IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

SPECIAL SECTION ON GNSS, LOCALIZATION, AND NAVIGATION TECHNOLOGIES

Received February 28, 2018, accepted May 21, 2018, date of publication June 22, 2018, date of current version July 6, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2842679

Visual Homing Navigation With Haar-Like Features in the Snapshot

CHANGMIN LEE AND DAEEUN KIM

School of Electrical and Electronic Engineering, Yonsei University, Seoul 03722, South Korea

Corresponding author: DaeEun Kim (daeeun@yonsei.ac.kr)

This work was supported by the National Research Foundation of Korea through the Korean Government (MSIT) under Grant 2017R1A2B4011455.

ABSTRACT Visual homing navigation has been a challenging issue in indoor localization and navigation. Inspired by insect navigation, the snapshot model was introduced for homing navigation, where a pair of snapshots at the current location and at the nest are compared to guide the homing direction. We investigate Haar-like features in vision to extract visual cues, based on the snapshot model. The Haar-like features consist of masks randomly generated over the snapshot image at the home location, and later, their matching scores at the snapshot available at the current location are calculated for the correspondence measure. We draw landmark vectors using the correspondence measure of Haar-like features at their angular positions. Interestingly, a collection of Haar-like features provide visual characteristics to reflect a pair of snapshot images, which can determine the homing direction. In this paper, we propose two types of homing methods based on the image difference using Haar-like features, the Haar-like landmark vector model and the Haar-like image distance model. We demonstrate the effectiveness of the methods in several environments.

INDEX TERMS Local visual navigation, Haar-like feature, snapshot model, landmark vector, homing navigation.

I. INTRODUCTION

Local visual homing has been a challenging problem. Vision plays a great role in localization, mapping and navigation. There have been studies involved with local visual navigation [1]–[6]. More advanced algorithms have been developed for SLAM [7]–[9] or graph model [10], [11]. There have been biomimetic models inspired by a variety of biological sensors [12]–[16]. We note that insects demonstrate very effective homing navigation even with a low resolution of visual images. A snapshot model has been suggested to model the bee's navigation by Cartwright and Collett [17], computing the home direction by comparing the current view and the stored views. For homing direction, only two images are sufficient, instead of the whole time course of images in the moving path. According to the snapshot model [17], agents need to move along the direction which decreases the difference between the current view and view from the origin or the nest. If there are visible landmarks available at the two locations, they can help guide homing efficiently.

Based on the insect-inspired snapshot model, a number of incarnations of the snapshot model have been suggested for local homing navigation, including Average Landmark Vector method (ALV), Center-of-Mass ALV (COMALV), Distance-Estimated Landmark Vector (DELV), Fourier encoded snapshots, Descent in Image Distance (DID), or warping methods with variations [18]–[22]. These methods compare visual features in a pair of visual images to compute the homing vector.

With the ALV method [18], landmark vectors are defined as vectors pointing towards landmarks on the retinal image, each of which has a unit distance under the assumption of equal distance. That is, the distance information is ignored but only the angular positions of landmarks are considered. Summing landmark vectors forms the averaged landmark vector called ALV (Averaged Landmark Vector). Subtracting the ALV at the nest from that at the current position can determine the homing direction. An extension of the ALV model, called Center-of-Mass ALV (COMALV) has been suggested to extract robust features from a pair of snapshot images, together with neural network learning over the visual features [19], [23]. The landmark vector is calculated using the image intensity value for each angular

2169-3536 © 2018 IEEE. Translations and content mining are permitted for academic research only.
Personal use is also permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

direction, to imitate ommatidia observed in insect vision. Also, more advanced ALV models have been tested with robotic experiments [24], [25] or the ALV was combined with invariant visual feature detection [26]. Alternatively, the DELV method considers the distance information for landmarks as well as their angular positions [20], [27]. It was reported that a holistic approach with depth information can robustly determine the homing direction [27].

The Descent in Image Distance (DID) method uses the whole pixel images to estimate relative position changes for the snapshot [21], [28]. The image difference becomes larger if a pair of positions for the snapshot image are farther away. The direct comparison of the whole images and calculation of the image difference can determine the homing direction. Generally, the image distance becomes small when a pair of positions are close. The DID method uses this property to guide homing. It was argued that the model explains well the insect navigation and has good homing performance [29]. According to the DID method [21], [30], computing the gradient of the distance function at any point needs a target image and three images around the current position, the current image and the two images observed at two displacements to the orthogonal direction from the current position.

Another snapshot model is the warping method [22], [31]–[33]. Different from the DID approach, the warping method estimates all possible changes in image pixels depending on every movement of a mobile agent, and also calculates the movement effect to each pixel. A pair of snapshot images can reversely allow to estimate the movement direction, ultimately homing direction when a snapshot is taken at the home location. Warping-based models have been studied using dynamic programming [34], [35]. Those warping methods are effective to run even under various conditions like illumination change and can be applied to a cleaning robot environment [35]. Especially, 2D-warping models have more robustness in matching image pixels compared to simple warping and as a variation of 2D-warping model, MinWarping [35] calculates warped images in advance and compares the current image with them, which can reduce the overall computing time. Furthermore, the warping method was extended into a novel approach combined with both Scale-Invariant Feature Transform (SIFT) features and mismatching elimination [33] to obtain robustness in accuracy and time consumption.

Various bio-inspired visual homing models have been compared in real environments [23], [36], [37]. With the skyline information and the intensity values around the horizontal line, local homing navigation can be achieved [36]. The gradient-ascent algorithm on the correlation between the target image and the current image can be applied. Also, the ALV model with those feature values as the length of the landmark vectors can show effective homing performance. It indirectly supports that ants might use the skyline features in their visual homing. Various bio-inspired visual homing algorithms have been compared with the navigation pattern of real crickets [23]. In the work, navigation algorithms including the Average Landmark Vector (ALV), the Center-of-Mass ALV (COMALV), the optic flow methods and the gradient-descent models have been tested for visual homing. According to the authors, the gradient-descent method and the optical flow method show good homing records and the COMALV method shows similar homing patterns with the observed ones of crickets. Visual navigation has also been applied to the route-following robots [37]–[39]. Based on the image familiarity, the optimal heading direction for the route following task can be searched. An agent collects and saves visual snapshots in the learning route and in the test phase, it finds the most similar image with the current input image to determine the heading direction.

The biological vision system has a complex structure, and both cone cells and rod cells are directly connected to the retinal image. Bipolar cells receive synaptic input from either rods or cones, or both rods and cones [40], and they are the basic units for visual perception about shape. Many insects demonstrate excellent homing navigation capability even though they have a low-level vision system. Even a simple model of visual features might be sufficient for good homing navigation. We argue that simple visual features such as Haar-like features can follow the bipolar cell operation in an abstract form. The concept of Haar-like features was introduced for object recognition in vision or face recognition [41]. There have been studies for visual perception using Haar-like features [42]–[44]. A Haar-like feature uses neighboring rectangular regions at a specific location, integrates the pixel intensities for each region or calculates the absolute difference between two region sums. There are several mask patterns depending on the number of rectangle regions or the distribution of a set of adjacent rectangles.

If a part of target image has a similar pattern to some mask, then the matching score of the Haar-like mask becomes high. Those Haar-like features have very simple patterns consisting of positive or negative sum of pixel intensities, which might be described by on and off bipolar cells by analogy. Studies show that a collection of those weak pattern classifiers can lead to a strong classifier for pattern recognition [42]. It was reported that the route-following task for a mobile robot can be achieved using Haar-like features [38]. Route following is the task in which an agent can follow a memorized path. The Haar-like features can visually guide the right direction to the goal position without waypoint setup. Despite the potential of Haar-like features, there has been no application with the snapshot model using those features.

In this paper, we develop new visual navigation methods with Haar-like features, which follow the snapshot model [17] in local homing navigation. In our approaches, a pair of snapshot images at the home location and at the current location are compared through a random set of Haar-like masks, assuming that each snapshot image is represented as a set of Haar-like features. We argue that the difference of matching score for a set of masks can indirectly estimate the distribution of landmarks in the environment, and the set of difference vectors can determine the homing direction. We propose two

(a) (b)

FIGURE 1. Experimental environment and mobile robot: indoor environment (left) a mobile robot with omnidirectional camera (right).

types of homing methods based on the image difference using Haar-like features, and we will call them, the Haar-like Feature Landmark Vector (HFLV) model and the the Haar-like Feature Image Distance (HFID) model. We also compare those methods with other conventional methods.

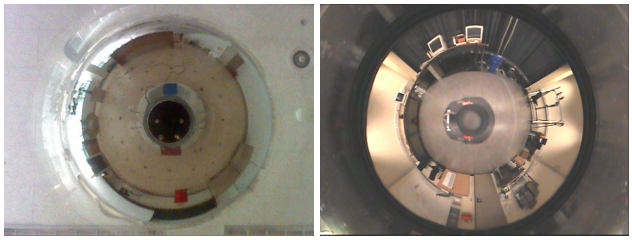
Here, we seek to answer the following questions: Do the Haar-like features provide the landmark distribution information in the environment? What characteristics can be found with comparison of Haar-like feature scores in a pair of snapshot images? Are the suggested models with Haar-like features better to find homing direction than other methods? How many Haar-like masks are needed for desired homing performance? Are the homing methods based on Haar-like features robust for a variety of environments? Which of the HFLV model and the HFID model is more effective in local homing navigation? The contribution of this paper is related to answering the above questions. The Haar-like features are weak classifiers to determine the homing direction, but a set of those classifiers can draw the landmark distribution and further estimate the homing direction accurately. Especially, the matching scores of Haar-like features characterize the landmarks. We will demonstrate that the suggested HFID model based on the image distance has stable homing performance for all the test environments, compared to other state-of-art homing algorithms. Also, we will show that an appropriate number of Haar-like features are sufficient to guide the robot in the homing direction.

II. METHODS

For homing navigation, the mobile agent needs to decide the homing direction at an arbitrary position. A sequence of moving steps can finally arrive at the goal position, if each step has a moving direction towards home. In our experiments, the agent has visual information about the surrounding environment. We first explain the environmental setup and introduce two novel algorithms based on Haar-like features for homing navigation. We assume that our models have a reference compass to align two coordinates in a pair of snapshots, often found in the conventional snapshot models.

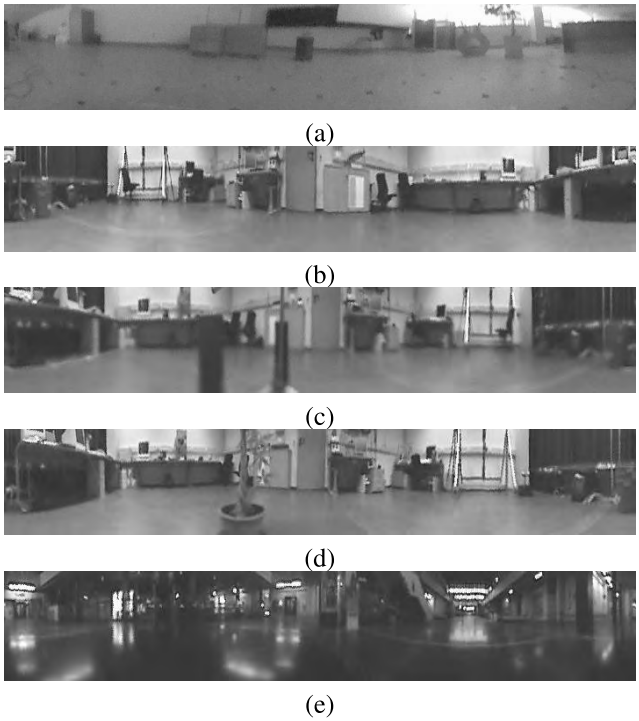
A. EXPERIMENTAL ENVIRONMENT

We test several indoor environments for homing navigation. The environment shown in Fig. 1 is our



(a) (b)

FIGURE 2. Omnidirectional images. (a) An image in our lab environment; (b) an image from a benchmark dataset .



(a) (b) (c) (d) (e)

FIGURE 3. Panoramic snapshot images: (a) our lab environment; (b) environment "a1original"; (c) environment "screen"; (d) environment "arboreal"; and (e) environment "hall1".

experimental setup which will be called "ourlab" whose arena size is about 2.2m × 2.2m. More environments have been collected from Vardy's dataset [45], [46], called "arboreal", "screen", "hall1" and "a1original" whose area size is about 2.7m x 4.3m.

Our robot platform uses iRobot ROOMBA which has two wheels, and it is mounted with an omnidirectional camera, as shown in Fig. 1. The camera consists of a convex mirror and webcam, Logitech Webcam E3500. The visual image is transferred to a notebook.

Fig. 2(a) shows an example of omnidirectional image from our robot. Another image in Fig. 2(b) is from Vardy's dataset. The omnidirectional image is converted into a panoramic image as shown in Fig. 3, since Haar-like masks consist of rectangular regions for image feature.

We test five indoor environments for local visual homing, our lab environment and Vardy's four dataset as shown

in Fig. 3. Datasets of our lab environment are of low-resolution and the lab has a large window area with sunlight. The sunlight in the window has high intensity, and strongly influences extraction of image features. Estimating the homing direction is affected by the landscape. One of Vardy's environments, ''a1original'' shows a large room with well-distributed objects near the walls. The environment ''screen'' and ''arboreal'' have an object, a screen or a plant at the center of the room which can make occlusion, which affects the homing performance. The environment ''hall1'' shows an open space, but it has a large glittering surface. Many lights get reflected on the surface depending on the view angle. We test the suggested methods and various state-of-art algorithms for the environments to see the homing performance under various conditions.

For the snapshot model, a pair of images are compared, in particular, the landmarks. The home snapshot image is given as the reference image, and the other image at an arbitrary position is checked to estimate the homing vector. If visible landmarks are available at the two positions, then the snapshot algorithm can easily guide the robot in the homing direction. Many navigation methods have a vision processing to extract landmark features from the image for similar purpose.

B. HAAR-LIKE FEATURES AND MATCHING SCORES

In our approach, 5000 Haar-like features are randomly generated and applied to the two snapshot images. The features have random sizes, patterns and locations in the image. Fig. 4 shows examples of Haar-like features and their matching in the snapshot. The matching score can be calculated as how closely a Haar-like feature match a patch in the snapshot image. Each Haar-like feature can be a landmark candidate at its angular position in the snapshot. We assume that the change of the snapshot image depending on the displacement of an observer's position can be reflected on a set of Haar-like mask scores. The matching score of each Haar-like feature relies on the pattern matching. If the feature has similar pattern to a patch in the snapshot image, the score becomes high. The difference of a Haar-like feature score measured in a pair of snapshot images can provide a sign of moving direction. That is, the homing direction can be analyzed as a moving direction to decrease the score differences obtained from a set of Haar-like features.

If there is a snapshot image I available, the matching score $S(I, M)$ for a given Haar-like mask M is calculated as

$$S(I, M) = \left| \sum_{i \in M+} I_i/m_1 - \sum_{i \in M-} I_i/m_2 \right| \quad (1)$$

where I_i indicates the i-th pixel of the image and $m_1 = \sum_{i \in M+} 1$, $m_2 = \sum_{i \in M-} 1$ are the size of white and black regions, respectively. The score calculates the difference of the mean values of pixel intensities of white and black regions. We can easily see that the shape, location and size of the mask is important to get high score.

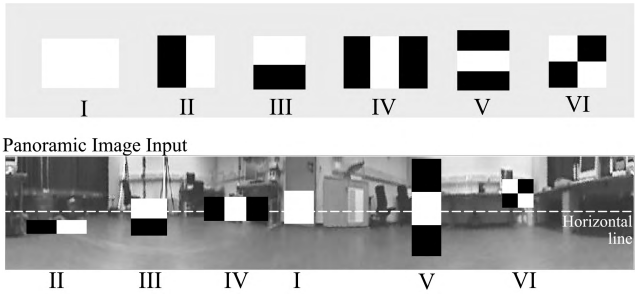


FIGURE 4. Examples of Haar-like masks and their matching in the image; six types of masks with one or more rectangles. Type I has one rectangle, Type II and III two regions, Type IV and V three regions, and Type VI four regions; lower picture shows a panoramic image and matching of those Haar-like masks.

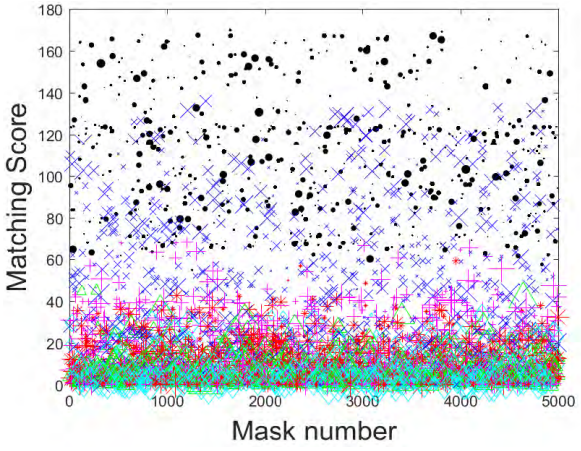


FIGURE 5. Matching scores for 5000 random Haar-like masks; six types of masks were randomly generated and their scores were investigated. Larger markers indicate larger sizes of masks (black dot for Type I, blue x for Type II, red star for Type III, magenta + for Type IV, cyan diamond for Type V, green triangle for Type VI).

Many interesting landmarks are observed near the horizontal line. Random Haar-like masks are generated around the horizontal line. The bottom image in Fig. 4 shows examples of well-matched cases for six types of masks. Type I can be matched well on a flat section or object. Type II-III are good for the boundary line matching, and Type IV-V can be better for isolated segmented objects. Type VI will be effective on the corner of objects.

Each mask is given its center position (θ, y) with its mask size $(w, h) = (\Delta\theta, \Delta y)$ in the panoramic image that has 720 pixels in width (a resolution of 0.5 degree) and 120 pixels in height. For a given snapshot image, 5000 Haar-like features were randomly generated. Each mask has a random center position around the horizontal line (within a window of ± 10 pixels of the horizontal line) Its mask type and the size are also randomly selected; the width ranges from six pixels to 120 pixels and the height ranges from 18 pixesl to 60 pixels. Fig. 5 shows a distribution of scores for the whole set of masks, and each mask has its own angular position in the snapshot image. We will see a large collection of masks contribute to see the visual characteristics of the snapshot

image, more strictly speaking, a population of the mask score differences in the two snapshots characterize dominant change in the visual landmarks.

C. METHOD 1: HAAR-LIKE FEATURE LANDMARK VECTOR (HFLV) MODEL

A set of Haar-like masks is applied to the two snapshot images, home (I_0) and the current view (I). Each type of mask has one or two rectangular regions, white ($M+$) or black ($M-$). Then a mask (M_j) can be represented by

$$M_j(\theta_j, y_j, w_j, h_j) = M_j+ \cup M_j-, \quad M_j+ \cap M_j- = \emptyset \quad (2)$$

where the j-th mask M_j has an angular position θ_j, y-axis position y_j, width w_j and height h_j. There is no overlapped area between white and black regions, and the center position of a mask is (θ_j, y_j).

We calculate the matching score $S(I, M_j)$ for each Haar-like mask M_j, using eq. (1), and a landmark vector $\overrightarrow{L_j}(I)$ is defined as a combination of the score and angular position of each mask. That is, its score is set as the vector length and the angular position as the vector direction. Thus, the landmark vector is written as

$$\overrightarrow{L_j}(I) = S(I, M_j)\hat{u}_j = (S(I, M_j)\cos\theta_j, S(I, M_j)\sin\theta_j) \quad (3)$$

where I is a snapshot image taken at a specific position, θ_j is the angular position for the center of mask M_j and $\hat{u}_j = (\cos\theta_j, \sin\theta_j)$ is a unit vector for direction θ_j.

Based on the snapshot model, we can estimate the homing vector as a difference between the landmark vectors at the home location and at the current position.

$$\overrightarrow{H} = \sum_{j=1}^{N} \overrightarrow{h_j} = \sum_{j=1}^{N} (v_j\hat{u}_j - w_j\hat{u}_j) \quad (4)$$

$$= \sum_{j=1}^{N} (v_j - w_j)\hat{u}_j = \sum_{j=1}^{N} \Delta L_j\hat{u}_j \quad (5)$$

where N is the number of masks, $w_j = S(I_0, M_j)$, $v_j = S(I, M_j)$ for the home snapshot image I_0 and the snapshot I at the current position, and $\Delta L_j = v_j - w_j$.

The above landmark model is similar to the weighted form of ALV model and the COMALV model [18], [19] or the DELV model [20], [47], [48]. This model is an abstract form of landmark vector. The matching score is not proportional to the landmark distance, but large magnitude of ΔL_j can be observed along the moving direction from the current position to the home position, since much change of the visual image along that direction can happen. A set of masks around the angular direction can greatly contribute to $\Delta L_j = S(I_0, M_j) - S(I, M_j)$, which will be effective to estimate the homing direction. If a given landmark feature observed at one snapshot is searched at another snapshot image, then it may resemble to calculate optical flow fields for a pair of images at short distances. Here, a Haar-like mask is applied at the same position over two snapshot images.

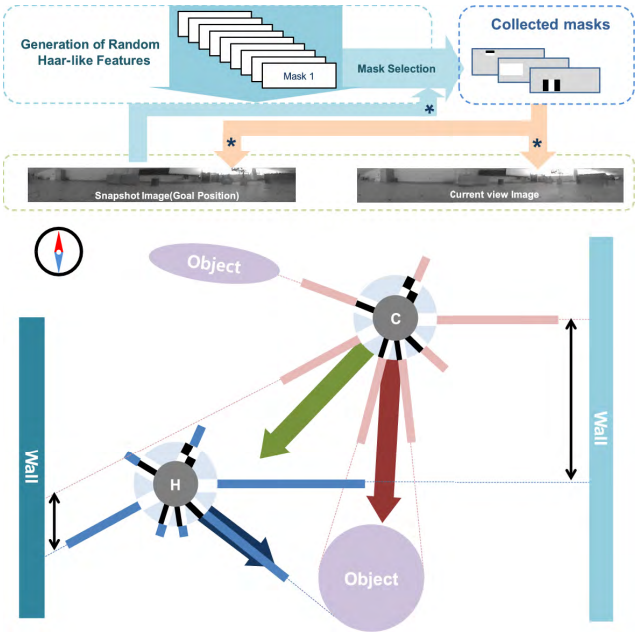


FIGURE 6. Overview of the Haar-like feature homing algorithm, the HFLV model. Initially random masks are generated and the matching score of each mask over the home snapshot image can be calculated. Then a set of landmark vectors are derived from the mask scores. The same set of masks are applied to the current snapshot and another landmark vectors are obtained. The homing vector can be estimated from the two sets of landmark vectors. Lower figure shows a diagram of landmark vectors at the home location (H) and at the current location (C). The matching score for each mask becomes the length of the mask landmark vector. The average of landmark vectors is represented as blue arrow at home location and red arrow at the current position. The difference of the two averaged vectors makes homing direction, displayed with green arrow.

Fig. 6 shows an overview of the HFLV model. Random masks with a variety of types are generated and their matching scores are calculated over two snapshot images at the home location and at the current location. Each mask determines a landmark vector whose direction is equal to the angular position of the mask and the matching score of the mask over a snapshot is denoted as the length of the landmark vector. A collection of landmark vectors represent a snapshot image, and possibly the averaged landmark vector can be an representative vector for the snapshot at a given location. The difference of the two averaged vectors at the home location and at the current position can determine the homing direction, similar to the DELV model [20]. Otherwise, each mask calculates the difference of the matching scores over the two snapshots and a collection of the differences for a set of masks can estimate the homing direction as similarly tested in the COMALV model [19].

D. METHOD 2: HAAR-LIKE FEATURE IMAGE DISTANCE (HFID) MODEL

The above landmark model (HFLV) describes how to estimate the homing vector using the matching scores of Haar-like masks. The method is largely affected by the image contrast or brightness. We note the relative score differences of masks rely on how far away the two snapshots are. The

score difference increases when the distance becomes larger. We separate the relative distance into two components, the x-axis or y-axis movement in Cartesian coordinate.

To handle the problem, the HFID model uses multiple snapshot images including the home image (I_0), an image near home along the x-axis (I_X) and another image near home along the y-axis (I_Y). Those images are compared with the image at the current position through Haar-like feature masks. The method first calculates the matching score difference of each mask M_j between a pair of snapshots. The sum of the matching score difference over a set of masks approximately estimates the distance between the two snapshot positions; small matching scores imply small distances. That is, the relative distance among the snapshot positions can be inferred from the image difference. Similar to the DID method which explains well the insect navigation [21], we consider multiple images as reference home images to find the homing direction. The DID method uses multiple snapshots near the current position and a single target image, but it needs a sequence of movements to catch multiple images. To determine the homing direction at an arbitrary spot, we take three images around the home location. The three reference images and the current snapshot can estimate the image distances between a pair of snapshots and the relative distances can derive the homing direction.

The homing direction can be estimated as follows:

$$X_{diff} = \sum_{j=1}^{N} |S(I_X, M_j) - S(I, M_j)| - \sum_{j=1}^{N} |S(I_0, M_j) - S(I, M_j)| \quad (6)$$

$$Y_{diff} = \sum_{j=1}^{N} |S(I_Y, M_j) - S(I, M_j)| - \sum_{j=1}^{N} |S(I_0, M_j) - S(I, M_j)| \quad (7)$$

$$\overrightarrow{H} = (X_{diff}, Y_{diff}) \quad (8)$$

where the homing vector ($\overrightarrow{H}$) is determined by the ratio between the x-direction difference and the y-direction difference. The vector magnitude becomes large when the current position is close to the target position, since the relative image difference depends on the distance. The homing direction can be estimated by the normalized home vector $\hat{H} = \overrightarrow{H} / \sqrt{X_{diff}^2 + Y_{diff}^2}$.

The sum of score differences over a set of Haar-like masks for a pair of snapshot images can derive the relative distance between the positions for the snapshot images. Equation (6)-(7) is a simple but reasonable measure to determine the relative difference of the distance along the two different directions. Fig. 7 shows an overview of our second model, HFID model. Masks with six types are randomly generated and their matching scores are calculated over the

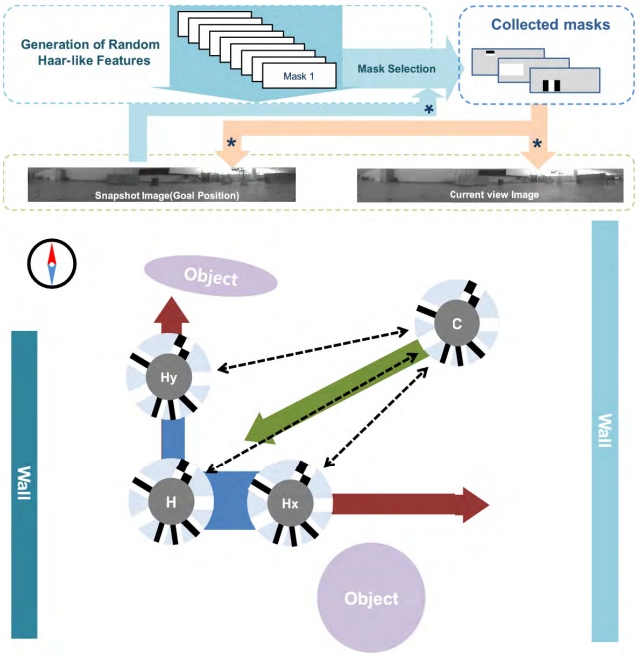


FIGURE 7. Overview of the Haar-like feature homing algorithm, the HFID model. Initially random masks are generated and the sum of matching scores of each mask over three reference images (H, H_X, and H_Y) can be calculated by applying the same set of masks. Each summation is directly proportional to geometric distance and the homing vector can be estimated from the differences of two terms. It shows that changes between pairs with the width of blue links and overall length as red arrows, respectively. The ratio between two pairs makes homing direction, displayed with green arrow.

snapshot images, three reference images around the home location and the snapshot at the current location. The differences in the sum of matching scores determine the image distances and the relative distance between each of three home images and the current image can be estimated, which derive the homing direction. Local gradient of the distance can be estimated with Haar-like features while the DID algorithm use the whole pixel image for it.

We propose two methods using Haar-like masks. One is to utilize the matching score of masks and the other is to estimate the relative distance of matching scores for the snapshot at the current position and three home images, through a set of masks. We will test the methods for several environments.

E. COMPARISON WITH OTHER ALGORITHMS

The snapshot models that we suggest are compared with other state-of-art techniques for visual homing navigation, the Center-of-Mass ALV (COMALV) method [19], the Descent in Image Distance (DID) method [21], [28], the MinWarping algorithm [32], [34], [35] and the SIFT-Warping (SIFTw) algorithm [33].

The COMALV [19] is a variant of the ALV model. A set of landmark vectors are estimated at both the home and the current location. A landmark vector consists of an angular direction and its magnitude. The magnitude is calculated as the mean of subsampled image intensities at each angular direction (a resolution of 4 degrees). The COMALV

thus models 90 biological ommatidia outputs for a snapshot image. The method originally use a learning procedure with artificial neural networks to estimate the averaged vector, but in our experiments, we directly calculate the averaged vector over a set of image intensity values.

Here, we convert a panoramic image with 720 by 100 pixels into 90 landmark vectors, covering 80 pixels for each angular direction, ± 40 pixels around the horizontal line in the vertical direction. A representative vector for a given snapshot image is calculated as the average of those landmark vectors. Then the home vector is estimated by the subtraction of the two average landmark vectors for a pair of snapshot images.

The DID method [21] uses three reference images, and based on the image familiarity, it calculates local gradient of image distance by using these three images to determine the homing direction. The original DID model finds homing direction by comparing images at the neighboring positions around the current position to the home snapshot. The most similar image with the home image can determine the best moving direction towards home. In this paper, we revised the method by choosing three fixed reference images (I_X, I_Y and I_0) around the home location (x, y) instead of the three images near the current position. The images H_X and H_Y are collected at $(x + \delta x, y)$ and $(x, y + \delta y)$, respectively. That is, the three snapshot images are compared with the snapshot image at the current position. Here, a panoramic snapshot has 720 by 100 pixels with an angular resolution of 0.5 degree and ± 50 pixels around the horizontal line in the vertical direction.

To understand the effect of the suggested HFID model, a variation of the DID method, called the COM-ID (Center-of-Mass Image Distance) was tested. That is, the image distance approach is applied, instead of the image intensity difference. Its basic structure for the snapshot is similar to the COMALV, and each of 90 subsamples of a snapshot image is compared with that of another snapshot image. We allow three reference images around the home location and the pre-defined set of subsamples are one by one compared for a pair of snapshots in order to calculate the differences of the matching score for a set of 90 ommatidia. It can calculate the relative image distance between the current snapshot and each of the three reference images. That is, the COM-ID model uses three target images and the current snapshot like the DID method and the HFID model. The method is similar to the HFID model in that the image distances for a set of features are calculated and each feature has a vector representation. Here, a uniform distribution of 90 features are available as in the COMALV method and the sum of feature vectors can determine the relative image distance, while the HFID model use a random set of Haar-like features. In contrast, the DID method uses the whole picture for the image difference. Similar with the HFID model, we apply the same 90 ommatidia features available in the current snapshot to three target images and calculate three differences with the current one. The homing direction is designated by calculating the ratio between three differences.

The MinWarping method is a local visual homing approach as a variant of warping [32]. For a pair of two panoramic images at different locations, it computes a bearing angle from one capture point to another and an orientation angle related to the azimuth rotation between the two images. That relative pose can be estimated by a systematic search on the two parameters, the bearing angle and the rotation angle with the best matching score. The score evaluates how well hypothetical movements of the robot agree with the transformation between the target snapshot and the current image. A snapshot image is scaled around the horizon using a discrete set of scale factors. Various scale and horizontally shifted combinations are pre-calculated in a stack and the parameters with the best matching score are searched to find the homing direction. MinWarping searches for best-matching image columns over a range of parameters by assuming that an image column of the panoramic images have the same ground distance from the camera. The method shows very robust performance even in real environments [34], [35]. Here, we use a snapshot image with 720 by 100 pixels with an angular resolution of 0.5 degree.

The SIFT-Warping (SIFTw) method is another variant version of warping [33]. It applies SIFT features widely used in computer vision to the warping method. The SIFT features are robust scale-invariant HoG-like local visual features and the approach uses the features as a set of landmarks instead of pixels often used in the warping methods. The SIFT feature matching acts like object matching and it can be applied to the warping model with equal distance assumption. With matched outputs through the warping equation, it additionally uses mismatching elimination algorithm with constraints about feature distribution. The approach is one of sophisticated homing algorithms developed recently. The SIFT features greatly contributes to improving the performance. It provides a stable homing performance in the environment involved with scaling, occlusion, rotation, translation and illumination. The approach as a state-of-art homing algorithm will be tested.

The suggested methods, the HFLV model and the HFID model using Haar-like features, will be compared with the above approaches, the DID, the COMALV, the COM-ID, the MinWarping and SIFTw for homing performance.

III. EXPERIMENTS

We test five indoor environments, ''ourlab'' from our lab environment and Vardy's dataset environments ''a1original'', ''screen'', ''arboreal: and ''hall1''. The first landmark model uses only one home snapshot while the second model uses three snapshots around the home location, including the home snapshot image. Each environment has panoramic snapshot images in the grid with regular intervals. We assume that the snapshots are pre-aligned with a reference compass.

A. HFLV MODEL

We first test HFLV model based on Haar-like feature scores, where one home image is used. The whole set of Haar-like features has 5000 randomly generated masks with varying sizes, random locations and types.

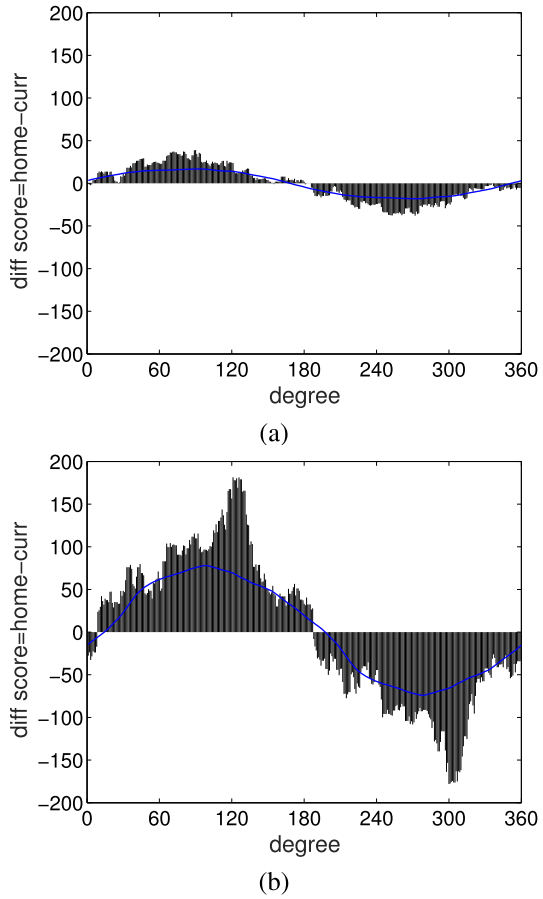


FIGURE 8. An example of histogram of the matching score differences over a pair of snapshot images in environment "a1original" (blue curve indicates a low-pass filtered result over the histogram) (a) test position in a close distance from home; home (5,8) and test position (5,7) (b) test position in a far distance from home; home (5,8) and test position (5,1) .

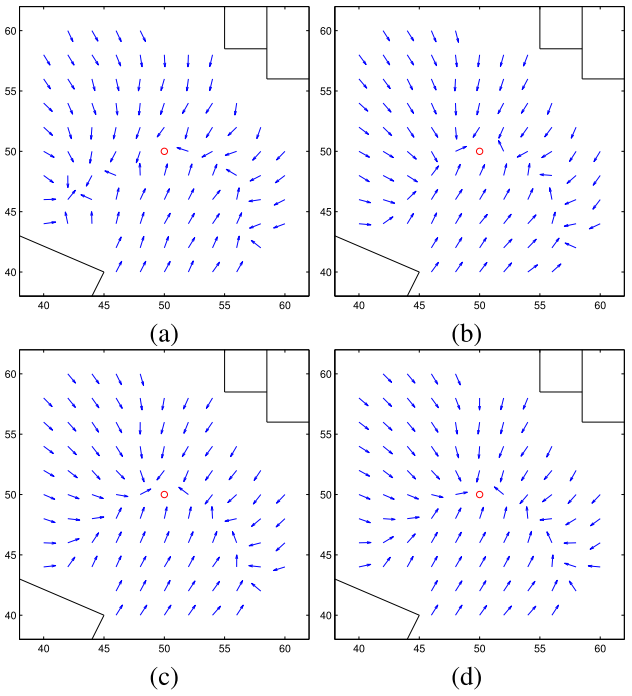


FIGURE 9. Homing performance in "ourlab" environment using the HFLV model with varying number of Haar-like masks. (a) 100 masks, (b) 500 masks, (c) 1000 masks, and (d) 5000 masks (red circle indicates the home position and each arrow the homing direction); the above environment shows a topview of arena, the arena size is 2.2 × 2.2 meter (unit 10 in the figure indicates 1 meter) and the landmark vector is calculated at a given (x, y) position.

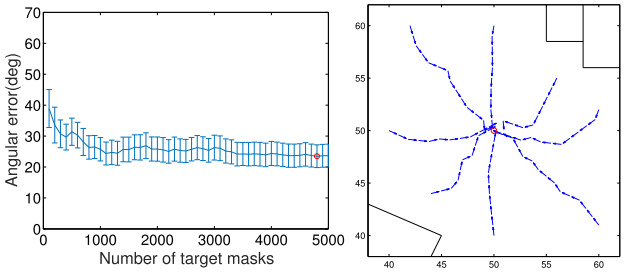


FIGURE 10. Angular errors with the HFLV model depending on varying number of Haar-like masks for environment 'ourlab'; left: error bars indicate 95% confidence intervals assuming t-distribution and red circle on the left indicates the minimum averaged error, right: homing trajectories.

An example of the matching score differences with a collection of random masks for environment "a1original" is shown in Fig. 8. The score differences within a window size of 90 degrees are accumulated into each bin and it roughly estimates the best homing direction (90°, the weighted average of histogram as a population coding). If the test position (5,7) is close to the home location (5,8), more symmetric distribution is observed. Even at a far position (5,1), the distribution of matching score difference can determine the homing direction. We also note that the sum of bin heights has information of geometric distance to home.

Fig. 9 shows the homing performance, depending on the number of random Haar-like masks. Even 100 masks can roughly estimate the homing direction at most of positions, although there are large angular errors for homing direction at some positions, especially at the bottom left area. With 500 masks, the homing performance was much improved. More masks were assigned to reflect well the environmental situation in that area. A sufficient number of random masks can guide homing well and the angular errors can be reduced greatly. As the number of random masks increases, the homing performance can be improved but we observed that it cannot reach zero angular error.

The angular errors for homing can be calculated as the difference between the desired homing direction and the estimated homing direction. The desired direction is directly drawn from the current position to the home location. Fig. 10 shows the angular errors depending on varying number of masks. The set with 1000 masks or more shows stabilized homing performance as we observed the same result in Fig. 9. The limit of homing performance may be related to the scope of the location, size or type of varying masks.

Next, we applied our landmark model to two different environments, "a1original" and "screen". In Fig. 11, the model

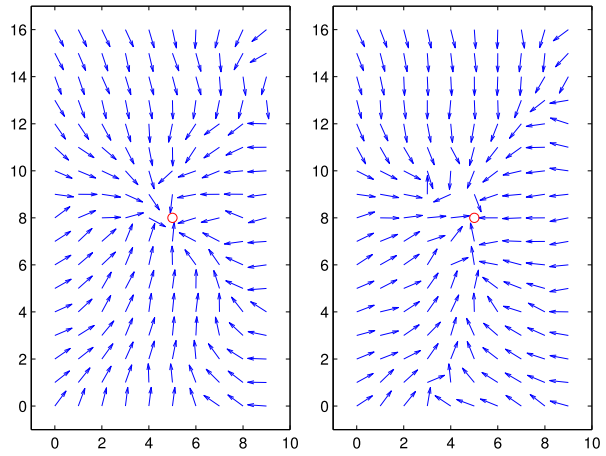


FIGURE 11. Homing performance in environments "a1original" and "screen" using the HFLV model with 1000 random masks; red circle indicates the home position and each arrow the homing direction (unit 1 indicates 0.3 meter).

shows a reasonable pattern of homing directions in the environment ''a1original'', but the flow of homing directions is not successful at the left side of the environment ''screen''. The ''screen'' environment has some occlusion of visual landmarks, which makes the problem difficult. The suggested landmark model with Haar-like masks assumes that all the landmarks should be commonly observed in any position. Violating that condition may degrade severely the homing performance.

We measured the sum of matching score differences given in equation (1) for a pair of snapshots to determine the homing direction with the HFLV model. Fig. 12 shows the sum of the score differences over a set of 1000 Haar-like masks for a pair of snapshots taken at the home location and at each grid position in a given environment, 'ourlab' or ''a1original''. The measure is roughly proportional to the geometric distance to the home location from each grid position (that is, distance between the two image-capturing points). The image difference depends on the relative distance of the capturing points. If a target position is fixed, the geometric distance between the target position and a given position can be roughly estimated by the matching score difference. As shown in Figure 12, the sum of score differences changes within a variance for a fixed distance and four spots in the grid have the same distance from the home location (four x marks are available for a fixed distance in Fig. 12.); see grid positions in Fig. 9. For small distances, the variance of the measure is small, and at a far distance, the variance tends to become large. It indirectly supports that neighboring positions, or positions close each other have more similar image distances. We observe that the matching score difference has the information of the geometric distance between the two capturing points. This idea will be used in the HFID model.

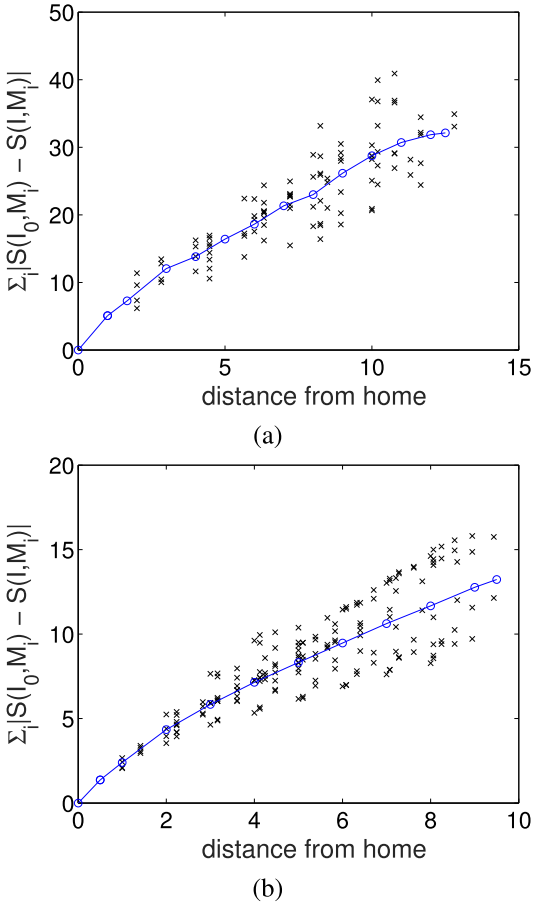


FIGURE 12. Sum of matching score differences over a pair of snapshots vs. distance between the snapshot-capturing points (one snapshot is the home snapshot image); blue line is the curve fitted to data samples: (a) environment 'ourlab' and (b) environment ''a1original''.

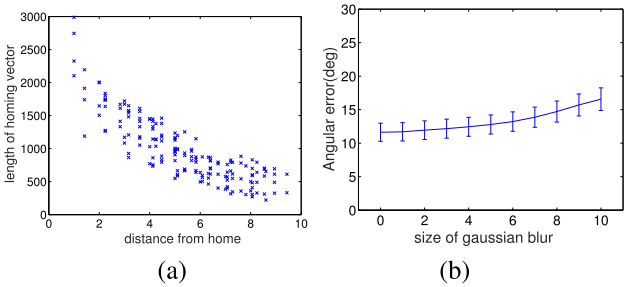


FIGURE 13. The HFID model test with 1000 random Haar-like masks for the environment "a1original": (a) length of homing vector (from eq. (8)) vs. distance from the home location, (b) effect of Gaussian blur (σ: blurring parameter used) .

B. HFID MODEL

In a similar way to the above experiments with the HFLV model, we tested the HFID model based on the relative distance of matching scores. This method uses three snapshot images around the home location and thus three comparisons with the snapshot taken at the current position are required. The matching score difference between each pair of snapshots can determine the homing direction.

Determining the homing direction at a given location is involved with comparing the snapshot at the position and three similar snapshots around the home location. Those

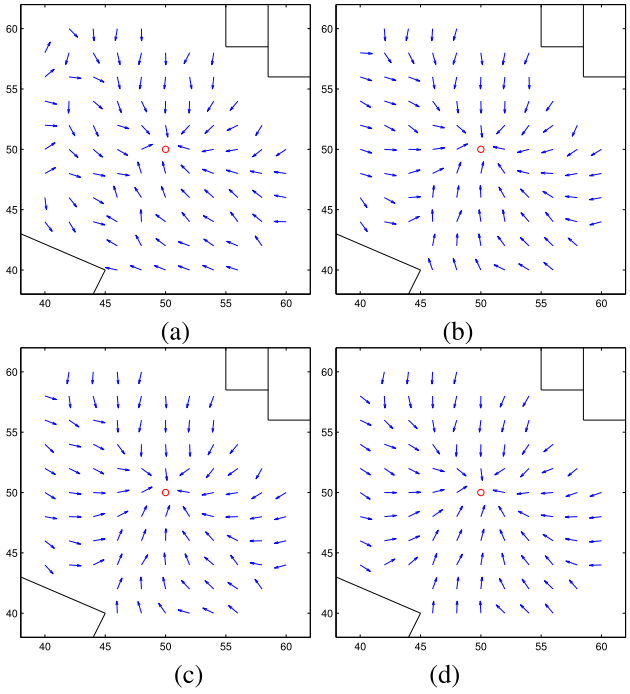


FIGURE 14. Homing performance in 'ourlab' environment using the HFID model with varying number of Haar-like masks: (a) 100 masks; (b) 500 masks; (c) 1000 masks; (d) 5000 masks; red circle indicates the home position and each arrow the homing direction.

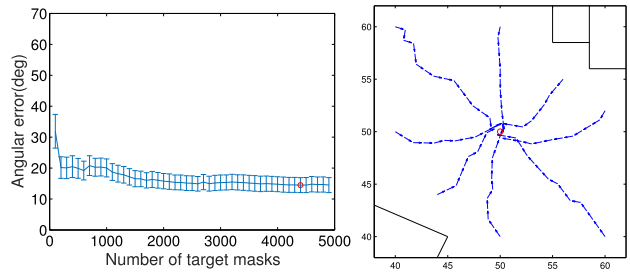


FIGURE 15. Angular errors with the HFID model depending on varying number of Haar-like masks for environment 'ourlab'; left: error bars indicate 95% confidence intervals assuming t-distribution and red circle on the left indicates the minimum averaged error, right: homing trajectories.

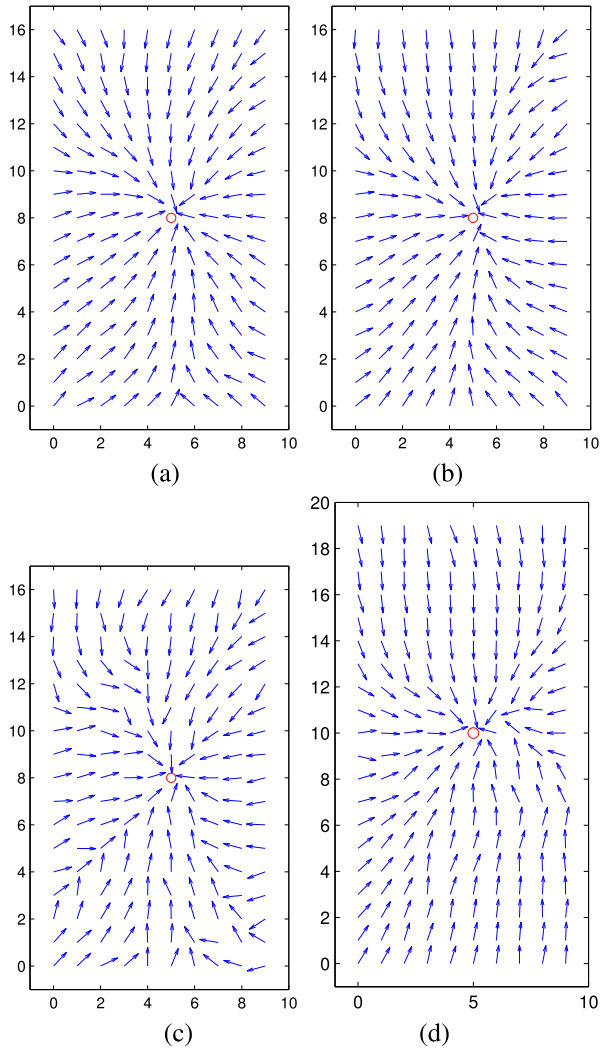


FIGURE 16. Homing performance with the HFID model in four different environments with 1000 random masks; (a) "a1original," (b) "screen," (c) "arboreal," and (d) "hall1"; red circle indicates the home position and each arrow the homing direction.

image differences through a set of Haar-like masks can more precisely estimate the homing direction.

For the HFID model, the homing vector is estimated by computing the relative image distance to three target images from the current view as shown in equation (8). The vector maps only show the homing direction and the vector magnitude becomes larger when the current position is closer to the home location. Fig. 13(a) shows the length of homing vector depending on the distance from the target location with the HFID model. Large vector length has a potential for more accurate estimation of the homing direction. If the robot takes a snapshot at a far distance from the home, the image distance between each of the three target images and the current view becomes similar. Similar image distances derive a small magnitude of home vector, since the home vector is calculated with the difference of image distances.

Also, we investigated the effect of blurring images for the HFID model. More blurred images tend to degrade the homing performance as shown in Fig. 13(b), but not much change. The resolution of images can influence the performance. The HFID method needs two extra target images near the home location in addition to the home snapshot, and the distance between those targets and the home also affects the homing performance; large distances tend to have larger angular errors (data not shown here).

Fig. 14 shows the homing directions with the HFID model according to the number of masks. With 100 random masks, many scattered direction patterns are observed. The set with 500 masks or more show stabilized homing performance, while there are large angular errors at a few positions near the bottom left corner. The HFID model shows more direct direction patterns in homing, compared to the HFLV model. When we tested varying number of random masks, the angular errors with the HFID model are significantly lower than those

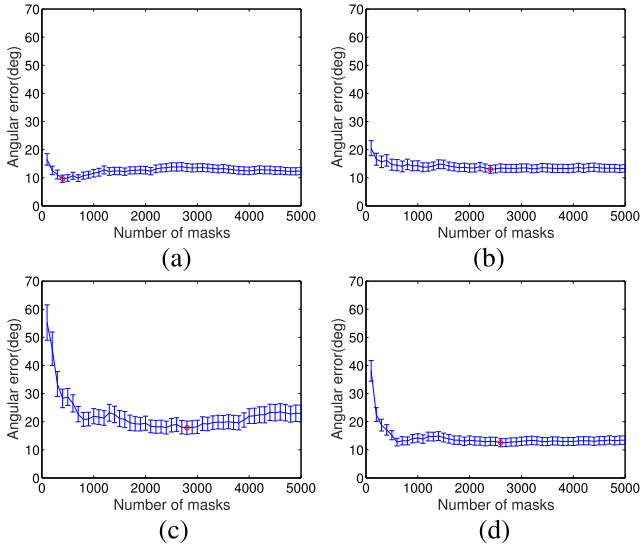


FIGURE 17. Angular errors with the HFID model depending on varying number of Haar-like masks for four environments (a) "a1original," (b) "screen," (c) "arboreal," and (d) "hall1" (error bars indicate 95% confidence intervals assuming *t*-distribution, and red circle indicates the minimum averaged error).

with HFLV model as shown in Fig. 15. More random masks can improve the homing performance, but the performance seems to be stabilized at some angular errors. The homing route is also demonstrated with various starting positions and smaller angular errors are observed near the home location.

We applied the HFID model to the other four environments. Figure 16 shows the results for homing performance, and the HFID model more robustly operate to estimate the homing directions in most of positions. The method calculates the matching score differences with three snapshots near the home location. The relative difference over the three snapshots is not much influenced even when there are occlusion points or significant change of landmarks at the current position. The surrounding background images will more contribute to the sum of matching score differences.

Fig. 17 shows the homing errors depending on varying number of random masks in four different environments. We observe commonly that a small number of masks, for example, 100 masks are insufficient to read the environmental information. The set with 1000 random masks or more can sketch the visual cue needed for good homing performance. For the environment ''arboreal'', much change of angular errors are observed when the number of masks increases. In contrast, the environment ''hall1'' experiences very smooth change of angular errors when the number of masks increases, and a relatively small number of masks are sufficient to derive homing directions.

Fig. 18 shows the result for homing in the environment ''a1original'' and ''hall1'' with varying number of masks. With a small set of masks, for instance, 100 masks, there are large angular errors at many positions. Too sparse masks in the angular space have difficulty in catching the landmark characteristics. As expected, more masks helps better homing performance. A large number of masks can read the environmental features better. It is notable that more direct paths to the goal position from an arbitrary position are observed using the HFID model than the HFLV model.

C. PERFORMANCE COMPARISON

We tested various local homing approaches for several environments and evaluated their performances to compare them with the suggested methods, the HFLV and the HFID model. Fig. 19 shows vector maps with our approaches as well as other four different algorithms for the environment ''a1original''. The HFID model shows better homing performance than the HFLV model. With the HFLV method, angular errors can be observed at some positions. Especially, the COMALV has some homing patterns depending on the test area – see Fig. 19(d). The COM-ID has more direct homing direction at many points, but large angular errors are shown at specfic positions. The MinWarping method guides a direct route to the home at most of positions. The HFLV has better homing performance than the COMALV, and similarly the HFID has better performance than the COM-ID. It implies that a random distribution of Haar-like features contribute to the homing decision effectively. The HFID has better performance than even the MinWarping and SIFTw in those environments.

Table 1 shows homing performances for various environments with the HFLV model, the HFID model, the DID, the MinWarping, the COMALV, the COM-ID and the SIFTw method. The HFID model produces small homing errors for all the test environments. It shows that the HFID model has a more direct route to the target position than the other approaches. Except for the environment ''arboreal'', the HFID model has the best homing performance in terms of the averaged angular errors and the catchment area. The DID method also uses the image distance with three reference images around the home location, but instead uses the whole pixels for snapshot comparison, while the HFID model uses Haar-like features. It indirectly supports that a distribution of Haar-like features characterizes well the landmarks in the environment. A simple form of the image pixel bands used in the COMALV and the COM-ID seem to have a limitation for desired homing capacity.

The HFLV model shows good results in the environment ''a1original'', but it has worse performance in the environment ''hall1''. The environment ''hall1'' includes many lights reflected on the surface, which change the image intensities a lot depending on observing points. If an environment is not isotropic, the HFLV model and the COMALV experience much degradation of homing performance. The HFID method and the COM-ID method using the relative image distance to the three target images are very effective in the non-isotropic environment or the environment with occlusions or many pixel-level changes. In the environment ''ourlab'' with intense illumination condition in the window, the HFID and the SIFTw have an outstanding performance. We can confirm that the SIFT features extract landmarks effectively in the environment with the illumination

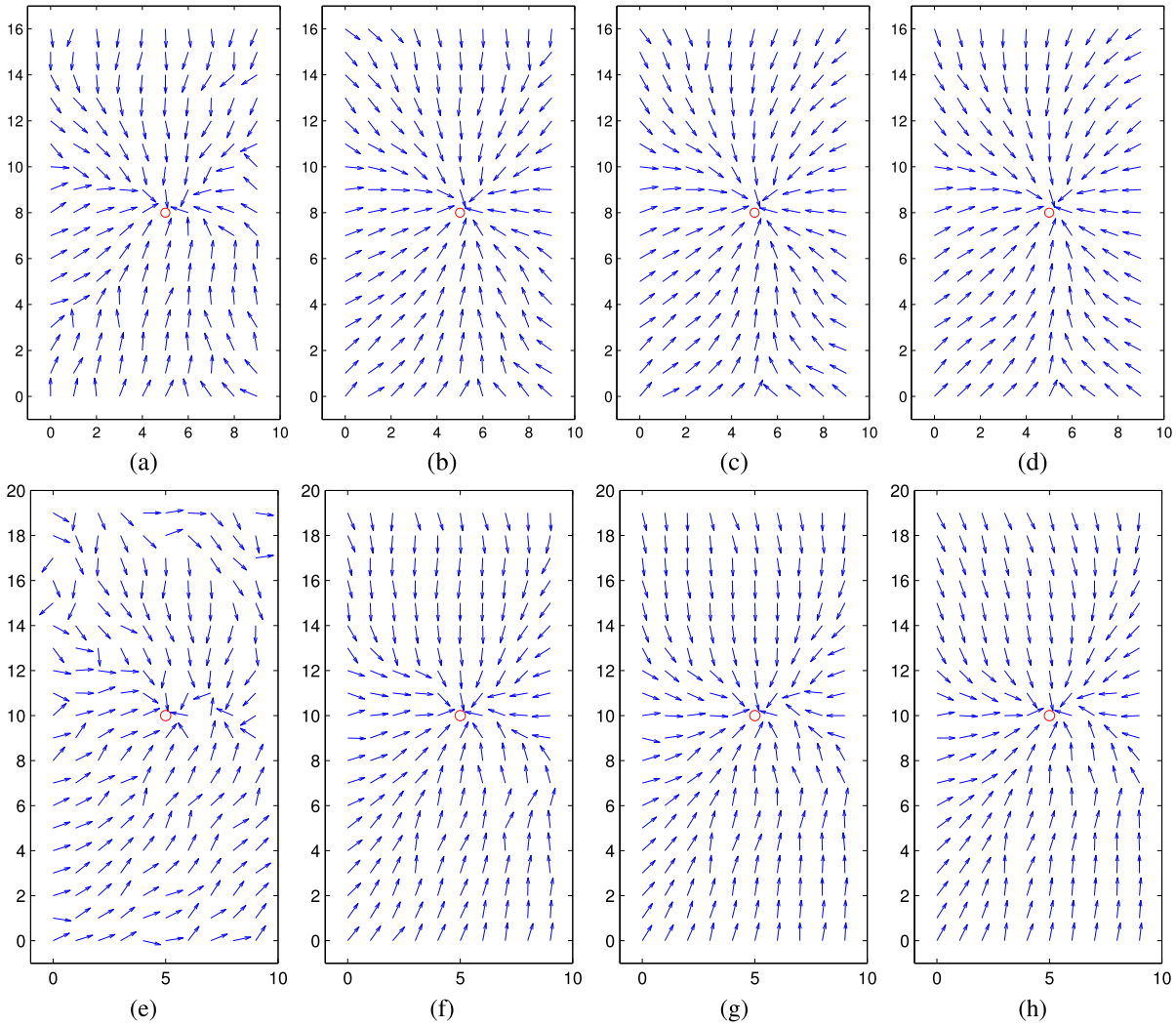


FIGURE 18. Homing performance with the HFID model in the environments "a1original" (a)-(d) and "hall1" (e)-(h) (varying number of Haar-like masks tested); env A and env H represent "a1original" and "hall1" environments, respectively, and red circle indicates the home position and each arrow the homing direction. (a) 100 masks, env A. (b) 500 masks, env A. (c) 1000 masks, env A. (d) 5000 masks, env A. (e) 100 masks, env H. (f) 500 masks, env H. (g) 1000 masks, env H. (h) 5000 masks, env H.

condition. The MinWarping method shows a superior performance in the environment ''arboreal''. Generally, the HFID method shows stable performance for all the environments we tested. The approach is comparable to two state-of-art homing algorithms, the MinWarping and the SIFTw.

IV. DISCUSSION

From the experiments, we observed that the suggested two methods effectively estimate the homing vector. The second method, the HFID method, shows more robust homing performance while the first landmark model, HFLV method may be sensitive to the environmental change, since some landmark features at a specific angular position may not be identified easily with Haar-like masks. The HFLV method has weak points in the environment with many occlusions or much pixel-intensity change, for example, ''arboreal'' and ''hall1''. Similar results are observed with the COMALV method. However, the first method needs only a single target image to determine the homing direction. The two suggested methods may degrade the performance when many landmark occlusions occur or the visual landscape changes a lot in the two snapshots.

Our visual navigation algorithms using Haar-like features are based on the snapshot model. The model comparing a pair of snapshots at the goal and at the current position are useful for local navigation. It assumes the majority of landmarks are commonly observed in the two snapshots as isotropic environments. Complex environments with many occlusions can lead to large homing errors, due to the violation of the assumption. A long range of homing model expecting a lot of environmental changes may need a series of waypoints [49]. Each waypoint can be a target position for local navigation. An appropriate collection of waypoints can lead to the final goal position.

In the HFID model with matching score differences based on Haar-like masks, we use three snapshots around the home

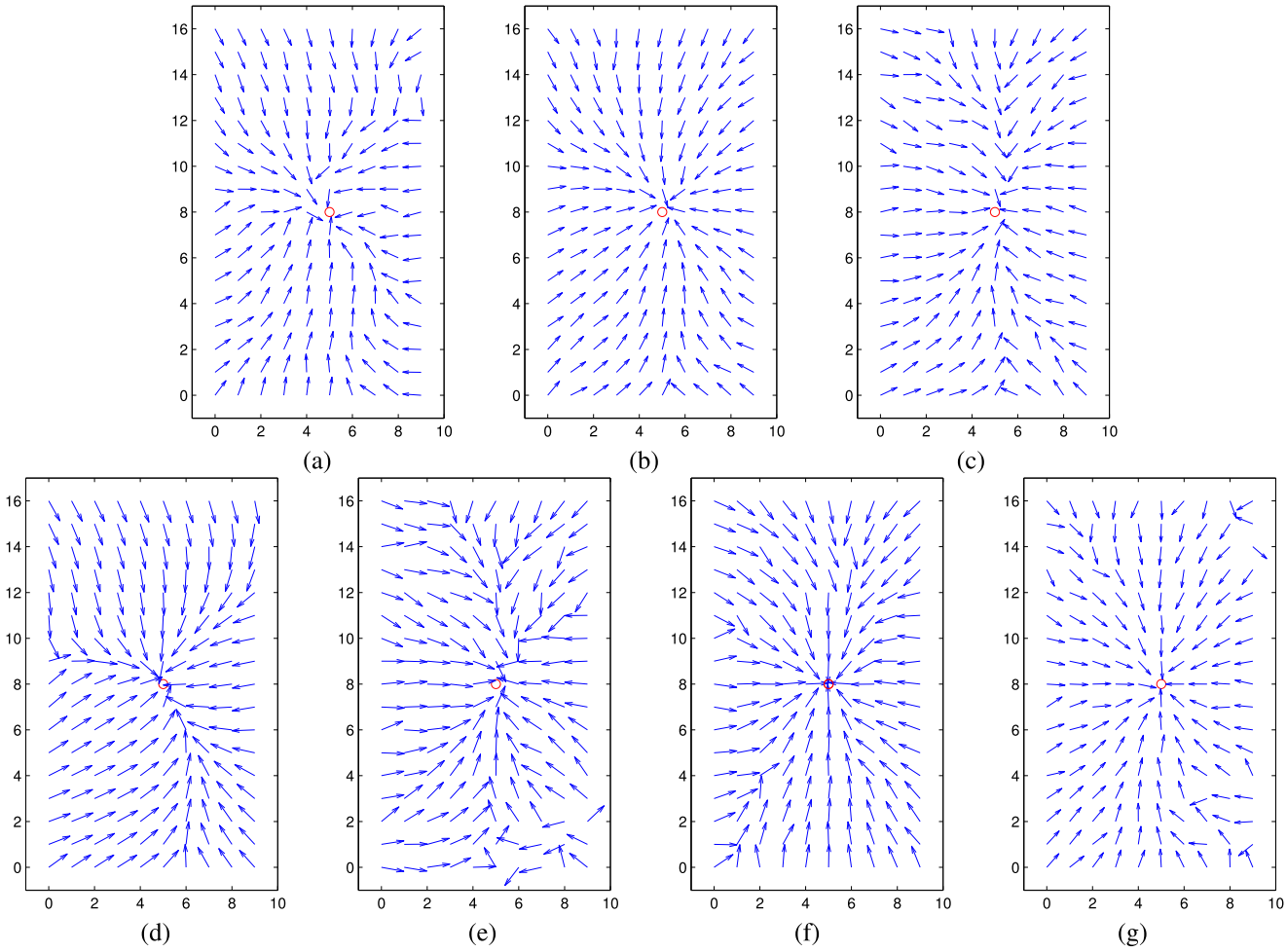


FIGURE 19. Comparison of various methods in vector map for the environment "a1original." (a) The HFLV model; (b) the HFID model; (c) the DID method; (d) the COMALV method; (e) the COM-ID method; (f) the MinWarping method; and (g) the SIFT-Warping method; red circle indicates the home position and each arrow the homing direction.

location to be compared with the snapshot at the current position. The comparisons can estimate the relative distances to the current position. As an alternative, temporal update of the matching score difference as the robot moves can determine the homing direction. That is, the robot can calculate the matching score differences between the snapshots at a series of points and the snapshot at the home location.

A collection of landmarks at the two snapshots are compared for matching. A landmark's angular position can be shifted depending on the observer's position. The reference compass is an essential component to find the correspondence of landmarks. If the coordinate frame is rotated, all landmarks in view are moved to another angle. If there is a reference compass available, the change of landmark positions can easily be monitored. It is reported that biological animals use a reference compass [16], [50]–[52]. In our snapshot model experiments, a reference compass is available and the coordinate is aligned. In robotic experiments, the alignment process of landmarks may be needed, if there is no reference compass. The homing navigation approaches, ALV, warping, and DID methods also need the alignment process if there is no reference compass. The Distance Estimated Landmark Vector (DELV) method [47], [48] uses a landmark arrangement method to change the coordinate for the whole set of landmarks. Also, visual compass method is available to align a pair of snapshot images [21], [53]. The visual compass approach can be applied to align the two snapshot images before Haar-like features are used.

The COMALV uses a uniform distribution of the image subregions to estimate the image intensity values by modeling biological ommatidia. In contrast, the HFLV uses a set of randomly distributed Haar-like features each of which has a varying size at an arbitrary location. In the homing experiments, the HFLV is significantly better than the COMALV, except for the environment "hall1". Similarly, the HFID is significantly better in homing performance than the COM-ID method using a uniform distribution of image subsamples. It seems that a collection of Haar-like masks, that is, a diverse form of masks or bipolar-like regions contribute to capturing salient features in the snapshot to guide homing. Possibly

TABLE 1. Angular errors with various methods including ours in five environments (ϵ_θ is the absolute angular error and N is the number of test points; 1000 masks used, μ is the mean of angular errors and σ is 95% confidence interval by assuming t-distribution).

		ourlab	a1original	screen	arboreal	hall1
	N	95	169	169	169	199
HFLV model	error μ ($\pm\sigma$)	25 ($\pm$4.4)	16 ($\pm$1.8)	22 ($\pm$2.9)	27 ($\pm$3.0)	59 ($\pm$6.2)
	$0 \le \varepsilon_\theta < 45°$	77%	96%	89%	78%	42%
	$45 \le \varepsilon_\theta < 90°$	22%	4%	9%	21%	30%
	$90 \le \varepsilon_\theta < 180°$	1%	0%	2%	1%	28%
	catchment area (%)	100%	100%	77%	68%	39%
HFID model	error μ ($\pm\sigma$)	20 ($\pm$2.6)	**9 ($\pm$1.2)**	**13 ($\pm$1.4)**	18 ($\pm$2.4)	**13 ($\pm$1.4)**
	$0 \le \varepsilon_\theta < 45°$	88%	99%	100%	96%	99%
	$45 \le \varepsilon_\theta < 90°$	12%	1%	0%	4%	1%
	$90 \le \varepsilon_\theta < 180°$	0%	0%	0%	0%	0%
	catchment area (%)	100%	100%	100%	100%	100%
DID	error μ ($\pm\sigma$)	25 ($\pm$3.0)	23 ($\pm$2.6)	19 ($\pm$2.0)	37 ($\pm$5.4)	21 ($\pm$1.7)
	$0 \le \varepsilon_\theta < 45°$	94%	92%	96%	75%	95%
	$45 \le \varepsilon_\theta < 90°$	6%	7%	4%	14%	5%
	$90 \le \varepsilon_\theta < 180°$	0%	1%	0%	11%	0%
	catchment area (%)	100%	100%	100%	88%	100%
MinWarping	error μ ($\pm\sigma$)	61 ($\pm$9.5)	11 ($\pm$1.5)	20 ($\pm$2.9)	**16 ($\pm$2.4)**	31 ($\pm$4.2)
	$0 \le \varepsilon_\theta < 45°$	45%	99%	91%	92%	81%
	$45 \le \varepsilon_\theta < 90°$	25%	1%	8%	8%	13%
	$90 \le \varepsilon_\theta < 180°$	26%	0%	1%	0%	6%
	catchment area (%)	67%	100%	100%	100%	95%
COMALV	error μ ($\pm\sigma$)	49 ($\pm$7.4)	19 ($\pm$2.0)	34 ($\pm$4.4)	44 ($\pm$6.3)	49 ($\pm$6.6)
	$0 \le \varepsilon_\theta < 45°$	56%	98%	71%	64%	58%
	$45 \le \varepsilon_\theta < 90°$	29%	2%	25%	21%	21%
	$90 \le \varepsilon_\theta < 180°$	15%	0%	4%	15%	21%
	catchment area (%)	12%	100%	64%	51%	59%
COM-ID	error μ ($\pm\sigma$)	35 ($\pm$5.4)	25 ($\pm$3.3)	19 ($\pm$2.9)	26 ($\pm$4.1)	24 ($\pm$2.2)
	$0 \le \varepsilon_\theta < 45°$	72%	86%	93%	83%	90%
	$45 \le \varepsilon_\theta < 90°$	25%	13%	6%	12%	9%
	$90 \le \varepsilon_\theta < 180°$	3%	1%	1%	5%	1%
	catchment area (%)	96%	98%	99%	96%	99%
SIFTw	error μ ($\pm\sigma$)	**14 ($\pm$2.2)**	12 ($\pm$2.5)	20 ($\pm$4.2)	21 ($\pm$4.2)	18 ($\pm$2.6)
	$0 \le \varepsilon_\theta < 45°$	99%	96%	91%	88%	92%
	$45 \le \varepsilon_\theta < 90°$	1%	3%	6%	8%	7%
	$90 \le \varepsilon_\theta < 180°$	0%	1%	3%	4%	1%
	catchment area (%)	100%	100%	99%	97%	100%

some saliency measure can be used to search for the location of Haar-like masks. A mask for a specific landmark can be positioned and better matching masks can be collected. We need further study to check if those masks are more helpful to estimate the homing direction.

The basic idea with Haar-like masks is to use a set of low-level weak classifiers. A set of those classifiers can become a strong classifier or provide a cue for the homing direction. Here, collecting the classifiers does not follow a sophisticated procedure in order, but simply distribute the weak classifiers at random positions to catch landmark features in the snapshot image. In the HFID model, Haar-like masks were randomly distributed. We also tested a uniform distribution of masks, 720 masks where each mask was selected for each angular direction (a resolution of 0.5 degree) as the mask with highest matching score on the home snapshot; 100 random masks were generated for each angle and the best matching mask were chosen. However, this process did not improve the HFID model, but the homing accuracy becomes worse. A uniform distribution of the best matching masks is not a solution to achieve the desired homing performance. Weak classifiers in a random set can contribute to covering various environmental landscape. We need more rigorous analysis to explain this, and leave it as a future work.

V. CONCLUSION

In this paper, we handle visual homing navigation with the snapshot model. The snapshot model is to compare the snapshot images at the current location and at the home location to determine the homing direction. Following the snaphot model suggested by Cartwright and Collett [17], our approach calculates the matching scores with Haar-like features over the two snapshot images. We propose two types of homing vectors based on the image difference.

The first method (the HFLV model) calculates the matching score differences with a set of Haar-like masks and then estimates the homing vector. When each mask is expressed as a landmark vector whose length is set to the matching score, the averaged landmark vector can represent the corresponding snapshot at a given location. The difference of the two averaged landmark vectors can determine the homing direction, which is equivalent to the net effect of the matching score difference with a set of masks.

Another method (the HFID model) that we suggest is to use the matching score difference between each one of three snapshots near the home location and the snapshot at the current location. The matching score difference approximately estimates the distance between a pair of the image-capturing spots, and a set of relative image distances for three snapshots

near the home location can determine the homing direction. The results with the HFID model show more accurate estimation of homing direction even in the environments with occlusions or with intense change in the image intensity. Our results show that a collection of random Haar-like masks represent well the characteristics of a snapshot image. Sophisticated feature extraction is not required for visual homing, but a set of simple masks are sufficient to find the visual features, especially to derive the homing direction.

Our visual homing navigation method can be applied to cleaning robots that need to return home for a battery recharge or exploration robots which are supposed to move towards a goal position. Also, the visual navigation can be useful in the application of guiding a mobile robot towards a given target, compensating for odometry information, for example, at a loop-closing positon in localization and mapping. We tested five indoor environments and some environments include occlusions. In the complex cluttered environment, more occlusions or environmental change can occur. We need further study to develop more advanced model with Haar-like features. We can also study a series of waypoint searches leading to the final goal position for real application with a long travel of exploration.

REFERENCES

[1] E. Huber and D. Kortenkamp, "Using stereo vision to pursue moving agents with a mobile robot," in *Proc. IEEE Int. Conf. Robot. Autom.*, vol. 3, May 1995, pp. 2340–2346.

[2] J. A. Méndez-Polanco, A. Muñoz Meléndez, and E. F. Morales, "People detection by a mobile robot using stereo vision in dynamic indoor environments," in *MICAI 2009: Advances in Artificial Intelligence*. Springer, 2009, pp. 349–359.

[3] R. Möller, "Local visual homing by warping of two-dimensional images," *Robot. Auto. Syst.*, vol. 57, no. 1, pp. 87–101, 2009.

[4] D. Churchill and A. Vardy, "An orientation invariant visual homing algorithm," *J. Intell. Robot. Syst.*, vol. 71, no. 1, pp. 3–29, 2003.

[5] A. Denuelle, R. Strydom, and M. V. Srinivasan, "Snapshot-based control of UAS hover in outdoor environments," in *Proc. IEEE Int. Conf. Robot. Biomimetics (ROBIO)*, Dec. 2015, pp. 1278–1284.

[6] A. Denuelle, S. Thurrowgood, F. Kendoul, and M. V. Srinivasan, "A view-based method for local homing of unmanned rotorcraft," in *Proc. 6th Int. Conf. Autom., Robot. Appl. (ICARA)*, Feb. 2015, pp. 443–449.

[7] S. Thrun and M. Montemerlo, "The graph SLAM algorithm with applications to large-scale mapping of urban structures," *Int. J. Robot. Res.*, vol. 25, nos. 5–6, pp. 403–429, 2006.

[8] A. J. Davison, Y. G. Cid, and N. Kita, "Real-time 3D SLAM with wide-angle vision," *IFAC Proc. Volumes*, vol. 37, no. 8, pp. 868–873, 2004.

[9] M. W. M. G. Dissanayake, P. Newman, S. Clark, H. F. Durrant-Whyte, and M. Csorba, "A solution to the simultaneous localization and map building (SLAM) problem," *IEEE Trans. Robot. Autom.*, vol. 17, no. 3, pp. 229–241, Jun. 2001.

[10] W. Hübner, "From homing behavior to cognitive mapping: Integration of egocentric pose relations and allocentric landmark information in a graph model," Ph.D. dissertation, Fachbereich Math. Inform., Univ. Bremen, Bremen, Germany, 2005.

[11] C. S. Jensen, H. Lu, and B. Yang, "Graph model based indoor tracking," in *Proc. 10th Int. Conf. IEEE Mobile Data Manage., Syst., Services Middleware*, May 2009, pp. 122–131.

[12] M. Srinivasan, "Honey bees as a model for vision, perception, and cognition," *Annu. Rev. Entomol.*, vol. 55, pp. 267–284, Jan. 2010.

[13] M. Collett and T. Collett, "How do insects use path integration for their navigation?" *Biol. Cybern.*, vol. 83, no. 3, pp. 245–259, 2000.

[14] M. A. Willis, "Chemical plume tracking behavior in animals and mobile robots," *Navigation*, vol. 55, no. 2, pp. 127–135, 2008.

[15] A. Garm, M. Oskarsson, and D.-E. Nilsson, "Box jellyfish use terrestrial visual cues for navigation," *Current Biol.*, vol. 21, no. 9, pp. 798–803, 2011.

[16] S. F. Reid, A. Narendra, J. M. Hemmi, and J. Zeil, "Polarised skylight and the landmark panorama provide night-active bull ants with compass information during route following," *J. Exp. Biol.*, vol. 214, no. 3, pp. 363–370, 2011.

[17] B. A. Cartwright and T. S. Collett, "Landmark learning in bees," *J. Comparative Physiol.*, vol. 151, no. 4, pp. 521–543, 1983.

[18] D. Lambrinos, R. Möller, T. Labhart, R. Pfeifer, and R. Wehner, "A mobile robot employing insect strategies for navigation," *Robot. Auto. Syst.*, vol. 30, pp. 39–64, Jan. 2000.

[19] V. V. Hafner, "Adaptive homing—Robotic exploration tours," *Adapt. Behav.*, vol. 9, pp. 131–141, Sep. 2001.

[20] S.-E. Yu, C. Lee, and D. Kim, "Analyzing the effect of landmark vectors in homing navigation," *Adapt. Behav.*, vol. 20, pp. 337–359, 2012.

[21] J. Zeil, M. I. Hofmann, and J. S. Chahl, "Catchment areas of panoramic snapshots in outdoor scenes," *J. Opt. Soc. Amer. A, Opt. Image Sci.*, vol. 20, no. 3, pp. 450–469, 2003.

[22] M. O. Franz, B. Schölkopf, H. A. Mallot, and H. Bülthoff, "Where did I take that snapshot? Scene-based homing by image matching," *Biol. Cybern.*, vol. 79, no. 3, pp. 191–202, 1998.

[23] M. Mangan and B. Webb, "Modelling place memory in crickets," *Biol. Cybern.*, vol. 101, no. 4, pp. 307–323, 2009.

[24] L. Smith, A. Philippides, P. Graham, B. Baddeley, and P. Husbands, "Linked local navigation for visual route guidance," *Adapt. Behav.*, vol. 15, no. 3, pp. 257–271, 2007.

[25] A. Goldhoorn, A. Ramisa, R. L. de Mántaras, and R. Toledo, "Using the average landmark vector method for robot homing," *Frontiers Artif. Intell. Appl.*, vol. 163, pp. 331–338, 2007.

[26] A. Ramisa, A. Goldhoorn, D. Aldavert, R. Toledo, and R. L. de Mántaras, "Combining invariant features and the ALV homing method for autonomous robot navigation based on panoramas," *J. Intell. Robot. Syst.*, vol. 64, pp. 625–649, Dec. 2011.

[27] C. Lee, S.-E. Yu, and D. Kim, "Landmark-based homing navigation using omnidirectional depth information," *Sensors*, vol. 17, no. 8, p. 1928, 2017.

[28] R. Möller and A. Vardy, "Local visual homing by matched-filter descent in image distances," *Biol. Cybern.*, vol. 95, no. 5, pp. 413–430, 2006.

[29] J. Zeil, N. Boeddeker, and W. Stürzl, "Visual homing in insects and robots," in *Flying Insects and Robots*. Berlin, Germany: Springer, 2009, pp. 87–100.

[30] F. Labrosse, "Short and long-range visual navigation using warped panoramic images," *Robot. Auto. Syst.*, vol. 55, no. 9, pp. 675–684, 2007.

[31] M. Franz, "Minimalistische visuelle navigation," Ph.D. dissertation, Univ. Tübingen, Tübingen, Germany, 1999.

[32] R. Möller, M. Krzykawski, and L. Gerstmayr, "Three 2D-warping schemes for visual robot navigation," *Auto. Robots*, vol. 29, pp. 253–291, Nov. 2010.

[33] Q. Zhu, C. Liu, and C. Cai, "A novel robot visual homing method based on SIFT features," *Sensors*, vol. 15, no. 10, pp. 26063–26084, 2015.

[34] M. Horst and R. Möller, "Visual place recognition for autonomous mobile robots," *Robotics*, vol. 6, no. 2, p. 9, 2017.

[35] D. Fleer and R. Möller, "Comparing holistic and feature-based visual methods for estimating the relative pose of mobile robots," *Robot. Auto. Syst.*, vol. 89, pp. 51–74, Mar. 2017.

[36] K. Basten and H. A. Mallot, "Simulated visual homing in desert ant natural environments: Efficiency of skyline cues," *Biol. Cybern.*, vol. 102, no. 5, pp. 413–425, 2010.

[37] A. Wystrach, A. Dewar, A. Philippides, and P. Graham, "How do field of view and resolution affect the information content of panoramic scenes for visual navigation? A computational investigation," *J. Comparative Physiol. A*, vol. 202, no. 2, pp. 87–95, 2016.

[38] B. Baddeley, P. Graham, A. Philippides, and P. Husbands, "Holistic visual encoding of ant-like routes: Navigation without waypoints," *Adapt. Behav.*, vol. 19, no. 1, pp. 3–15, 2011.

[39] B. Baddeley, P. Graham, P. Husbands, and A. Philippides, "A model of ant route navigation driven by scene familiarity," *PLoS Comput. Biol.*, vol. 8, no. 1, p. e1002336, 2012.

[40] S.-B. Paik and D. L. Ringach, "Retinal origin of orientation maps in visual cortex," *Nature Neurosci.*, vol. 14, no. 7, pp. 919–925, 2011.

[41] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 1, Dec. 2001, pp. I-511–I-518.

[42] P. Viola and M. J. Jones, "Robust real-time face detection," *Int. J. Comput. Vis.*, vol. 57, no. 2, pp. 137–154, 2004.
[43] T. Mita, T. Kaneko, and O. Hori, "Joint haar-like features for face detection," in *Proc. 10th IEEE Int. Conf. Comput. Vis. (ICCV)*, vol. 2, Oct. 2005, pp. 1619–1626.
[44] R. Lienhart and J. Maydt, "An extended set of haar-like features for rapid object detection," in *Proc. Int. Conf. Image Process.*, vol. 1, Sep. 2002, pp. I-900–I-903.
[45] A. Vardy and R. Möller, "Biologically plausible visual homing methods based on optical flow techniques," *Connection Sci.*, vol. 17, pp. 47–89, Mar. 2005.
[46] R. Möller, A. Vardy, S. Kreft, and S. Ruwisch, "Visual homing in environments with anisotropic landmark distribution," *Auto. Robots*, vol. 23, no. 3, pp. 231–245, 2007.
[47] S.-E. Yu and D. Kim, "Landmark vectors with quantized distance information for homing navigation," *Adapt. Behav.*, vol. 19, no. 2, pp. 121–141, 2011.
[48] S.-E. Yu and D. Kim, "Image-based homing navigation with landmark arrangement matching," *Inf. Sci.*, vol. 181, no. 16, pp. 3427–3442, 2011.
[49] A. Vardy, "Long-range visual homing," in *Proc. IEEE Int. Conf. Robot. Biomimetics (ROBIO)*, Dec. 2006, pp. 220–226.
[50] R. Wehner, "The ant's celestial compass system: Spectral and polarization channels," in *Orientation and Communication in Arthropods*. Basel, Switzerland: Birkhäuser, 1997, pp. 145–185.
[51] T. Kimchi, A. S. Etienne, and J. Terkel, "A subterranean mammal uses the magnetic compass for path integration," *Proc. Nat. Acad. Sci. USA*, vol. 101, no. 4, pp. 1105–1109, 2004.
[52] T. S. Collett and J. Baron, "Biological compasses and the coordinate frame of landmark memories in honeybees," *Nature*, vol. 368, pp. 137–140, Mar. 1994.
[53] F. Labrosse, "The visual compass: Performance and limitations of an appearance-based method," *J. Field Robot.*, vol. 23, no. 10, pp. 913–941, 2006.

CHANGMIN LEE received the B.E. degree from the Department of Electrical and Electronic Engineering, Yonsei University, South Korea, where he is currently pursuing the Ph.D. degree. His research interests are in the areas of visual navigation, artificial intelligence, and neural networks.

DAEEUN KIM received the B.E. degree from the Department of Computer Science and Engineering, Seoul National University, the M.S. degree from the University of Michigan at Ann Arbor, and the Ph.D. degree from the University of Edinburgh in 2002. From 2002 to 2006, he was a Research Scientist with the Max Planck Institute for Human Cognitive and Brain Sciences. He is currently an Associate Professor with Yonsei University, South Korea. His research interests are in the areas of biorobotics, autonomous robots, artificial life, neural networks, and neuroethology.