Scale-only Visual Homing from an Omnidirectional Camera

Ming Liu, Cédric Pradalier, François Pomerleau, Roland Siegwart
Autonomous Systems Lab, ETH Zurich, Switzerland
[ming.liu, cedric.pradalier, francois.pomerleau]@mavt.ethz.ch, rsiegwart@ethz.ch

***Abstract*—Visual Homing is the process by which a mobile robot moves to a Home position using only information extracted from visual data. The approach we present in this paper uses image keypoints (e.g. SIFT) extracted from omnidirectional images and matches the current set of keypoints with the set recorded at the Home location. In this paper, we first formulate three different visual homing problems using uncalibrated omnidirectional camera within the Image Based Visual Servoing (IBVS) framework; then we propose a novel simplified homing approach, which is inspired by IBVS, based only on the scale information of the SIFT features, with its computational cost linear to the number of features.**

This paper reports on the application of our method on a commonly cited indoor database where it outperforms other approaches. We also briefly present results on a real robot and allude on the integration into a topological navigation framework.

I. Introduction

HOMING is defined as the navigation of a robot from an arbitrary position towards a previously specified Home position [1]. In this paper, we consider the case where the control is achieved by extracting visual features and matching them with the features extracted at the home position for mobile robots in 2D. Visual Homing is considered to be one of the important abilities of a mobile robot and also one of the most important components of visual topological navigation[2], [3]. In visual topological navigation, the homing method is utilized to perform the transition between topological nodes.

As the key to visual topological navigation, the main challenge in the Visual Homing problem is the estimation of the *homing vector*, defined as the direction in which the robot has to move to reach the home position. Our method solves this problem by taking inspiration from the generic framework of Visual Servoing and using an omnidirectional camera as the only sensor. "Visual Servoing" [4], [5] has been widely cited in the area of motion control of robotic platforms, such as industrial arms or mobile helicopters. The originality of our approach is that we use and only use the variation of scale information of the SIFT features to compute our simplified control law for visual homing under the image based visual servoing framework. We will show in section IV that the scale of the features provides sufficient information to build an image based visual servoing control law. Since a camera also provides naturally the bearing to the features, we will also show how this information can be used in the control, and compare the control performance using scale and bearing, only bearing and only scale.

This work was supported by the EU FP7 project NIFTi (contract # 247870) and EU project Robots@home (IST-6-045350).

In general, visual servoing approaches require the computation of the pseudo-inverse of a matrix whose size is proportional to number of features. For systems with limited resources, this can quickly become intractable. To alleviate that, our approach implements the homing task with a cost linear in the number of features. At the same time, our approach doesn't depend on calibration parameters. We show in section V that the resulting control is stable and is able to converge exponentially.

The major contributions of this paper are

1) We study four visual homing approaches under IBVS framework for an uncalibrated omnidirectional camera;
2) A fast homing framework based on the scale information of keypoints is developed;
3) Evaluations on datasets and a mobile robotic platform.

In the following, we first give an overview of related work in Section II. We then formulate the visual homing problem using scale information in the visual servoing framework. Sections V describes the algorithm and control strategy of our approach. Sections VI and VII include the simulation and experiment results.

II. Related Work

Visual homing is often implemented using a bearing-only method. An initial work was presented by Cartwright and Colletti [6] as the 'snapshot' model. Franz et al. [7] continued this direction by analyzing the error and convergence properties. In our previous work [8], we gave a proof of the convergence of a simplified bearing-only method, based on Lyapunov stability theory. In this paper, we extend the bearing-only problem to the classical IBVS framework. Some recent work shows that homing can be efficiently achieved with known landmarks [9].

Our novel approach is stimulated by the work of Corke et al. [10], where the authors used the ALV[11] (Average Landmark Vector) principle to implement a visual servoing task. The ALV method converts the homing problem to a vector operation process, by summing up the bearing vectors to a number of keypoints at the reference and current position. The difference between these two sums is then used to compute the homing vector. However this approach depends

on knowing the position of the landmarks used as keypoints. This in turns assumes that it is possible to estimate the distance to the keypoints. In comparison our approach takes advantage of the scale information attached to the keypoints to calculate the homing vector without distance estimation and with a computational requirement as low as possible.

According to Goedeme et al. [12], knowing the structure of environment and in particular the landmark position is not necessary in visual homing. These informations can be recovered by estimating the ratio of the distances to the matching keypoints by triangulation using an Extend Kalman Filter. Using the features scale, we can avoid this estimation step and use the scale error as a proxy for the distance error.

Among other recent results in the field of homing algorithms, the work of Lim et al[13] divides the 2D plane into four regions and estimates the current robot position by measuring the bearings of landmarks. The theory was proved geometrically but the approach requires more test in a dynamic environment. Cherubini et al [14] proposed a redundancy framework for visual homing problem, which allows online obstacle avoidance.

Note that there are many other works aimed at the visual homing problem, but using different strategies such as [15] and later [16], which relied on the 1D trifocal tensor from the omnidirectional camera. Further, [17] used a sliding-mode control law to exploit the epipolar geometry; [18] directly calculates the homographies from raw images and so on. The comparison with these works is not considered in this report, since the basic strategies and premises are significantly different.

Some related early work using SIFT as main features for visual homing was proposed in [19] [20]. They considered the epipolar geometries as well as the orientation and scale of SIFT features for monocular cameras, following the similar framework proposed in [4]. The work of Vardy et al. [21] is the closest to our simplified approach using scale information. Their first work developed a scale invariant local image descriptor for visual homing, based on the optical flow of unwrapped panoramic image from an omnidirectional camera. This work was continued by Churchill et al. in [22], which presents results of real-time homing experiment using the scale difference field in panoramic images, computed from SIFT matches. In comparison to their work, we stress the following two main differences: firstly, we reason on the effect of the change of scales in a more dedicated way, by embedding the scale measures inside visual servoing framework. Secondly, we give a mathematical proof of the convergence of the controller. We will refer to their method as "scale space homing" in the following and show our approach outperforms theirs in terms of precision.

III. Problem Definition

The visual homing problem can be defined as shown as figure 1, where $\boldsymbol{p}_1, \boldsymbol{p}_2, \ldots, \boldsymbol{p}_n$ are n keypoints, which are extracted by SIFT, SURF [23] or other method providing the scale information of these keypoints. It is assumed that all the listed keypoints can be seen from the current position C and the home position O. The objective is to guide the robot from C to O only by knowing the observed scale s_i and bearing angle β_i associated with keypoint $\boldsymbol{p}_i$. Negre et al. [24] showed that the intrinsic scale can be a measurement to the time to collision. [20] showed a direct relation between the scale and the distance to the feature point. However, for different setup and different environment, the absolute distances to the features cannot be mapped directly. We believe that the variation of the scale of a keypoint can be seen, in first approximation, as a proxy for the variation of its distance.

One fundamental reason is that the scale of a keypoint, in pixels, corresponds to the physical scale of some object imaged from the camera. Using a standard pin-hole model of the camera, the scale is actually inversely proportional to the distance between the camera and the imaged object, assuming the object scale and focal length are constants. As a result, the error between the observed scale and the reference scale gives us an indication of the error between the respective distances and can be used to control the robot towards the home position.

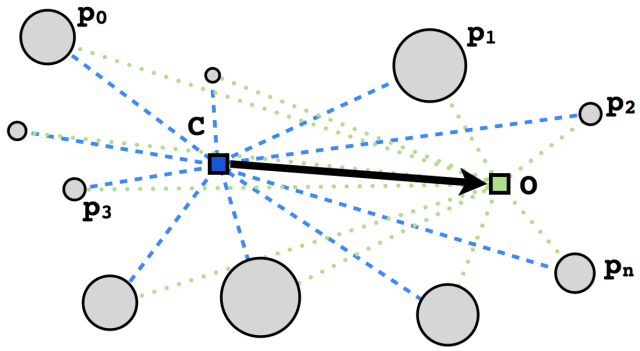

Fig. 1. Abstracted problem of homing. Keypoints $\boldsymbol{p}_1$ to $\boldsymbol{p}_n$ can be observed at the current position C and at the *Home* position O. The variant sizes indicate the differences of keypoints in scale.

IV. Image Based Visual Servoing

In this paper, we intend to use the visual servoing framework to build a robot controller taking advantage of the scale and bearing to the keypoints, instead of their coordinates. In this section, we introduce how could the interaction matrix $\boldsymbol{L_e}$ be designed using these features for panoramic images. We assume that the keypoints are extracted from an omnidirectional camera and we can convert image coordinates to bearing angles. We also assume that we are controlling a ground robot whose configuration can be summarised by its position (x, y) and its heading θ.

A. Definitions

The error of the system is made of two components: the scale errors and the bearing angle error. Therefore the vector of the error can be written as:

$$\boldsymbol{e} = (\boldsymbol{s} - \boldsymbol{s}^*, \boldsymbol{\beta} - \boldsymbol{\beta}^*)^T \tag{1}$$

where $\boldsymbol{s} = (s_1, \ldots, s_n)$ is the vector of observed scale of the keypoints and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_n)$ is the vector of their bearing angles. The variables with '*' superscripts are reference values.

Before computing the derivative of the error, we need to derive some relations between the scale of a feature s_i and the distance to the corresponding object l_i. Let us denote f the focal length of the camera[1] and S the physical scale of the keypoint. Using simple triangulation and the camera pin-hole model, we have

$$s_i = \frac{S\,f}{l_i} \quad \text{and} \quad s_i^* = \frac{S\,f}{l_i^*} \tag{2}$$

which leads to

$$s_i = s_i^* \frac{l_i^*}{l_i} \tag{3}$$

If we assume that the physical keypoint i is at the 2D coordinates (x_i, y_i) in the same frame as the robot coordinates, we can also infer the relation between (l_i^g, β_i^g) and the robot coordinates:

$$l_i^g = \sqrt{(x_i - x)^2 + (y_i - y)^2} \tag{4}$$

$$\beta_i^g = \text{atan2}(y_i - y, x_i - x) - \theta \tag{5}$$

B. Error derivative

To compute the error derivate $\dot{\boldsymbol{e}}$, we derive independently the scale and bearing derivative, by considering them as function of the robot pose. Using equation 5 and 6,

$$\begin{aligned} \frac{d}{dt}\left[s_i(x,y,\theta) - s_i^*\right] &= s_i^* l_i^* \left[v_x \frac{d}{dx}\frac{1}{l_i} + v_y \frac{d}{dy}\frac{1}{l_i} \right] \\ &= -\frac{s_i^* l_i^*}{l_i^2}\left[v_x \cos\beta_i^g + v_y \sin\beta_i^g \right] \end{aligned} \tag{6}$$

with $v_x = \frac{dx}{dt}$ and $v_y = \frac{dy}{dt}$.

Similarly, the bearing error can be derived by equation 5 as:

$$\begin{aligned} \frac{d}{dt}\left[\beta_i^g(x,y,\theta) - \beta_i^*\right] &= -\frac{y_i - y}{l_i^2} v_x - \frac{x_i - x}{l_i^2} v_y - \omega \\ &= -\frac{1}{l_i}\left[v_x \sin\beta_i^g + v_y \cos\beta_i^g \right] - \omega \end{aligned} \tag{7}$$

with $\omega = \frac{d\theta}{dt}$.

Note that β_i^g's are in the global frame instead of robot local frames. In order to transform them to robot local frame, the heading difference needs to be considered, such that the real bearing observations by the robot is:

$$\beta_i = \beta_i^g + (\theta - \theta^*) \tag{8}$$

where we consider $\theta^* = 0$.

Combining equations 6, 7 and 8, we can write the error dynamics as follows:

$$\frac{d}{dt}\boldsymbol{e} = \boldsymbol{L_e} v \tag{9}$$

$$\frac{d}{dt}\begin{pmatrix} s_1 - s_1^* \\ \vdots \\ s_n - s_n^* \\ \beta_1 - \beta_1^* \\ \vdots \\ \beta_n - \beta_n^* \end{pmatrix} = \begin{pmatrix} -\frac{s_1^* l_1^*}{l_1^2}\cos\beta_1 & -\frac{s_1^* l_1^*}{l_1^2}\sin\beta_1 & 0 \\ \vdots & \vdots & \vdots \\ -\frac{s_n^* l_n^*}{l_n^2}\cos\beta_n & -\frac{s_n^* l_n^*}{l_n^2}\sin\beta_n & 0 \\ -\frac{1}{l_1}\sin\beta_1 & -\frac{1}{l_1}\cos\beta_1 & -1 \\ \vdots & \vdots & \vdots \\ -\frac{1}{l_n}\sin\beta_n & -\frac{1}{l_n}\cos\beta_n & -1 \end{pmatrix} \begin{pmatrix} v_x \\ v_y \\ \omega \end{pmatrix}$$

As mentioned earlier, the interaction matrix in equation 9 is enough to implement a visual servoing controller. One remaining problem is neither the distances l_i nor l_i^* can be quantified easily using a single camera. Based on [25] and the analysis on the errors in [5], these values can be approximated by constants due to the low sensitivity of the controller to these parameters.

A direct way to reduce the complexity is to notice that either the upper part or the lower part of equation 9 are enough to implement a visual servoing task. As it is trivial to rotate the robot on the spot once the translational error has been corrected, a two-stage controller can be considered [2]. We consider only the first stage here - the translation to the home position, because it is the key issue for homing. In practice, this means that we can either implement a scale-only visual servoing or a bearing-only visual servoing.

The interaction matrix for scale-only visual servoing is shown as follows, where $\alpha_i = -\frac{l_i^*}{l_i^2}$.

$$\frac{d}{dt}\begin{pmatrix} s_1 - s_1^* \\ \vdots \\ s_n - s_n^* \end{pmatrix} = \begin{pmatrix} \alpha_1 s_1^* \cos\beta_1 & \alpha_1 s_1^* \sin\beta_1 \\ \vdots & \vdots \\ \alpha_n s_n^* \cos\beta_n & \alpha_n s_n^* \sin\beta_n \end{pmatrix} \begin{pmatrix} v_x \\ v_y \end{pmatrix} \tag{10}$$

A similar method which uses the lower part of equation 9 is the bearing-only approach. The error dynamics can be derived as follows, where $\gamma_i = -\frac{1}{l_i}$:

$$\frac{d}{dt}\begin{pmatrix} \beta_1 - \beta_1^* \\ \vdots \\ \beta_n - \beta_n^* \end{pmatrix} = \begin{pmatrix} \gamma_1 \sin\beta_1 & \gamma_1 \cos\beta_1 \\ \vdots & \vdots \\ \gamma_n \sin\beta_n & \gamma_n \cos\beta_n \end{pmatrix} \begin{pmatrix} v_x \\ v_y \end{pmatrix} \tag{11}$$

Looking back to equation 8, we could see that the estimation of the heading θ is crucial for the calculation of the interaction matrices. It implies that a robot may need to have absolute references such as a magnetic compass or a reliable visual compass for better accuracy. Regarding the dataset that we use in section VI, where the robot is well aligned, this problem is trivial. However, this matter needs to be considered in other applications.

V. Fast Visual Homing

In this section, we will describe a scale-based visual homing approach based on the local observations from robot frame. We will then prove that the resulting control law converges

[1] In the case of an unwarped catadioptric image, we abuse the notation by relating the focal length to the vertical field of view α and the image height h: $\frac{\alpha}{2} = atan2(h, f)$.

[2] In general cases, the calibration of the camera to the rotation center need also to be considered. Here we only consider the case in which the camera is aligned.

to the Home position. [20] provided an indication that how the scale of visual features relate to the distance to the certain physical feature point. It stimulated us the possibility to perform homing via only scale information.

A. Scale-based Control for a 1-D Robot

Recall equation 6:

$$\frac{d}{dt}(s_i - s_i^*) = -\frac{s_i^* l_i^*}{l_i^2}\left[v_x \cos\beta_i + v_y \sin\beta_i\right]$$

For the sake of the argument, let us consider a 1-D robot, only able to move towards keypoint i. Because the right side of the above equation can be seen as the projection of the robot velocity on direction towards the keypoint, naming $e_i = s_i - s_i^*$ and $v_i = v_x \cos\beta_i + v_y \sin\beta_i$, we have:

$$\frac{d}{dt}e_i = -\frac{s_i^* l_i^*}{l_i^2} v_i \tag{12}$$

Following the basic strategy of visual servoing, we would like to ensure an exponential decoupled decrease of the error [26]. The following trivial control would achieve this goal (λ is a positive constant).

$$v_i = \lambda_i e_i \tag{13}$$

B. Generic Scale-based Control

If we now, come back to the 2-D case, we can decide to intuitively combine the velocities that would be given by the individual 1-D controllers:

$$\begin{pmatrix} v_x \\ v_y \end{pmatrix} = \sum_{i=1}^{n} \lambda_i (s_i - s_i^*) \begin{pmatrix} \cos\beta_i \\ \sin\beta_i \end{pmatrix} \tag{14}$$

However, even if the convergence was obvious in the 1-D case, there is no guarantee that this sum of control contributions would lead to a stable controller. To prove this, we will resort to the Lyapunov theory. Let us define the following non-negative energy function (Lyapunov candidate function):

$$E = \frac{1}{2}\sum_{i=1}^{n}\left(\frac{s_i - s_i^*}{s_i^*}\right)^2 \tag{15}$$

In this autonomous system with n-dimensional states $\boldsymbol{s}$, the only still point is where $\boldsymbol{s} = \boldsymbol{s}^*$ in the feature space; and physically it's the *home* position. According to Lyaponov theory, we need to show that,

$$\begin{cases} \frac{d}{dt}E(t) = 0, \text{only when all } s_i = s_i^* \\ \frac{d}{dt}E(t) < 0, \text{otherwise} \end{cases} \tag{16}$$

Based on the calculation in Eq. 6, the derivative of the energy function is:

$$\begin{aligned} \frac{dE}{dt} &= \sum_{i=1}^{n} \frac{s_i - s_i^*}{s_i^*}\frac{ds_i}{dt} \\ &= -\sum_{i=1}^{n} \frac{s_i - s_i^*}{s_i^*}\frac{s_i^* l_i^*}{l_i^2}\left[v_x \cos\beta_i + v_y \sin\beta_i\right] \end{aligned} \tag{17}$$

(18)

We plug in the control law of robot for 2-D plane in equation 14 at this step, the derivative is formulated as:

$$\frac{dE}{dt} = -\left[v_x \sum_{i=1}^{n} \frac{s_i^* l_i^*}{l_i^2}\frac{s_i - s_i^*}{s_i^*}\cos\beta_i + v_y \sum_{i=1}^{n} \frac{s_i^* l_i^*}{l_i^2}\frac{s_i - s_i^*}{s_i^*}\sin\beta_i\right]$$

By letting

$$\lambda_i = \frac{l_i^*}{l_i^2}, \tag{19}$$

the equation above simplifies to

$$\frac{dE}{dt} = -\left[v_x^2 + v_y^2\right] = -\sum \lambda_i^2 (s_i - s_i^*)^2 \leq 0 \quad \square \tag{20}$$

where the equality is only valid when all $s_i = s_i^*$.

Assuming the distance l_i and l_i^* are known, the above reasoning shows that the control law given in equation 14 with λ_i given in equation 19 is stable and converges to $s_i = s_i^*$, which is the Home position. However, l_i and l_i^* are not known in practice. According to the error analysis in [5], here we approximate λ_i's by constants, since it does not affect the convergence. Comparing to [22], we not only manage to calculate a more precise homing direction, the amplitude of the velocity is also indicated.

The intuition behind this concept is that each feature point provides a direction and amplitude which the robot should follow, then the output of the controller will be a joined decision of all the observed features. As the number of features is big, even if some of the features can not be matched with the reference, the performance and convergence of the controller is not jeopardized.

VI. Results on an indoor dataset

In order to compare our approach with other methods under similar conditions, we tested our approach on a widely cited dataset [27]. The dataset is a collection of omni-images and unwrapped images in an indoor environment, plus the calibration information and pose information. All the images are 561x81 resolution, the actual intervals between two nearest nodes are 30cm. An instance is shown as figure 2.a.

According to the comparison in [22], the TAAE (total average angular error) and AAE (average angular error) are two important factors to evaluate the homing ability. We compare our results with the same metric.

Figure 2.b shows the homing vector computed by each controller over the whole dataset, assuming the homing position is at the center position (5,8). The color of the filled circles shows the different number of matching features. It is interesting to see that our simplified controller exhibits a very clean behaviour pointing towards the home position, even when the matching result is not high. The AAE for each position of the database in given in figure 2.c.

The generic statistic result of the test is shown as figure 2.d. The blue curve indicates the processing time over 28900 matches (cross matching of 17*10 image array). It shows that the processing time is in average around 2Hz to 7Hz, depending on the resolution level. This is defined as the starting octave level for the image pre-processing. Generally speaking, a resolution level of 0 means the original size of

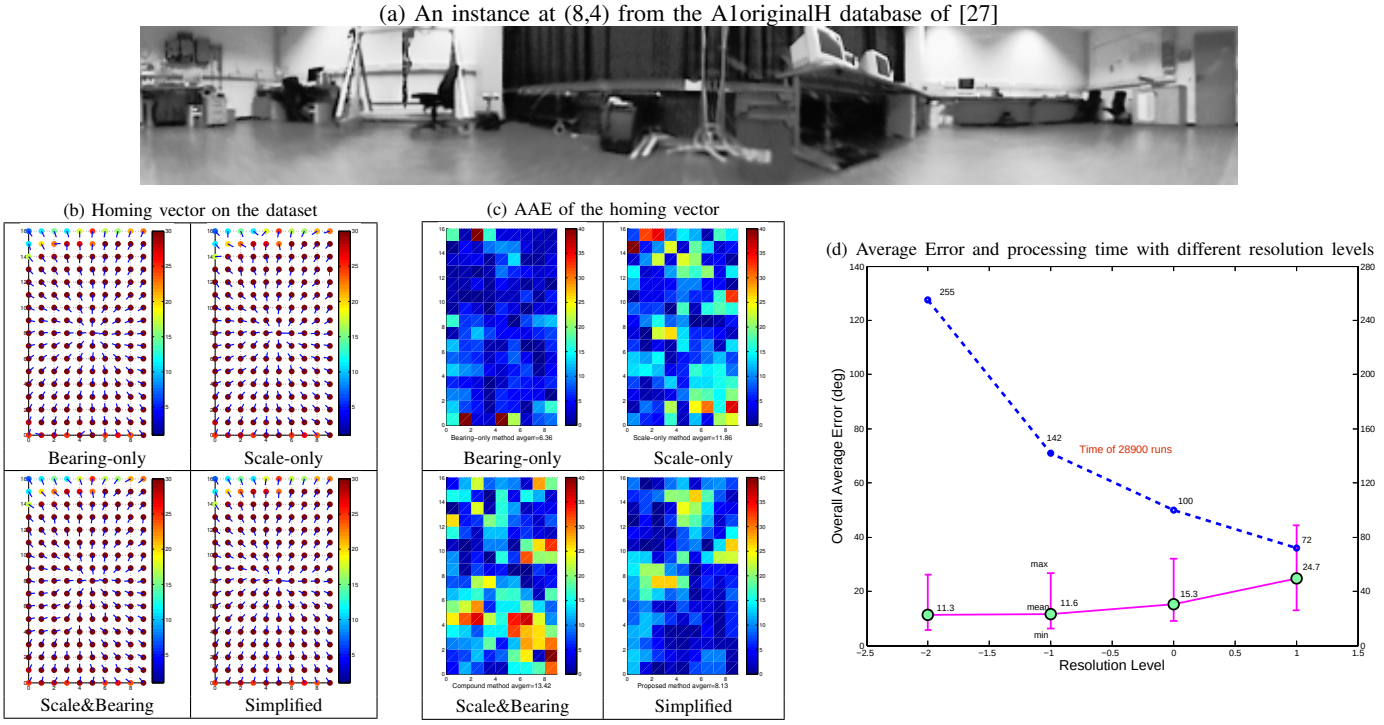

Fig. 2. Performance analysis on the standard dataset from [27].

the image, -1 means an up-sample of the original image to its double size etc. The effect of this parameter is trivial for different implementations of feature extractions. Nevertheless, we would like to utilize this mean to easily see the influence of the number of features. Intuitively, a bigger image can provide more features. Note that more than 99% of the computation time is for the SIFT extraction and feature matching. Taking the simplified approach as an instance, we plot the TAAE (total average angular error) under different resolution level with error bars, using the purple curve in figure 2.d.

A detailed analysis on the effect of different resolution levels on AAE, and the computation time are shown in table I. It shows that at a higher resolution (lower resolution level), all methods can work better in general. A primary reason is that the higher resolution leads to more feature points, which provides more constraints for the error correction. Note that the simplified method can provide faster and more consistent results than others. The bearing-only visual servoing approach provides best results at node (5,8), in the sense of low AAE, at the cost of 4 to 10 times computation time than the simplified method. However, this result doesn't suggest the bearing-only method is the best for all the nodes: according to our further test, the TAAE of the bearing-only method is 11.53 degrees, with a bigger variance than the simplified approach. In this report, we chose to focus on the simplified approach for its lower computation requirements. Further comparisons are not included due to limited space.

In real cases, we used resolution level -1, with nearly 4.0 Hz processing speed. It is sufficient for real-time applications. The *TAAE* of our simplified approach (11.01°), over the entire database, outperforms scale space homing method (12.4°) [22] and warping method (snapshot model)(46.6°) [7] [3].

TABLE I
EFFECT OF IMAGE RESOLUTION REFERING TO (5,8).

Reso. level	Bearing-only	Scale-only	Compound	Simplified
-2	6.37[a]	10.48	13.87	7.75
Time(μs)	25531	21579	27584	6528
-1	6.36	11.86	13.42	8.13
Time(μs)	18421	14594	20459	4804
0	6.26	12.74	14.69	9.90
Time(μs)	18017	9591	12303	2464
1	8.39	16.37	17.86	12.24
Time(μs)	11044	5875	6231	1037

[a]All the AAE values are in degrees.

VII. HOMING EXPERIMENT

The simplified controller was tested on a differential robot with an omni-cam mounted on the top, above the rotation center. The experiment was carried out in an office environment, using an external laser tracker as position ground truth. The Home position was set at the center of the room, and the robot was manually driven to random starting positions. Using the homing process, the robot managed to drive back to the Home position. Due to the limited physical space, this specified experiment covered only a small area of the environment. Recorded trajectories are plotted in figure 3.

[3]Extended information: TAAE of other methods are bearing-only(11.53°), scale-only(13.51°), compound method(15.57°).

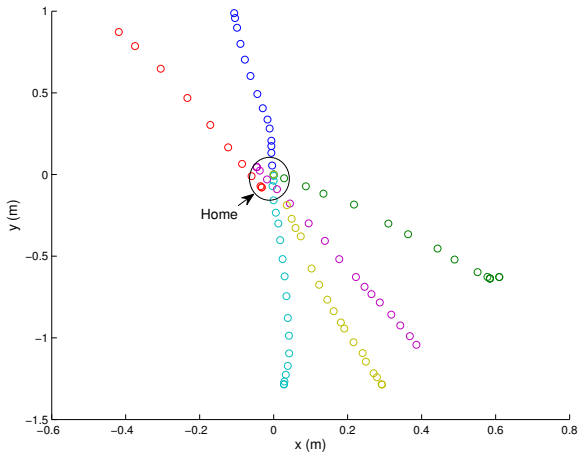

Fig. 3. Homing trajectories

VIII. Conclusion

In this paper, we presented a visual homing framework by visual servoing, based on the scale and/or bearing measurement of popular visual features such as SIFT and SURF. After showing how these measurements could be used in the standard visual servoing framework, we proposed a simplified controller with complexity linear with the number of observed features. We've demonstrated the usability of scale-based visual servoing and we have shown that our simplified approach is stable and offer similar performances as the fully-fledged scale-based visual servoing. When tested against a standard dataset, our controller outperformed state-of-the-art approaches in terms of average angular error.

Although space is missing to provide details in this paper, our controller was integrated into a visual navigation system, combining a topological map of the environment and our homing controller to align the robot with each topological node. the controller could successfully move between randomly selected node in the graph for close to one hour[4]. For further information, please refer to the report [28]. Based on our test results in real apartments and office environments, it also can be considered as a successful algorithm in practical use-cases.

References

[1] R. Basri, E. Rivlin, and I. Shimshoni, "Visual homing: Surfing on the epipoles," *International Journal of Computer Vision*, vol. 33, no. 2, pp. 117–137, 1999.

[2] M. Franz and H. Mallot, "Biomimetic robot navigation," *Robotics and autonomous Systems*, vol. 30, no. 1-2, pp. 133–154, 2000.

[3] A. Angeli and J.-a. Meyer, "Incremental vision-based topological SLAM," in *IEEE/RSJ international conference on Intelligent robots and systems*, 2008, pp. 22–26.

[4] B. Espiau, F. Chaumette, and P. Rives, "A new approach to visual servoing in robotics," *IEEE Transactions on Robotics and Automation*, vol. 8, no. 3, pp. 313–326, 1992.

[4]the limiting factor was the battery life of the laptop running the GPU-based SIFT extraction and matching

[5] F. Chaumette and E. Malis, "2 1/2 D visual servoing: a possible solution to improve image-based and position-based visual servoings," in *Robotics and Automation, 2000. Proceedings. ICRA'00. IEEE International Conference on*, vol. 1. IEEE, 2002, pp. 630–635.

[6] B. Cartwright and T. Collett, "Landmark maps for honeybees," *Biological Cybernetics*, vol. 57, no. 1, pp. 85–93, 1987.

[7] M. Franz, B. Scholkopf, H. Mallot, and H. Bulthoff, "Where did I take that snapshot? Scene-based homing by image matching," *Biological Cybernetics*, vol. 79, no. 3, pp. 191–202, 1998.

[8] M. Liu, C. Pradalier, Q. Chen, and R. Siegwart, "A bearing-only 2D / 3D-homing method under a visual servoing framework," in *IEEE International Conference on Robotics and Automation, 2010*, Anchorage Convention District, 2010, pp. 4062–4067.

[9] S. Yu and D. Kim, "Image-based homing navigation with landmark arrangement matching," *Information Sciences*, 2011.

[10] P. Corke, "Mobile robot navigation as a planar visual servoing problem," in *Robotics Research: The Tenth International Symposium*, vol. 6. Springer, 2001, pp. 361–372.

[11] D. Lambrinos, R. Moller, R. Pfeifer, and R. Wehner, "Landmark navigation without snapshots: the average landmark vector model," in *26th Goettingen Neurobiology Conference*, vol. 1, 1998, p. 221.

[12] T. Goedeme, T. Tuytelaars, L. Van Gool, D. Vanhooydonck, E. Demeester, and M. Nuttin, "Is structure needed for omnidirectional visual homing?" in *Computational Intelligence in Robotics and Automation, 2005. CIRA 2005. Proceedings. 2005 IEEE International Symposium on*, 2005, pp. 303–308.

[13] J. Lim and N. Barnes, "Robust Visual Homing with Landmark Angles," in *Proceedings of Robotics: Science and Systems*. Citeseer, 2009.

[14] A. Cherubini and F. Chaumette, "A redundancy-based approach for obstacle avoidance in mobile robot navigation," in *Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on*. IEEE, 2010, pp. 5700–5705.

[15] H. Becerra, G. López-Nicolas, and C. Sagüés, "Omnidirectional visual control of mobile robots based on the 1d trifocal tensor," *Robotics and Autonomous Systems*, vol. 58, no. 6, pp. 796–808, 2010.

[16] M. Aranda, G. Lopez-Nicolas, and C. Sagues, "Omnidirectional visual homing using the 1D trifocal tensor," in *2010 IEEE International Conference on Robotics and Automation (ICRA)*, 2010, pp. 2444–2450.

[17] H. Becerra, G. Lopez-Nicolas, and C. Sagüés, "A sliding-mode-control law for mobile robots based on epipolar visual servoing from three views," *Robotics, IEEE Transactions on*, vol. 27, no. 1, pp. 175–183, 2011.

[18] G. López-Nicolis, J. Guerrero, and C. Sagüés, "Multiple homographies with omnidirectional vision for robot homing," *Robotics and Autonomous Systems*, vol. 58, no. 6, pp. 773–783, 2010.

[19] T. Nierobisch, J. Krettek, U. Khan, and F. Hoffmann, "Optimal large view visual servoing with sets of sift features," in *Robotics and Automation, 2007 IEEE International Conference on*. IEEE, 2007, pp. 2092–2097.

[20] F. Hoffmann, T. Nierobisch, T. Seyffarth, and G. Rudolph, "Visual servoing with moments of sift features," in *Systems, Man and Cybernetics, 2006. SMC'06. IEEE International Conference on*, vol. 5. IEEE, 2006, pp. 4262–4267.

[21] A. Vardy and F. Oppacher, "A scale invariant local image descriptor for visual homing," *Lecture notes in computer science*, vol. 3575, p. 362, 2005.

[22] D. Churchill and A. Vardy, "Homing in scale space," in *IEEE/RSJ International Conference on Intelligent Robots and Systems, 2008. IROS 2008*, 2008, pp. 1307–1312.

[23] H. Bay, T. Tuytelaars, and L. Van Gool, "Surf: Speeded up robust features," *Lecture notes in computer science*, vol. 3951, p. 404, 2006.

[24] A. Negre, C. Braillon, J. Crowley, and C. Laugier, "Real-time time-to-collision from variation of intrinsic scale," *Algorithmic Foundations of Robotics VI*, p. 75, 2005.

[25] E. Malis, F. Chaumette, and S. Boudet, "2 1/2 D visual servoing," *IEEE Transactions on Robotics and Automation*, vol. 15, no. 2, pp. 238–250, 1999.

[26] F. Chaumette. and S. Hutchinson, "Visual Servo Control, Part I: Basic Approaches," *IEEE Robotics & Automation Magazine*, vol. 13, no. 4, pp. 82–90, 2006.

[27] A. Vardy, "Panoramic image database." [Online]. Available: http://www.ti.uni-bielefeld.de/html/research/avardy/index.html

[28] M. Liu, "The role of homing in visual topological navigation." [Online]. Available: http://www.asl.ethz.ch/people/lium/personal/navigation