Robotica (2013) volume 31, pp. 479–491. © Cambridge University Press 2012
doi:10.1017/S0263574712000434

Multi-waypoint visual homing in piecewise linear trajectory

Yu Fu†, Tien-Ruey Hsiang‡,* and Sheng-Luen Chung†

†*National Taiwan University of Science and Technology, Electrical Engineering, Taiwan*
‡*Department of Computer Science and Information Engineering, National Taiwan University of Science and Technology, Taiwan*

(Accepted July 18, 2012. First published online: August 16, 2012)

SUMMARY
This paper proposes an image sequence-based navigation method under the teaching-replay framework for robots in piecewise linear routes. Waypoints used by the robot contain either the positions with large heading changes or selected midway positions between junctions. The robot applies local visual homing to move between consecutive waypoints. The arrival at a waypoint is determined by minimizing the average vertical displacements of feature correspondences. The performance of the proposed approach is supported by extensive experiments in hallway and office environments. While the homing speed of robots using other approaches is constrained by the speed in the teaching phase, our robot is not bounded by such limit and can travel much faster without compromising the homing accuracy.

KEYWORDS: Multi-waypoint visual homing; Piecewise linear route.

1. Introduction

Navigation is one of the most important abilities of a service robot. Navigation with visual sensors[1] is widely investigated due to the availability of fast processors, low-cost, and high-quality cameras.

This paper presents an image sequence-based navigation method. The image sequence-based navigation method consists of two phases: teaching phase and replay phase. In the teaching phase, the robot moves along a route with manual guidance or other exploration methods. Meanwhile, a reference image sequence is recorded to represent the route. The positions where the reference images are captured are used as waypoints. With the reference image sequence, the robot is capable of navigating automatically between any two consecutive waypoints with local visual homing (or visual servoing[2,3]) in the replay phase.

Prior work on image sequence-based navigation[4–21] provides solutions for the selection of waypoints, the best route when a destination is given, the navigation between consecutive waypoints, and the moment to change the targeted waypoint to the next, and to start another local visual homing. Different methods for detecting the arrival at a waypoint are discussed in Section 2.2.

Local visual homing approaches [22–33] that robots use to travel between consecutive waypoints can be classified into two types: whole image-based approaches and correspondence-based approaches. A whole image-based approach derives motion vector to the targeted waypoint from the sum square image distance or Fast Fourier Transform[21] between the robot's view and the closest reference image.[32] On the other hand, a correspondence-based approach first finds correspondences between images, then computes the motion vector from the correspondences. Some common features are SIFT (Scale-Invariant Feature Transform),[34] SURF (Speeded Up Robust Feature),[35] corners, and lines. Different computations for the motion vector in correspondence-based approaches are discussed in Section 2.1. Both whole image-based approaches and correspondence-based approaches aim to reduce the difference between the robot's current pose (X_C, Z_C, θ_C) and the pose of the targeted waypoint (X_W, Z_W, θ_W). In this paper, the difference $X_W - X_C$, $Z_W - Z_C$, and $\theta_W - \theta_C$ are called lateral difference, forward difference, and heading difference, respectively.

1.1. Summary of the proposed approach

This paper proposes an image sequence-based navigation method under the teaching-replay framework for robots in piecewise linear routes. The robot rotates at the departing waypoint and moves directly toward the next waypoint in the replay phase. Since the robot arrives at the waypoint in a different heading to the reference image and heading difference leads to a large horizontal displacement of correspondences, the proposed approach detects the arrival at a waypoint by minimizing vertical displacement of correspondences. While the homing speed of robots in prior work[36] is constrained by its speed in the teaching phase, the robot in the proposed approach is less bounded by such speed limit and can travel faster without compromising the homing accuracy.

1.2. Paper organization

The rest of this paper is organized as follows: Section 2 introduces related work on correspondence-based local visual homing approaches and different detection of the arrival at a waypoint; Section 3 explains the construction of reference image sequence in a piecewise linear route and the proposed approach of multi-waypoint visual homing; Section 4 demonstrates the feasibility of the proposed approach in simulations and experiments. Besides, prior works on local visual homing and image sequence-based navigation are compared to the proposed approach; Section 5 concludes this paper.

* Corresponding author. E-mail: trhsiang@mail.ntust.edu.tw

2. Related Work

This section briefly introduces prior work on image sequence-based navigation in two aspects: how a robot navigates with local visual homing approach and how the arrival at a waypoint is detected.

2.1. Motion in correspondence-based local visual homing

Most published work decides the motion vector in local visual homing from multiple-view geometry[6,9–11,13–15,17,22,23,26–31,33], relative bearing angle from features to the robot,[5,24] 3D positions of features,[12,18,19] or the displacement of correspondences on image plane.[4,7,16] Cherubini *et al.*[8] compare the navigation performance between some of the above approaches in multi-waypoint routes.

1. Multiple-view geometry: Local visual homing approach based on multiple-view geometry[37] decides robot motion from essential matrix, epipole,[11] homography[17,27] or trifocal tensor.[23,30] The motion vector can be derived from the translation matrix and rotation matrix which are decomposed from essential matrix.[6,13–15] A curved trajectory to the waypoint is designed by moving epipoles gradually[22,26,31] to desired locations. The heading difference can be corrected from elements in homography.[30,33] Courbon *et al.* derive the angular velocity for navigation based on lateral difference and heading difference, which can be obtained from either homography[9] or epipolar geometry.[10] Lopez-Nicolas *et al.*[28] compute motion vector from epipolar geometry or homography under different circumstances in order to avoid the degeneracy of epipolar geometry due to short baseline or a planar scene.
2. Bearing vectors from features to the robot:[5,24] For every pair of features, the average bearing direction to the robot is considered. The robot's new heading is the weighted average over these directions, where the weight of each bearing direction is defined as the difference between the relative bearing angles to the waypoint and to the robot. This approach reduces the computational costs by requiring few feature correspondences and eliminating the use of image Jacobian matrix and multiple view geometry. However, the robot is only guaranteed to reach a waypoint located in the convex hull of features, therefore an omnidirectional camera is required to ensure a beneficial distribution of features.
3. 3D reconstruction of features: Local visual homing based on 3D reconstruction estimates the 3D positions of features globally or locally. Segvic *et al.*[19] compute the 3D positions of features between consecutive waypoints. The occluded or temporarily invisible features can be retracked by projecting features. Royer *et al.*[18] reconstruct the taught route and features globally before navigation by bundle adjustment.[37] Therefore, a position-based visual servo control is designed by comparing the taught trajectory and the robot's current position. Fontanelli *et al.*[12] localize the robot and estimate the 3D positions of features between waypoints.
4. Displacement of correspondences: The horizontal displacement of correspondences is used to decide the angular velocity or the direction of motion vector. Ido *et al.*[16] determine the direction of motion vector from horizontal displacement of a matched template. In the early work of Chen and Birchfield,[36] every correspondence votes for the direction of motion vector according to its horizontal displacement. This approach is later improved[7] by combining the average horizontal displacement of correspondences and the desired robot heading from odometry recorded in the teaching phase.

2.2. Detection of the arrival at a waypoint

Most research detects the arrival at a waypoint from image similarity,[6,13] displacement of correspondences,[9,15,17,20] traveled distance between waypoints,[20] or the magnitude of the motion vector.[20] Ido *et al.*[16] record the image similarity when the robot is brought to the last waypoint in the teaching phase. Thus, a precise navigation to the last waypoint is achievable by comparing the current image similarity and the recorded similarity. Argyros *et al.*[5] detect the arrival at a waypoint when a zero motion vector is found. Chen and Birchfield[7] define a criterion based on the displacement of correspondences, traveled distance, and the robot's heading to detect the arrival of a waypoint.

Depending on a robot's capability and its desired trajectory, there are various realizations of detecting the arrival using image correspondences. Some works[10,26,31] design a trajectory to minimize lateral deviation and heading deviation first, where the arrival at a waypoint is detected by checking if the longest or average displacement of correspondence is smaller than a threshold. The camera in Erinc's work[11] is able to rotate independently to the robot. In this case, the robot's heading can be corrected first and the arrival at a waypoint is detected only by examining the horizontal displacement of correspondence.

3. Construction of Reference Images and Visual Homing in Indoor Environments

This section first describes the construction of reference images, then explains the local visual homing between consecutive waypoints, especially on how to detect the arrival at a waypoint. An overview of the proposed multi-waypoint visual homing is provided at the end of this section.

3.1. The teaching phase: construction of reference images

Most works on image sequence-based navigation construct the reference images by comparing image similarity between the robot's view and the reference images.[11] The image similarity is defined as the number of tracked features[5,7,9] or matched features.[6,10,13,15,16] When the similarity is low, the robot's view is taken as a new reference image. On the other hand, when a robot navigates in a hybrid (metric/topological) map, Fontanelli *et al.*[12] decide a new waypoint and take a reference image when the uncertainty of 3D positions for features is small; Vardy[20] takes a reference image by comparing the difference between visual odometry and odometry; Zhang and Kleeman[21] sample reference images for fixed distance or angle.

The navigation approach in this paper teaches the robot piecewise linear routes. Waypoints used by the robot contain

https://doi.org/10.1017/S0263574712000434 Published online by Cambridge University Press

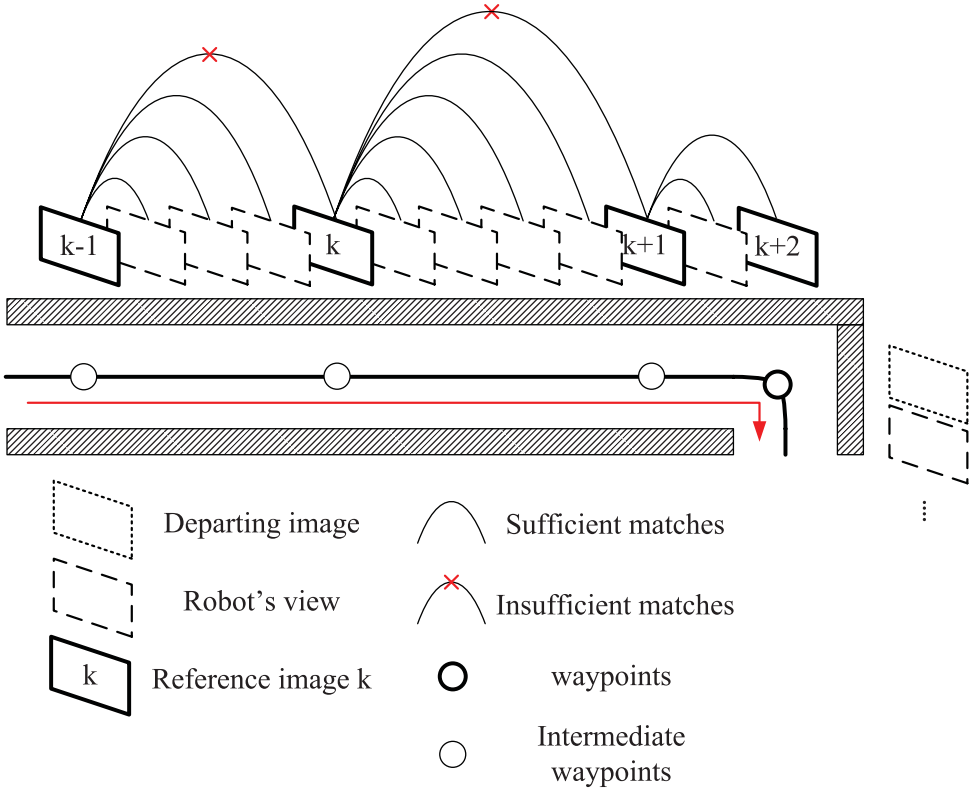

Fig. 1. (Colour online) The procedure for construction of reference images.

either the positions with large heading changes or selected midway positions between junctions. In the teaching phase, reference images are captured when the robot moves toward and arrives at a waypoint while no reference image is taken during robot's rotation on a waypoint. When the robot moves, correspondences between the robot's current image and the last reference image are found to decide when a new reference image should be taken. The pose where the reference image is captured due to few correspondences is defined as an intermediate waypoint. The reference image captured when the robot arrives at a waypoint is a clue for the robot to stop at a waypoint for a potential rotation. After the rotation, a departing image is captured as a temporarily reference image which is ready to be matched to the robot's image. An example for the construction of reference images is illustrated in Fig. 1.

In this paper, correspondences are found by matching SIFT features[34] between two images. Incorrect correspondences are removed with RANSAC (RANdom SAmple Consensus)[38] by estimating epipolar geometry.[37]

3.2. The replay phase: local visual homing between consecutive waypoints

The proposed local visual homing approach reduces forward difference, lateral difference, and heading difference between the robot's pose and a targeted waypoint. In the beginning, the robot rotates until a similar scene to the targeted reference image is found. With the matches between the robot's view and the reference image, the robot moves in a motion vector computed from epipolar geometry to the waypoint in the second step. After the waypoint is reached, the robot rotates to correct heading as the last step. The last step is performed only when the robot arrives at the last waypoint. In the case of arriving at intermediate waypoints, the last step is skipped and local visual homing repeats by targeting the next reference image.

The proposed local visual homing method is designed for a nonholonomic robot with a forward-looking camera to navigate in a flat environment, such as most indoor environments. The camera is installed so its optical axis is parallel to the robot heading. This specific installation simplifies calibration between the camera position and the robot center.

The three steps in the local visual homing are described as follows:

1. Rotation to find similar scene: A similar scene is found when the number of matches between the current view and the reference image is more than a threshold. The robot rotates every fixed degree in order to take new image to match the reference image. Since the forward-looking camera has limited Field Of View(FOV), the fixed degree is selected to be smaller than $FOV_{\max}$.
2. Navigation to waypoint: The robot approaches the waypoint straight by minimizing the forward difference and lateral difference between the robot position and the waypoint.
   (a) Motion vector: The direction of the motion vector is the relative direction from the robot position to the waypoint. The relative direction is computed from epipolar geometry between the robot current view and the reference image.[6,13–15,37] Given correspondences, such as matched SIFT features, as $(u_k, v_k) \leftrightarrow (u'_k, v'_k)$, $k = 1, 2, \cdots$, fundamental matrix

https://doi.org/10.1017/S0263574712000434 Published online by Cambridge University Press

F representing the epipolar geometry between two images is calculated in

$$\begin{pmatrix} u'_k \\ v'_k \\ 1 \end{pmatrix}^T F_{3\times 3} \begin{pmatrix} u_k \\ v_k \\ 1 \end{pmatrix} = 0, k = 1, 2, \cdots \quad (1)$$

If the camera is calibrated, the essential matrix E is derived from F and calibration matrix K as

$$E = K^T F K. \quad (2)$$

By decomposing E, relative translation and orientation between two positions where two images are taken are obtained as

$$E = R[t]_x = R \begin{pmatrix} 0 & -T_Z & T_Y \\ T_Z & 0 & -T_X \\ -T_Y & T_X & 0 \end{pmatrix}, \quad (3)$$

where R is the rotation matrix for relative orientation and $[t]_x$ is the skew symmetric matrix of the translation matrix. The relative direction φ from the robot position to the waypoint is computed in

$$\varphi = \arctan\left(\frac{T_X}{T_Z}\right) \quad (4)$$

as the direction of motion vector. The estimated relative direction becomes more accurate when the robot gradually approaches the waypoint, because the estimated epipolar geometry is more precise with more matches.

(b) Detection of the arrival at a waypoint: The detection of the arrival at a waypoint is used to stop the robot on the waypoint without 3D reconstruction during local visual homing, since the relative distance from the robot to the waypoint is unknown. Detection of the arrival at a waypoint based on image similarity, displacement of correspondences, traveled distance, or magnitude of motion vector is adopted in prior works, as discussed in Section 2.2. However, detection of the arrival at a waypoint with image similarity may not be metrically precise; heading difference must be minimized first when the displacement of correspondences is used to detect the arrival at a waypoint since a large heading difference leads to large horizontal displacement of correspondences; the methods based on traveled distance may be infeasible when the total traveled distance is so long that the accumulated error is large; the methods based on zero motion vector are feasible only with holonomic robots and panoramic view.

The proposed method detects the arrival at a waypoint when a small enough local minimum of the vertical displacement of corresponding features is found instead of the displacement of corresponding features, since the heading difference in the proposed local visual homing is not corrected first. Different heading induces a larger horizontal displacement of corresponding features than the vertical displacement of corresponding features when the robot is on the waypoint. In the following, this property is first derived, then the feasibility on detection of the arrival at a waypoint by checking vertical displacement of correspondences is discussed.

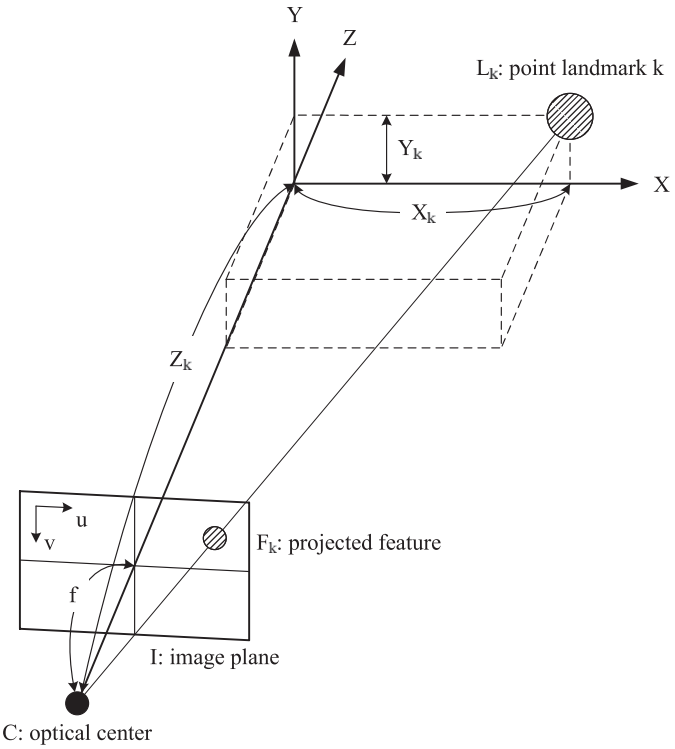

Fig. 2. Camera perspective model.

Given a landmark L_k and its relative position (X_k, Y_k, Z_k) to the camera optical center C, as shown in Fig. 2, a simplified camera perspective model is used to calculate its projected feature F_k on the image plane I as

$$F_k = \begin{pmatrix} u_k \\ v_k \end{pmatrix} = \begin{pmatrix} \frac{fX_k}{Z_k} \\ \frac{fY_k}{Z_k} \end{pmatrix}, \quad (5)$$

where f is the focal length. An azimuth angle α and an elevation angle ϵ from C to L_k are computed in

$$\begin{cases} \alpha = \arctan\frac{X_k}{Z_k} \\ \epsilon = \arctan\frac{Y_k}{Z_k}. \end{cases} \quad (6)$$

After the camera rotates θ along Y-axis, the corresponding feature F'_k is calculated in

$$F'_k = \begin{pmatrix} u'_k \\ v'_k \end{pmatrix} = \begin{pmatrix} \frac{f(X_k \cos\theta - Z_k \sin\theta)}{X_k \sin\theta + Z_k \cos\theta} \\ \frac{fY_k}{X_k \sin\theta + Z_k \cos\theta} \end{pmatrix}. \quad (7)$$

The horizontal and vertical displacements of corresponding features caused by different heading

https://doi.org/10.1017/S0263574712000434 Published online by Cambridge University Press

are written as

$$\begin{aligned}|\Delta u| &= |u_k - u'_k| \\ &= f\left|\frac{X_k}{Z_k} - \frac{X_k\cos\theta - Z_k\sin\theta}{X_k\sin\theta + Z_k\cos\theta}\right| \\ &= f|X_k\sin\theta + Z_k\cos\theta||X_k^2\sin\theta + Z_k^2\sin\theta|\end{aligned} \tag{8}$$

and

$$\begin{aligned}|\Delta v| &= |v_k - v'_k| \\ &= f\left|\frac{Y_k}{Z_k} - \frac{Y_k}{X_k\sin\theta + Z_k\cos\theta}\right| \\ &= f|X_k\sin\theta + Z_k\cos\theta||X_kY_k\sin\theta \\ &\quad + Y_kZ_k\cos\theta - Y_kZ_k|.\end{aligned} \tag{9}$$

By replacing X_k and Y_k with Eqs. (6), (8) and (9) become

$$\begin{aligned}|\Delta u| &= fZ_k^2|X_k\sin\theta + Z_k\cos\theta||\tan^2\alpha\sin\theta + \sin\theta| \\ &= fZ_k^2|X_k\sin\theta + Z_k\cos\theta||\sin\theta||\tan^2\alpha + 1|\end{aligned} \tag{10}$$

and

$$\begin{aligned}|\Delta v| &= fZ_k^2|X_k\sin\theta + Z_k\cos\theta||\tan\epsilon\sin\theta\tan\alpha \\ &\quad + \tan\epsilon\cos\theta - \tan\epsilon| \\ &= fZ_k^2|X_k\sin\theta + Z_k\cos\theta||\tan\epsilon\sin\theta||\tan\alpha \\ &\quad + \cot\theta - \csc\theta|.\end{aligned} \tag{11}$$

The camera used in this paper is calibrated so that $|\alpha_{\max}| = 31.1^\circ$ and $|\epsilon_{\max}| = 23.2^\circ$. In order for two views to overlap, $|\theta_{\max}| = 2|\alpha_{\max}|$. We note that for general off-the-shelf cameras $|\alpha|$ and $|\epsilon|$ are ranged between $\pm 27.3^\circ$ and $\pm 18.9^\circ$ respectively and the focal length is 35mm.

In Eq. (10), $|\tan^2\alpha + 1| \geq 1$. On the other hand, in Eq. (11), $|(\tan\alpha + \cot\theta - csc\theta)| < 1.5$ and $|\tan\epsilon| < 0.5$ for general parameters of front-facing cameras given above, and $|\tan\epsilon||(\tan\alpha + \cot\theta - \csc\theta)| < 0.75 < 1$. Therefore, $|\Delta u| \geq fZ_k^2|X_k\sin\theta + Z_k\cos\theta||\sin\theta| > |\Delta v|$, when $\theta \neq 0$. If $\theta = 0$, both $|\Delta u|$ and $|\Delta v|$ are 0.

Since the vertical displacement of corresponding feature is less affected by the different heading direction than the horizontal displacement of corresponding feature when two images are captured on a same location but in different direction, the proposed method uses the vertical displacement Δv as a signal to detect the arrival at a waypoint.

The positions where the robot can find zero Δv are examined. Given a waypoint W as origin, the top view of Fig. 2 is shown in Fig. 3. When $\Delta v_k = 0$, the robot position must fit

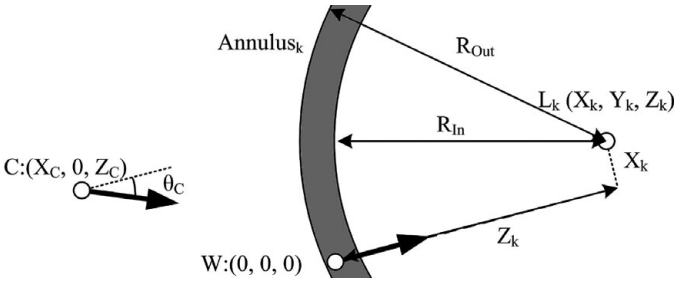

Fig. 3. The positions where Δv_k is 0.

$$(X_k - X_C)\times\sin\theta_C + (Z_k - Z_C)\times\cos\theta_C = Z_k. \tag{12}$$

By considering the limited FOV of a forward-looking camera, the positions where zero Δv_k is found in

$$\begin{aligned}&\left|\arccos\frac{(X_k - X_C)\times\sin\theta_C + (Z_k - Z_C)\times\cos\theta_C}{\sqrt{(X_k - X_C)^2 + (Z_k - Z_C)^2}}\right| \\ &<= \frac{FOV_{\max}}{2}.\end{aligned} \tag{13}$$

By solving Eqs. (12) and (13), all positions where zero Δv_k form an annulus $Annulus_k$, as the gray region shown in Fig. 3. The inner radius R_{In} of the annulus is found when the robot is facing to L_k directly. On the other hand, the outer radius R_{Out} is obtained when the robot bearing angle to L_k is $\frac{FOV_{\max}}{2}$. R_{In} and R_{Out} are computed in

$$\begin{cases} R_{In} = Z_k \\ R_{Out} = \frac{Z_k}{\cos\frac{FOV_{\max}}{2}}. \end{cases} \tag{14}$$

In general, multiple landmarks are observed within FOV and multiple annuli are constructed, as Fig. 4. All annuli intersect at a region around the waypoint, as the black small region shown in Fig. 4. Therefore, the robot can detect the arrival at a waypoint by checking if the average Δv becomes zero during the navigation. In the proposed method, due to possible incorrect matches and noise, the robot is aware of the arrival at a waypoint by finding a small enough local minimum of the average $|\Delta v|$.

(c) Moving velocity: The moving velocity is proportional to the average $|\Delta v|$. In this design, the robot tends to move quickly when departing from previous waypoint and gradually slow down when approaching the targeted waypoint.

3. Rotation to correct heading: The heading difference is corrected by rotating the robot after the robot stops on the waypoint. The robot rotates in a fixed angular velocity to minimize the average $|u_k - u'_k|$.

https://doi.org/10.1017/S0263574712000434 Published online by Cambridge University Press

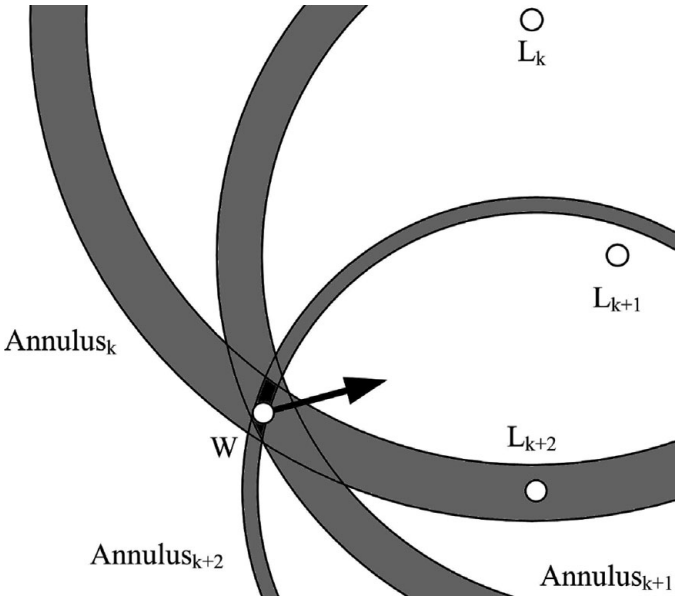

Fig. 4. Multiple annuli.

3.3. *Summary of multi-waypoint visual homing*

The multi-waypoint visual homing approach is summarized as a state machine in Fig. 5. The robot starts to rotate in state 1 until the robot observes a similar scene to the targeted reference image. In state 2, the robot computes the motion vector and moves accordingly. The motion stops when a small enough local minimum of the average $|\Delta v|$ is found since the robot has arrived at the waypoint. When the robot arrives at the final destination, it aligns its heading to the direction where the last reference image is taken. Otherwise, the robot changes its targeted reference image to the next and transits to state 1.

4. Performance Evaluations

The feasibility and performance of the proposed multi-waypoint visual homing are evaluated in this section. Section 4.1 provides the variation of $\overline{|v_k - v'_k|}$ in a simulation in order to demonstrate the effectiveness of the detection of the arrival at a waypoint; Section 4.2 describes the used robot and platform for experiments; Section 4.3 demonstrates experiments of the proposed multi-waypoint visual homing in long routes of multiple waypoints; Section 4.4 compares the navigation accuracy of the proposed approach of local visual homing to other prior work[26] which computes motion vector from epipolar geometry and detects the arrival at a waypoint based on the average displacement of correspondence $\overline{\sqrt{(u_k - u'_k)^2 + (v_k - v'_k)^2}}$; Section 4.5 compares the performance of proposed multi-waypoint visual homing in a piecewise linear route and the performance of other image sequence-based navigation method[36] in a smooth trajectory.

4.1. *Variation of vertical displacement of correspondences*

A simulation shows $\overline{|v_k - v'_k|}$ when the robot is in a neighbor region of a waypoint in order to demonstrate the effectiveness of the detection of the arrival at a waypoint. Given a waypoint

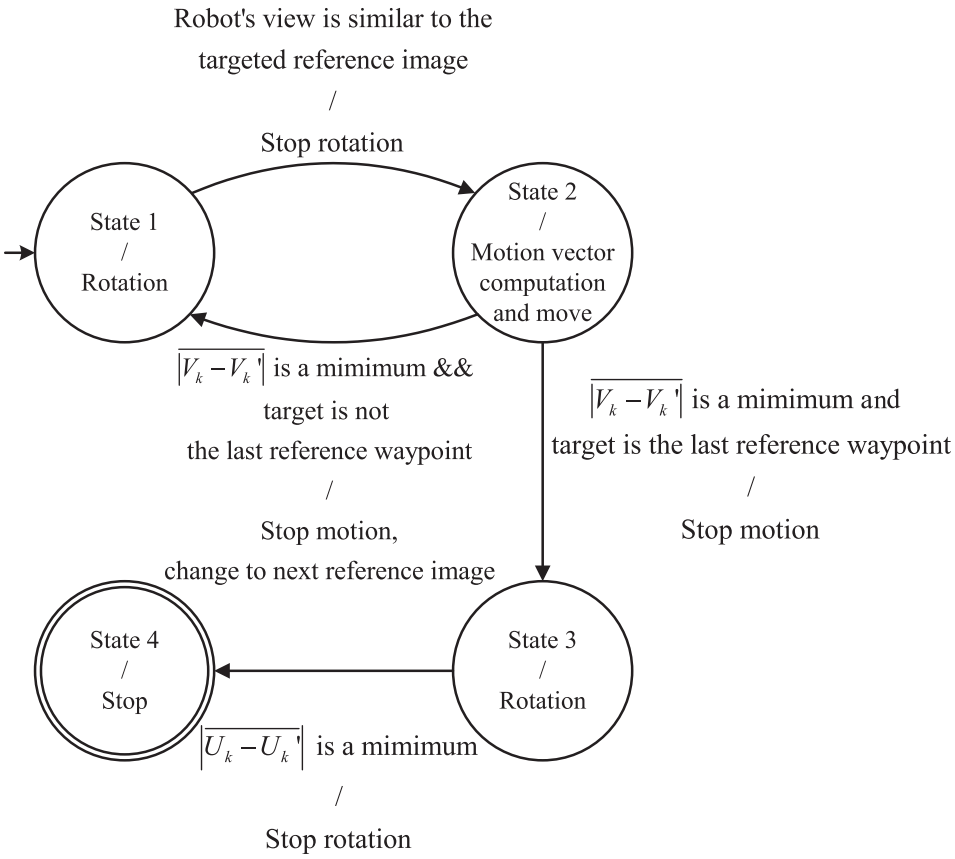

Fig. 5. State machine of multi-waypoint visual homing.

https://doi.org/10.1017/S0263574712000434 Published online by Cambridge University Press

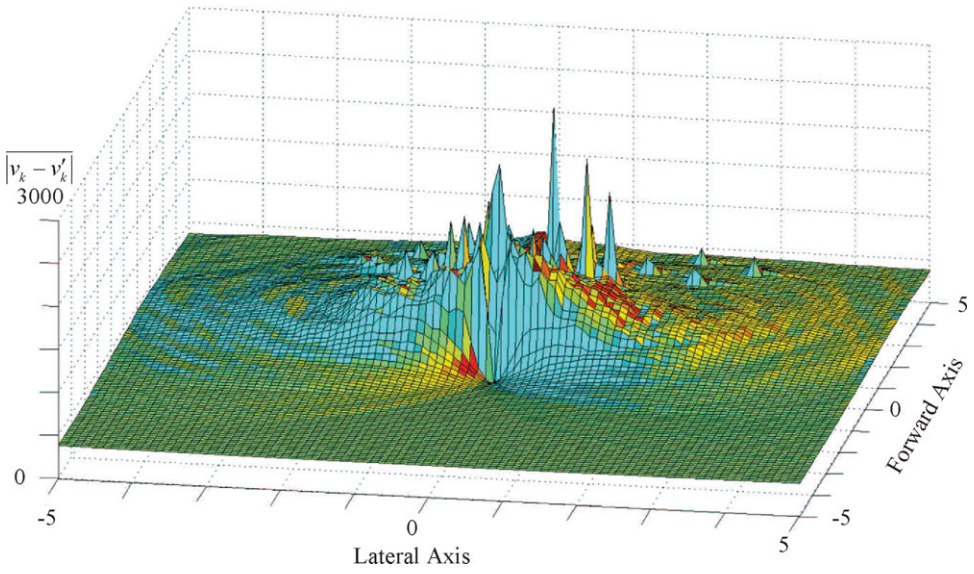

Fig. 6. (Colour online) The simulation of $\overline{|v_k - v'_k|}$ around waypoint (0, 0).

and landmarks, the simulation calculates $\overline{|v_k - v'_k|}$ at each position in the neighbor region around the waypoint. The landmarks are generated randomly inside the FOV when the robot is on the waypoint in order to produce observable landmarks on the reference image. Figure 6 shows the variation of $\overline{|v_k - v'_k|}$. The size of the neighbor region around the waypoint is 100 m^2 and $\overline{|v_k - v'_k|}$ is calculated every 10 cm. The heading of the robot is the direction from the robot to the waypoint. The waypoint is set as origin and the heading direction is 0°, which is in the same direction to the positive forward axis. Figures 7(a) and (b) are the observations of Fig. 6 from the lateral and forward axes, respectively. The simulation shows that $\overline{|v_k - v'_k|}$ reduces when the robot approaches the waypoint and drastically increases after the robot passes the waypoint.

4.2. Robot and platform for experiments

The robot for experiments is called U-BOT[39] which is designed by Industrial Technology Research Institute(ITRI).[40] U-BOT is a four-wheeled robot with two-wheeled drive in the front. The dimension of U-BOT is $596 \times 495 \times 373(mm^3)$ and the maximal payload is 23 kg. The maximal moving velocity and acceleration of U-BOT is 1200 mm/sec and 1000 mm/sec^2, respectively. The commands of moving velocity and the direction of motion vector are transmitted through serial communication to U-BOT, and a Digital Signal Processor(DSP) inside U-BOT controls U-BOT accordingly. A usb webcam, QuickCam Ultra Vision[41] from Logitech, is mounted on the top of U-BOT as the forward-looking camera. A laptop with Intel Core 2 Duo P7450 (2.13 GHz) is used to perform the navigation approach. The graphics card of the laptop is NVIDIA GeForce G 105M so that the SIFT feature matching can be processed by GPU. Figure 8 shows U-BOT and the forward-looking camera. In the teaching phase, the forward-looking camera is used to take reference images on waypoints. In the replay phase, the robot motion for navigation is decided by the proposed approach of local visual homing with the forward-looking camera. The sonars, infrared distance sensors, the laser range finder from Hokuyo, and the side-looking camera are not used.

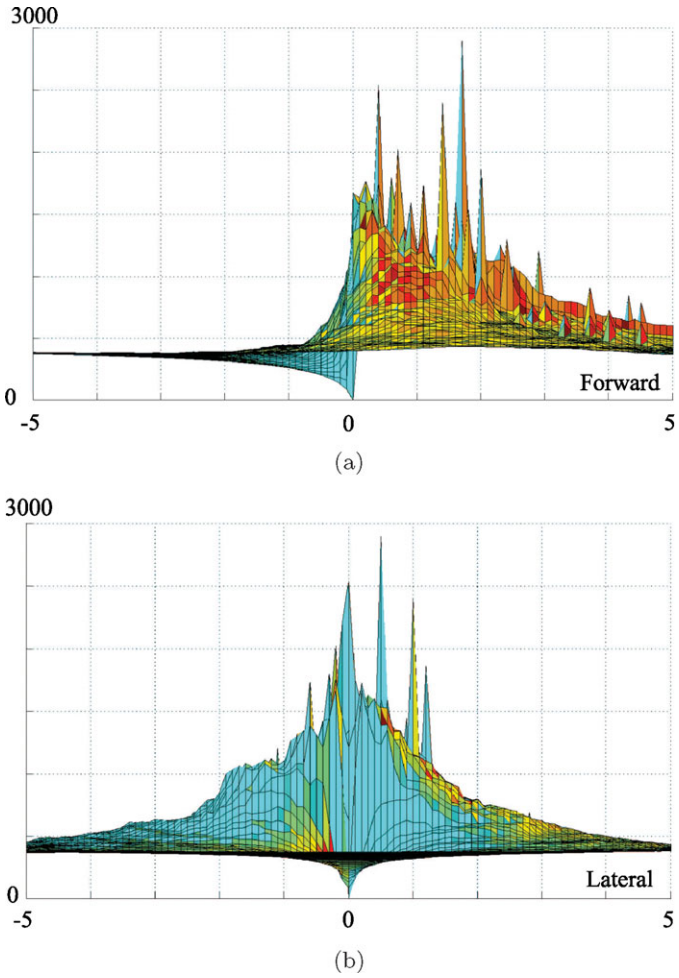

Fig. 7. (Colour online) Observations of Fig. 6. (a) Observations from lateral axis; (b) Observations from forward axis.

https://doi.org/10.1017/S0263574712000434 Published online by Cambridge University Press

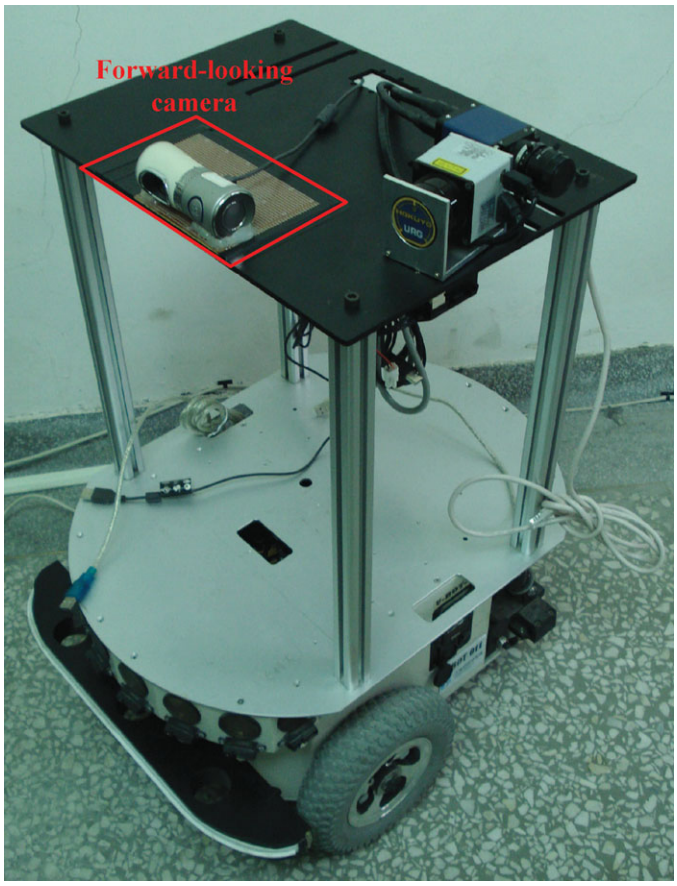

Fig. 8. (Colour online) U-BOT and the mounted forward-looking camera.

4.3. Experiments of multi-waypoint visual homing

Three experiments are conducted in different environments in order to evaluate the feasibility of the proposed multi-waypoint visual homing. The first environment is a rectangular route in a hall. The other two environments are l-shaped hallways on different floors of a building. Figure 9 shows one of the reference images in each environment. The length, the shape, and the average runtime for navigation in each route are summarized in Table I. The navigation error is smaller than 17 cm for all waypoints. For all experiments, a same set of parameters is adopted in the proposed local visual homing. In the first step, due to the similarity in the third test environment, the robot view is considered as a similar scene to a reference image when at least 20 matched SIFT features are found, while 8 matched SIFT features are sufficient to calculate epipolar geometry in the second step. In the second step, the robot detects the arrival at a waypoint when the average magnitude of vertical displacement of correspondences $\overline{|v_k - v'_k|}$ is a local minimum and smaller than 4. In the last step, the robot stops heading correction when the magnitude of average horizontal displacement of correspondences $\overline{|u_k - u'_k|}$ is a local minimum and smaller than 10.

Table I. The length, shape, and runtime in each test environment.

Environment	Shape	Length (m)	Average runtime (sec)
1	Rectangle	25	217
2	L shape	28	204
3	L shape	35	245

The experiment in the third environment is discussed in detail, because of the similarity and dynamic objects in the environment. Figure 10 provides the positions of waypoints in the third environment and 18 required reference images. The length of the route from S to 18th waypoint is about 35 meters. Figure 11 shows the views when the robot arrives at all waypoints in one trial.

In each local visual homing, the proposed approach tends to increase the number of matched SIFT features and to decrease the average magnitude of vertical displacement for correspondences $\overline{|v_k - v'_k|}$. Figure 12 shows the number of matched SIFT features and the average magnitude of vertical displacement for correspondences $\overline{|v_k - v'_k|}$. The robot takes image and does local visual homing at each timestep. The scales on horizontal axis are the timesteps at which the robot arrives at waypoints. After the robot arrives at each waypoint and changes its target to the next waypoint, the number of matched SIFT features decreases and $\overline{|v_k - v'_k|}$ increases steeply.

When the robot moves, people move in the hallway and several doors are opened. These environmental changes or occlusions may reduce the number of matched SIFT features

(a) (b) (c)

Fig. 9. (Colour online) One of reference images in the test environments. (a) Reference image in test environment 1 (hall); (b) Reference image in test environment 2 (hallway); (c) Reference image in test environment 3 (hallway with similarity).

https://doi.org/10.1017/S0263574712000434 Published online by Cambridge University Press

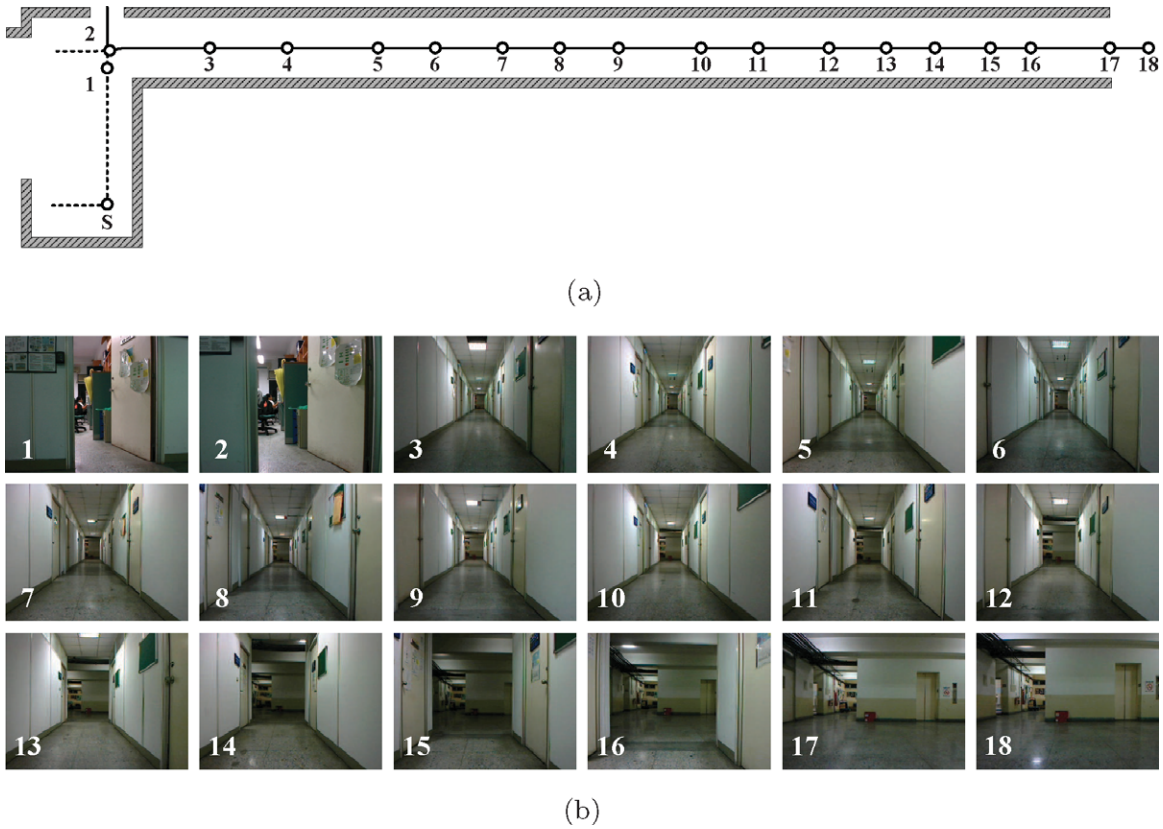

Fig. 10. (Colour online) Waypoints in the third environment where similarity and moving objects appear. (a) Route and waypoints (circles) in the test environment; (b) Reference images on waypoints.

during navigation but does not affect local visual homing, as long as the number of matched SIFT features is sufficient to calculate the motion vector. Figure 13(a) shows the robot views in which a person appears when the robot moves from 7th waypoint to 8th waypoint. Figure 13(b) shows an open door as an environmental change when the robot moves from 14th waypoint to 15th waypoint.

4.4. Navigation accuracy between consecutive waypoints

Since the navigation error does not accumulate in the image sequence-based approach, the navigation accuracy in a multi-waypoint route depends on the navigation accuracy between consecutive waypoints. In order to evaluate the navigation accuracy, the proposed approach is compared to prior work[26] which designs local visual homing based on epipolar geometry and kinematics of the robot. In the rest of this section, epipole-based visual homing(EVH) is used to represent this prior work.[26]

Epipole-based visual homing (EVH) designs a control law for local visual homing by linearizing the visual servoing system with input-output linearization, while the input is the robot's motion command and the outputs are the locations of epipoles. The robot's trajectory to the waypoint depends on a specific reference trajectories defined for both epipoles in the robot's image and the reference image. By following the trajectory, the robot minimizes the lateral difference and

Fig. 11. (Colour online) The robot views when the robot arrives at each waypoint in the third environment.

https://doi.org/10.1017/S0263574712000434 Published online by Cambridge University Press

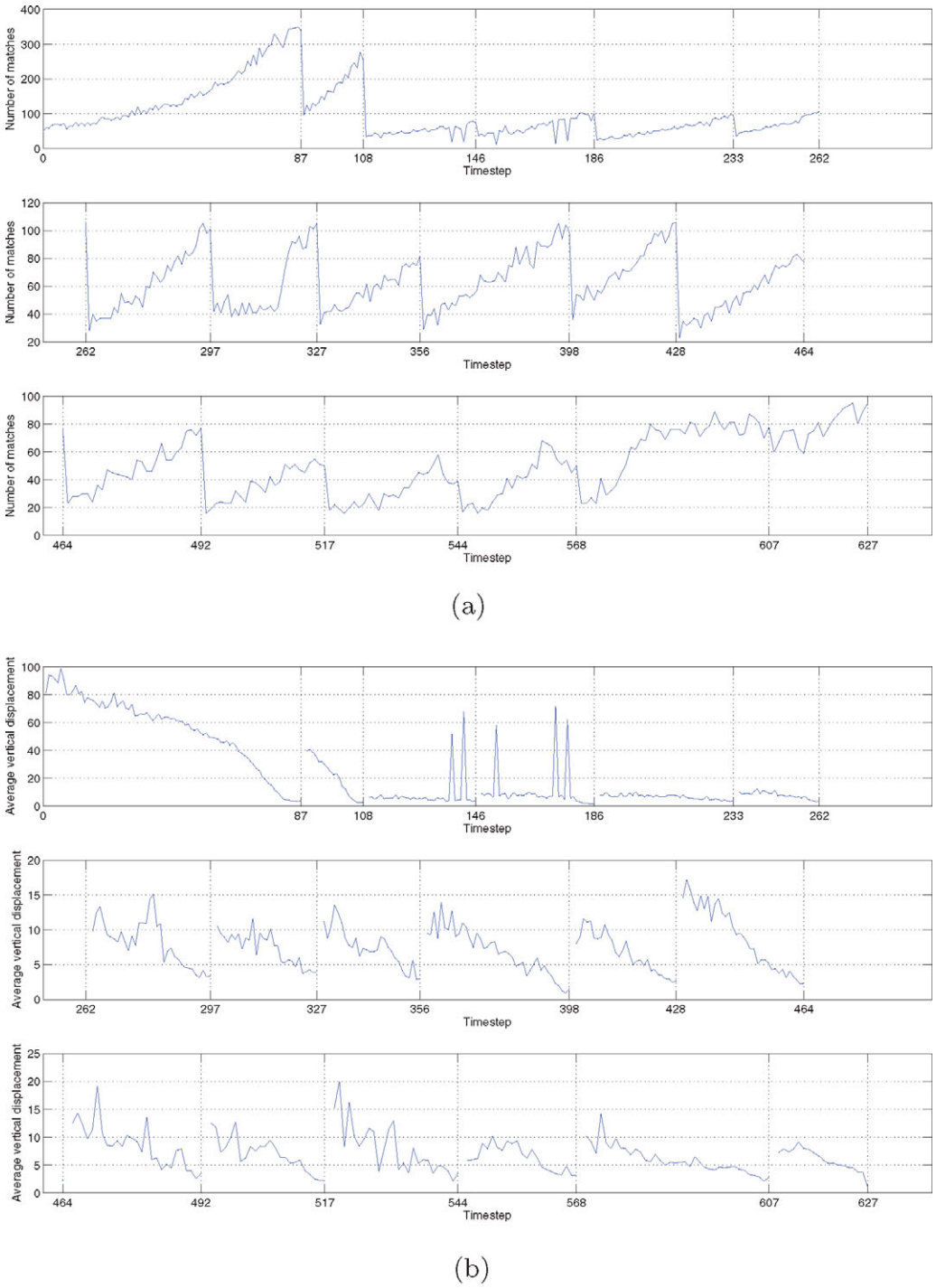

Fig. 12. (Colour online) Number of matched SIFT features and $\overline{|v_k - v'_k|}$ during navigation. (a) Number of matched SIFT features; (b) Variation of $\overline{|v_k - v'_k|}$.

Fig. 13. (Colour online) Environmental changes during navigation. (a) Moving person and opened door; (b) Opened door.

https://doi.org/10.1017/S0263574712000434 Published online by Cambridge University Press

Table II. Navigation accuracy of local visual homing.

Applied approach	Average lateral error	Average forward error	Detection the arrival at a waypoint
Proposed approach	5.76 cm	7.36 cm	average $\|\Delta v\|$
EVH	1.63 cm	5.96 cm	average $\sqrt{\Delta u^2 + \Delta v^2}$
Proposed trajectory	failed	failed	average $\sqrt{\Delta u^2 + \Delta v^2}$

heading difference, then reduces the forward difference. The arrival at the waypoint is detected by finding a minimal average displacement of correspondences.

To compare the navigation accuracy, 30 tests of the proposed approach and EVH are done under different robot's initial location and location distribution of features in indoor environments. In each test, objects in the scene change their location and the robot is randomly placed to start local visual homing. Between the robot's initial location and the waypoint, the lateral difference is within 50 cm and the forward difference is between 100 cm and 200 cm. The features are at least 150 cm away from the waypoint. The same set of parameters and control gains is used during these 30 tests.

The average navigation errors of the proposed approach and EVH are shown in the first two rows of Table II. On average, the navigation of EVH is more accurate, especially in the lateral part. The forward motion at the last step in EVH contributes to the smaller lateral error. However, the design of the trajectory for the robot requires more parameters or control gains such as the coefficients for the reference trajectories of epipoles.

The last row in Table II shows the necessity of observing the average magnitude of vertical displacement of correspondences to detect the arrival at a waypoint in the proposed approach. If the robot moving along a piecewise linear trajectory detects the arrival at each waypoint by observing the displacements of correspondences, it is unable to stop at the waypoint by minimizing the magnitude of the displacement of correspondence unless no lateral difference is found between the robot's initial location and the waypoint, because the displacements of correspondences often do not converge steadily when the robot approaches the waypoint.

4.5. Navigation efficiency in long routes

The proposed multi-waypoint visual homing approach is compared to other work on image sequence-based navigation which decides the motion vector from horizontal displacement of correspondences.[4,7,16,36] As mentioned in qualitative visual navigating method,[36] the robot's velocity in the replay phase must be smaller by a factor of 1.5∼2.5 in order to avoid going off course due to the delay caused by computational and response times. This section provides experiments that confirm the speed limit of prior work and shows the proposed method is not bounded by such limit.

Qualitative visual navigation(QVN)[36] is implemented because its visual input configuration and computational requirement are similar to our settings. In the teaching phase, QVN takes reference images along a smooth route every 0.8 s in order to divide the route into segments. In each segment, up to 50 corners are tracked as features, to which correspondences are found in the replay phase. In the replay phase, the robot moves in a fixed velocity and the direction of motion vector is decided by a simple algorithm based on voting. Each correspondence between the robot's image and the closest reference image is compared in its locations on both images and votes for the direction: right, left, or unchanged. In the experiment, corners are detected and tracked by functions provided in OpenCV.[42] The velocity and the angular velocity in the teaching phase are 10 cm/s and 4°/s, respectively.

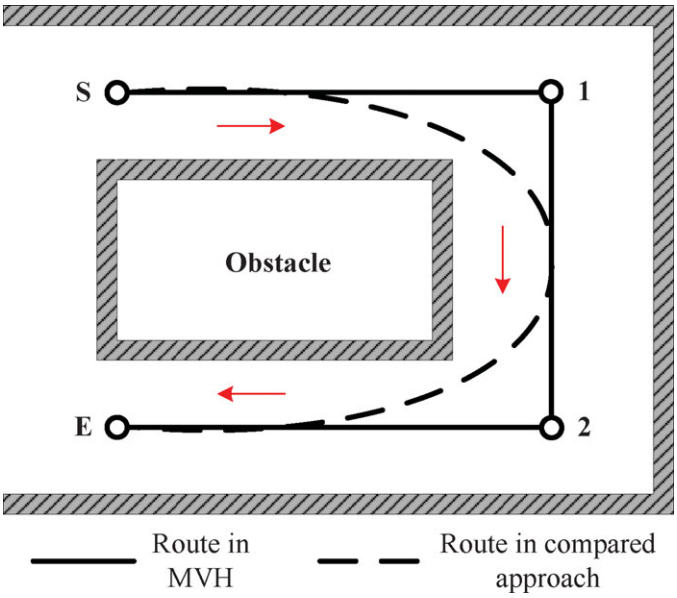

Fig. 14. (Colour online) Test environment for comparison between MVH and QVN.

The proposed approach of multi-waypoint Visual Homing (MVH) and QVN are tested in U-shaped routes in an office, as shown in Fig. 14. The endpoints of the routes, S and E, are two positions in the office. The length of the route of MVH and QVN are 6.4 m and 5.3 m, respectively. Both approaches are tested under different moving velocities: 4 cm/sec, 8 cm/sec, 10 cm/sec, 20 cm/sec, and 50 cm/sec. QVN adopts each moving velocity as its fixed moving velocity in the replay phase, while MVH adopts the moving velocity as a maximal moving velocity. If the calculated velocity in second step of MVH is higher than the maximal velocity, the robot moves in the maximal velocity. The same parameters in Section 4.3 are used by MVH in this comparison. Table III shows the results of experiments.

Our results show the claimed limit on velocity in QVN[36] and demonstrate that the moving velocity in MVH is not

https://doi.org/10.1017/S0263574712000434 Published online by Cambridge University Press

Table III. Comparison between the proposed approach of MVH and QVN.

Method	Velocity (cm/sec)	Lateral error (cm)	Forward error (cm)	runtime (sec)
MVH	4	$-1 \sim 1.5$	$-0.5 \sim 1.5$	$174 \sim 189$
MVH	8	$-2.5 \sim 1$	$1 \sim 3.5$	$108 \sim 121$
MVH	10	$-1 \sim 1.5$	$1 \sim 3$	$96 \sim 101$
MVH	20	$-3 \sim 2.5$	$-4.5 \sim 1.5$	$70 \sim 74$
MVH	50	$-1.5 \sim 2$	$2 \sim 4.5$	$61 \sim 64$
QVN	4	$-2.5 \sim 1$	$0.5 \sim 1$	$135 \sim 141$
QVN	8	$-4.5 \sim 3$	$2 \sim 4.5$	$68 \sim 71$
QVN	10	$-9.5 \sim -6.5$	$7 \sim 11$	$55 \sim 57$
QVN	20	failed	failed	failed
QVN	50	failed	failed	failed

constrained by such speed limit. When the velocity in the replay phase is below the velocity in the teaching phase, both MVH and QVN work well and QVN spends less time on navigation than MVH. Increasing the velocity in the replay phase in QVN causes deviation from the taught route and this deviation sometimes leads to a failed navigation, especially when the robot makes sharp turns in the curvilinear part of the route. On the other hand, because the accurate motion vector reduces the lateral difference and large change of heading is performed on waypoints, the deviation in MVH is not as large as the deviation in QVN at high speed.

Two main reasons in MVH to break the limit on velocity are the accurate motion vector and the variable velocity when the robot approaches a waypoint. The computed motion vector in MVH not only minimizes the displacement of correspondences but also points to the waypoint, however the motion vector in QVN is only to minimize the displacement of correspondences and does not point to the waypoint mostly. A visual navigation with less accurate motion vector requires higher frame rate in order to correct the robot's motion more often so that the moving velocity is limited. Besides, the robot with slower velocity when approaching a waypoint allows more corrections of the robot's motion and denser sampling of the displacement of correspondences in order to avoid a premature or late detection of the arrival at the waypoint.

5. Conclusion

This paper proposes an image sequence-based navigation method under the teaching-replay framework for robots in piecewise linear routes. Waypoints used by the robot contain either the positions with large heading changes or selected midway positions between junctions. Between consecutive waypoints, local visual homing is designed to rotate on the departing waypoint and to move directly toward the targeted waypoint. The arrival at a waypoint is determined by minimizing the average vertical displacements of feature correspondences. The performance the proposed approach is evaluated through extensive experiments in hallway and office environments. Robots employing the proposed approach can break the speed limit claimed by other studies and travel at higher speeds without compromising the navigation accuracy.

Robot navigation in highly dynamic environments is still a challenging work. We plan to refine the proposed framework to be adoptable in these environments. Different clues such as virtual reality[43] and 3D environment models may also be used for navigation. Also, fusing visual information with probabilities in pose[44,45] are under investigation.

Acknowledgment

This research was supported in part through grants NSC 99-2221-E-011-134 and NSC 100-2221-E-011-051 of the National Science Council of Taiwan.

References

1. F. Bonin-Font, A. Ortiz and G. Oliver, "Visual navigation for mobile robots: A survey," *J. Intell. Robot. Syst.* **53**(3), 263–296 (2008).
2. P. Corke, *Visual Control of Robots: High-Performance Visual Servoing* (Research Studies Press Ltd, 1996).
3. S. Hutchinson, G. Hager and P. Corke, "A tutorial on visual servo control," *IEEE Trans. Robot. Autom.* **12**, 651–670 (1996).
4. Y. Matsumoto, M. Inaba and H. Inoue, "Visual Navigation Using View-Sequence Route Representation," *Proceedings of the IEEE International Conference on Robotics and Automation*, Minneapolis, MN, USA (Apr. 22–28, 1996) vol. 1, pp. 83–88.
5. A. A. Argyros, K. E. Bekris, S. C. Orphanoudakis and L. E. Kavraki, "Robot homing by exploiting panoramic vision," *Auton. Robots* **19**(1), 7–25 (2005).
6. O. Booji, B. Terwijn, Z. Zivkovic and B. Krose, "Navigation Using an Appearance-Based Topological Map," *Proceedings of the IEEE International Conference on Robotics and Automation*, Rome, Italy (Apr. 10–14, 2007) pp. 3927–3932.
7. Z. Chen and S. T. Birchfield, "Qualitative vision-based path following," *IEEE Trans. Robot.* **25**(3), 749–754 (2009).
8. A. Cherubini, M. Colafrancesco, G. Oriolo, L. Freda and F. Chaumette, "Comparing Appearance-Based Controllers for Nonholonomic Navigation from a Visual Memory," *Proceedings of the ICRA 2009 Workshop on Safe Navigation in Open and Dynamic Environments: Application to Autonomous Vehicles*, Kobe, Japan (May 12, 2009).
9. J. Courbon, Y. Mezouar and P. Martinet, "Indoor navigation of a non-holonomic mobile robot using a visual memory," *Auton. Robots* **25**(3), 253–266 (2008).
10. J. Courbon, Y. Mezouar and P. Martinet, "Autonomous navigation of vehicles from a visual memory using a generic camera model," *IEEE Trans. Intell. Transp. Syst.* **10**(3), 392–402 (2009).
11. G. Erinc and S. Carpin, "Image-Based Mapping and Navigation with Heterogenous Robots," *Proceedings of the IEEE International Conference on Intelligent Robots and Systems*, St. Louis, MO, USA (Oct. 11–15, 2009) pp. 5807–5814.
12. D. Fontanelli, A. Danesi, F. A. W. Belo, P. Salaris and A. Bicchi, "Visual servoing in the large," *The Int. J. Robot. Res.* **28**(6), 802–814 (2009).
13. F. Fraundorfer, C. Engels and D. Nister, "Topological Mapping, Localization and Navigation using Image Collections," *Proceedings of the IEEE International Conference on Intelligent Robots and Systems*, San Diego, CA, USA (Oct. 29–Nov. 2, 2007) pp. 3872–3877.
14. Y. Fu, T.-R. Hsiang and S.-L. Chung, "Robot Navigation Using Image Sequences," *Proceedings of the 6th International Conference on Ubiquitous Robots and Ambient Intelligence*, Gwangju, Korea (Oct. 29–31, 2009) pp. 163–167.

https://doi.org/10.1017/S0263574712000434 Published online by Cambridge University Press

15. T. Goedeme, M. Nuttin, T. Tuytelaars and L. Van Gool, "Omnidirectional vision based topological navigation," *Int. J. Comput. Vis.* **74**(3), 219–236 (2007).
16. J. Ido, Y. Shimizu, Y. Matsumoto and T. Ogasawara, "Indoor navigation for a humanoid robot using a view sequence," *The Int. J. Robot. Res.* **28**(2), 315–325 (2009).
17. A. Remazeilles and F. Chaumette, "Image-based robot navigation from an image memory," *Robot. Auton. Syst.* **55**(4), 345–356 (2007).
18. E. Royer, M. Lhuillier, M. Dhome and J.-M. Lavest, "Monocular vision for mobile robot localization and autonomous navigation," *Int. J. Comput. Vis.* **74**(3), 237–260 (2007).
19. S. Segvic, A. Remazeilles, A. Diosi and F. Chaumette, "Large Scale Vision-Based Navigation Without an Accurate Global Reconstruction," *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition*, Minneapolis, MN, USA (Jun. 18–23, 2007) vol. 0, pp. 1–8.
20. A. Vardy, "Long-Range Visual Homing," *Proceedings of the IEEE International Conference on Robotics and Biomimetics*, Kunming, China (Dec. 17–20, 2006) pp. 220–226.
21. A. M. Zhang and L. Kleeman, "Robust appearance based visual route following for navigation in large-scale outdoor environments," *Int. J. Robot. Res.* **28**(3), 331–356 (2009).
22. H. M. Becerra and C. Sagues, "A Sliding Mode Control Law for Epipolar Visual Servoing of Differential-Drive Robots," *Proceedings of the IEEE International Conference on Intelligent Robots and Systems*, Nice, France (Sep. 22–26, 2008) pp. 3058–3063.
23. H. M. Becerra, G. Lopez-Nicolas and C. Sagues, "Omnidirectional visual control of mobile robots based on the 1d trifocal tensor," *Robot. Auton. Syst.* **58**(6), 796–808 (2010).
24. M. Liu, C. Pradalier, Q. Chen and R. Siegwart, "A Bearing-Only 2d/3d-Homing Method Under a Visual Servoing Framework," *Proceedings of the IEEE International Conference on Robotics and Automation*, Anchorage, AK, USA (May 3–7, 2010).
25. S. G. Loizou and V. Kumar, "Biologically Inspired Bearing-Only Navigation and Tracking," *Proceedings of the IEEE International Conference on Decision and Control*, London, UK (Apr. 15–17, 2007) pp. 1386–1391.
26. Lopez-Nicolas, C. Sagues, J. J. Guerrero, D. Kragic and P. Jensfelt, "Switching visual control based on epipoles for mobile robots," *Robot. Auton. Syst.* **56**(7), 592–603 (2008).
27. G. Lopez-Nicolas, N. R. Gans, S. Bhattacharya, C. Sagues, J. J. Guerrero, and S. Hutchinson, "Homography-based control scheme for mobile robots with nonholonomic and field-of-view constraints," *IEEE Trans. Syst. Man Cybern. Part B: Cybern.* **99**, 1–13 (2009).
28. G. Lopez-Nicolas, J. J. Guerrero and C. Sagues, "Visual control of vehicles using two-view geometry," *Mechatronics* **20**(2), 315–325 (2010).
29. G. Lopez-Nicolas, J. J. Guerrero and C. Sagues, "Visual control through the trifocal tensor for nonholonomic robots," *Robot. Auton. Syst.* **58**(2), 216–226 (2010).
30. G. Lopez-Nicolas, J. J. Guerrero and C. Sagues, "Multiple homographies with omnidirectional vision for robot homing," *Robot. Auton. Syst.* **58**(6), 773–783 (2010).
31. G. L. Mariottini, G. Oriolo and D. Prattichizzo, "Image-based visual servoing for nonholonomic mobile robots using epipolar geometry," *IEEE Trans. Robot.* **23**(1), 87–100 (2007).
32. R. Moller, "Local visual homing by warping of two-dimensional images," *Robot. Auton. Syst.* **57**(1), 87–101 (2009).
33. C. Sagues and J. J. Guerrero, "Visual correction for mobile robot homing," *Robot. Auton. Syst.* **50**, 41–49 (2005).
34. D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.* **60**(2), 91–110 (2004).
35. H. Bay, T. Tuytelaars and L. Van Gool, "Surf: Speeded up Robust Features," *Proceedings of the 9th European Conference on Computer Vision*, Graz, Austria (May 7–13, 2006) vol. 3951, pp. 404–417.
36. Z. Chen and S. T. Birchfield, "Qualitative Vision-Based Mobile Robot Navigation," *Proceedings of the IEEE International Conference on Robotics and Automation*, Orlando, FL, USA (May 15–19, 2006) pp. 2686–2692.
37. R. I. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed. (2004). Cambridge University Press, ISBN: 0521540518.
38. M. A. Fischler and R. C. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM* **24**(6), 381–395 (1981).
39. "U-bot vendor," From the World Wide Web: http://www.atechsystem.com.tw/.
40. "Mechanical and systems research laboratories in industrial technology research institute," From the World Wide Web: http://www.itri.org.tw/eng/MSL/.
41. "Logitech," From the World Wide Web: http://www.logitech.com/en-us/435/238.
42. G. Bradski, "The opencv library," *Dr. Dobb's J. Softw. Tools* 2000.
43. G. Klein and D. Murray, "Parallel Tracking and Mapping for Small AR Workspaces," *Proceedings of the Sixth IEEE and ACM International Symposium on Mixed and Augmented Reality*, Nara, Japan (Nov. 13–16, 2007).
44. J. Kwon, M. Choi, F. C. Park, and C. Chun, "Particle filtering on the euclidean group: Framework and applications," *Robotica* **25**, 725–737 (2007).
45. W. Park, Y. Liu, Y. Zhou, M. Moses and G. Chirikjian, "Kinematic state estimation and motion planning for stochastic nonholonomic systems using the exponential map," *Robotica* **26**, 419–434 (2008).