 and the scores of Vardy's data sets from 0.67 to 0.85. This shows that the method performs almost as well in the different rooms and with the different types of panoramas.

Finally some comparison to other work can be made, however in most works other error measurements are used such as the distance at which it stops from home. In this work however no such experiments have been done yet. Hafner (2001) also did experiments in an office environment in a grid. After off-line learning the average error was smaller than 90° in 92% of the cases and smaller than 45° in more than 69%. This is comparable to the results in the *robotics laboratory* for the DoG feature points, and our results for using MSER feature points were even better.

The experiments by Franz et al (1998) were done in a 118 cm × 110 cm environment but the catchment area was relatively smaller than the catchment area of the IIIA data sets. Their algorithm performed robustly up to an average distance of 15 cm. They also mention experiments done in an office environment in which the algorithm performed robustly until about 2 m.

5 Conclusion and Future Work

In this work a method to complement the global localization system of Ramisa et al (2008) is proposed, where feature regions computed in a panoramic image were used to characterise a room. The purpose of that method was directing the robot from one of the map nodes to the next with the minimal cost (i.e. no matches have to be established between visual feature points of the images). The method does not rely in any form of artificial landmarks contrarily to previous works, and uses only natural landmarks extracted from the panoramic images acquired with the robot visual system, therefore it can be applied in unprepared environments.

Also, when the robot has several likely hypotheses about its current location (i.e. room) then homing can be used to return to the position where the most likely panorama from the database was made. If the hypothesis with the highest probability was correct, then the panorama at that location should be more similar to the panorama from the database, making it the significantly best hypothesis. In the case where an incorrect panorama was chosen from the database, then the same steps should be taken as before in order to find out in which room the robot actually is.

Although there are several methods to do homing, such as the 1D method of Hong et al (1991), warping (Franz et al 1998) or snapshots (Lambrinos et al 2000), the *ALV*

homing method (Lambrinos et al 1998, 2000) has been used mainly because of its simplicity and low computational complexity.

In order to evaluate the proposed method, initial experiments using a simulated environment were conducted and later it was tested in a real world scenario. The real world experiments were done with panoramas acquired in three different rooms at the IIIA research center.

The locations at which the robot acquired the panoramas were measured manually and used to calculate the ground truth homing directions, which were then used to verify the homing method results. The panoramas were created with the camera on a pan tilt unit which rotates around a fixed point to get images from all different directions. Next, these images were combined to create the panorama. Feature points from these images were extracted to be used by the homing method. Two types of invariant feature detectors were tested: the Difference of Gaussians extrema (DoG) of Lowe (2004) and the Maximally Stable Extremal Regions (MSER) of Matas et al (2002). Only the horizontal location of the feature points was used, i.e. the cylindrical angle, and not height, nor depth.

The ALV homing was found to be a good working method, however the method performed worse in rooms where the width and length differ greatly. This has been explained by the way the feature points are projected on the panorama and by the *equal distance assumption* (Franz et al 1998).

Vardy (2005) discusses biologically based homing methods in his thesis and also did experiments with several of them. In his work he used panoramas that were made with a camera pointed to a parabolic mirror. The advantage of creating a panorama like this is the speed of creation. Whereas with our method first images from several angles had to be retrieved and then stitched to create a high resolution panorama. In order to compare both methods of panorama acquisition additional experiments using Vardy's data sets were performed using the SIFT and MSER feature points.

When comparing the results of IIIA data sets and Vardy's data sets we can see that the ALV homing method performs slightly better on the IIIA data sets, but the difference is not significant. There might be several reasons to explain this, such as the difference in resolution, the camera or the environment. For these reasons it cannot be concluded that having a panorama with a higher resolution made by rotating a camera or a camera ring is much better for ALV homing than using a camera and a parabolic mirror. On the other hand, the faster creation of the parabolic panoramas might be more important than the slightly better performance.

Regarding the feature types, in our experiments MSER significantly outperformed SIFT. Also Mikolajczyk et al (2005) confirmed that MSER is one of the most robust feature detectors. The artificial landmarks in the *robotics laboratory* were used to compare the local feature approach with the more traditional artificial landmarks. As expected, the results with the artificial landmarks were significantly better than with the invariant feature points since they are less affected by occlusions and viewpoint changes. The error with the artificial landmarks was about 7° less than using the MSER detector (in only the lower half of the panorama). However this difference seems low enough to justify the applicability of the presented homing method, because does not require setting up the environment by placing artificial landmarks.

5.1 Future work

Clearly, the next step is the evaluation of the method in real navigation experiments combined with Ramisa et al (2008) global localization system.

Hafner (2001) mentioned that a magnetic compass does not work very well inside buildings, therefore she used camera information to compensate for that. Extra experiments should be done to verify the stability of the compass orientation (required for the ALV method). Other options to recover orientation from the visual sensors include that of Lambrinos et al (2000), who used a polarised-light compass which worked good, but needed sunlight from all directions above it and glass windows depolarise the light, therefore it cannot be used inside buildings. Vardy (2005) proposed to use the coherence of flow fields as an indicator of correct orientation, and finally Zeil et al (2003) suggested to use the difference between images to align them by using one-dimensional gradient descent.

For the homing method to work in real-time the panoramas should be created faster and therefore the use of a camera and parabolic mirror is a good option. This however should also be tested with the used global localization method of Ramisa et al (2008).

Another possible improvement could be using a machine learning method to discard spurious feature points, for example by tracking feature points in a training sequence and modeling those that do not survive long. These feature points are a source of noise for the homing method, and discarding them could improve significantly the results.

Annex

#	dataset	type	detector	mean	median	std. dev.	score	best home	n
1	square room	upper half	MSER	6,83	4,10	5,33	0,9621	138	3
2	square room	not filtered	MSER	9,65	12,03	7,84	0,9464	138	3
3	square room	not filtered	DoG	13,78	12,00	11,31	0,9234	138	3
4	robot lab	not filtered	Landmarks	14,88	10,16	14,86	0,9173	110	38
5	square room	lower half	DoG	14,94	14,52	10,75	0,9170	138	3
6	square room	lower half	MSER	20,62	25,27	8,49	0,8855	138	3
7	square room	upper half	DoG	20,91	18,96	6,64	0,8838	138	3
8	robot lab	lower half	MSER	21,96	11,09	30,05	0,8780	117	38
9	day	20	MSER	26,18	18,73	27,62	0,854	17	170
10	robot lab	lower half	DoG	26,90	13,05	34,74	0,8506	117	38
11	robot lab	not filtered	MSER	27,84	16,03	35,51	0,8454	117	38
12	screen	20	MSER	28,64	18,42	31,04	0,840	95	170
13	doorlit	15	MSER	30,69	19,35	33,38	0,829	15	170

Continued

#	dataset	type	detector	mean	median	std. dev.	score	best home	n
14	arboreal	20	MSER	34,89	23,39	35,49	0,806	50	170
15	doorlit	20	MSER	35,41	21,27	38,52	0,803	50	170
16	robot lab	not filtered	DoG	35,60	22,85	38,67	0,8022	117	38
17	arboreal	15	MSER	37,83	25,31	37,20	0,789	17	170
18	day	15	MSER	39,78	29,30	36,49	0,779	17	170
19	hall1	15	MSER	42,61	31,55	38,32	0,763	159	200
20	original	20	MSER	43,18	32,23	38,49	0,760	50	170
21	screen	15	MSER	45,71	33,75	40,43	0,746	0	170
22	hall1	10	MSER	45,81	35,12	39,05	0,745	41	200
23	twilight	20	MSER	46,21	34,90	39,72	0,743	50	170
24	doorlit	10	MSER	48,45	33,95	43,90	0,730	14	170
25	hall	lower half	MSER	48,83	39,02	41,63	0,7287	203	6
26	arboreal	10	MSER	49,97	35,14	44,80	0,722	153	170
27	robot lab	upper half	MSER	50,62	39,33	42,47	0,7188	117	38
28	winlit	20	MSER	52,39	39,58	44,07	0,708	50	170
29	corridor	not filtered	MSER	52,66	35,71	44,89	0,7074	200	6
30	robot lab	upper half	DoG	56,14	45,77	43,84	0,6881	117	38
31	corridor	not filtered	DoG	56,26	44,58	43,64	0,6874	203	6
32	corridor	lower half	DoG	56,45	49,69	42,39	0,6864	203	6
33	corridor	upper half	DoG	57,08	38,19	45,65	0,6829	203	6
34	twilight	15	MSER	57,63	44,59	46,70	0,679	153	170
35	hall1	20	MSER	58,49	48,71	44,53	0,675	99	200
36	original	15	MSER	58,53	45,11	47,56	0,674	17	170
37	chairs	20	MSER	58,92	45,23	47,55	0,672	84	170
38	corridor	upper half	MSER	59,15	42,56	46,02	0,6714	199	6
39	hall2	20	MSER	59,51	49,20	43,09	0,669	18	200
40	hall2	15	MSER	61,00	50,90	45,30	0,661	18	200
41	day	10	MSER	62,50	51,30	47,17	0,652	17	170
42	hall1	5	MSER	62,69	53,82	44,78	0,651	39	200
43	screen	10	MSER	63,43	54,30	45,02	0,647	153	170
44	screen	5	MSER	66,71	52,61	50,03	0,629	102	170
45	hall2	5	MSER	68,57	57,22	49,19	0,619	19	200
46	original	10	MSER	72,09	62,67	50,85	0,599	14	170
47	winlit	15	MSER	72,39	63,58	50,84	0,597	50	170
48	day	5	MSER	72,67	62,94	50,31	0,596	169	170
49	hall2	10	MSER	72,72	62,01	50,68	0,596	19	200
50	twilight	10	MSER	73,55	65,58	51,23	0,591	18	170
51	hall1	20	DoG	77,16	71,90	45,43	0,571	0	200
52	winlit	10	MSER	78,69	72,12	51,96	0,562	16	170
53	chairs	15	MSER	79,61	72,79	51,91	0,557	16	170
54	doorlit	5	DoG	80,15	75,59	52,15	0,554	169	170
55	doorlit	20	DoG	83,07	80,38	49,17	0,538	151	170
56	doorlit	10	DoG	83,79	81,13	50,27	0,534	169	170

Continued

#	dataset	type	detector	mean	median	std. dev.	score	best home	n
57	chairs	10	MSER	84,42	82,22	50,74	0,531	14	170
58	doorlit	15	DoG	84,44	81,34	49,75	0,530	152	170
59	chairs	20	DoG	86,15	82,29	49,48	0,521	3	170
60	screen	5	DoG	86,87	85,63	51,82	0,517	135	170
61	screen	15	DoG	88,42	85,67	51,75	0,508	135	170
62	screen	10	DoG	88,76	88,36	52,02	0,506	135	170
63	screen	20	DoG	89,25	86,89	51,88	0,504	3	170
64	hall1	15	DoG	89,27	85,64	45,78	0,504	0	200
65	chairs	15	DoG	90,33	87,09	51,16	0,498	4	170
66	arboreal	20	DoG	90,33	88,88	50,27	0,498	4	170
67	original	20	DoG	91,36	88,78	49,70	0,492	3	170
68	twilight	20	DoG	91,66	89,34	49,59	0,490	5	170
69	day	10	DoG	92,99	94,81	51,31	0,483	152	170
70	day	15	DoG	93,00	94,20	50,95	0,483	135	170
71	day	20	DoG	93,05	93,13	50,42	0,483	135	170
72	day	5	DoG	93,10	94,15	51,72	0,482	152	170
73	chairs	10	DoG	93,46	92,24	51,88	0,480	4	170
74	chairs	5	DoG	93,51	92,00	50,99	0,480	5	170
75	arboreal	15	DoG	95,20	95,05	51,65	0,471	4	170
76	twilight	15	DoG	96,44	96,35	50,29	0,464	5	170
77	winlit	5	MSER	97,11	103,58	54,36	0,460	136	170
78	original	15	DoG	97,93	97,66	49,81	0,455	4	170
79	winlit	20	DoG	98,86	99,65	44,11	0,450	0	170
80	arboreal	10	DoG	99,07	101,86	51,64	0,449	135	170
81	arboreal	5	DoG	100,55	104,77	51,59	0,441	135	170
82	twilight	10	DoG	101,25	102,97	50,17	0,437	6	170
83	hall1	10	DoG	101,86	99,59	44,55	0,434	40	200
84	original	10	DoG	101,98	105,32	50,19	0,433	4	170
85	twilight	5	DoG	102,16	105,37	49,95	0,432	6	170
86	doorlit	5	MSER	102,79	110,41	51,90	0,429	14	170
87	winlit	5	DoG	103,01	109,13	46,78	0,427	134	170
88	original	5	DoG	103,19	108,47	50,36	0,426	50	170
89	winlit	15	DoG	103,28	105,54	45,33	0,426	34	170
90	winlit	10	DoG	104,93	109,14	46,08	0,417	135	170
91	hall1	5	DoG	108,85	109,20	46,09	0,395	80	200
92	arboreal	5	MSER	112,73	122,76	48,51	0,373	14	170
93	chairs	5	MSER	116,47	126,06	46,16	0,352	14	170
94	hall2	5	DoG	116,47	130,14	49,96	0,352	198	200
95	original	5	MSER	118,50	128,27	44,93	0,341	153	170
96	hall2	20	DoG	122,09	132,21	44,38	0,321	20	200
97	twilight	5	MSER	122,21	133,26	44,17	0,321	136	170
98	hall2	10	DoG	124,42	137,90	45,31	0,308	199	200
99	hall2	15	DoG	125,99	137,57	43,33	0,300	61	200

Table 4 This table shows the results of all real world experiments (IIIA in white and Vardy in light gray) sorted by score. For the IIIA data set, the type column shows which part of the panorama has been used: all feature points (*not filtered*), only the feature points at the *lower half* of the panorama or only at the *upper half* (see Section 4.2.4). For Vardy data set, the type column shows the number of degrees above and below the horizon of the image which were used. The detector column shows which feature detector has been used to perform homing: DoG, MSER or artificial *landmarks* which were only available in the *robot laboratory*. The next three columns: mean, median and std. dev. (standard deviation) show information about the direction error of the home vector in degrees. The calculation of the score is shown in Eqn. 7; 1 being best and 0 being worst. The *best home* column shows the ID of the location of the home where to the mean error is smallest. Finally the n column shows the number of samples, i.e. different panoramas, for the data set.

References

Busquets D (2003) A multiagent approach to qualitative navigation in robotics. PhD dissertation, Universitat Politècnica de Catalunya

Busquets D, Sierra C, de Mantaras RL (2003) A multiagent approach to qualitative landmark-based navigation. Autonomous Robots 15(2):129–154

Carwright BA, Collet TS (1983) Landmark learning in bees: Experiments and models. Journal of Comparative Physiology 151:521–543

Csurka G, Bray C, Dance C, Fan L (2004) Visual categorization with bags of keypoints. Workshop on Statistical Learning in Computer Vision, ECCV pp 1–22

Edelman S, Intrator N, Poggio T (1997) Complex cells and object recognition, unpublished manuscript, University of Cornell

Franz MO, Schölkopf B, Mallot HA, , Bülthoff HH (1998) Where did i take that snapshot? scene-based homing by image matching. Biological Cybernetics 79:191–202

Goldhoorn A (2008) Solving ambiguity in global localization of autonomous robots. Master's thesis, University of Groningen

Goldhoorn A, Ramisa A, de Mantaras RL, Toledo R (2007a) Robot homing simulations using the average landmark vector method. Tech. Rep. RR-IIIA-2007-03, IIIA-CSIC, Bellaterra

Goldhoorn A, Ramisa A, de Mantaras RL, Toledo R (2007b) Using the average landmark vector method for robot homing. In: 19th International Conference of the ACIA, IOS Press, Frontiers in Artificial Intelligence and Applications, vol 163, pp 331–338

Hafner V, Moller R (2001) Learning of Visual Navigation Strategies. In: European Workshop on Learning Robots (EWLR), Prague

Hafner VV (2001) Adaptive homing: Robotic exploration tours. Adaptive Behavior 9(3/4):131–141

Hong J, Tan X, Pinette B, Weiss R, Riseman E (1991) Image-based homing. In: Proceedings of the 1991 IEEE International Conference on Robotics and Automation, pp 620 – 625

Lambrinos D, Möller R, Pfeifer R, Wehner R (1998) Landmark navigation without snapshots: the average landmark vector model. In: Elsner N, Wehner R (eds) Proc 26th Göttingen Neurobiology Conference, Thieme-Verlag

Lambrinos D, Möller R, Labhart T, Pfeifer R, Wehner R (2000) A mobile robot employing insect strategies for navigation. Robotics and Autonomous Systems 30(1-2):39–64

Lindeberg T, Gårding J (1997) Shape-adapted smoothing in estimation of 3-D shape cues from affine deformations of local 2-D brightness structure. Image Vision Comput 15(6):415–434

Lowe DG (1999) Object recognition from local scale-invariant features. In: ICCV '99: Proceedings of the International Conference on Computer Vision-Volume 2, IEEE Computer Society, Washington, DC, USA, pp 1150–1157

Lowe DG (2004) Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision 60(2):91–110

Matas J, Chum O, Martin U, Pajdla T (2002) Robust wide baseline stereo from maximally stable extremal regions. In: Rosin PL, Marshall D (eds) Proceedings of the British Machine Vision Conference, BMVA, London, UK, vol 1, pp 384–393

McMillan L, Bishop G (1995) Plenoptic modeling: an image-based rendering system. In: SIGGRAPH '95: Proceedings of the 22nd annual conference on Computer graphics and interactive techniques, ACM Press, New York, NY, USA, pp 39–46

Mikolajczyk K, Tuytelaars T, Schmid C, Zisserman A, Matas J, Schaffalitzky F, Kadir T, Gool L (2005) A comparison of affine region detectors. International Journal of Computer Vision 65:43–72(30)

Möller R (2000) Insect visual homing strategies in a robot with analog processing. Biological Cybernetics, special issue: Navigation in Biological and Artificial Systems 83:231–243

Möller R, Lambrinos D, Roggendorf T, Pfeifer R, Wehner R (2001) Insect strategies of visual homing in mobile robots. In: Webb B, Consi TR (eds) Biorobotics. Methods and Applications, AAAI Press / MIT Press, pp 37–66

Murphy-Chutorian E, Trivedi M (2006) N-tree disjoint-set forests for maximally stable extremal regions. In: Proceedings of the British Machine Vision Conference

Ramisa A (2006) Qualitative navigation using panoramas. Master's thesis, Universitat Autònoma de Barcelona

Ramisa A, Tapus A, Lopez de Mantaras R, Toledo R (2008) Mobile robot localization using panoramic vision and combinations of feature region detectors. In: Robotics and Automation, 2008. ICRA 2008. IEEE International Conference on, pp 538–543

Shum H, Szeliski R (1997) Panoramic image mosaics. Tech. Rep. MSR-TR-97-23, Microsoft Research

Szeliski R, Shum HY (1997) Creating full view panoramic image mosaics and environment maps. In: SIGGRAPH '97: Proceedings of the 24th annual conference on Computer graphics and interactive techniques, ACM Press/Addison-Wesley Publishing Co., New York, NY, USA, vol 31, pp 251–258

Tapus A, Siegwart R (2005) Incremental robot mapping with fingerprints of places. In: Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Edmonton, Canada, pp 2429–2434

Usher K, Ridley P, Corke P (2003) Visual servoing of a car-like vehicle-an application of omnidirectional vision. In: Robotics and Automation, 2003. Proceedings. ICRA'03. IEEE International Conference on, vol 3

Valgren C, Lilienthal AJ (2008) Incremental spectral clustering and seasons: Appearance-based localization in outdoor environments. In: Proc. IEEE Int. Conf. on Robotics and Automation, pp 1856–1861

Vardy A (2005) Biologically plausible methods for robot visual homing. PhD dissertation, Carleton University

Vinyals M, Ramisa A, Toledo R (2007) An Evaluation of an Object Recognition Schema Using Multiple Region Detectors. FRONTIERS IN ARTIFICIAL INTELLIGENCE AND APPLICATIONS 163:213

Wehner R (1987) Spatial organization of foraging behavior in individually searching desert ants, cataglyphis (sahara desert) and ocymyrmex (namib desert). Experientia Supplementum 54:15–42

Wehner R (1994) The polarization-vision project: championing organismic biology. Progr Zool (Fortschr Zool) 39:103–143

Zeil J, Hofmann M, Chahl J (2003) Catchment areas of panoramic snapshots in outdoor scenes. JOSA-A 20(3):450–469