

# Navigation by Image-Based Visual Homing

Matthew Szenher

University of Edinburgh, UK

## INTRODUCTION

Almost all autonomous robots need to navigate. We define navigation as do Franz & Mallot (2000): “Navigation is the process of determining and maintaining a course or trajectory to a goal location” (p. 134). We allow that this definition may be more restrictive than some readers are used to - it does not for example include problems like obstacle avoidance and position tracking - but it suits our purposes here.

Most algorithms published in the robotics literature localise in order to navigate (see e.g. Leonard & Durrant-Whyte (1991a)). That is, they determine their own location and the position of the goal in some suitable coordinate system. This approach is problematic for several reasons. Localisation requires a map of available landmarks (i.e. a list of landmark locations in some suitable coordinate system) and a description of those landmarks. In early work, the human operator provided the robot with a map of its environment. Researchers have recently, though, developed simultaneous localisation and mapping (SLAM) algorithms which allow robots to learn environmental maps while navigating (Leonard & Durrant-Whyte (1991b)). Of course, autonomous SLAM algorithms must choose which landmarks to map and sense these landmarks from a variety of different positions and orientations. Given a map, the robot has to associate sensed landmarks with those on the map. This data association problem is difficult in cluttered real-world environments and is an area of active research.

We describe in this chapter an alternative approach to navigation called visual homing which makes no explicit attempt to localise and thus requires no landmark map. There are broadly two types of visual homing algorithms: feature-based and image-based. The feature-based algorithms, as the name implies, attempt to extract the same features from multiple images and use the change in the appearance of corresponding features to navigate. Feature correspondence is - like data association - a difficult, open problem in real-world environments. We argue that image-based homing algorithms, which

provide navigation information based on whole-image comparisons, are more suitable for real-world environments in contemporary robotics.

## BACKGROUND

Visual homing algorithms make no attempt to localise in order to navigate. No map is therefore required. Instead, an image  $I_S$  (usually called a snapshot for historical reasons) is captured at a goal location  $S = (x_S, y_S)$ . Note that though  $S$  is defined as a point on a plane, most homing algorithms can be easily extended to three dimensions (see e.g. Zeil et al. (2003)). When a homing robot seeks to return to  $S$  from a nearby position  $C = (x_C, y_C)$ , it takes an image  $I_C$  and compares it with  $I_S$ . The home vector  $\mathbf{H} = S - C$  is inferred from the disparity between  $I_S$  and  $I_C$  (vectors are in upper case and bold in this work). The robot's orientation at  $C$  and  $S$  is often different; if this is the case, image disparity is meaningful only if  $I_C$  is rotated to account for this difference. Visual homing algorithms differ in how this disparity is computed.

Visual homing is an iterative process. The home vector  $\mathbf{H}$  is frequently inaccurate, leading the robot closer to the goal position but not directly to it. If  $\mathbf{H}$  does not take the robot to the goal, another image  $I_C$  is taken at the robot's new position and the process is repeated.

The images  $I_S$  and  $I_C$  are typically panoramic grayscale images. Panoramic images are useful because, for a given location  $(x, y)$  they contain the same image information regardless of the robot's orientation. Most researchers use a camera imaging a hemispheric, conical or paraboloid mirror to create these images (see e.g. Nayar (1997)).

Some visual homing algorithms extract features from  $I_S$  and  $I_C$  and use these to compute image disparity. Alternatively, disparity can be computed from entire images, essentially treating each pixel as a viable feature. Both feature-based and image-based visual homing algorithms are discussed below.

## FEATURE-BASED VISUAL HOMING

Feature-based visual homing methods segment  $I_S$  and  $I_C$  into features and background (the feature extraction problem). Each identified feature in the snapshot is then usually paired with one feature in  $I_C$  (the correspondence problem). The home vector is inferred from - depending on the algorithm - the change in the bearing and/or apparent size of the paired features. Generally, in order for feature-based homing algorithms to work properly, they must reliably solve the feature extraction and correspondence problems.

The Snapshot Model (Cartwright & Collett (1983)) - the first visual homing algorithm to appear in the literature and the source of the term “snapshot” to describe the goal image - matches each snapshot feature with the current feature closest in bearing (after both images are rotated to the same external compass orientation). Features in (Cartwright & Collett (1983)) were black cylinders in an otherwise empty environment. Two unit vectors, one radial and the other tangential, are associated with each feature pair. The radial vector is parallel to the bearing of the snapshot feature; the tangential vector is perpendicular to the radial vector. The direction of the radial vector is chosen to move the agent so as to reduce the discrepancy in apparent size between paired features. The direction of the tangential vector is chosen to move the agent so as to reduce the discrepancy in bearing between paired features. The radial and tangential vectors for all feature pairs are averaged to produce a homing vector. The Snapshot Model was devised to explain the behaviour of nest-seeking honeybees but has inspired several robotic visual homing algorithms.

One such algorithm is the Average Landmark Vector (ALV) Model (Möller et al. (2001)). The ALV Model, like the Snapshot Model, extracts features from both  $I_C$  and  $I_S$ . The ALV Model, though, does not explicitly solve the correspondence problem. Instead, given features extracted from  $I_S$ , the algorithm computes and stores a unit vector  $ALV_S$  in the direction of the mean bearing to all features as seen from  $S$ . At  $C$ , the algorithm extracts features from  $I_C$  and computes their mean bearing, encoded in the unit vector  $ALV_C$ . The home vector  $\mathbf{H}$  is defined as  $ALV_C - ALV_S$ . Figure 1 illustrates home vector computation for a simple environment with four easily discernible landmarks.

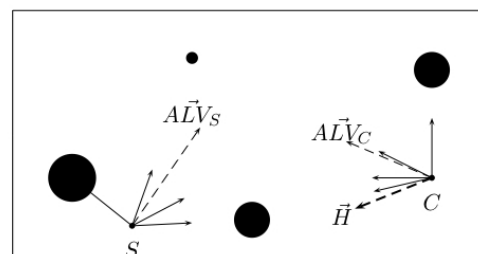
Several other interesting feature-based homing algorithms can be found in the literature. Unfortunately, space constraints prevent us from reviewing them here.

Two algorithms of note are: visual homing by “surfing the epipoles” (Basri et al. (1998)) and the Proportional Vector Model (Lambrinos et al. (2000)).

The Snapshot and ALV Models were tested by their creators in environments in which features contrasted highly with background and so were easy to extract. How is feature extraction and correspondence solved in real-world cluttered environments? One method is described in Gourichon et al. (2002). The authors use images converted to the HSV (Hue-Saturation-Value) colour space which is reported to be more resilient to illumination change than RGB. Features are defined as image regions of approximately equal colour (identified using a computationally expensive region-growing technique). Potential feature pairs are scored on their difference in average hue, average saturation, average intensity and bearing. The algorithm searches for a set of pairings which maximise the sum of individual match scores. The pairing scheme requires  $O(n^2)$  pair-score computations (where  $n$  is the number of features). The algorithm is sometimes fooled by features with similar colours (specifically, pairing a blue chair in the snapshot image with a blue door in the current image). Gourichon et al. did not explore environments with changing lighting conditions.

Several other methods feature extraction and correspondence algorithms appear in the literature; see e.g. Rizzi et al. (2001), Lehrer & Bianco (2000) and Gaussier et al. (2000). Many of these suffer from some of the same problems as the algorithm of Gourichon et al. described above. The appearance of several competing feature extraction and correspondence algorithms in recent publications indicates that these are open and difficult problems; this is why we are advocating image-based homing in this chapter.

Figure 1. Illustration of Average Landmark Vector computation. See Section titled “Feature-based Visual Homing” for details



4 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: [www.igi-global.com/chapter/navigation-image-based-visual-homing/10390](http://www.igi-global.com/chapter/navigation-image-based-visual-homing/10390)

## Related Content

---

### Cybersecurity in Remote Work

S. Ida Evangeline (2025). *Global Work Arrangements and Outsourcing in the Age of AI* (pp. 227-240). [www.irma-international.org/chapter/cybersecurity-in-remote-work/378544](http://www.irma-international.org/chapter/cybersecurity-in-remote-work/378544)

### Intelligent Agents for E-Learning

Ralf Brunsand Jürgen Dunkel (2008). *Intelligent Information Technologies and Applications* (pp. 25-42). [www.irma-international.org/chapter/intelligent-agents-learning/24258](http://www.irma-international.org/chapter/intelligent-agents-learning/24258)

### AI-Driven Regenerative Intelligence for Ecological Restoration and Climate Action

K. Aravinthan, Anorgul Ashirova, S. Aarthi, R. N. Ravikumar, Babamuratov Bekzodand Ulugbek Vosiqov (2026). *Sustaining Climate Action With AI* (pp. 87-116). [www.irma-international.org/chapter/ai-driven-regenerative-intelligence-for-ecological-restoration-and-climate-action/411374](http://www.irma-international.org/chapter/ai-driven-regenerative-intelligence-for-ecological-restoration-and-climate-action/411374)

### Supporting Text Retrieval by Typographical Term Weighting

Lars Wernerand Stefan Böttcher (2007). *International Journal of Intelligent Information Technologies* (pp. 1-16). [www.irma-international.org/article/supporting-text-retrieval-typographical-term/2415](http://www.irma-international.org/article/supporting-text-retrieval-typographical-term/2415)

### Trust based Dynamic Multicast Group Routing Ensuring Reliability for Ubiquitous Environment in MANETs

Shobha Tyagi, Subhranil Somand Qamar Parvez Rana (2017). *International Journal of Ambient Computing and Intelligence* (pp. 70-97). [www.irma-international.org/article/trust-based-dynamic-multicast-group-routing-ensuring-reliability-for-ubiquitous-environment-in-manets/176714](http://www.irma-international.org/article/trust-based-dynamic-multicast-group-routing-ensuring-reliability-for-ubiquitous-environment-in-manets/176714)