A dynamic model for autonomous vehicle navigation

Alexander Kick, Ashley Walker, Robert Fisher
Department of Artificial Intelligence
University of Edinburgh
5 Forrest Hill
Edinburgh EH1 2QI,
Scotland

Abstract
In recent years, much progress has been made in robotics research by employing ideas taken from biology. In particular, inspiration has been drawn from ethologist's study of insect navigatory capabilities. The long distance foraging behavior of honey bees (Apis mellifera) has been well studied and models of bee navigation have been suggested and simulated. The most popular of these models (due to Cartwright and Collett), proposes that honey bees navigate via a static image comparison process wherein motion vectors towards a target are calculated based on the differences between an image currently projected on a bee's retina and a remembered snapshot – taken from a similar place in the environment. This work describes our implementation of such a static localization model on a mobile robot and discusses our conclusion that a dynamic model of bee navigation, based on optical flow, is more plausible. We propose a dynamic model based on comparison of apparent speeds of edges.

1 Introduction
Recent trends in robotics research suggest that characterizations of biological systems may inspire useful engineering insights to the designers of artificial systems [2]; [15]. An example of this is active vision – a technique which employs the known motion of the viewing system to make tractable certain visual problems which are mathematically ill-conditioned or underconstrained for a stationary observer. Franceschini et. al [10] have built a robotic system with a vision sensor based on the compound eye of the hoverfly which exploits motion parallax in an effort to explore how motion detecting (artificial) neurons enable collision avoidance. In previous work, one of the authors has employed the motion parallax effects induced by the movement of a camera to build up (in real-time) a depth map of visual scenes [11]. Recently this work was extended to robotic world model construction through exploitation of the constrained optic flow arising from controlled forward motion of a robot vehicle [16].

In search of a relatively simple spatial representation technique and a primitive spatial information processing algorithm for use in our autonomous mobile robot, we too took an example from biology. The navigational systems of central place foraging insects have been a source of wonder to scientists for more than a century. Honey bees (Apis mellifera) are capable of traveling up to 10000 meters on food foraging journeys, after which they can return home along highly efficient (i.e., bee-line) routes [21]. These insects also have precision short range navigational abilities which allow them to pinpoint the location of specific targets, e.g., the small opening to their hive, and to orient precisely toward the food bearing portion of a flower [4]. The behavior of central place foraging honey-bees has been extensively studied by ethologists and computational models have been suggested and tested in simulation (e.g., [1], [4], [5], [6]). We developed and implemented a bee-like robot-navigation system [12], similar to a model suggested by Cartwright and Collett [4], which employs a static image comparison process as the basis of precision navigation localization. Before our work, this kind of model had only been tested in computer simulation and it was generally considered capable of predicting behavior encouragingly similar to real bees. During our implementation on a robot, however, we found that employing only static representations of space for precision localization has various disadvantages.

In response to these concerns, we devised a dynamic model of navigation based on optical flow which seeks to overcome the problems we encountered with a static model. Our dynamic model could account for the ethological findings regarding bee navigation which motivated the Cartwright and Collett model and incorporates knowledge about bees acquired since then.

This paper is structured as follows. We will briefly discuss the features of bee navigation in Section 2. Section 3 reviews a static model of bee localization and our
implementation of such a model on a mobile robot. Disadvantages of models employing only static images are discussed in this section. In Section 4, a dynamic model of localization is presented. In Section 5 we draw conclusions and discuss our future work.

2 Bee navigation

It is commonly believed that bees navigate due to the interplay of two distinct means: dead reckoning and localization. During dead reckoning, bees use a path integration process (which receives proprioceptive input) to continuously update a vector pointing back to their hive. After having explored new food sources, bees reverse their flight direction and travel home along this vector without having to re-trace the (often) circuitous outbound journey. However, this dead reckoning mechanism only facilitates coarse-grained navigation because integration errors accumulate. Therefore, bees must employ a second mechanism for fine-grained (i.e., precision) navigation.

This secondary navigation facility is vision-based. During precision localization, bees exploit visual information describing landmarks en route to their destination. It is widely held that bees compare the memorized characteristics of landmarks surrounding their target to the characteristics of those landmarks as seen from the bee’s current position in order to calculate motion vectors which will steer them towards the desired position precisely. It is uncertain as to exactly what information insects obtain from an array of landmarks and how they use it. Various theories exist, one of the earliest and still most widely cited of which is described below.

3 A static model for localization

3.1 Cartwright and Collett’s static model for localization

Cartwright and Collett conducted several experiments to investigate how bees use vision to perform localization. Among these, we will summarize two which motivated their static model.

In the first experiment, bees were trained to find food at a fixed distance to one landmark of a certain size. During the subsequent test phase, the food source was removed, the original landmark replaced by one of greater or lesser size, and the search pattern of a hungry, navigating bee recorded. Results showed that bees search closer (than the position of the original food source) to small substitute landmarks and farther from large substitute landmarks. From this Cartwright and Collett concluded that the size of the landmark, rather than distance to them, is memorized.

In a second type of experiment, bees were trained with a configuration of three landmarks. A similar test procedure was run, but this time the size and configuration of landmarks was varied. It turned out that “a simple rule predicts the bee’s search area ... the bee always searches where the compass bearing of the landmarks on its retina were the same as they had been when it was stationed at the food source [during training] [4].” (Cartwright and Collett also determined that the bearings of landmarks on the retina were determined with respect to external compass bearings.)

The idea that insects can remember relatively unprocessed images was introduced by Wehner’s [18] study of pattern recognition in honey bees. Later work by Wehner and Flatt [20] showed that bees learn and inspect particular regions of the visual pattern using particular regions of the eye. Collett and Land [7] first suggested that insects (i.e., hoverflies in this case) learn a representation of 3-D space consisting of these 2-D “snapshots” and use the difference between a current and remembered snapshot in order to steer their return to a preferred hovering spot.

Cartwright and Collett’s findings corroborated this earlier work and, in 1983, they suggested a computer model [4] of landmark guidance based on eidetic images. According to their theory, a bee memorizes a snapshot (i.e., an intensity image which has received minimal amounts of perceptual processing) of its nearby surroundings as perceived from some position (e.g., a food source or the hive) to which it may wish to return. In subsequent navigation trials to this target, the bee could continuously compare its retinal image with the snapshot and adjust its flight path so as to lessen the discrepancy between the two.

In their most successful computer simulation, Cartwright and Collett [4] assumed that both the memorized and current image are segmented into light and dark areas, and that an area in the snapshot is paired with the closest retinal area of the same type (dark, light). “Each of these pairings gives rise to two unit vectors... One is perpendicular to the bisector of the retinal area and points so as to align the two areas. The other is radial to the retinal bisector and acts to lessen size discrepancy between the two areas. Flight direction is given by the sum of these tangential and radial vectors.” This model simulates the behavior of honey bees in this experimental scenario quite well. However, it was not motivated by considerations of how most of its features might actually be implemented in an insect’s brain. For example, the model makes the strong assumption that snapshot orientation is constant and stabilized by an independent compass.
3.2 Our implementation of a bee-like robot navigation system based on a static bee model

Our work is the first (known to us) to implement (on a mobile robot) a static localization algorithm inspired by Cartwright and Collett's model. Although our goal was to engineer a navigation control system for a mobile robot, we believe that our results are relevant to ethologists because our implementation provided a more realistic test environment for a biological navigation model - i.e., our robot, like a navigating bee, must embrace the full complexity of operation in the real world (including real-time processing constraints, sensor uncertainty, vision problems such as illumination variations and occlusion, and actuator imprecision). As we will see below, we faced unforeseen difficulties and came to view problems differently.

Like the bee, our robot begins navigating by exploring its environment in order to build up a memory of important locations and the associated motion vectors which link them. When the robot desires to go to a memorized location, it either follows a path traveled during the exploration phase or plans a new path by reasoning about routes which it has memorized (i.e., by performing vector addition as suggested by Cartwright and Collett suggested [5]).

Our first concern in implementing this localization technique was centered around the desire to re-create (using CCD camera with a 90 degree field of view) a sensor whose field or view matched the 270 degrees of the honey bee's compound eye. In order to do this, we built up a 270 degree field of view place description (i.e., snapshot) from three individual images - taken to the robot's left, front, and right viewing quadrants.

Upon return (via dead reckoning) to a location where a description was previously memorized, the robot collects new images and compares them to the corresponding snapshot. In the Cartwright and Collet simulation, image features were extracted, matched, and the differences in image positions of corresponding features were compared. In our implementation, we simplified this procedure by instead correlating corresponding images. By comparing pixels, more information is retained and the comparison computations required are easier. The algorithm correlates previously stored views with the current views and estimates the orientation and translation (forwards/backwards) errors. This is repeated for three views at 90 degrees to each other. From this information, a motion vector is calculated which corrects for the robot's dead reckoning error and thus brings it closer to the target.

We tested our artificial bee at various locations and with various (introduced) dead-reckoning errors in our laboratory. Since we stored the approximate distance of observed objects (derived through motion parallax calculations), the robot can improve its position by an iterative algorithm which takes into account its current position estimate and relative target positions. The iterative correction of the dead-reckoning error stops when the position improvement with respect to the previous iteration is below a given threshold.

The iterative version of the algorithm reduced errors on the average by a greater amount than the simple version, especially if errors were large (Table 1). The various dead reckoning error ranges we introduced and tested were:

- small: sideways, towards/away errors smaller than 21 mm; rotational errors less than 2.1 degrees,
- medium: sideways, towards/away errors from 21 to 40 mm; rotational errors from 2.1 to 4.0 degrees, and
- big: sideways, towards/away errors above 40 mm; rotational errors above 4.0 degrees.

Thus, using this procedure, the robot is able to maintain its dead-reckoning error within a certain limit. Guidance correction vectors gleaned from localization can be used as part of a target acquisition process as we have shown above. Alternatively (or additionally) this information could be used to recalibrate a dead reckoning system. In the latter capacity, it is a valuable tool for roboticists. (Artificial calibration techniques (i.e., those involving decorating an environment with artificial beacons and bar codes providing calibration constants) are not always feasible and more sophisticated techniques which rely on external sensing of naturally occurring environmental cues (e.g., Kalman filtering) are computationally expensive relative to this technique.

Table 1: Average percentage of dead reckoning error left after application of simple and iterative version of the global localization algorithm

<table>
<thead>
<tr>
<th>d. r. error</th>
<th>simple version</th>
<th>iterative version</th>
</tr>
</thead>
<tbody>
<tr>
<td>small errors</td>
<td>62.6</td>
<td>39.7</td>
</tr>
<tr>
<td>medium errors</td>
<td>57.4</td>
<td>51.1</td>
</tr>
<tr>
<td>big errors</td>
<td>94.8</td>
<td>35.4</td>
</tr>
<tr>
<td>total</td>
<td>65.6</td>
<td>45.1</td>
</tr>
</tbody>
</table>

---

1. The robot used is a cylindrically shaped (30 cm diameter and 50 cm high) machine whose head unit (i.e., the external sensing and computational facilities) translates and rotates atop a B/12 RealWorld Interface omnidirectional mobile robot base. The head electronics include: sensors and their associated control and processing systems (including a CCD camera with frame grabber), a switch interface board, the stand-alone TransTech transputer array (i.e., four T800 transputers), and an Electronics Flight 68k board. The transputer system can send data to and receive data from a remote terminal via a radio modem. (For more detail see [9].)
3.3 Discussion of static model, disadvantages, questions

Using a static localization algorithm based upon comparison of image data, our bee-like robot navigation system enabled a robot to navigate towards targets more precisely than possible via dead reckoning alone. We believe that these results could be improved further while still employing static images. However, during the process of implementation, we began to hypothesize that static image comparison alone may not provide enough information to guide an autonomous agent’s return toward a precise location. We believe that dynamic information (i.e., optical flow descriptions) are used as well. This hypothesis is motivated by three noted shortcomings in the static localization model we implemented:

- It is not clear how valuable a snapshot taken from a fixed position would be for an autonomous mobile agent. Such a snapshot would be only a tiny temporal fraction of the world projected on the agent’s retina and thus not contain much information about a location. Furthermore, a single perspective on a scene rarely provides enough information to disambiguate its 3-dimensional structure (e.g., shadows and occluded regions complicate the image). From these considerations, it would seem that information acquired during the process of motion certainly provides a more rich and relevant description of the world than one gathered from a particular vantage point.

- Noise in an image matching system is also a critical consideration. Zeil argues: “Consider a snapshot taken at the nest entrance and covering half of the full solid angle, i.e., 6.28 steradians. A small cylindrical landmark (2.2 cm x 6.3 cm) at 6 cm from the nest entrance would cover 0.386 steradians, i.e., 6% of the snapshot. If one allows for noise, it is difficult to see how such a small part of the snapshot could generate a reliable signal in an image matching procedure” [23].

- Image matching by static image correlation degrades rapidly as distance from the observed features deviates from the true distance. We attempted to correct for this by estimating a conversion factor and approximate target distance; however, we still found the effect to be significant.

These concerns and the abundance of recent attention ethologists have focused on motion cues in insect navigation systems inspired the navigation model suggested in the next section.

4 A dynamic navigation model

Flying animals use visual motion cues to stabilize their flight (e.g., opto-motor response in hover flies [7]). In addition to optomotor pathways, it is apparent that insects use visual motion information in more sophisticated ways. Due to the small interocular separation in the insect visual system, stereo vision can not be used to measure the range of objects at distances greater than a few centimeters. Therefore, visual motion cues must be employed to recover this information. Scanning behavior of insects (i.e., honey bees), as described by Lehrer et al. [14], demonstrates that bees do induce motion of images on their retina during navigation. Srinivasan [17] has shown that honey bees use optical flow cues to determine the distances of surfaces, discriminate between objects at different distances, land on a contrast edge, and distinguish objects from background.

There is mounting evidence that dynamic visual information plays a vital part of an insect’s spatial representation. Additional experiments, similar to those performed by Cartwright and Collett [4] which follow the search pattern of navigating honey bees, indicate (i) that they weight landmarks close to the target more heavily than distant ones [6], and (ii) that they frequently search at the appropriate distance from a landmark regardless of its apparent size [3]; [6] especially when no other cues are available [13]. Both findings suggest that bees use patterns of motion cues in their localization procedures. The retinal position of motion vectors in a parallax field produced during an orientation flight, their size and direction all contain information on the spatial layout of the nest surroundings [22].

In the next section we propose a way by which an autonomous mobile agent might learn and use dynamic descriptions of the world in its localization procedure.

4.1 Analysis of optical flow

4.1.1 Speed of edges in the optical flow

The motion of an autonomous mobile agent induces apparent motion of objects on its imaging surface. Regions on the agent’s eyes or camera(s) that move with different speeds represent different objects. The further the object from the agent and the closer to the focus of expansion (the pole towards which the observer is moving), the smaller the speed of the projection of the object on its imaging surface will be. From this dynamic information, it is possible to segment the image by analyzing the optical flow field.

Suppose the agent moves with velocity \( \mathbf{v} \) and at time \( t = 0 \) has distance \( d \) to the object. Furthermore, the agent’s motion direction and \( r, s, \) and \( \alpha \) are as in Figure 1. Then we can calculate the speed of the induced motion of the object on the imaging surface \( a(t) \):

\[
\tan \alpha = \frac{s}{r - vt}
\]
4.1.2 Dependence of relative speed on agent’s relative motion

Formula 3 shows dependence of induced speed $a$ for certain position of the agent ($t = 0$) and the object on $r$, $s$, and $v$. How does $a$ depend on the orientation of the agent with respect to the object, for fixed distance between agent and object.

Looking at Figure 2 we can deduce with formula (3):

$$ s = \frac{d \sin \alpha}{a(0)} $$

$$ a(0) = \frac{r \sin |\alpha|}{d} $$

The graph of the apparent speed $a$ is depicted in Figure 3. It shows that, for a given distance of agent to object, apparent motion speed $a$ has its maximum if the agent moves perpendicular to the line of sight to the object and $a$ decreases as agent orients itself towards the object, i.e.

Figure 1: Calculating apparent motion

Figure 2: Dependence of apparent motion on orientation towards the object

Figure 3: Dependence of apparent motion on the angle between motion direction and object.

when the smaller angle between motion direction of agent and line between agent and object decreases.

4.1.3 Effects of changing bearing

This fact can be used when comparing apparent speeds of objects on the left half of an agent’s viewing surface with the right (Figure 4). If the agent turns towards object 1 (changed motion direction) the speed of object 1 will be decreased and speed of object 2 increased. Turning right has the opposite effect. Thus, the motion parallax field is a function of not only the spatial layout of the scene, but also of ego-parameters such as the agent’s motion direction and agent position. This complexity challenged the original Cartwright and Collett model. They resolved that snapshots (i.e., static snapshots in his case) are defined with respect to compass bearings and may be fixed to a bee’s retina, but only activated when the bee is oriented in the direction in which the snapshot was taken. (Since bee’s can return to a particular location from many directions, they would need several snapshots, each taken when the bee was facing in a different direction and used when the bee is flying approaching along that heading.)

In the following subsection, we outline a localization model which employs motion parallax information. In doing so, we employ the same set of assumptions used in previous bee localization models regarding orientations effects, i.e., the agent maintains its scene description (e.g., static snapshots and/or motion parallax fields) with respect to the particular compass bearing orientation it employed when the description was recorded.
Figure 4: Turning reduces apparent motion on one hemisphere of the agent’s viewing system and increases it in the other hemisphere (We assume that apparent motion induced by turning is corrected for.).

4.1.4 Localization by comparing left and right speeds

How does the difference of object (e.g., an edge) velocities on the left half of the imaging surface compare with those on the right half (subtraction of speed on left half from speed on right half) look for a given scenario? Figure 5 shows a scenario (seen from above) where the agent moves towards edges 1 and 2 with motion direction parallel to the y-axis. The distances are given in arbitrary units.

Figure 6 depicts the difference (z-axis) between a of edge 2 (a2) and edge 1 (a1) for various (x,y)-positions. The shape of the graph can be used to devise a localization algorithm: at a target position (which can be any position on the graph in Figure 6) the agent memorizes the apparent speeds of the edges a1, and a2. When the agent returns to a target (starting with a y value of 0) it can perform gradient ascent/descent so as to experience a difference between a2 and a1 (d) similar to the difference between the stored values a2, and a1, (ds).

For example, suppose the agent stored the target position (0,7). Starting from (2,0), the agent experiences a small difference between a2 and a1. In order to increase this difference the agent will position itself more towards edge 1 and thus increase speed of edge 2 and decrease speed of edge 1 (as shown in Section 4.1.3). As a consequence, the difference between a2 and a1 increases.

Turning to the right would do the opposite and worsen the similarity between stored and currently experienced difference. The agent stops its surface ascent/descent (on the 3D-shape described by |d_e(x, y) - d_s(x, y)| for two given edges e1 and e2) when moving forwards in either direction reduces the similarity between d_e and d_s.

In the example shown, d_e is identical for various (x,y) positions. The more objects in the motion parallax field, the more interesting and complicated this surface will become. It is possible that target position can be reached more accurately on a motion parallax field produced by a more cluttered, realistic scene. (Alternatively, some perceptual filtering may be necessary.) Therefore

\[ f(x, y) = \sum_{e1, e2} d_e(e1, e2, x, y) - d_s(e1, e2, x, y) \]

could be the function on which gradient ascent/descent is performed, where e1 are edges on left part of the eye, e2 edges on the right part. However, there are other ways to combine this information which might create a more desirable surface.

4.2 A dynamic localization model

Based on this analysis of optical flow, we suggest the following representation of the world in an agent’s memory and its use in localization. When an agent initially moves to a target location, it stores the speeds of edges
experienced through motion parallax.

4.2.1 Feature disambiguation and orientation correction using distant features

When the agent approaches a memorized location another time, it matches object features (e.g., edges) from current motion parallax and/or static descriptions with the corresponding stored features.

The amount that the agent is off-course (due to imprecision in the dead reckoning mechanism) will affect the scene description of distant objects only negligibly (in contrast to close objects). That is, far away objects will have the same characteristics, especially retinal image size and velocity, as the memorized ones. We hypothesize (similar to what has been stated in [5]) that the agent exploits this fact and first matches far away objects (which are easily recognized since they move very slowly on the imaging surface during the motion of the agent) with memorized distant landmarks. This kind of matching is fairly unambiguous because of the similar characteristics.

The matching of far away objects will reduce the ambiguities of the remaining regions so that the subsequent matching of closer objects is facilitated considerably. Furthermore, the agent can use this initial phase in the localization process to correct for deviations in its current orientation from the original orientation so that the later, precision phases of orientation need only work in correcting for position misalignments.

4.2.2 Inducing similar optical flow field by comparing speeds

Now, the characteristics of close objects that define the position of a goal accurately are compared to memorized ones in order to achieve a fine correction. The agent now moves so as to induce the flow field experienced when first moving to the goal. This is done by performing gradient ascent/descent on the differences between speeds on the left and the right half of the viewing surface as described above.

5 Conclusions, future work

Collett and Kelber [8] found that distant landmarks and other contextual cues ensure that bees retrieve the correct memory of a constellation of local landmarks while the bees are still some distance away from their goal. The outlined model suggested above incorporates this finding: the navigating agent first tries to match currently experienced distant landmarks with those in memory. Having found the correct match for the distant surroundings, the retrieved context will determine the searching behavior of the agent for the close landmarks. In particular, it will enable the agent to correctly orient itself so that the difficult position corrections can then be made independently. This hierarchical approach to localization (i.e., wherein coarse-scale deviations and orientation errors are corrected before the agent fine-tunes its position) reduces the computational complexity of the algorithm and allows the agent to pay most attention to the rich information provided by near landmarks [6].

Our implementation of a static bee-like navigation system on a robot has furthered our understanding of the behaviors and underlying mechanisms that allow autonomous mobile agents to adapt and survive in their environments. Although we were mainly interested in building a robust robot navigation system, our implementation on a robot changed our beliefs about the organizational principles which induce the bee’s adaptive behavior. Thus, our work has implications for both natural and artificial animals.

We believe that a dynamic model of localization could be very effective since it provides a more subjectively relevant representation of a mobile agent’s world. It has been shown that information about dynamic changes (e.g., the apparent motion of objects) is a useful environmental feature for a navigating autonomous agent. Furthermore, the level of scene necessary analysis still remains low: instead of creating a precise geometric description of objects and employing complicated reasoning schemes, directly available information about objects (their apparent motion) is stored within a context (the location).

Clearly, we have much interesting work ahead of us in implementing a dynamic localization model.

Acknowledgements

Thanks for assistance with the research and paper to Andrew Fitzgibbon and Martin Westhead. This research was supported by the Hanns-Seidel-Stiftung und Univer-
References


