THE UNIVERSITY
of EDINBURGH

MECHANISMS OF PLACE RECOGNITION AND PATH INTEGRATION BASED ON THE INSECT VISUAL SYSTEM

THOMAS J. STONE

Doctor of Philosophy
School of Informatics
University of Edinburgh

2017

Thomas J. Stone:
Mechanisms of place recognition and path integration based on the insect visual system
Doctor of Philosophy, 2017

SUPERVISORS:
Barbara Webb
Stanley Heinze

Dedicated to the loving memory of John L. Stone.

1926 – 2014

LAY SUMMARY

Biorobotics is a scientific field that involves building robots or computer agents based on biological systems, which serves two main purposes. Firstly, to gain insight into how animal behaviour arises. Secondly, to learn biological solutions to challenging tasks and apply them to human problems. In this thesis both approaches are applied through the study of two species of visually navigating insects.

In part i, the desert ant *Cataglyphis velox* was investigated. To learn how they might be navigating, their sensory input was mimicked by filming with a camera sensitive to the same wavelengths of light: green and ultraviolet (UV). The ant's natural environment was photographed to search for prominent landmarks in this colour space that could be used by the ant for homing. Although no landmarks were identified, what became evident was the huge contrast between ground and sky, creating a prominent skyline. This lead to the hypothesis that ants are indeed using the skyline contour for navigation, and that it could also prove useful for robotic navigation. Certain advantages that result from navigating using the shape of the sky include stability over time and invariance to lighting and weather conditions. Another key finding is that, if stored in the correct way, the sky shape can reliably be used to identify a previously visited location, despite facing in a different direction or being tilted. This could be advantageous for ants that routinely cross uneven terrain when foraging, but also proved useful for robots travelling at high speeds on bumpy terrain.

In part ii, a nocturnal species of bee, *Megalopta genalis*, was studied by creating a computer model representing the central complex (CX), a brain region found in all species of insects. While foraging, this bee performs a task called path integration: the bee flies out in a winding path looking for food, then upon feeding returns to the nest in a straight line. To achieve this, the bee needs to be constantly storing how far it has travelled in all directions by monitoring its speed and heading. A set of newly identified neurons were shown by Stanley Heinze to respond in a proportionate manner to visual input corresponding to forwards and backwards motion, allowing speed to be estimated. In addition, physically close to these speed-sensitive cells were previously identified cells which are known to capture the current heading of an insect relative to the position of the sun, acting as a compass. Both cell populations are located in the CX and by creating a computational model of the neural circuit, inferred by overlap between these cell populations, a hypothesis was developed on how path integration and homing occurs in insects.

ABSTRACT

Animals are often able to solve complex navigational tasks in very challenging terrain, despite using low resolution sensors and minimal computational power, providing inspiration for robots. In particular, many species of insect are known to solve complex navigation problems, often combining an array of different behaviours (Wehner et al., 1996; Collett, 1996). Their nervous system is also comparatively simple, relative to that of mammals and other vertebrates.

In the first part of this thesis, the visual input of a navigating desert ant, *Cataglyphis velox*, was mimicked by capturing images in ultraviolet (UV) at similar wavelengths to the ant's compound eye. The natural segmentation of ground and sky lead to the hypothesis that skyline contours could be used by ants as features for navigation. As proof of concept, sky-segmented binary images were used as input for an established localisation algorithm SeqSLAM (Milford and Wyeth, 2012), validating the plausibility of this claim (Stone et al., 2014). A follow-up investigation sought to determine whether using the sky as a feature would help overcome image matching problems that the ant often faced, such as variance in tilt and yaw rotation. A robotic localisation study showed that using spherical harmonics (SH), a representation in the frequency domain, combined with extracted sky can greatly help robots localise on uneven terrain. Results showed improved performance to state of the art point feature localisation methods on fast bumpy tracks (Stone et al., 2016a).

In the second part, an approach to understand how insects perform a navigational task called path integration was attempted by modelling part of the brain of the sweat bee *Megalopta genalis*. A recent discovery that two populations of cells act as a celestial compass and visual odometer, respectively, led to the hypothesis that circuitry at their point of convergence in the central complex (CX) could give rise to path integration. A firing rate-based model was developed with connectivity derived from the overlap of observed neural arborisations of individual cells and successfully used to build up a home vector and steer an agent back to the nest (Stone et al., 2016b). This approach has the appeal that neural circuitry is highly conserved across insects, so findings here could have wide implications for insect navigation in general. The developed model is the first functioning path integrator that is based on individual cellular connections.

ACKNOWLEDGEMENTS

It has been a privilege to spend several years of my life studying navigation, a topic that is of great personal interest to me. Studying on my PhD has been thoroughly enjoyable and I hope that the work described in this thesis will provide some useful contributions to the field. My experience was hugely enhanced by the fantastic people around me, especially my supervisors Barbara Webb and Stanley Heinze, who both put in a great amount of time and effort. I am also very grateful to the fantastic collaborators that I have been able to work with: Michael Mangan, Paul Ardin, Michael Milford, Dario Differt, Andrea Adden and Nicolai Weddig. The informatics technicians were always extremely helpful in constructing the custom hardware necessary to carry out experiments described in this thesis. I'd also like to say thanks to everyone from the Insect Robotics group, Neuroinformatics Doctoral Training Centre and other academic staff who have been supportive to me, in particular Antoine Wystrach, Benjamin Risse, Matthew Graham, David Wood, Sander Keemink, Maciej Pajak, Scott Lowe, Helen Ramsden, Mark van Rossum, Margarita Chli, Matt Nolan and Iain Murray. I am very grateful to my family and friends, who often provided support and even proofread several of my papers, so thank you Liz, Keith, Matt, Sam and Rebekah Stone, Per Magnus Veierland, Jason O'nions and David Barkley. Finally, a huge thanks to Helen Longford for being supportive, very patient and feeding me baked goods.

DECLARATION

I declare that this thesis was composed by myself, that the work contained herein is my own except where explicitly stated otherwise in the text, and that this work has not been submitted for any other degree or professional qualification except as specified.

Edinburgh, 2017

Thomas J. Stone,
May 27, 2017

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACRONYMS

ACO ant colony optimisation

ANN artificial neural network

BoW bag-of-words

CBL lower division of the central body

CBU upper division of the central body

CNN convolutional neural network

CX central complex

DRA dorsal rim area

DSLR digital single-lens reflex camera

EB elipsoid body

EKF extended Kalman filter

EM electron microscopy

FB fan-shaped body

FD Fourier descriptors

fMRI functional magnetic resonance imaging

FOV field of view

GPS global positioning system

ICN international congress of neuroethology

IMU inertial measurement unit

LAL lateral accessory lobe

MB mushroom bodies

MCFO multicolor flip-out

NO noduli

PB protocerebral bridge

POL polarisation-sensitive

RGB-D red, green, blue, depth

SH spherical harmonics

SIFT scale-invariant feature transform

SURF speeded up robust features

SLAM simultaneous localisation and mapping

UAV unmanned aerial vehicle

UV ultraviolet

V1 primary visual cortex

ZM Zernike moments

1 INTRODUCTION

1.1 GENERAL INTRODUCTION

Navigation is an important human skill that has been developed and refined over millennia to solve tasks ranging from sailing around the world to driving rovers on Mars. It has become particularly interesting lately, due to the advent of autonomous vehicles, such as self-driving cars and unmanned aerial vehicles (UAVs). The umbrella term *navigation* really describes a set of different techniques that can help to either determine where an entity is located (localisation) or how to reach a desirable location through, e. g., mapping, planning or homing. These navigational tasks are problems that living organisms have repeatedly needed to solve across different ecosystems and nature has given rise to a diverse array of, often highly efficient, solutions. Understanding and drawing inspiration from how animals navigate is an important endeavour, as they often solve complex navigational tasks in very challenging terrain, despite using only low resolution sensors and minimal computational power. When studying navigation, one of the most interesting classes of the animal kingdom to focus on are insects. Many species of insect are known to solve complex navigation problems, often combining an array of different behaviours (Wehner et al., 1996; Collett, 1996). Despite this, their nervous systems are comparatively simple, relative to that of mammals and other vertebrates (Haberkern and Jayaraman, 2016).

This thesis is concerned with two main research questions, both of which concern navigational mechanisms of insects that rely upon visual input:

1. Can the spectral sensitivity of insect eyes provide a clue about which features of the environment they use to visually localise? If so, do these features provide any advantages in overcoming environmental challenges that could also be exploited by navigating robots?

2. Given the neural connectivity of two cell groups in the insect midbrain that act as a visual odometer and compass respectively, can we infer how insects might be carrying out complex navigational behaviours such as path integration?

A biorobotics approach was taken to attempt to answer both of these questions and is described in detail in subsequent chapters.

First, the current chapter summarises the various fields that this thesis is concerned with. In the following section, the field of biorobotics is introduced. The challenges of both robotic and insect navigation are described and the concept of neural modelling is introduced. Finally, the main contributions and overall structure of this thesis are outlined.

1.1.1 *Biorobotics*

Biorobotics is a subfield of robotics, in which biological organisms are emulated by robots or simulated through agents. Robots can often provide an interesting and useful model for natural animal behaviour by acting as a platform to learn about how the animal itself functions. Webb (2001) argues that to learn about animals, the most relevant models are those that can test or create hypotheses which are applicable to biology. To accomplish this, a robot must model the animal at an appropriate level of abstraction and face problems similar to the animal it is mimicking, preferably in a similar environment. Observing where the robot succeeds or fails can provide a good indication on the soundness of any hypothesis about the solution used by the animal to perform a task, e. g., visual navigation.

Another role of biorobotics is to apply biologically inspired solutions to human problems, also known as biomimetics (Vincent et al., 2006). Animals have often evolved efficient ways of tackling challenging problems, and copying them can lead to the development of algorithms that prove to be useful in the modern world. Examples include ant colony optimisation (ACO) by Dorigo et al. (1996), where the ants' process of laying a chemical trail to reinforce the shortest paths to food has successfully been used to efficiently solve graph traversal problems, e. g., in telecommunications networks (Schoonderwoerd et al., 1997). Learning about the insect visual system has already helped to develop methods for navigating through corridors and improved robotic visual odometry solutions (Srinivasan et al., 1999).

Combining these two roles of biorobotics can provide a feedback loop; the more knowledge that is accumulated about a species, the easier it is to copy and apply its features to engineering problems. Similarly, as more sophisticated models of the animal can be engineered, the ability to test more behavioural hypotheses leads to a greater understanding of how the species functions.

1.1.2 *Robotic navigation*

The term navigation covers a number of techniques that are used to determine or maintain a course or trajectory to a goal location (Franz and Mallot, 2000). Navigation

is an area of robotics that is particularly well-suited to bioinspired approaches, as the navigational problems that animals face are often similar to those faced by robots. It follows that behaviours seen in insects and other animals to solve these problems could likewise prove useful for robots. An example robot task, that is also observed in the insect world, is known as visual teach and repeat, where a robot needs to retrace a route it has previously seen (Furgale and Barfoot, 2010). Ants are known to perform a similar task, visually recognising scenes along long foraging routes between their nest and a food source, allowing them to retrace their path (Wehner et al., 1996). Another example is the task of visual homing, where a robot walks towards a goal by minimising the difference between the current view and a previously stored one (Labrosse, 2007), which is similar to how an insect might home towards its nest (Zeil, 2012).

However, in the field of robotics, being able to navigate does not exclusively mean knowing where you are or where to go, but often also involves techniques for building and maintaining consistent spatial representations or maps of the surroundings. Autonomous robots often rely on detailed maps to reliably navigate in their environment, an example being autonomous cars (Guizzo, 2011). Maps can be metrically accurate, where each place is assigned to a location, or topological, where known places are linked in a relative manner. Robots can utilise pre-made maps by matching features from the sensed environment to a location on the map, however a harder problem, common to autonomous robots, is to construct a map on-the-fly. In autonomous robotics, many navigational approaches incorporate some form of simultaneous localisation and mapping (SLAM), where both tasks are carried out iteratively (Durrant-Whyte and Bailey, 2006). To reliably localise in this manner, robots need to solve a number of tricky problems. When adding more information to a map, the positional data is often inferred through noisy self-motion sensor measurements. As the map grows, more errors are introduced, resulting in a gradual build up of relative drift between areas at various extremities of the map (Fig. 1.1). Loop closure is achieved when the robot recognises a previously visited location and in doing so is able to correct the map, making it globally more consistent.

Due to the importance of loop closure, robust place recognition is one of the most important navigational tasks for robots, as they need to accurately identify previously visited locations to build consistent maps. However, place recognition can often be hard due to the large variability of scenes that occur in the natural world. Discrepancies can be due to time of day, weather, seasonal changes, but also because of traffic, construction, and other human activities. To localise reliably, the key is to identify and extract stable, non-moving features from the scene that can reliably be used to

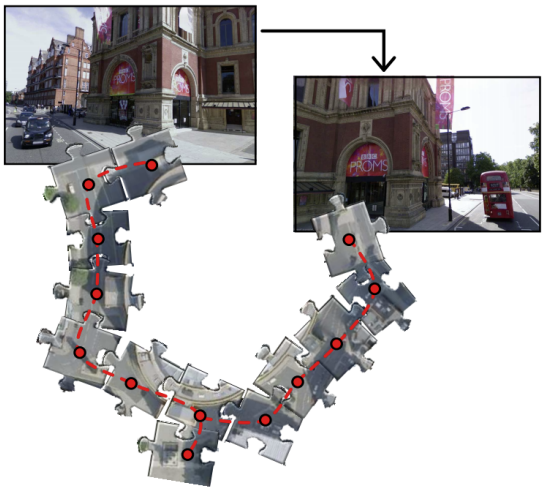


Figure 1.1: **Loop closure.** As a robot explores the environment, cumulative distance errors build up. Loop closure is achieved when the robot recognises a previously visited location and in doing so is able to correct the relative distances between places, making the map globally more consistent. Figure from Chli (2010)

identify a place. There are a number of different approaches to this problem (Sec. 2.1.1) and chapters 3 and 4 propose a novel insect-inspired solution.

Besides recognising a location directly from the immediate surroundings, a robot can also infer its location by working back from recently recognised places. As the robot moves, its estimated position can be constantly updated by relying on odometric sensors, a process known as dead reckoning (Sec. 2.1.2). A crude example is the location estimate of your car satnav system while you drive through a tunnel, based on the car's velocity and it's most recently known position. Additional sensors measuring, e. g., angular velocity can be used to maintain more accurate positional estimates, despite convoluted pathways over extended periods with no place recognition occurring.

This thesis will not cover the mapping aspect of robotic navigation, but rather focus on two forms of visual localisation performed by both robots and insects: visual place recognition and dead reckoning, known in animal studies as path integration, through visual orientation and odometry.

1.1.3 *Insect navigation*

Insects display a wide range of navigational behaviours, ranging from the long distance migrations seen in locusts, butterflies and moths (Merlin et al., 2012) to the short straight sprints that dung beetles make when rolling dung balls away from rivals at a dung pile (Byrne et al., 2003). One of the major differences from robotic navigation is the current consensus that insects do not create a cognitive map of their

surroundings (Wehner and Menzel, 1990; Collett and Collett, 2006). This is considered unlikely, both, due to the lack of neural resources, and because computational models, e. g., Cruse and Wehner (2011), show that maps are unnecessary to recreate a range of observed complex navigational behaviours. Examples of visual navigation include feature recognition, heading determination based on allocentric cues and optic flow to determine speed. Both egocentric and allocentric visual navigation strategies are employed by insects, meaning motion and positioning are measured and processed either with respect to the self or some external system of reference such as a solar compass (Wehner et al., 1996).

Most of what is known about insect navigation is inferred through behavioural studies. However, in many behavioural assays it is impossible to know which navigational strategies are actually being employed, with animals likely combining a wide range of navigational behaviours (Collett, 1996), resulting in emergent, complex and often highly redundant behaviour. A core challenge in studying animal navigation is to break down behaviour into simple heuristics. However, simply analysing behavioural data to infer the underlying algorithms can be extremely challenging (Wystrach and Graham, 2012). The strategies that insects employ in similar situations can differ greatly within the same species or even the individual and it is entirely possible that the importance of various strategies depends on behavioural state, e. g., if the animal is in a novel or familiar location (Wystrach et al., 2012). Insects could be using visual information for a number of strategies (Collett, 1996), making it hard to know how they interact and are weighted by simply restricting visual input.

Previous work on insect navigation has approached these problems by trying to categorise high-level behaviour in to clear types, e. g., visual homing, route following and path integration (Collett, 1996). Computer models have also been used to understand the underlying algorithms of behaviour and act as a proof of concept for proposed navigation algorithms in a semi-realistic environment, e. g. Baddeley et al. (2012). Many of these models however fall prey to problems such as being too far removed from biology, requiring computational power that is unlikely to be found in insects, or navigating in a brittle manner in practice.

A new tool that has recently become available, but has yet to be fully exploited is the access to neural data from insect brains (Webb and Wystrach, 2016). Examining the neural circuitry of animals can provide clues about their underlying mechanisms, e. g., place cells in mammals (Moser et al., 2008), however the networks are often too large to understand completely. Brain areas are sometimes less accessible in certain species due to the lack of genetic tools available or anatomical hurdles such as thick exoskeletons, making it especially hard to record neural activity in behaving animals. In chapter 5, this thesis makes use of newly available data on neural connections and

activity to implement a biologically plausible model of path integration, a common navigational behaviour across insects.

1.1.4 *Neural modelling*

Neural models are computational tools that allow us to understand more about the dynamics of biological networks that they approximate. A good model should clarify how the nervous system is achieving a particular function through proof of principle. They should also make testable predictions, allowing the validity of the model to be verified. Models often involve some representation of individual neurons, which can be modelled at many different levels of abstraction (Herz et al., 2006). Low-level approaches can involve detailed modelling of the morphology and chemical properties in each cell, e. g., the Human Brain Project (Markram, 2012). On the other extreme, entire brain areas can be grouped and simply given abstract activation values based on imaging data such as functional magnetic resonance imaging (fMRI) (Bullmore and Sporns, 2009). The choice of model is often constrained by computational resources, electrophysiological and anatomical data, assumptions about activity external to the modelled area, but also time limitations related to building, maintaining and verifying the correctness of a model with a certain level of complexity. Even with strong constraints, it can be hard to know which approach is most suitable to a particular problem and experts often disagree.

1.2 CONTRIBUTIONS

In the following section, the main contributions for all the studies in this thesis are listed, grouped by model organism. In part i of this thesis, the visual input of a navigating desert ant *Cataglyphis velox* was mimicked by developing a camera that was able to capture images in ultraviolet (UV), allowing the skyline to be easily extracted. Main contributions to the field include:

- The development of a novel combination of sensor and segmentation algorithm to reliably extract the skyline by using UV light, despite changes in condition.
- Supporting evidence through proof of concept that ants could use the skyline to navigate.
- Use of extracted skyline features as input to an established localisation algorithm, matching the performance of greyscale images using similar number of bits per image.

- The application of spherical harmonics (SH) to the extracted sky features, showing that global descriptors can be used in a condition and pose invariant manner.
- Demonstration of improved performance to state of the art point feature localisation methods on fast bumpy tracks, where motion blur interferes with point feature extraction.
- Proof of concept that ants could localise under foliated areas using sky versus ground features.
- Pilot study, showing that simulated ants can improve homing by using frequency-based features.

In part ii, an approach to understand how insects perform a navigational task called path integration was undertaken by modelling the central complex (CX) of the sweat bee *Megalopta genalis*. The main contributions are:

- Development of the first plausible model of path integration that is anatomically grounded on individual neural connections.
- Expansion of existing theoretical models of path integration by the addition of realistic holonomic motion.
- The suggestion of new neural roles based on this model, including $CPU4$ columns acting as a compound memory consisting of two subpopulations and Pontin cells performing a crucial normalisation role.
- Potential wide implications for understanding insect navigation in general, as this approach has the appeal that neural circuitry is highly conserved across insects.

1.3 ORGANISATION OF THE THESIS

This thesis consists of three studies: two published robotics conference papers and a journal article that is currently under review. Chapter 2 provides a broad literature review covering the necessary background for all the papers. The following three chapters each briefly introduce the area of interest for the corresponding paper, and a justification for the chosen approach. Supplementary results and a brief discussion follow all papers, putting findings in to perspective with the current state of the art, outlining problems and suggesting potential future work.

The three papers are grouped into two parts. Part i covers two related studies on place recognition using UV sensors and segmented sky/ground images. Chapter 3 describes how UV can be good for skyline detection, which, in turn, can serve as a robust feature for localisation. Chapter 4 is a follow-up study, where rotation invariant localisation is achieved by storing the UV-extracted sky shape in the frequency domain. Part ii contains only chapter 5 and describes a computational model of the CX, a region of the insect brain which is highly implicated in the navigational task of path integration. Finally, chapter 6 is a broader discussion, summarising and linking the results of all three studies and hopefully providing some insight into the biorobotics approach, especially concerning the study of insect navigation.

2

BACKGROUND

This chapter provides some background on both robot and insect localisation and attempts to outline where the two overlap, focusing especially on visual place recognition and path integration/dead reckoning. The possible neural underpinnings of these visual navigation mechanisms are introduced. The final section explores what is known about the structure and function of the central complex (CX), how it is related to visual navigation and provides a description of existing models.

2.1 LOCALISATION IN ROBOTICS

Localisation can be described as any technique to determine ones current position. A familiar and ubiquitous example of a localisation method is the global positioning system (GPS), which works anywhere on earth with a clear view of the sky and can determine a position to around 5 metres accuracy. [1] Although GPS can be useful in robotic applications, it is often limited due to power consumption, timing or precision constraints and can catastrophically fail in areas such as urban canyons, meaning other localisation systems are necessary.

Many localisation approaches involve place recognition, where observed features of the environment are matched to a stored representation. Features can be simple snapshots of the current state of one or more sensors. However, often sensory input is first processed, allowing it to be more discriminatory or provide more efficient storage and lookup. Extracted features can be stored as vectors matching a single location or can contribute to a generated map or model of the environment. Place recognition does not necessarily have to be a view-based, but can also be achieved by measuring other sensory input, such as the local Wi-Fi fingerprint (Biswas and Veloso, 2010) or magnetic field (Li et al., 2012). The shape of the environment can also be used to determine a familiar location by using sonar (Drumheller, 1987) or LIDAR (Bosse and Zlot, 2009). Finally, localisation can occur, not only by recognising the current location, but by estimating displacement from a known reference location, through a process known as dead reckoning. This thesis covers two localisation processes in particular: visual place recognition and dead reckoning, both of which are discussed in more detail in the context of robotics in the following sections (Sec. 2.1.1 and 2.1.2).

1 http://www.gps.gov/systems/gps/performance/accuracy/

2.1.1 *Visual place recognition*

Due to the wide availability of cameras and the natural relationship to our dominant sense, vision-based approaches to place recognition are some of the most prevalent. To fully exploit visual input, it can be beneficial to use multiple cameras (Lemaire et al., 2007), or even red, green, blue, depth (RGB-D) sensors (Fallon et al., 2012) which provide more accurate information about the position in 3D space of the observed pixels. However, accurate localisation is possible using just a single camera (Davison et al., 2007), which is a common constraint in robotics.

(a) Point features (b) Global features

Figure 2.1: **Features for place recognition. (a)** SIFT features are based on the local image gradient. Shortcomings of this approach to place recognition are the lack of global shapes that are taken in to account, such as the spatial proportions of the building. Also, some features are erroneously detected as useful such as the clouds. Image from OpenCV website (Bradski et al., 2000). **(b)** Scene descriptors by McManus et al. (2014) are an example of trying to identify reliable features that work across conditions by choosing environment-specific salient images patches.

To perform place recognition, one or more images is taken of the current view, which can cover a small visual angle or even be fully panoramic. These are then matched against stored representations of previously taken images. Simple pairwise pixel comparisons usually do not work well when matching an observed image to a large database of stored images. The procedure can be extremely expensive due to the comparison of many high dimensional image vectors, especially when taking into account the additional operations needed for image registration, to ensure each pair of images are correctly aligned. Whole image matching is also not robust to changes in condition, such as lighting, movement of non-stationary objects in the scene or even just seasonal changes such as foliation on trees. Finally, whole images are also very brittle to changes in pose, where images have been taken from a slightly different position or angle. To more robustly compare images, useful features need to be

extracted that solve these problems. The following sections will briefly describe some of the common solutions to visual place recognition. For a comprehensive review see Lowry et al. (2016).

Local feature descriptors

Engineered solutions to view-based place recognition have historically relied on point features, such as scale-invariant feature transform (SIFT) (Lowe, 1999) (Fig. 2.1a) and speeded up robust features (SURF) (Bay et al., 2008), where local patterns in the image are found, often corresponding to corners or sinusoidal gratings. These features can be robust to orientation, and to some extent also to affine transformations and changes in light intensity. After features have been extracted from an image, they are commonly stored using a visual bag-of-words (BoW) (Sivic and Zisserman, 2003), by mapping each feature to a codeword, chosen by the midpoints of k-means clusters in feature space. A place, can then be recognised by the histogram representing codewords that it contains. A major advantage of using point features is pose invariance. If the spatial relationship between the local features is ignored, a place can often be recognised despite significant changes in pose. An example localisation algorithm that successfully uses point features is FAB-MAP, which has been shown to perform loop-closure correctly across long routes in a number of outdoor environments (Cummins and Newman, 2008). However, methods that rely on point features typically have several shortcomings, including a lack of robustness to changes in condition, such as differences in lighting caused by the time of day, changing weather or seasonal changes such as the amount of foliage on trees.

Global descriptors

Comparisons have been made between the sinusoidal response characteristics of simple cells in the mammalian primary visual cortex (V1) and SIFT features (Serre et al., 2005). However, storing and matching up big lists of point features is almost certainly not the same solution to place recognition that we and other animals have evolved. Intuitively, we often remember and describe familiar scenes by their overall shape and higher level features, rather than a long list of small details. A plausible alternative approach are global features, where higher level shapes in the scene are also taken in to account. Attempts at global descriptors include capturing the distribution of colour in the scene (Ulrich and Nourbakhsh, 2000). Gist is another global descriptor method where images are convolved with Gabor filters at different scales and angles, which has been successfully used for localisation (Oliva and Torralba, 2001; Murillo and Kosecka, 2009). Milford and Wyeth (2012) achieved good condition invariant localisation by matching sequences of whole images at low resolution, where

local contrast is normalised to overcome problems due to lighting conditions. This approach proved to be very robust at matching images taken at the same location across extreme weather changes. Recently, McManus et al. (2014) attempted to solve condition variability by learning salient parts of the image using broad region detectors and training on the same route in different conditions. A bioinspired solution to this problem is suggested by Stone et al. (2014) in chapter 3.

Learnt features

Visual place recognition has many commonalities with visual object recognition, a field that has recently experienced a renaissance due to fast-paced developments in the uses of convolutional neural networks (CNNs) (LeCun et al., 1998; Krizhevsky et al., 2012). Instead of handcrafting features, a CNN essentially learns a hierarchy of salient features through being exposed to large sets of training data. This method has lately also been applied to localisation. Sünderhauf et al. (2015) used features to localise that had been pre-trained on object recognition tasks. Although this is computationally cheap and provides good performance, features that are optimised for cat or dog recognition tasks are almost certainly suboptimal for place recognition. Gomez-Ojeda et al. (2015) trained image-specific features using a triplet learning rule that is the current state of the art algorithm for facial recognition (Schroff et al., 2015). One of the major disadvantages of learnt features is the inability to know exactly what characteristics of the image are being used for recognition, although some studies have recently explored visualisation of higher level features in CNNs to try to understand them better (Zeiler and Fergus, 2014; Yosinski et al., 2015). Although this approach is still in its infancy, it has been making big inroads in many other areas of pattern recognition, meaning it will likely have a large impact on visual place recognition soon.

Choosing features

For visual localisation tasks, an important decision is which type of feature to use. Disadvantages of certain approaches include the susceptibility to changes in pose, which can be costly due to uneven terrain, or failures caused by changes in lighting due to weather, time of day and seasonal changes (Milford and Wyeth, 2012). Typically, features that have addressed one of these problems well have not addressed other issues. Point features tend to be robust to changes in pose but deal weakly with changes in condition Furgale and Barfoot (2010). Global features tend to deal well with changes in condition, yet are susceptible to viewpoint changes. An attempt to address this problem using a novel bioinspired approach (Stone et al., 2016a) is outlined in chapter 4.

Landmarks

A subset of visual place recognition is landmark recognition, where one or more prominent features of the environment are matched, allowing the current relative location to be inferred. Examples in day to day life include familiar features in the environment such as recognisable church towers or hills. These can be used to obtain a positional estimate through triangulation, or simply as a homing beacon. Recently localisation with prominent landmarks has been achieved through both point features from a large mapping database (Zheng et al., 2009) and by using CNNs to fingerprint buildings (Sünderhauf et al., 2015; Weyand et al., 2016). This section is mentioned for completeness, as landmark recognition in ultraviolet (UV) is one of the hypotheses that drove the experiments in chapter 3.

2.1.2 *Dead reckoning*

Dead reckoning, also know as path integration in animals, is carried out by starting from a known location and measuring displacement through odometry, using sensors to measure changes in course and speed. Speed can be measured in robots by rotary encoders on wheels or step counters for legged robots or by measuring acceleration using an inertial measurement unit (IMU). Displacement can also be estimated visually by extracting features in the environment and measuring optic flow (Nistér et al., 2004). Bearing also needs to be estimated, which can be done in a egocentric manner through optic flow or rotational inertial information using an IMU. An allocentric measure, e. g., by matching known features in environment or measuring the surrounding magnetic field is often better as not susceptible to drift.

Engineered solutions to localisation rely heavily on probabilistic approaches, where rather than keeping a single location estimate, a probability distribution over the map is propagated using, e.g., extended Kalman filters (EKFs) (Thrun et al., 2001; Julier and Uhlmann, 2004) or particle filters (Gustafsson et al., 2002). Bioinspired solutions to storing and updating the current location estimate include RatSLAM by Milford et al. (2004), which uses the mammalian place and grid cell system found in the hippocampus and entorhinal cortex (O'Keefe and Dostrovsky, 1971; Hafting et al., 2005) as inspiration.

2.2 LOCALISATION IN INSECTS

While the field of robotic navigation has gained a lot of interest over the past decades, the manner in which insects navigate has been studied for well over a century. Insects

have brains that are several orders of magnitude smaller than those of mammals, yet carry out complex tasks, making them a good candidate for investigating how a nervous system can drive navigation (Haberkern and Jayaraman, 2016). For robotics, they act as a proof of concept that expensive feature extraction or object recognition does not need to occur to navigate well.

Almost all of the robot localisation methods described in section 2.1 are employed by animals in some form. Animals can perform place recognition through olfaction (Ioalè et al., 1990), vision (Sec. 2.2.2), including the use of landmarks, and even echolocation (Ulanovsky and Moss, 2007). They are also known to path integrate (Sec. 2.2.3). Animals are very good at exploiting the salient features in their surroundings; e. g., to determine their orientation, different species are known to employ a variety of solutions ranging from magnetoreception (Kimchi et al., 2004) to orientation against the milky way (Dacke et al., 2013). All insects display navigational behaviour and although they use many sensory cues, one of the most prominent tools is vision. The following sections will focus exclusively on the visual navigation techniques used by these insects.

Two good visual navigators, that feature prominently in the subsequent chapters of this thesis, are the desert ant *Cataglyphis velox* and the sweat bee *Megalopta genalis* (Fig. 2.2).

(a) Compound eye of Megalopta genalis

(b) Cataglyphis velox

Figure 2.2: **Insects up close (a)** The compound eye of the sweat bee, *Megalopta genalis*, is made up of thousands of individual ommatidia. The part of the eye that is visible in this picture is ± 1 mm wide. Photograph by Bruno Garcia **(b)** The desert ant *Cataglyphis velox* lives in southern Spain in a shrubby environment. It has long legs, keeping it further from the hot ground, which also make it a fast runner Photograph by Michael Mangan and Hugh Pastoll

Desert ants are excellent navigators that include several genera, such as the honey carrying *Melophorus bagoti* and the various species of *Cataglyphis* (Fig. 2.2b). They are solitary scavengers that forage during the hottest part of the day, allowing them to feed on other dead insects that have succumbed to the heat (Cerdá et al., 1998). Surviving in this environment has caused them to adapt by developing navigation strategies without the use of chemical trails (Van Oudenhove et al., 2012), meaning they primarily use vision (Wehner, 2003). Ants can travel far, often managing to find their way back home from foraging runs 100 000 times their own body length, through terrain with no obvious landmarks (Wehner, 2003). *Cataglyphis* typically only live for around one week (Ziegler and Wehner, 1997), meaning that visual learning needs to happen quickly and much of their navigational toolkit needs to be innate. The navigation strategies of the desert ant have previously been modelled using both agents, e. g., route following in virtual environments by Baddeley et al. (2012) and robots, e. g., path integration, visual piloting and systematic search by Lambrinos et al. (2000).

Megalopta genalis are a species of bee that live in the tropical forests of Central and South America. They search for food during the morning twilight, making long-distance foraging and homing flights, and have evolved extra sensitive compound eyes (Greiner et al., 2004), using low-light vision to navigate in the forest (Warrant and Dacke, 2011).

2.2.1 *Sensory system*

Insects have compound eyes, made up of many photoreceptors known as ommatidia (Fig. 2.2a). The number of ommatidia can highly vary depending on the species. *Cataglyphis* have around 1000 per eye (Menzel and Wehner, 1970), while *Megalopta*, a nocturnal bee, has just under 5000 in total (Greiner et al., 2004). In comparison to mammals, insect visual acuity is relatively low, making their ability to visually navigate quite remarkable. Insects typically have a much larger field of view (FOV) than mammals, with well studied visually navigating ants having a visual field of around 300° (Schwarz et al., 2011; Zollikofer et al., 1995). Although there is evidence for stereopsis in insect prey tracking (Nityananda et al., 2016), most work assumes place recognition can be achieved without needing depth information (Wehner et al., 1996).

The spectral sensitivity of insect eyes is also different to that of humans, peaking in the green, blue and UV regions (Menzel et al., 1986). The specificity of preferred wavelengths differs across species, indicating that insect vision has evolved to deal with particular environments (Briscoe and Chittka, 2001). The *Cataglyphis* species, desert ants known for their visual navigation abilities, are believed to be dichromatic,

seeing in UV and green (Mote and Wehner, 1980; Labhart, 1986), however recent work suggests some species also posses a third receptor (Ogawa et al., 2015). Möller (2002) suggests that UV and green vision could be helpful for landmark identification. An example of where this applies in nature is the high UV reflectivity of certain flowers, enabling near optimal detection by bee photoreceptors (Chittka et al., 1994). A good way to test this hypothesis would be to visualise the environment using cameras that are sensitive to the same frequencies as the ant eye. Chapter 3 describes the construction of a multispectral camera to achieve this. Besides different preferred frequencies, many insects also have ommatidia in the dorsal rim area (DRA) of their compound eyes that are sensitive to light at certain polarisation angles (Wehner, 1976), allowing them to be used as a solar compass (Sec. 2.2.3).

In addition to compound eyes, most insects also have three dorsal ocelli, which are light-sensitive simple eyes, each made up of a single lens element and a number of photoreceptors. The ocelli are believed to play a role in steering behaviour (Stange and Howard, 1979; Taylor, 1981) and recent work has suggested they are also able to provide form vision in some insects (Berry et al., 2007). Ocelli have provided inspiration for biomimetic solutions to attitude stabilisation on unmanned aerial vehicles (UAVs) (Stange et al., 2007; Fuller et al., 2014).

2.2.2 *Visual place recognition*

It is unclear whether insects are able to form the concept of a known place, in the sense of storing visual features linked to a spatial representation. Regardless, it would still be useful for them to extract visual features from the environment and associate actions with similar features, and substantial evidence exists that insects do this. Unlike robotics, where the details of each place recognition algorithm are known, for insect studies these must be inferred through observing and categorising navigational behaviour. Several types of visual navigation strategies used by insects are outlined in the following sections.

Visual homing

Insects are known to home towards their nest using vision. Sometimes, this could occur by using a recognised landmark, like a tree, as a beacon (Collett, 1996). The importance of landmarks for navigation is what lead to the study in chapter 3, where the ant environment was viewed in the same visual spectrum as the peak sensitivity of the ant ommatidia to potentially reveal hidden landmarks that they use. Cartwright and Collett (1983) suggested that insects are able to home towards known locations by taking retinotopic snapshots and gradually reducing the mismatch between the

whole image of the view at their current location and their target. Behaviours that support this are the learning walks and flights that insects make around their nest, which could be used to capture snapshots of the nest from a number of vantage points (Müller and Wehner, 2010).

Visual homing in this way has been modelled using panoramic images and a pixelwise mean squared difference function to determine the basin of attraction (Franz et al., 1998), also known as the catchment area, in natural environments. Zeil et al. (2003) were able to estimate the catchment area resulting from snapshots taken at a target location in an outdoor environment, showing that a modest area around the nest could in theory lead to successful homing. Möller (2009, 2012) used a predictive model for ant homing, similar to the robotic algorithm developed by Franz et al. (1998), where a panoramic image is warped to predict what it would look like after displacement in a certain direction. A significant shortcoming of this model is that images need to be retinotopically aligned. This can be somewhat mitigated by storing all images relative to an external frame of reference, e. g., heading north (however, for an example failure case see Fig. 4.6). An alternative solution to alignment, known as the visual compass (Labrosse et al., 2006), is to compare an image at different yaw rotations using a simple pixel-wise difference function to find the best heading, exploiting the autocorrelation between pixels in natural scenes.

Finally, Ardin et al. (2015, 2016b) showed that ants are able to home, despite their views not being retinotopically aligned both due to pitch variation from uneven terrain or through backwards motion when dragging food to a nest. A proof of principle for successful rotation-invariant homing is introduced in chapter 4, section 4.3.

Route following

In cluttered environments, desert ants maintain idiosyncratic routes on their homeward path (Wehner et al., 1996; Wehner, 2003; Mangan and Webb, 2012), with primarily visual information about the path used to follow the same route (Kohler and Wehner, 2005). The use of vision has been tested by placing zero-vector ants that have just returned to the nest back at their feeding site, so they are unable to rely on path integration (Sec. 2.2.3). Route following behaviour could function by aiming at a series of beacons along the way, or by taking a series of snapshots and associating a preferred direction with each visual memory (Collett, 1996).

Image features

Recent computational models have shown that the entire image does not need to be stored for an agent to retrace a known route. Futhermore, image features corresponding to discrete locations need not be stored to carry out route following (Wystrach

et al., 2013). It is sufficient for the agent to simply remember typical features of the route that correspond to actions. An implementation is the model by Baddeley et al. (2012), which uses an artificial neural network (ANN) with an Infomax (Linsker, 1988) learning rule to learn scene familiarity and can carry out route following in simulated ant environments. The model essentially works by estimating the likelihood that the current view was drawn from a distribution that describes the route images. A more recent model proposes that the insect mushroom bodies (MB) are able to store and subsequently recognise the co-occurrence of activity from photoreceptor input, driving steering behaviour (Ardin et al., 2016a), effectively a form of feature detection.

The use of skylines for scene recognition by navigating insects was first proposed by Fukushi (2001) in a study where wood ants, *Formica japonica*, were shown to alter their homeward path when portions of the visual panorama were covered. Towne and Moscrip (2008) observed how honeybees behaved when displaced between two sites with similar global features and suggested that the skyline may be the preferred reference when navigating. By constructing an artificial skyline around a feeder, Graham and Cheng (2009a) showed that ants could be using features of the skyline to navigate (Fig. 2.3). A complementary study, where certain parts of the visual field of the ant were obscured, supported the argument that skyline contours provide a high amount of information for view-based navigation (Graham and Cheng, 2009b). Philippides et al. (2011) extracted skylines from images taken in the ant environment, and confirmed that skylines can work well when used as input for both visual compass and visual homing algorithms described in the previous section. Möller (2002) suggests that, in fact, the sensory system of ants may be optimised for skyline extraction through UV-green colour opponency. These findings suggest that the skyline contour is a stable cue for localising, and could perhaps be exploited by robots, a conclusion which is similar to independent findings described in chapter 3 and which led to two studies exploiting this feature for robotics applications (Ch. 3 and 4).

2.2.3 *Path integration*

Path integration in animals is equivalent to dead reckoning (Sec. 2.1.2) in robotics. Some animals are able to keep track of where they have travelled by maintaining a home vector, pointing back to a reference location, which could be a nest or other salient location such as a feeding site or landmark. This ability was first suggested by Darwin (1873) and has since been observed in many birds (Mittelstaedt and Mittelstaedt, 1980), mammals (Etienne and Jeffery, 2004) and several insect species, including bees and ants (Pieron, 1904; Collett and Collett, 2000a; Wehner, 2003).

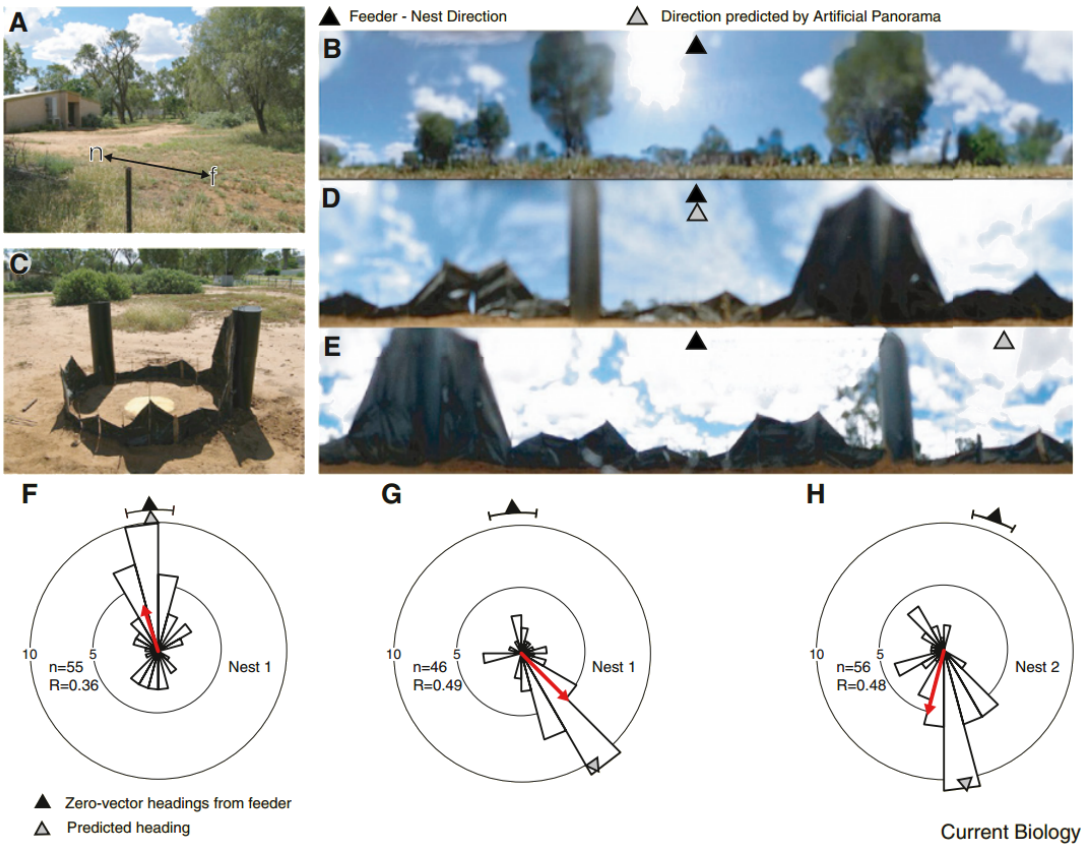
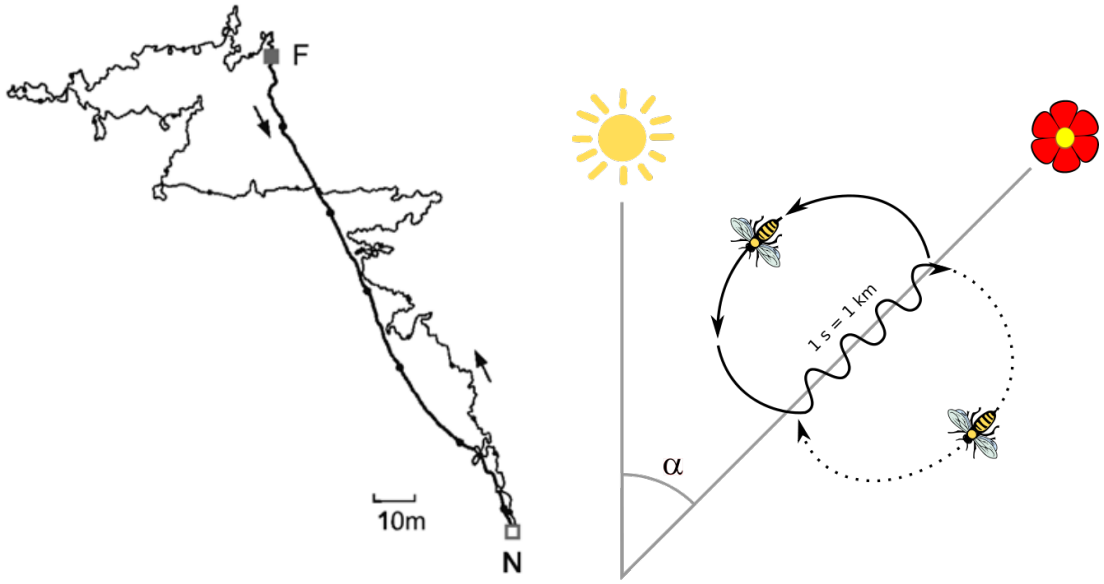


Figure 2.3: **Evidence that ants navigate using the skyline. (a-c)** Graham and Cheng (2009a) created an artificial panorama based on the natural skyline around a feeding site near an ant nest. In a displacement study, ants were caught and subsequently released inside the artificial skyline arena. **(d-e)** The arena could be manually rotated, **(f-h)** and the angle of rotation strongly predicted the heading direction of the ants, as they tended to home towards outlines matching the shape of natural features in the direction of the nest. These results suggest that the skyline provides a strong cue for visual navigation in certain species of ant. Figure from Graham and Cheng (2009a).

An example of behaviour demonstrating path integration is shown in figure 2.4a, where a desert ant forages on a highly tortuous outbound route, then returns in a straight line towards the nest. It is remarkable that accurate homing can occur despite the exploratory paths by desert ants often exceeding 100 m, which is tens of thousands of times their body length (Wehner, 2003). Displacement studies, where an ant with a built-up home vector is moved to a new location, can accurately predict where the ant will home to, demonstrating that the ant is not visually homing (Collett and Collett, 2000b). Other evidence for path integration comes from the decoding of bee waggle dances (Von Frisch, 1967; Kirchner and Braun, 1994), a method used by bees to communicate good sources of food through polar coordinates (Fig. 2.4b).



(a) Path integration (b) Waggle dance

Figure 2.4: **Evidence of path integration. (a)** After foraging for food (F) on a tortuous outbound route (thin line), the desert ant *Cataglyphis fortis* heads back to the nest (N) in a relatively straight path (thick line). Figure from Wehner (2003). **(b)** Bees perform a waggle dance, using a figure-of-eight pattern to communicate where a good source of food can be found to other members of the hive. The direction and length of the waggling section indicate the distance and angle of the food source with respect to the direction of the sun. This provides a strong clue about the internal representation of their path integration system. Figure by Emmanuel Boutet, Creative Commons

To path integrate successfully, a compass, an odometer and some form of memory are needed. The insect implementation of these three features is covered in the following sections.

Compass

Insects have access to an allocentric celestial compass, which is hugely beneficial for path integration, as it prevents the build-up of error through drift when accumulating

angular motion. The compass works by determining heading with respect to the sun (Santschi, 1911). This can be done by measuring the spectral input from different directions, indicating the position of the sun directly (Dacke et al., 2014). However, insects can also use the polarisation patterns across the sky to infer where the sun is (Wehner, 1997). During the daytime the sky has a particular polarisation pattern, described by the Rayleigh polarisation model (Coulson, 1988) (Fig. 2.5). Insects can

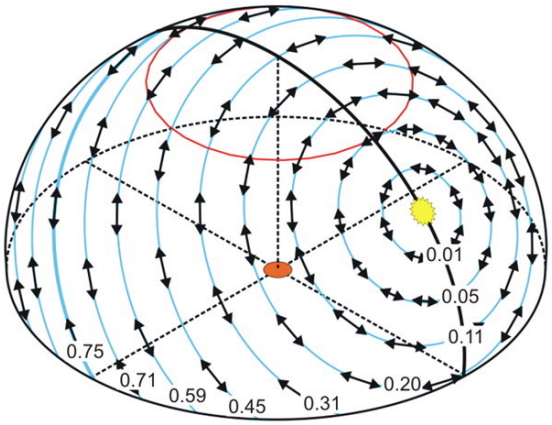


Figure 2.5: **Rayleigh sky model.** Polarisation patterns across the sky depend on the position of the sun. The degree and direction of polarisation can be expressed as e-vectors, with a particular direction and magnitude. When mapped, e-vectors form concentric circles around the sun. By observing the zenith (area inside red circle), the direction of the sun can be inferred by perpendicular e-vectors. Image taken from Homberg et al. (2011).

use the pattern of polarisation in the sky around the zenith as an allocentric frame of reference, as experiments have demonstrated by rotating linear polarisers above the insect. Besides the sun, insects have also been shown to use the moon (Dacke et al., 2003) and even the Milky Way (Dacke et al., 2013) for directional navigation. A bioinspired compass using polarised light has also been implemented on robots as proof of principle by Lambrinos et al. (1997).

Odometry

Bees use optic flow to judge speed (Srinivasan, 2014) while walking (Schöne, 1996) and flying (Srinivasan and Gregory, 1992). It is possible to verify this through tunnel experiments (Srinivasan et al., 1996). Bees are tricked into misjudging the distance they have travelled on outward foraging routes by forcing them to fly through tunnels with moving bars projected on the walls. Visual odometry has also been found in species of desert ant (Ronacher and Wehner, 1995).

There is evidence that ants can count their steps to estimate distance. By artificially lengthening or shortening the legs of an ant in a homing state the under- or

overshoot with respect to their nest can be predicted (Wittlinger et al., 2006). Other proposed methods of odometry include energy expenditure, however evidence for this is weaker and more susceptible to miscalculations through changes in wind direction or a change of mass due to carried food.

Representation

Perhaps the least known aspect of path integration is how insects store their accumulated path. The cellular mechanism for accumulating distance could be implemented in a variety of manners, including chemical changes within cells, synaptic strengthening or persistent neural activity. It is possible that changes in location occur too quickly for path integration to depend on synaptic plasticity. An example of a chemical mechanism that could potentially adapt to faster changes is integration by calcium build-up (Loewenstein and Sompolinsky, 2003). Alternatively, some form of recurrent connections, resulting in persistent activity could preserve a home vector.

The waggle dance used by bees to communicate the location of a food source provides a clue that insects could store their position using polar coordinates; however, the dance could simply be a translation from a different internal representation. A major obstacle to discovering how insects represent home vectors internally is knowing where inside the brain to look. In mammals, the discovery of grid cells (McNaughton et al., 2006) has sparked many theoretical models explaining how space is represented Moser et al. (2014). However, it is unlikely that a similar system exists inside the insect brain, simply due to the comparatively small amount of neural resources. The model outlined in chapter 5 proposes a candidate group of cells, the CPU4 cells in the CX as a probable area for memory accumulation in insects.

Systematic search

A commonly observed navigational behaviour, found in several insect types, is a systematic search, where the insect will loop outwards, gradually exploring more terrain from a central point (Wehner and Srinivasan, 1981; Müller and Wehner, 1994). A search is one of the simplest navigational strategies possible, yet provides a very useful tool for the insect.

Although systematic searching is not intrinsic to path integration, or a visually driven behaviour, it is worth mentioning, as a search can indicate where an insect believes their goal is. Searching is often encountered in displacement experiments, where the insect has run down its path integrator or has visually homed to a location and searches around the nest. For the insect, it can serve as an important strategy to deal with errors during integration or displacement by external factors such as the wind. The search area can increase depending on the uncertainty of the location

estimate (Merkle et al., 2006a). Interestingly, the path integration model described in chapter 5 exhibits searching behaviour automatically due to the manner in which the memory and steering interact.

2.2.4 *Neural underpinnings of visual navigation*

The neural substrate for navigation has historically been a popular and important topic for neuroscientists, with key findings leading to the 2014 Nobel prize in Physiology or Medicine for the discovery of grid and place cells in mammals by i.a. O'Keefe and Dostrovsky (1971); Hafting et al. (2005). Neural models now exist to explain a wide range of navigational behaviours, however, the focus in this section will be on models that closely relate to this body of work. Mammalian models are also briefly introduced, to provide the reader with some perspective on how theory differs between the two related fields.

Place recognition

Not much is known about the neural underpinnings of visual place recognition in mammals. Simple and complex cells in V1 are used to detect edge-like patterns (Hubel and Wiesel, 1968). A hierarchical system, where gradually more complicated patterns are detected forms the basis of most models of recognition in the visual cortex (Poggio and Serre, 2013).

The synaptic organisation in part of the MB of visually navigating desert ants (Kühn-Bühlmann and Wehner, 2006; Stieb et al., 2010) changes as they age and are exposed to visual input. Visual place recognition could be occurring there, a hypothesis put forward by Ardin et al. (2016a) by drawing parallels to the MB and an associative memory network by Willshaw et al. (1969).

Head direction

Head direction cells have been identified in mammals that keep an allocentric representation of orientation (Taube et al., 1990). A simple model of how these cells interact, that also makes good predictions, is the ring attractor theory, proposed by Zhang (1996) (Fig. 2.6). Cells are topologically connected in a ring, in order of direction preference. A single bump of self-exciting activity inhibits neurons that are topologically further away, acting as a winner-takes-all representation of heading. This activity bump can be shifted around the ring by external visual cues or self-motion input. Similar ring attractor models provide explanations for other functions in the mammalian brain, including the organisation of orientation columns in V1

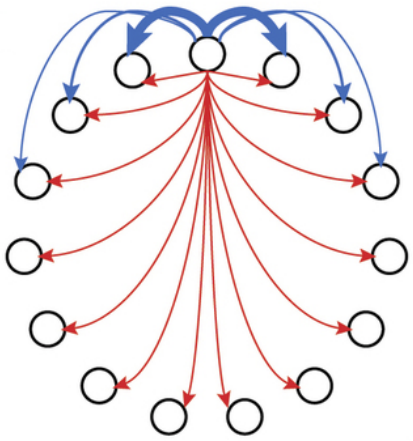


Figure 2.6: **Ring attractor.** Head direction cells of a rat can be modelled as a ring attractor. Cells are topologically arranged in a ring, in order of direction preference. Cells excite neighbours (blue) and inhibit all other cells (red), causing a stable activity bump to form. The bump acts as a winner-takes-all representation of direction. Figure from Schultheiss and Redish (2015).

(Ben-Yishai et al., 1995) or as a mechanism for oculomotor control (Seung, 1996). A similar ring structure that also represents heading is identified in chapter 5, through the mutually inhibitory connections of TB1 cells in the insect CX .

Path integration

In mammals, place cells have been identified that fire when the animal is in a certain location. Also, grid cells that fire in a regular hexagonal spatial pattern are believed to encode a spatial position. More complex models, often also based on attractor networks, have attempted to explain how mammalian grid and place cells could be giving rise to path integration (Samsonovich and McNaughton, 1997; McNaughton et al., 2006; Moser et al., 2008; Burak and Fiete, 2009).

Despite the advances mentioned above, the size and complexity of the mammalian brain has made it very challenging to fully understand the underlying neural mechanisms of navigation. In contrast to the mammalian brain, insect brains are simpler and increasingly possible to measure in high detail. Due to recent breakthroughs in imaging and genetics, cells can be individually labelled by techniques such as multicolor flip-out (MCFO) (Nern et al., 2015) and cellular activity can be observed during behaviour by e.g. two-photon Calcium imaging (Seelig et al., 2010). Historically however, the detail and realism of insect navigational models has lagged behind mammalian ones.

For insects, Hartmann and Wehner (1995) hypothesised an attractor network along a neural chain could be used to represent angular and distance measures. A different approach, the bicomponent model by Mittelstaedt (1962), suggests a way for insects

to steer using the sine and cosine of the angles relative to the sun and their desired heading, which when projected onto each other could output a steering signal indicating whether to turn left or right. Using evolutionary algorithms, Vickerstaff and Di Paolo (2005) created a leaky integrator circuit that was able to perform path integration using similar methods for steering. Wittmann and Schwegler (1995) describe a hypothetical neural circuit, where a ring of attached neurons forms an array, with each element making up part of a discrete sinusoid, where the phase indicates the direction and the amplitude the distance of the home vector. Haferlach et al. (2007) implemented a reduced version of this model with arrays of only three abstract neurons and neural properties determined by an evolutionary algorithm, a coarse coding approach that was similarly used by Kim and Lee (2011).

An area that modellers have historically disagreed on is the type of coordinate system used to store the animals position. Some researchers advocate egocentric models (Benhamou et al., 1990; Merkle et al., 2006b), where the nest is stored in coordinates relative to the heading of the animal, whereas other have favoured allocentric models (Mittelstaedt, 1962; Müller and Wehner, 1988), where the current position of the animal is stored relative to some external reference. Likewise, the difference between polar and Cartesian coordinates has been debated. Theoretical reasons exist to prefer certain representations over others, including advantages due to additive qualities of coordinate systems, and robustness against drift when making noisy motion estimates. In particular, geocentric Cartesian coordinate systems have the advantage of being additive, while remaining less susceptible to drift. For a review on the advantages and disadvantages of different coordinate systems see Vickerstaff and Cheung (2010).

Maurer and Séguinot (1995) argue that useful path integration models must be based on neurological knowledge rather than mathematical theory. Although this statement was made over twenty years ago, it remains relevant today and one of the driving factors for the model developed in chapter 5 of this thesis. Most insect models of path integration focus on theory or behaviour rather than neural activity or connections. The CX model described in chapter 5 is an attempt to address this problem. In insects, the CX is an area of the brain which is highly implicated in a wide range of navigational functions, including path integration and is discussed in more detail in the following section.

2.3 CENTRAL COMPLEX

The CX is a conglomerate of intricately structured brain components in the insect midbrain. It is believed to be at the centre of many navigational tasks, including orientation relative to celestial and other visual cues.

2.3.1 *Anatomy*

The CX comprises of four main bodies of neuropil: the protocerebral bridge (PB), upper division of the central body (CBU), lower division of the central body (CBL) and noduli (NO) (Fig. 2.7). Due to the large number of research groups that inde-

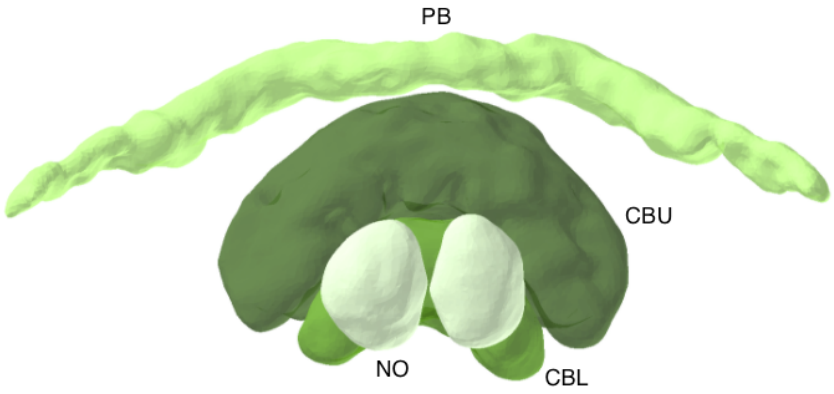


Figure 2.7: **Central complex anatomy.** The CX of the sweat bee *Megaloptis genalis*. The four major bodies of neuropil are the protocerebral bridge (PB), upper division of the central body (CBU) also known as the fan-shaped body, lower division of the central body (CBL) also known as the elipsoid body and the two noduli (NO) Image generated on the Insect Brain Database website (Heinze, 2015).

pendently study the insect brain, published works often refer to anatomical regions using different names. In the CX, most studies on locusts, moths, beetles, butterflies use a single term, while those on the fruit fly, *Drosophila melanogaster* use another. To avoid confusion, the corresponding names are shown in table 2.1.

Drosophila	**Other insects**
Protocerebral bridge (PB)	
Fan-shaped body (FB)	Upper central body (CBU)
Ellipsoid body (EB)	Lower central body (CBL)
Noduli (NO)	

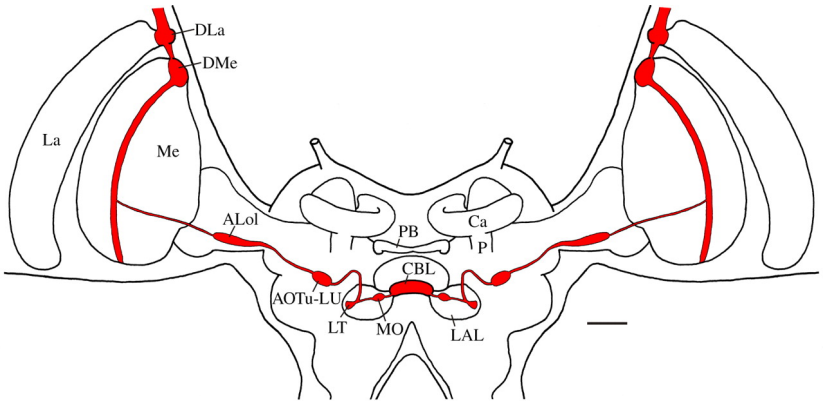
Table 2.1: Nomenclature of central complex anatomy used by different labs.

Cells have been identified at the individual level by intracellular dyeing (Heinze and Homberg, 2008) or, more recently, by genetic labelling (Wolff et al., 2015), providing a detailed picture of the CX neuroarchitecture. The CX is made up of sets of distinctive columns, with fibres of cells that project in very specific and targeted ways between different layers of the bodies. Individual cells have been matched across animals and species, indicating the connectivity pattern between individual neurons is important. Recording from cells or observing them during behaviour has shed some light on their functionality.

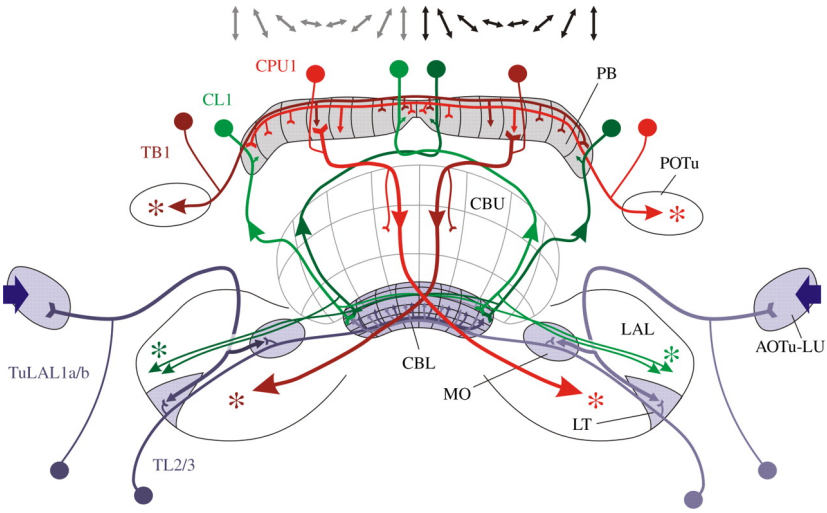
2.3.2 *Possible roles*

Polarisation-sensitive (POL) neurons have been found in the brain of many insects. Typically, they are identified by recording from individual cells, while rotating a linear polariser above the animal and looking for an oscillatory response in the firing rate, indicating a preferred e-vector tuning. Identified POL cells form a pathway from the DRA of the insect compound eyes all the way to the CX in their midbrain (Heinze and Homberg, 2009) (Fig 2.8a). POL cells enter the CBL via the lateral accessory lobes (LALs), before propagating POL information via CL1 cells to the PB (Fig. 2.8b). Heinze and Homberg (2007) showed by recording from cells in the PB that there is a map-like representation of polarised light e-vector orientations, with an ordered array formed across the columns. Neuser et al. (2008) demonstrated that disrupting neurons in the elipsoid body (EB) causes *Drosophila melanogaster* to fail at tasks where visual features are used to orient. Likewise, Ofstad et al. (2011) found that disrupting EB neurons also prevents them from learning locations visually. It is highly plausible that the CX could be acting as a compass, integrating an array of senses to provide a stable estimate of the insect's heading. Experiments have shown that neurons between the CBL and PB are not just polarisation sensitive, but capture directional information with respect to other external light sources or landmarks (Seelig and Jayaraman, 2013, 2015; el Jundi et al., 2015). Varga and Ritzmann (2016) show that several cells in the CX encode head direction through both allocentric visual cues and cumulative angular motion.

Lesioning and electrically stimulating areas of the cockroach CX affects walking and turning behaviour (Harley and Ritzmann, 2010; Bender et al., 2010), suggesting it also has a strong links to motor output. Newly identified neurons, described in Stone et al. (2016b), show that the NO also contain cells that are linearly sensitive to forward and backward optic flow. This suggests that the CX may also be acting as a visual odometer. Pfeiffer and Homberg (2014) provides a comprehensive review on functional roles of the CX.



(a) Polarisation pathway through the insect brain



(b) POL neurons in the CX

Figure 2.8: **Polarisation pathways in the insect brain. (a)** Polarisation-sensitive (POL) neurons form a pathway from the compound eyes, all the way to the CX. **(b)** Information is propagated via CL1 neurons (green) to the PB, forming an ordered array of e-vector preferences (grey arrows) and a candidate neural substrate of the insect celestial compass. Figures from Homberg et al. (2011).

Due to the existence of two groups of cells that act as compass and visual odometer, the close links with motor output, and the highly conserved nature of the CX across species, it is a plausible candidate for the neural substrate of various navigation algorithms and could be critical for path integration, as previously suggested by Homberg et al. (2003).

2.3.3 *Modelling the central complex*

Although the CX is highly implicated in navigational behaviours, very few models exist to explain how it could be giving rise to them. Previous models have mostly focused on the functionality and implemented the model from a top-down approach, rather than looking at neural connectivity. Strauss and Berg (2010) proposed a model based on *Drosophila melanogaster* where visual targets are represented as activity on the PB, with connections to the pre-motor areas affecting step length, causing the agent to home in on a target. Arena et al. (2010) created a computational architecture for the *Drosophila* brain, where the CX is implemented as a network of spiking neurons. In their model the fan-shaped body (FB) is implemented as an abstract group of neurons that perform feature detection, extraction and storage.

Significant weaknesses of these models include their lack of grounding in realistic neural connectivity, but also their inability to test hypotheses about how the animal neural circuitry might operate or make predictions about real behaviour. More recent and realistic models include an integrate-and-fire model, acting as proof of principle that interactions between EB and PB in *Drosophila melanogaster* can function as a ring attractor (Kakaria and de Bivort, 2016). These findings are similar to those described in chapter 5, where the TB1 neurons in the PB act as a ring attractor, matching the theoretical model by Zhang (1996). Finally, a promising neurally-grounded explanation is being developed to explain how the activity bump shifts around the EB using the interaction between overlapping wedge and tile neurons (Wolff et al., 2015) and imbalances along the PB (Jayaraman, unpublished).

Electrophysiological and anatomical data on the CX are being generated at a phenomenal rate, providing an increasingly clear view of the function of individual cells. However, despite the high interest in this area from a behavioural point of view, current computational models are lacking in both quantity and quality. To understand how the CX is performing navigational tasks, it is crucial to form hypotheses about how known neurons could be interacting. This is somewhat addressed in chapter 5 by a model of the CX that proposes a circuit for path integration (Sec. 2.2.3) based on inferred connectivity of individual neurons in the CX.

Part I

ANTS

3

SKY SEGMENTATION WITH ULTRAVIOLET IMAGES CAN BE USED FOR NAVIGATION

3.1 INTRODUCTION

A good way to learn about how animals visually navigate is to improve our understanding of their sensory input. In the summer of 2013 an investigation was carried out to photograph the natural habitat of the Spanish desert ant *Cataglyphis velox* using a camera that mimics the spectral response of their compound eyes. In particular, the motivation was to look for landmarks that could be invisible to human eyes. To achieve this a consumer digital single-lens reflex camera (DSLR) was modified to capture images in both ultraviolet (UV) and green wavelengths as well as regular photographs. Example images taken with bandpass filters peaking at all three wavelengths are shown in figure 3.1.

After taking pictures of the environment, no ground-based features were observed that clearly stood out from the rest of the natural environment. In contrast, the ground appeared mostly dark due to low UV reflectivity and the sky homogeneously bright because of the high rate of UV light that passes through clouds (Calbo et al., 2005). These observations lead to the hypothesis that ants could be using the skyline contours to navigate, as claimed previously by Fukushi (2001); Graham and Cheng (2009a). Möller (2002) suggested that this could be achieved by using colour opponency from UV and green photoreceptors in the insect eye.

One method of confirming that skyline contours are indeed good enough for localisation is to demonstrate that robots or agents can successfully use them. A natural way to test this was to use the UV-extracted skyline features as input to an established visual localisation algorithm. SeqSLAM by Milford and Wyeth (2012), had previously been shown to work well for localisation using global features, such as low-resolution whole greyscale images. It incorporates temporal information by taking sequences, rather than just individual images, into account. Additional properties of UV-extracted skylines that can be important during navigation, such as robustness to different lighting conditions, were also investigated. The skyline was found to be a good feature to use for localisation, and robust to changes in lighting. Results of localising using UV-extracted skylines are described in the publication featured in this chapter.

(a) Visible light

(b) UV

(c) Green

Figure 3.1: **The desert ant environment in their preferred visual spectra.** Partial panorama of the natural environment of *Cataglyphis velox* in Seville, southern Spain. Photographs were taken using a converted DSLR with three different filter configurations to capture images in **(a)** light visible to the human eye **(b)** UV and **(c)** green.

3.2 PUBLICATION: ROBOTICS SCIENCE AND SYSTEMS 2014

This paper was published in the proceedings of Robotics: Science and Systems 2014 conference (Stone et al., 2014). Personal contributions include:

- Aiding Paul Ardin with design and construction of UV camera.
- Photography of ant and human environments in UV, which included work that was done as part of the authors Masters by Research degree.
- Conceptualisation of a UV-passing quartz tube combined with a mirror and camera to allow panoramic photographs to be taken by a DSLR in UV (during PhD).
- Full experimental design of all data collection tasks and analysis.
- All data collection for both route-based and stationary experiments.
- Full data analysis of all experiments.
- Re-implementation and expansion of localisation algorithm SeqSLAM in Python.
- Significant contribution to the manuscript and content, including creation of the main body of text and all figures.

Sky segmentation with ultraviolet images can be used for navigation

Thomas Stone *
School of Informatics
University of Edinburgh
Edinburgh, EH8 9LE, UK
t.j.stone@sms.ed.ac.uk

Michael Mangan *
School of Informatics
University of Edinburgh
Edinburgh, EH8 9LE, UK
mmangan@staffmail.ed.ac.uk

Paul Ardin
School of Informatics
University of Edinburgh
Edinburgh, EH8 9LE, UK
s8528002@sms.ed.ac.uk

Barbara Webb
School of Informatics
University of Edinburgh
Edinburgh, EH8 9LE, UK
bwebb@inf.ed.ac.uk

***Abstract*—Inspired by ant navigation, we explore a method for sky segmentation using ultraviolet (UV) light. A standard camera is adapted to allow collection of outdoor images containing light in the visible range, in UV only and in green only. Automatic segmentation of the sky region using UV only is significantly more accurate and far more consistent than visible wavelengths over a wide range of locations, times and weather conditions, and can be accomplished with a very low complexity algorithm. We apply this method to obtain compact binary (sky vs non-sky) images from panoramic UV images taken along a 2km route in an urban environment. Using either sequence SLAM or a visual compass on these images produces reliable localisation and orientation on a subsequent traversal of the route under different weather conditions.**

I. INTRODUCTION

The recent success of autonomous robotic vehicles [26] owes much to the availability of specific sensor technologies (LIDAR scanners, IMU, HD cameras, RADAR, GPS) and the continued development of computational resources capable of processing the data they generate. However, for the wider deployment of robots, particularly into price and energy critical markets, it is essential to bring down the hardware, computational, and energy costs of navigation. With this goal in mind, we take our inspiration from desert ants, which provide proof of principle that low-power and low-computation methods for navigation in natural environments are possible. Desert ants forage individually and without the use of chemical trails, yet can efficiently relocate their nest or a food source over long distances using visual cues alone [29]. The ant's compound eye provides very coarse resolution (on the order of 4° per pixel), and thus is unlikely to be capable of extracting and matching many distinct features, a capacity which is fundamental to many current visual Simultaneous localisation and mapping (SLAM) methods (e.g. [5]). Rather, their visual system appears fine-tuned to extract a key global feature – the panoramic skyline [16] – which behavioural observations have shown is sufficient for guidance in their complex visual habitat [7].

It has also been demonstrated that panoramic skylines are sufficient for robot localisation in urban environments [21, 14, 18]. The skyline (or more precisely, the horizon) has also been investigated as a visual cue for stabilisation in UAVs [19]. However, reliable visual segmentation of sky from terrain using conventional visual wavelength imaging appears difficult [18] particularly in changing weather conditions. Image parsing methods, e.g. [25], can be used to label sky amongst other regions, but the most successful current methods depend on advanced machine learning methods over large numbers of sample images to estimate the characteristics of sky. A range of approaches specific to detecting the sky region only are summarised in [20]. These include combining colour and texture characteristics, modelling the colour gradient, using connectedness or picture region constraints, and using alternative colour spaces and classification algorithms such as support vector machines. Shen and Wang [20] propose a greyscale method that uses gradient information and energy function maximisation, combined with several picture region assumptions, to obtain results on the order of 95-96% correct classification.

Far-infrared (IR) has been used to extract skylines, even after sundown [14], but is not robust for sky segmentation across weather conditions [4]. Clouds and sky appear very different in near-IR images (see supplementary material); indeed IR is often used explicitly for cloud detection [8].

By contrast, desert ants show sensitivity solely in the ultraviolet (UV) ($\sim 350nm$) and green ($\sim 510nm$) wavelengths [17, 11]. The effectiveness of these two wavelengths for sky detection was investigated in [16] using a custom sensor with one photodiode sensor for each. The results suggested using UV and green as opponent channels would be the most effective way to segment the sky under dynamic illumination conditions. In a follow up study [9] using five different wavelengths it was found that UV vs IR performed slightly better. However, because UV-imaging is limited by the filtering in standard camera optics, it has received little further investigation in robotics. In [3] a UV-receiving camera was paired with an omnidirectional mirror, but used to assess the directional information available in the polarisation of UV-light rather than to detect sky per se (sky segmentation by measuring the degree of polarisation of light is suggested in [19]). To our knowledge, there is only one recent investigation [23, 24] involving use of a UV-passing filter with a camera to enhance the difference between sky and ground regions, in the context of horizon determination for stabilisation. Neither the accuracy or consistency of UV vs visible wavelength imaging for detecting sky was reported in this work. UV has occasionally been used on robots for other purposes, including

* First author

object marking [10] and identification [28].

For proof of concept, we investigated the utility of UV vs visible light cues for sky segmentation using an adapted digital single-lens reflex (DSLR) camera system allowing wide spectrum imaging in high resolution. We demonstrate a simple thresholding algorithm applied to UV-only images is sufficient for accurate (compared to ground truth) and repeatable (comparing time-lapsed images across weather conditions) segmentation, and that this can be used to create panoramic segmented sky images that support navigation in an urban environment. The key features could be rapidly realised using off-the-shelf components (e.g. CCD without colour filters) providing a new class of low-cost outdoor navigation system.

II. Methodology

The aims of this study were: first, to assess the relative accuracy of sky segmentation using different wavelengths, over a range of different locations and conditions; second to assess the consistency of segmentation for UV vs visible wavelengths in the same location but at different times of day and weather conditions; and third to test the effectiveness of the segmented sky shape as a cue for robot navigation. Here we describe in detail the imaging device, the data sets collected for each test, and the segmentation and navigation methods used.

A. Full Spectrum Imaging

To sample images in both UV and visible spectra the internal cut filter of a standard DSLR camera (Nikon D600) was replaced by a fused quartz window (Advanced Camera Services Ltd, UK), and fitted with a high UV transmission wide angle lens (Novoflex Noflexar 3.5/35mm). Images could then be recorded with different filters: Thor Labs FGB37 bandpass filter (335 - 610 nm) to sample only visible light; Baader U-Filter (350 nm, half-width 50 nm) to approximate the ant's UV photoreceptor; and Knight Optical 520FCS5050 bandpass filter (515nm) to approximate the ant's green photoreceptor profile. The camera aperture was set to $f/16$, sensor sensitivity (ISO) to 200 and for each image an appropriate shutter speed was automatically selected to provide proper exposure given the chosen aperture. All images were saved using the uncompressed 14 bit Nikon Electronic Format (NEF) file format and subsequently demosaiced to a true colour 16 bit Tagged Image File Format (TIFF). Visible light images were stored in RGB format, the CMOS red channel only used for UV images, and a greyscale conversion of all three channels for green filtered images. To mimic the bichromaticity of the ant visual system, UV-green (UV-G) images were created by combining separate UV and Green intensity images into a 2 dimensional false colour space. All images were then scaled down by a factor of 10 along each dimension to give a resolution of 604×403 pixels prior to analysis.

B. Diverse Image Database

In July 2013, we took around 750 photographs in a variety of locations, including the natural habitat of desert ants (Seville, Spain) and wood ants (Sussex, UK) and a series of urban settings. UV-passing, green-passing and visible light-passing filters were used in turn. We selected a subset of 18 images representing a wide diversity of environments and conditions (see table I for example images) for which ground-truth sky segmentation was established. These included images expected to be difficult to segment due to surfaces reflecting the sky, such as water and windows. We used semi-automated labelling of sky and ground pixels through iterative applications of the GrowCut algorithm with user-correction [27]. We note that it was not possible to use existing 'benchmark' labelled datasets as these do not include UV information.

C. Timelapse Database

A time-lapsed image database was gathered from a single vantage point in January 2014 by mounting the camera on a secure platform and taking images across 4 days at 10 minute intervals in varying weather conditions (15:35-17:55 low light, 10:25-11:55 overcast, 10:50-11:00 sunny and 10:50-11:10 partial cloud). As before, all three filter types were used at each time point.

D. Urban Route Database

To allow the entire panorama to be recorded in a single image, the full spectrum camera was mounted above a panoramic mirror (Kugler, GmbH), fixed via a UV-passing quartz cylinder with 2mm wall (Fig. 1a), and images recorded using the UV-passing filter. A Sony Bloggie MHS-PM5, fitted with Bloggie 360 video lens kit attachment, was used to simultaneously capture panoramic images in the visible light spectrum. The combined system was mounted on a pole to elevate the horizon above a human carrier, with an approximate height of recording of 200 cm. Both cameras recorded video while the carrier walked a 2 km route that included narrow streets, city parks and squares, with moving traffic and pedestrians. True position was logged once per second using a GPS watch (Suunto Ambit). For testing navigation we used two videos of the route that were sampled on different days, at different times of year (January and April, introducing a change in the foliage), under different weather conditions (overcast and blue sky with clouds), and with different sun positions (11:40 and 16:47). Each video was converted to JPG images at 1 fps, providing two sets, consisting of 1337 and 1452 images respectively.

E. Sky Segmentation

To remove any distortion effects caused by camera movement when changing filters, images were aligned prior to segmentation by mutual information based image registration [13, 1]. K-means ($K = 2$, $r = 7$, sq. Euclidean distance) and Gaussian mixture model (GMM) clustering algorithms were implemented to assess both the accuracy and computational ease with which the sky could be automatically segmented from terrain. Pixels were clustered using image intensity only, ignoring any spatial connectivity information. These methods allowed comparison of performance despite the three image

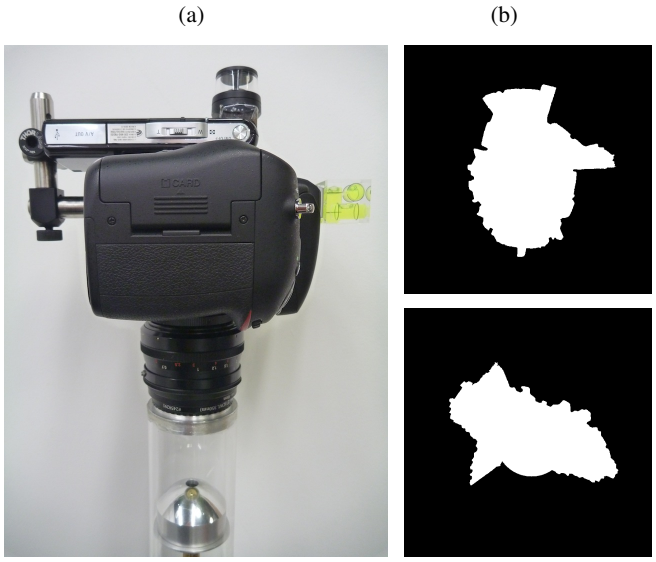


Fig. 1: **The panoramic camera set-up used for the urban route database.** (a) DSLR fitted with UV filter, quartz tube and parabolic mirror, with Sony Bloggie mounted on top. (b) Example of UV-segmented binary sky shape images used by our navigation algorithms. For each image the sky is shifted so that the centre of gravity of the shape is central in the image, to remove effects of tilt.

types having differing dimensions: UV=1D; UV-G=2D; and visible=3D (RGB).

We also tested a simpler segmentation algorithm tailored to the 1D UV intensity images. A histogram is computed from the intensity image, excluding the darkest and brightest 2 percent of pixels, and the adapted-watershed algorithm described in [30] is applied until only two segments of the histogram remain. The intensity value midway between the closest points of the two histogram segments is then set as the threshold for ground/sky segmentation. This method is robust to multiple maxima caused by a visible sun and ground reflectance, which were problematic for alternative algorithms.

F. Navigation Tests

OpenSeqSLAM [22], an open source Matlab implementation of SeqSLAM [15] , was used to test whether panoramic UV-segmented images contain sufficient information for localisation. Images were preprocessed by applying a mask to disregard sectors of the image that were not parts of the panorama in the parabolic mirror. The simple watershed segmentation algorithm described above was then applied to produce 90×90 binary images (sky vs non-sky). The sky region was isolated by boundary tracing from the centre of the image, and shifted by centralising the centre of gravity to reduce effects of tilt (Fig. 1b). Performance of our UV-segmented image set was compared to regular SeqSLAM using a set of visible light panoramic images recorded at the same locations and with a similar number of bits per image (16×16, 8 bit greyscale).

We tested localisation on the April route with the January reference images and vice versa. To distribute effects caused by the two sequences being particularly in or out of sync, we compared a reference set starting at the first frame of the route and sampled every 10 seconds against 10 test sets, starting with the first, second, third etc. frame of the other route and sampled every 10 seconds thereafter. SeqSLAM was run in both directions for each of the resulting 19 route pairs. SeqSLAM first performs an image difference function on the current image and all training images, stores the results in a matrix, and uses contrast enhancement to find the local best matches. A familiar sequence is then found by scoring trajectories across the matrix to look for those with the combined best match. The trajectory with the lowest score is considered a match if it is significantly smaller than all other scores outside a sliding window ($R_{window} = 10$).

We also tested an alternative navigation method that has been suggested as a possible mechanism used by ants, the visual compass algorithm [6, 12]. The basic assumption of this algorithm is that rotating a test image until the difference from a training image is minimised will indicate an approximately correct heading, under the assumption that captured training images are generally oriented in the direction of forward travel along the route. Although some versions of this algorithm require knowledge of the sequence to constrain which images to compare [12] it has also been shown to work effectively by taking the minimum over all rotations over all training images [2], i.e., assuming no sequence memory. This implicitly also returns a location estimate: the training image most similar to the test image after rotation.

Using the extracted skyline image sets at full resolution (691×691 pixels) a sub-sample of 1 image every 10 seconds from the test set was compared at 90 4° rotations against every image in the full 1fps training set, by summing the output of an *XOR* function. The location and rotation producing the smallest overall difference was chosen as best match. A comparison was made to both greyscale and colour visible light images, unwrapped and scaled to a similar number of bits; 180×28 and 90×14 pixels respectively.

III. RESULTS

A. Automatic Segmentation of Sky is Best Using UV Intensity Images

Table I shows the results of automatic sky segmentation, compared to manually labelled ground-truth, in the first image-database. Twelve example images are shown; the mean accuracy is based on 18 images. Segmentation using K-means clustering is most accurate using UV intensity images, with a mean accuracy of 96.8%. This is significantly better than visible light (mean accuracy 89.9%, paired Wilcox test $p = 0.0018$) and UV-G (mean accuracy 93.2%, paired Wilcox text $p = 0.025$). The accuracy for UV is also significantly less variable (standard deviation 3.6) than either visible light (s.d. 12.3, F-ratio test $p < .0001$) or UV-G (s.d. 9.5, F-ratio test $p = 0.0002$). Segmentation in the visible spectrum was not

TABLE I: **Accuracy of sky segmentation across locations.** A subset of our hand labelled image database is shown below to illustrate the diversity of images collected. Image triplets (Visible, UV-G and UV) were sampled locations chosen to assess the performance of sky segmentation in diverse surroundings. The percentage of correctly labelled pixels, when compared to a user defined ground truth, is shown for each image set. Mean scores (final row) are for the full database. Overall, for clustering methods performance is best using k-means on UV images, however similar results can be achieved with the computationally simpler adapted-watershed algorithm.

		K-means, $O(kDN)$			GMM, $O(iND^2)$			Watershed, $O(N)$
Image	**Feature**	**Visible**	**UV**	**UVG**	**Visible**	**UV**	**UVG**	**UV**
	**Desert scrub, Seville**	80.283	89.102	89.035	76.742	82.745	78.141	**91.107**
	**Forest floor, Sussex**	92.251	93.058	86.255	71.17	83.814	72.851	**93.28**
	**Pond reflecting sky**	93.245	97.15	95.459	95.63	92.911	90.489	**93.686**
	**Landscape with dark clouds**	52.42	97.731	59.771	**99.058**	98.198	58.491	98.563
	**Building hidden by trees**	**98.137**	97.817	98.126	91.452	94.665	92.284	97.689
	**Minaret with dark clouds**	75.549	97.75	90.272	**98.796**	98.155	98.538	98.511
	**Minaret with white clouds**	75.889	99.001	97.083	71.201	99.076	**99.171**	99.084
	**White building on overcast day**	**99.53**	99.353	99.336	91.805	91.334	99.33	99.512
	**City street with grey sky**	99.047	99.123	98.987	98.865	98.878	96.581	**99.263**
	**Building with metallic roof**	85.348	86.32	**87.298**	76.967	84.599	81.364	84.687
	**Building with metal wall**	96.366	98.595	98.218	61.211	**98.722**	91.661	98.901
	**Buildings with reflective windows**	97.612	**98.068**	97.662	71.548	81.109	76.985	97.754
	Mean accuracy (n=18)	89.867	96.778	93.229	85.962	92.36	87.448	**96.939**
	std. dev.	12.245	**3.631**	9.497	12.021	7.514	11.682	3.883

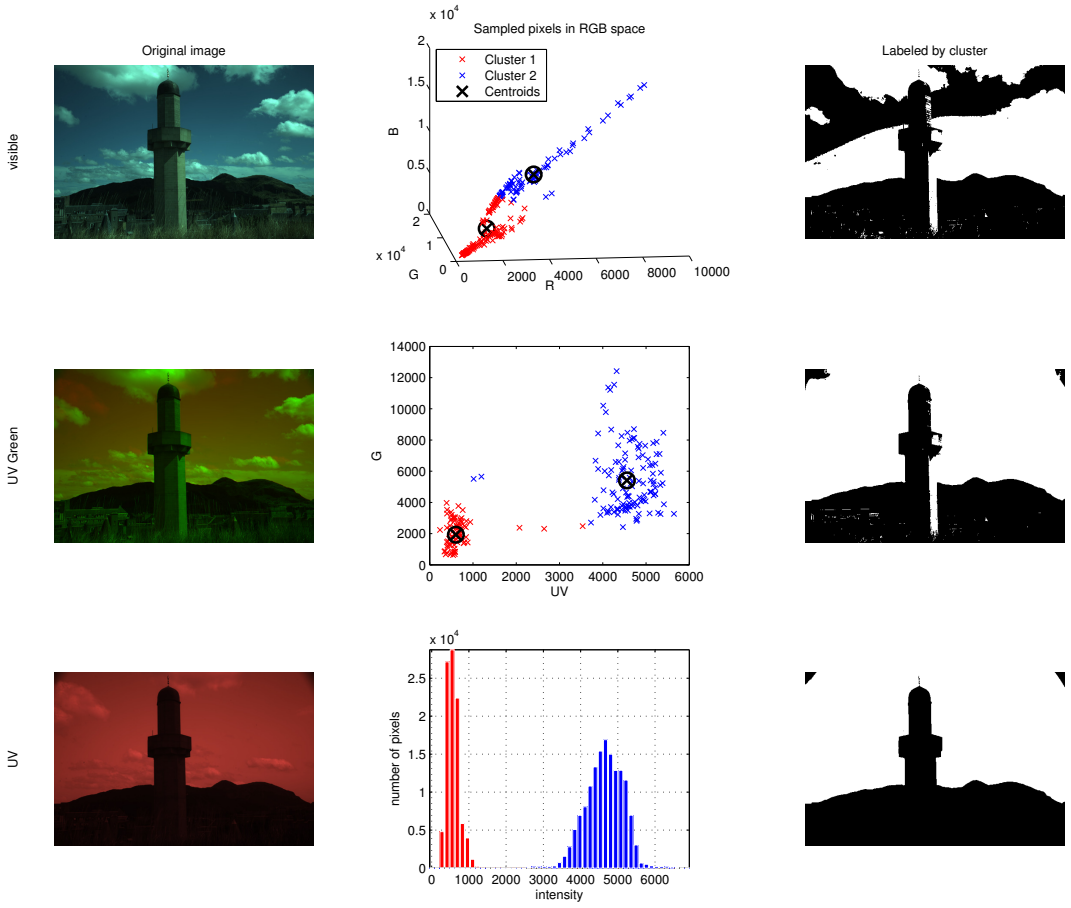


Fig. 2: **Typical example of conditions in which UV-intensity segmentation outperforms both visible and UV-G images. Performance is worst using visible light as sky and ground are not easily separable when clouds are present. In contrast, sky and ground naturally separate along the UV dimension giving robust segmentation in UV and UV-G images.** Raw images are shown in the left column. The central column shows the spread of image pixels plotted in their respective colour space, coloured with respect to the attempted automatic k-means segmentation (blue=sky, red=ground). In each scatter plot a random subsample of 200 pixel values was used. The right column shows the resulting segmented images with white representing sky and black representing ground. Note that the black marks in the corners of the the UV pictures are the result of vignetting due to an overlapping lens hood that is part of the Baader U-filter.

improved using the more robust Gaussian mixture model classifier (mean accuracy 86.0%). Visible light sky segmentation was also tested in other colour spaces, such as the A and B dimensions of LAB space, however no notable improvements were observed over RGB for any of our clustering algorithms. The simple watershed algorithm for UV performs just as well as K-means clustering (mean accuracy 96.9%) but at less computational cost ($O(N)$ instead of $O(kDN)$), where k is the number of clusters, D the dimensionality of our pixel space and N the number of pixels). The level of accuracy obtained is comparable with that reported in other state of the art sky segmentation methods, e.g. 96.05% reported in [20]. In fact a proportion of our remaining error was simply due to vignetting of the overlapping lens hood (see Fig. 2 middle row). We also note that the 'ground truth' labelling was performed on the visible images and so was if anything biased against UV.

Fig. 2 shows the underlying reasons for better performance of UV. In the visible light image (top left), the bright white clouds form a cluster. As the classifier is constrained to two groups, the blue sky is incorrectly labelled the same as the ground. This is the most common cause of error when using visible light, but reflectance from buildings and sun flare also cause segmentation irregularities. In contrast, in the UV-G image the clouds are labelled with the sky, and only reflection from the buildings causes some segmentation issues. The pixel-plot shows that the successful separation is attributable mostly to the UV channel, and in fact the segmentation is improved by using the UV channel only, which does not classify strong lighting on the buildings as sky.

The second image-database allows us to test both the accuracy (vs ground truth), and the reliability (vs images from the same location) of sky segmentation across weather and

lighting conditions. This is important as it may be acceptable for a navigating robot to misclassify some portion of the ground as sky (or vice versa) provided the classification remains consistent over time. Accuracy is compared against user labelled ground truth, and here again we find that segmentation is more accurate using UV intensity images (mean=99.5%, s.d=0.084) than visible (mean=95.2%, s.d.=5.327). Reliability was assessed by measuring the mean entropy and Pearson correlation coefficients of the binary labelled images. Specifically, the mean entropy per pixel location, $H_2(x_i)$, in labelled UV and visible images was calculated as

$$p(x_i) \approx \frac{\sum_{t=1}^{T}(x_{it})}{T} \tag{1}$$

$$H_2(x_i) = -p(x_i)\log_2 p(x_i) - (1-p(x_i))\log_2 (1-p(x_i)), \tag{2}$$

where x_{it} is the intensity value of a pixel at location i at time t, and T is the number of images in the set. The Pearson correlation is given by

$$r(\mathbf{x},\mathbf{y}) = \frac{cov^*(\mathbf{x},\mathbf{y})}{var^*(\mathbf{x})var^*(\mathbf{y})}, \tag{3}$$

where $cov^*(\cdot,\cdot)$ indicates the sample covariance between two images $\mathbf{x}$ and $\mathbf{y}$ and $var^*(\cdot)$ the sample variance of an image. Fig. 3 presents the resultant entropy and correlation plots for visible and UV image sets. It is clear that pixels in the visible domain are highly variable (mean entropy=0.1877) particularly in the crucial area where buildings meet the sky. There is also significant interference from sunlight on buildings and street lights. This variance results in low correlation coefficients between images throughout the data-set. Labelled images that correlate poorly in the heat map in Fig. 3c (images 5, 15 and 20) corresponded to the 3 images shown in the left column of the Fig. 3a, where cloud, sunlit buildings and low lighting caused discrepancies in segmentation. The causes of these problems are clearly visualised by viewing the plot showing average entropy per pixel (Fig. 3b). In contrast, UV images have a low mean entropy per pixel (0.0048), producing a highly correlated and thus reliable dataset across conditions, even at low light levels.

B. UV Sky Segmented Images Are Sufficient For Navigating Cities

A set of images recorded along a route can be used in several ways for subsequent navigation. We first tested whether binary sky/non-sky images derived from panoramic UV images recorded along a 2km urban route (see supplementary video) could be used for localisation (or loop-closing) using SeqSLAM as described in section II-F. Overall, using a window length of 10 images, we can obtain a matching precision of 100% (no false positives) with a recall of 81%; at 100% recall, the precision is 94%; these values compare well to previous results for this algorithm [15]. Fig. 4a shows where matching is unsuccessful, which appears to correspond to locations under trees, where the sky is either obscured or its

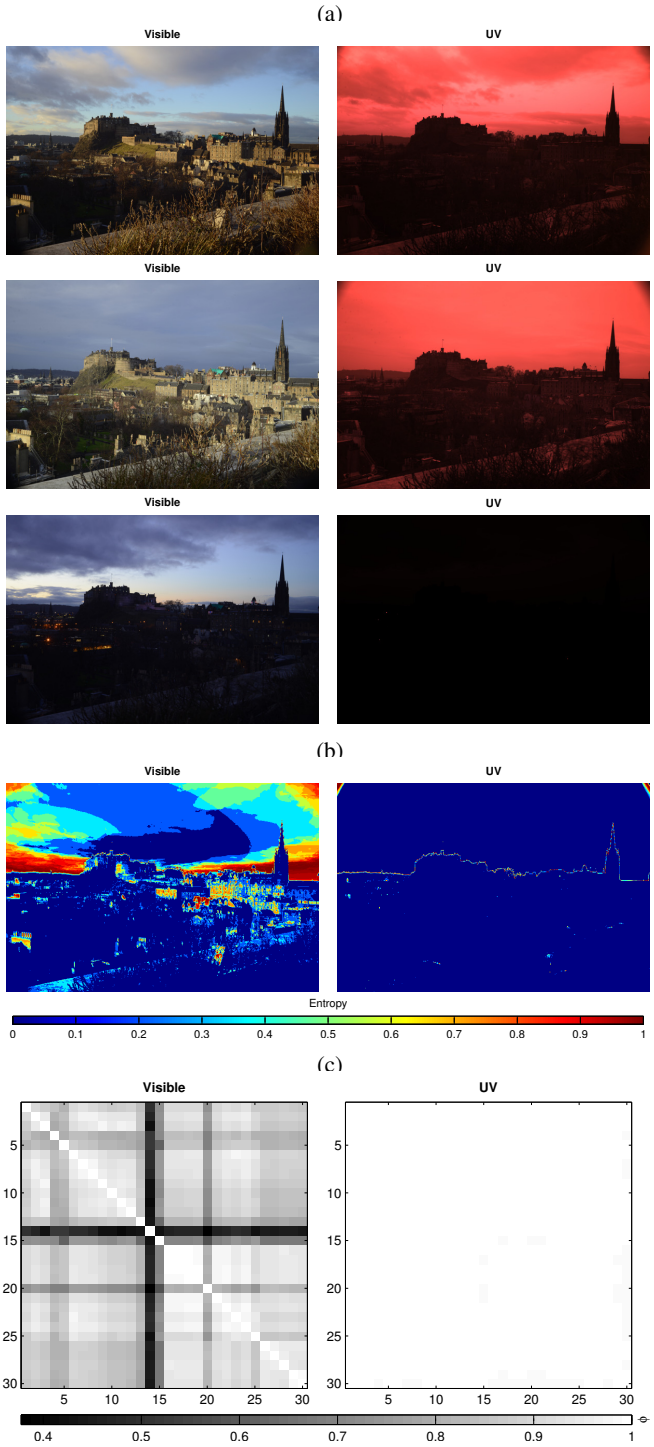


Fig. 3: **Reliability of sky segmentation across weather conditions.** (a) Time-lapse images sampled under differing weather conditions (note that the UV pixel intensities have been increased $5\times$ to make image more visible, but are still extremely low in the third example). (b) Pixel-wise entropy and (c) Pearson correlation coefficients computed across the image dataset. Both indicate high variance in visible light images but consistent labelling using UV images.

segmented shape changing too rapidly for the sparse temporal sampling used here. Our result for binary UV images is only a few percent below that obtained using standard SeqSLAM on the equivalent panoramic greyscale visible light images (precision 97%), which deals better with situations where there is little or rapidly changing sky shape information (under trees) as it can still exploit information in the ground region. On the other hand standard SeqSLAM tended to fail in areas where the ground region is relatively featureless but there are substantial changes in the sky region (clouds) between training and test sets.

While SeqSLAM is highly effective for localisation if following the same route, it does not provide a mechanism to stay on the route. Hence we also test the visual compass algorithm, as described in II-F which produces both a location match and an estimate of the heading direction needed to stay on route. The obtained precision using binary UV images was 84.7%. Fig. 4c shows the heading direction suggested by the visual compass at each point on the route; it can be seen that these align well with the required direction to stay on the route. Using the equivalent visible light images produced a significantly worse performance with a precision of only 44.1% and many heading direction choices that would lead away from the route (Fig. 4d) .

IV. Conclusions

It has been suggested that UV light sensing in desert ants is a specialisation that supports their impressive navigation by enhancing sky-ground contrast. We have demonstrated here that UV images are highly effective for sky segmentation, that the segmented sky shape is robust to changing light and weather conditions, and that the resulting binary sky/non-sky images can support effective navigation.

For a wide range of locations and light conditions, automated sky segmentation is significantly more effective for UV than visible light. Although only a limited set of images were hand-segmented for ground truth evaluation, visual inspection of UV segmentation in our larger collection of images suggests broad reliability, in particular, robustness against all kinds of cloud cover. The unidimensional signal also allows the application of a computationally cheap segmentation algorithm that is just as effective as K-means clustering.

There were a few cases where ground objects, such as particular kinds of metal roofing, had high UV reflectivity. However, for navigation, it is less important to find the skyline exactly than it is to find it consistently. By testing sky-ground segmentation for images taken from the same location on different days, at different times of day, and under different weather conditions, we found that UV images have an extremely high consistency in pixel classification, around 40 times better than visible light.

The navigation potential of a panoramic UV camera system was demonstrated by recording 2km routes in an urban environment on two different days with different weather conditions. Using either SeqSLAM or a visual compass, matching of UV-segmented binary images of the sky contour from one day against another supported successful localisation. For the visual compass a reliable indication of the correct heading direction to continue along the route was also obtained. Using non-segmented visible light images of a similar file size, SeqSLAM was still effective for localisation, but determining location or heading direction from the visual compass was far less effective. Failure of the visual compass is generally due to mismatch or aliasing, i.e., a training image far from the current location provides, at some arbitrary rotation, a closer match than the nearest image. This suggests its effectiveness could be improved (and its application sped up) by using SeqSLAM as a preprocessing step to find the best match, then finding the minimum over image rotations for this image only.

We note that the skyline is a particularly consistent aspect of urban (and some other) environments. Many changing features such as traffic and pedestrians, fall entirely below the skyline, and different lighting conditions such as sun falling on one or the other side of street make little difference to the segmented shape. Weather changes occur above the skyline. Consequently it appears that a substantial amount of information is retained by reducing images to binary sky vs non-sky: SeqSLAM performs almost as well on such images as on full grey scale images. The few situations in which the system failed were in locations where very little sky was visible (under scaffolding, or under trees). The considerable difference in foliage between our two datasets, and the 10 second time intervals between frames could cause a notable difference in sky shape directly above the camera between training and test sets in this situation. This could perhaps be rectified by smoothing over recent frames to filter out fast moving shapes. Alternatively, rather than using binary images, the UV threshold could be used to mask out the sky region (including uninformative features such as clouds) but information below the skyline, possibly from visible light channels, could be retained to improve robustness.

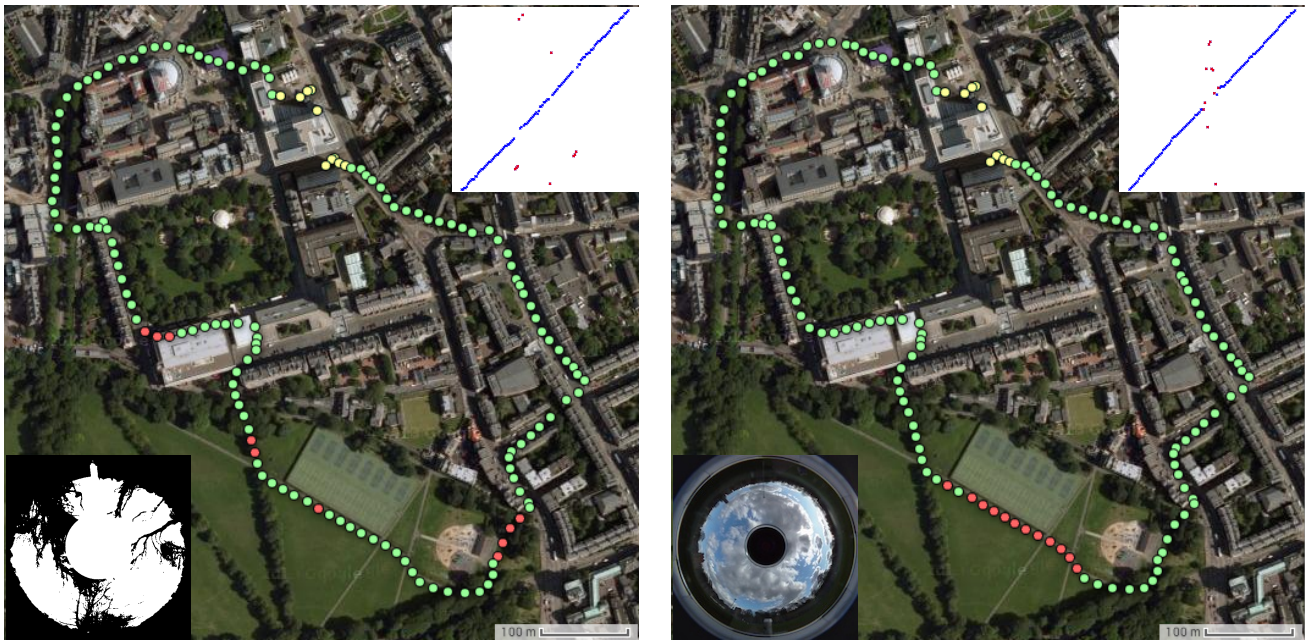
There are a number of potential applications for this approach. It could be used effectively in urban environments where GPS can be unreliable. The nature of the segmented image also makes it trivial to perform corrections to tilt by simply offsetting the image depending on the centre of gravity of the sky pixels. It could therefore also be potentially useful for robots in rugged environments, where SLAM needs to be reliably carried out on uneven terrain. Sensitivity to UV light at low light levels could be improved by using a CCD chip without colour filters, rather than the converted visible light camera used in these tests. The relatively simple computation to segment sky could be implemented in hardware, providing a compact and low-power sky detection sensor, which could offer cheap high precision localisation in many outdoor autonomous robot applications.

Acknowledgments

This work was supported in part by grants EP/F500385/1 and BB/F529254/1 for the University of Edinburgh School of Informatics Doctoral Training Centre in Neuroinformatics and Computational Neuroscience (www.anc.ac.uk/dtc), I014543/1 Bayesian Issues in Ant Navigation from the BBSRC and by the EPSRC and MRC. We are grateful to Paul Graham and IPAB for funding the camera equipment and to the Informatics technicians.

(a) UV sky segmented SeqSLAM

(b) Visible light full image SeqSLAM



(c) UV sky segmented Visual Compass

(d) Visible light full image Visual Compass

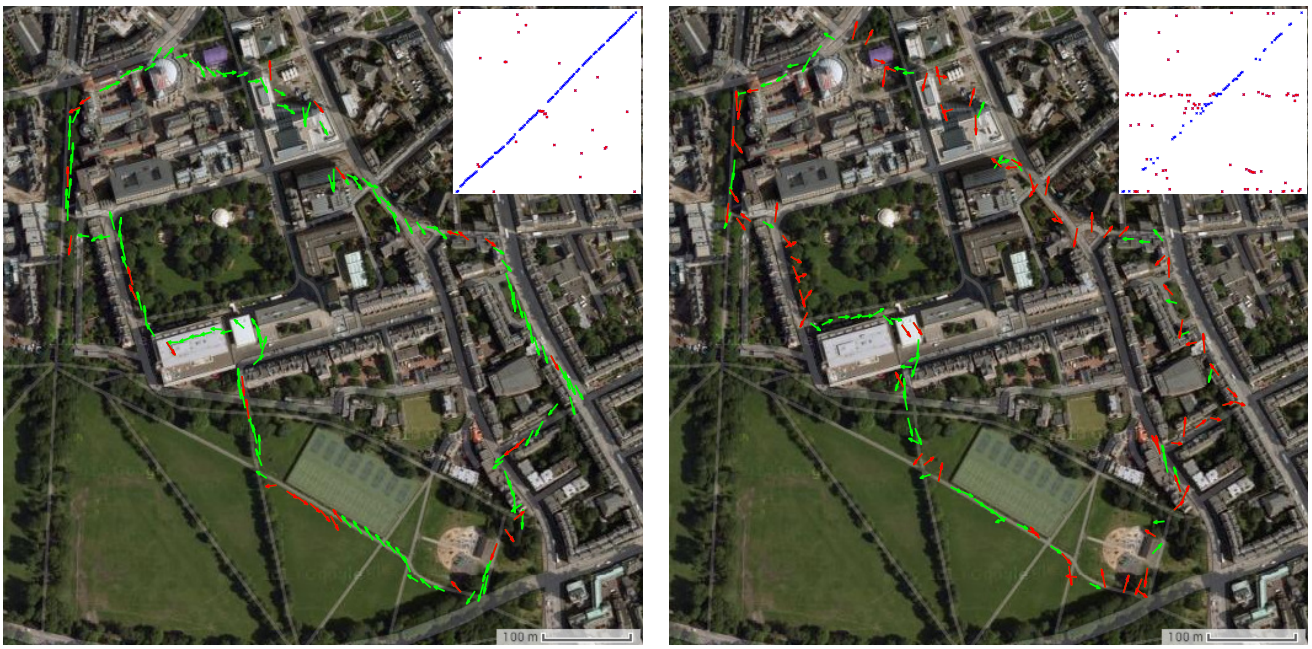


Fig. 4: **Using sky-segmented images to navigate a city.** (a) and (b) plot the example results of the SeqSLAM localisation trial taken from the worst test set for UV (left) and the worst for visible light (right). Green dots represent locations that were correctly recognised and red dots those that were not. Yellow dots are at the beginning and end of the route, where the sequence was too short to localise. In the top right corner of these figures the estimated location index of the training set is plotted against the location indices of this test set. In the bottom left corner an example panoramic image is shown, corresponding to an incorrectly localised image in this test set. Branches (for UV) and clouds (for visible light) were typical features that caused the algorithm to fail.
(c) and (d) show the heading direction that the visual compass algorithm would select in order to retrace the route (left for UV, right for visible light), with all headings corresponding to a correctly matched location coloured green. Performance is better for UV.
Map data: Google, Infoterra Ltd & Bluesky.

REFERENCES

[1] Artyushkova, K. Automatic Image Registration using (Normalized) Mutual Information for users of IP toolbox, 2006. URL http://www.R-project.org/.

[2] B. Baddeley, P. Graham, P. Husbands, and A. Philippides. A model of ant route navigation driven by scene familiarity. *PLoS Computational Biology*, 8(1):e1002336, January 2012. doi: 10.1371/journal.pcbi.1002336.

[3] N. Carey and W. Stürzl. An insect-inspired omnidirectional vision system including UV-sensitivity and polarisation. In *Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on*, pages 312–319. IEEE, 2011. doi: 10.1109/ICCVW.2011.6130258.

[4] L. Chapman, J.E. Thornes, J-P. Muller, and S. McMuldroch. Potential applications of thermal fisheye imagery in urban environments. *Geoscience and Remote Sensing Letters, IEEE*, 4(1):56–59, 2007. doi: 10.1109/LGRS.2006.885890.

[5] M. Cummins and P. Newman. FAB-MAP: Probabilistic Localization and Mapping in the Space of Appearance. *The International Journal of Robotics Research*, 27(6):647–665, June 2008. doi: 10.1177/0278364908090961.

[6] M.O. Franz, B. Schölkopf, H.A. Mallot, and H.H. Bülthoff. Where did I take that snapshot? Scene-based homing by image matching. *Biological Cybernetics*, 79(3):191–202, 1998. doi: 10.1007/s004220050470.

[7] P. Graham and K. Cheng. Ants use the panoramic skyline as a visual cue during navigation. *Current Biology*, 19(20):R935–7, November 2009. doi: 10.1016/j.cub.2009.08.015.

[8] D.I. Klebe, R.D. Blatherwick, and V.R. Morris. Ground-based all-sky mid-infrared and visible imagery for purposes of characterizing cloud properties. *Atmospheric Measurement Techniques*, 7(2):637–645, 2014. doi: 10.5194/amt-7-637-2014.

[9] T. Kollmeier, F. Röben, W. Schenck, and R. Möller. Spectral contrasts for landmark navigation. *Journal of the Optical Society of America. A, Optics, Image Science, and Vision*, 24(1):1–10, January 2007. doi: 10.1364/JOSAA.24.000001.

[10] S. Komai, T. Kuroda, and M. Takano. Development of Invisible Mark and Its Application to a Home Robot. In *Service Robotics and Mechatronics*, pages 171–176. 2010. doi: 10.1007/978-1-84882-694-6_30.

[11] T. Labhart. The electrophysiology of photoreceptors in different eye regions of the desert ant, Cataglyphis bicolor. *Journal of Comparative Physiology A*, 158(1):1–7, 1986. doi: 10.1007/BF00614514.

[12] F. Labrosse. The Visual Compass : Performance and Limitations of an Appearance-Based Method. *Journal of Field Robotics*, 23(10):913–941, 2006. doi: 10.1002/rob.

[13] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens. Multimodality image registration by maximization of mutual information. *IEEE transactions on medical imaging*, 16(2):187–98, April 1997. doi: 10.1109/42.563664.

[14] J. Meguro, T. Murata, Y. Amano, T. Hasizume, and J. Takiguchi. Development of a Positioning Technique for an Urban Area Using Omnidirectional Infrared Camera and Aerial Survey Data. *Advanced Robotics*, 22(6-7):731–747, January 2008. doi: 10.1163/156855308X305290.

[15] Michael J Milford and Gordon Fraser Wyeth. Seqslam: Visual route-based navigation for sunny summer days and stormy winter nights. In *Robotics and Automation (ICRA), 2012 IEEE International Conference on*, pages 1643–1649. IEEE, 2012.

[16] R. Möller. Insects could exploit UV-green contrast for landmark navigation. *Journal of theoretical biology*, 214(4):619–31, February 2002. doi: 10.1006/jtbi.2001.2484.

[17] M.I. Mote and R. Wehner. Functional characteristics of photoreceptors in the compound eye and ocellus of the desert ant, Cataglyphis bicolor. *Journal of comparative physiology*, 137(1):63–71, 1980. doi: 10.1007/BF00656918.

[18] S. Ramalingam, S. Bouaziz, P. Sturm, and M. Brand. Geolocalization using skylines from omni-images. In *Computer Vision Workshops (ICCV Workshops), 2009 IEEE 12th International Conference on*, pages 23–30. IEEE, 2009. doi: 10.1109/ICCVW.2009.5457723.

[19] A.E.R. Shabayek, C. Demonceaux, O. Morel, and D. Fofi. Vision Based UAV Attitude Estimation: Progress and Insights. *Journal of Intelligent & Robotic Systems*, 65(1-4):295–308, 2012. doi: 10.1007/s10846-011-9588-y.

[20] Y. Shen and Q. Wang. Sky Region Detection in a Single Image for Autonomous Ground Robot Navigation. *International Journal of Advanced Robotic Systems*, 10:362, 2013. doi: 10.5772/56884.

[21] F. Stein and G. Medioni. Map-based localization using the panoramic horizon. *IEEE Transactions on Robotics and Automation*, 11(6):892–896, 1995. doi: 10.1109/70.478436.

[22] N. Sünderhauf, P. Neubert, and P. Protzel. Are We There Yet? Challenging SeqSLAM on a 3000 km Journey Across All Four Seasons. In *Proc. of Workshop on Long-Term Autonomy, IEEE International Conference on Robotics and Automation (ICRA), Karlsruhe, Germany.*, 2013.

[23] M.H. Tehrani, M.A. Garratt, and S. Anavatti. Gyroscope offset estimation using panoramic vision-based attitude estimation and Extended Kalman Filter. In *CCCA12*, pages 1–5. IEEE, December 2012. doi: 10.1109/CCCA.2012.6417863.

[24] Mohsen H. Tehrani, M. A. Garratt, and S. Anavatti. Horizon-based attitude estimation from a panoramic vision sensor. In Ghose Debasish, editor, *IFAC-EGNCA 2012*, pages 185–188, February 2012. doi: 10.3182/20120213-3-IN-4034.00035.

[25] J. Tighe and S. Lazebnik. Superparsing. *International Journal of Computer Vision*, 101(2):329–349, October

2012. doi: 10.1007/s11263-012-0574-z.

[26] Urmson, C. The self-driving car logs more miles on new wheels, August 2012. URL http://googleblog.blogspot.hu/2012/08/the-self-driving-car-logs-more-miles-on.html.

[27] V. Vezhnevets and V. Konouchine. GrowCut: Interactive multi-label ND image segmentation by cellular automata. In *Proc. of Graphicon*, pages 150–156, 2005.

[28] H. Wajima, T. Makino, Y. Takahama, T. Sugiyama, and A. Keiichi. An Autonomous Inspection Robot System for Runways Using an Ultraviolet Image Sensor. In *Intelligent Autonomous Systems 7*, pages 357–364, 2002.

[29] R. Wehner. The architecture of the desert ant's navigational toolkit (Hymenoptera: Formicidae). *Myrmecol News*, 12:85–96, 2008.

[30] N.S. Zghal and D.S. Masmoudi. Improving watershed algorithm with a histogram driven methodology and implementation of the system on a virtex 5 platform. *International Journal of Computer Applications*, 9(11): 29–35, 2010. doi: 10.5120/1435-1934.

3.3 CONCLUSION

This chapter showed that spectral characteristics of the ants' visual system are well tuned for extracting the skyline. It also demonstrated that skylines can effectively be used for place recognition. Soon after this paper was published, our findings on skyline extraction using UV only were confirmed by Differt and Möller (2015, 2016). Recent studies have also re-affirmed the crucial role of UV for skyline extraction in insects (Schultheiss et al., 2016).

Major disadvantages of this method include scenes where overhead views change quickly, such as under dense foliage. Combined with a lack of clear skyline in those scenarios, this raises the question of whether ants could rely on this mechanism to navigate in highly foliated areas, where a large part of the horizon is blocked (although see section 4.4.1 for further discussion on this topic). Problems for robots, that ants are not susceptible to, include camera-specific issues, e.g. lens flare. Advantages of this method included robustness to changing conditions and relatively cheap storage and feature extraction compared to using point features, with no pre-training needed. The ability to do a simple binary comparison could also be advantageous for making quick comparisons with previously stored views.

3.3.1 *Future work*

One of the disadvantages of using binary images of sky and ground as features for robotic navigation is the inability to index or sort them in a natural manner. The solution described in this chapter could benefit from a different representation to store the image for fast lookup. From a biological point of view, the pixel-by-pixel manner of comparison between images would perhaps not fit in well with misaligned views caused by uneven terrain and non-matching body rotation that ants and other insects often must deal with (Ardin et al., 2015). From a computer vision perspective, this could perhaps be improved by storing higher-level features that capture the shape of the sky, rather than each individual pixel. The above points are all addressed in the next chapter, where a follow-up study investigates the viability of representing extracted skylines in a more suitable manner.

4

SKYLINE-BASED LOCALISATION FOR AGGRESSIVELY MANOEUVRING ROBOTS USING UV SENSORS AND SPHERICAL HARMONICS

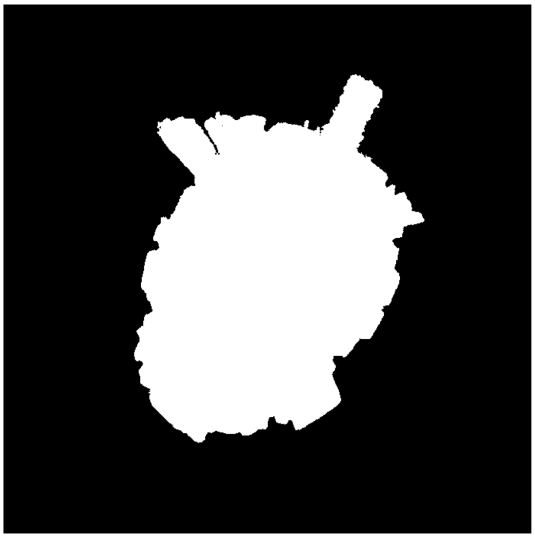
4.1 INTRODUCTION

As discussed in the previous chapter, capturing scenes in ultraviolet (UV) is useful for separating sky from ground, which in turn can be used as a stable feature for localisation. A natural follow-up study is to find what other immediate benefits emerge from using the skyline to navigate. To guide this research, some navigational challenges that insects face were considered and an investigation was made into how skylines could be well adapted to localise despite these problems.

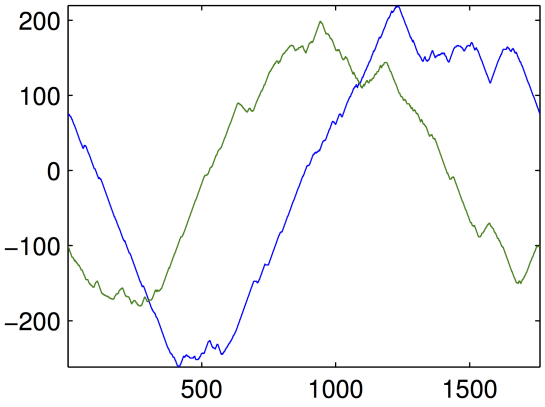
The ability of ants to visually navigate by place recognition, despite not being retinotopically aligned with previously observed scenes (Ardin et al., 2015) is still unexplained. Ants are able to make correct navigational decisions when in a previously visited location, despite facing in a different direction, or their field of view (FOV) being retinotopically misaligned due to bumpy terrain. In fact, there is evidence that ants are able to visually navigate, while dragging heavy food in a backwards motion, where the visual field will almost never align with previously seen scenes (Ardin et al., 2016b). A solution to this problem could be mental rotation of the scene, allowing the best alignment to be found, which has previously been proposed to be possible by humans for shape recognition (Tarr and Pinker, 1989). However, even for just yaw rotation this is highly unlikely due to the high computational power required to both rotate and store an image, whilst comparing it to all previously seen images and keeping track of the best overlap. It would be desirable to find a solution to this problem that is more biologically plausible.

4.1.1 *Localisation by matching sky shapes*

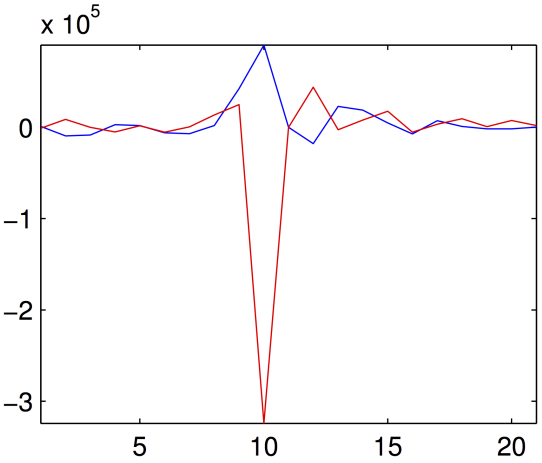
When sky is segmented from ground and stored as a binary image, the shape of the sky forms a single blob or group of blobs with a characteristic shape (Fig. 4.1). Matching a segmented sky-image to another, despite yaw rotations, is essentially a form of shape recognition, a well established field (Latecki et al., 2000). Various algorithms exist for matching shapes, despite variation in translation, scale and rotation. Common



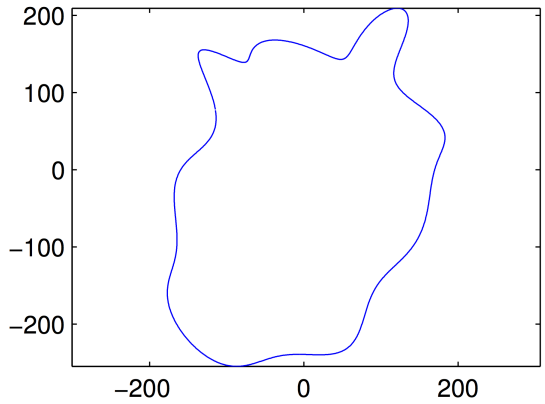
(a) Segmented sky shape



(b) Pixel coordinates



(c) Frequency components



(d) Reconstructed sky using 10 coefficients

Figure 4.1: **Fourier descriptors to match skylines. (a)** Segmented sky of an urban scene can be thought of as a closed shape. **(b)** x and y coordinates of the shape edge create a continuous signal. **(c)** The frequency components of this signal can describe the outline of the shape up to an arbitrary amount of detail. **(d)** Reconstruction of sky shape using only the first ten Fourier descriptors shows the rough shape signature, allowing high compression of skylines.

solutions include capturing properties of the contour (Persoon and Fu, 1977; Kaneko and Okudaira, 1985; Mokhtarian et al., 1997) or the region itself, typically through use of geometric moments (Hu, 1962; Khotanzad and Hong, 1990).

The first approach to match rotated skylines involved representing the images collected in the previous chapter using Fourier descriptors (FD) (Persoon and Fu, 1977). The outline of the shape is traced to form a complex signal representing x and y coordinates (Fig. 4.1b), which then undergoes a Fourier transform to represent it in the frequency domain (Fig. 4.1c). Ignoring phase and storing the amplitude provides a shape signature that is invariant to yaw rotation. An additional benefit of this encoding is that the level of detail can be adjusted by only retaining the coefficients representing lower frequencies (Fig. 4.1d), allowing for easy scaling of storage and fast lookup of similar skylines. Contour-based methods such as FD worked well for place recognition in many locations, but particularly struggled where the sky did not clearly consist of a single area, such as under foliage or even in areas with overhead cables.

To mitigate these problems, a region-based shape recognition approach was applied. Zernike moments (ZM) (Khotanzad and Hong, 1990) allow functions on a disk to be captured in the frequency domain. Like FD, this provides a yaw-rotation-invariant representation of an image and worked well, marginally improving localisation results on the urban dataset from chapter 3.

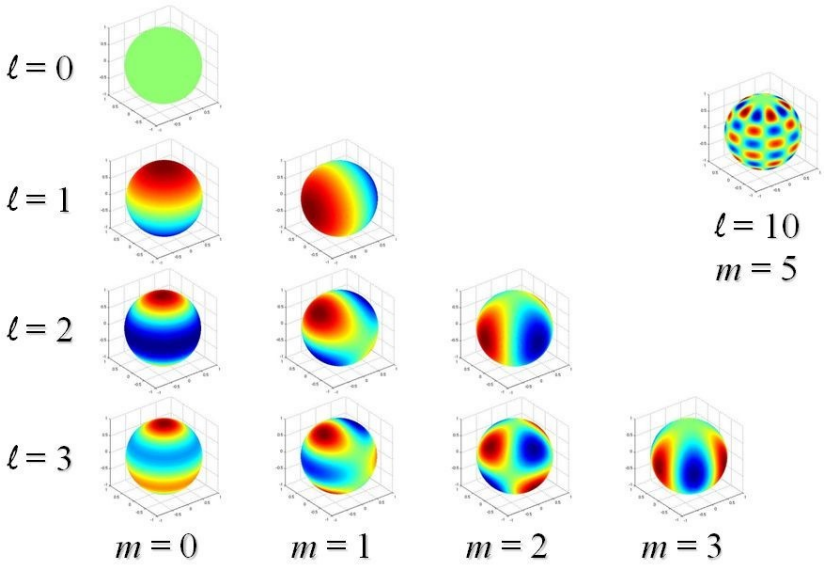


Figure 4.2: **Spherical harmonics.** Any function on the surface of a sphere can be captured to an arbitrary level of detail using a weighted sum of spherical harmonics, which are a complete set of orthogonal functions on the sphere. In this figure l is the degree of the polynomial and m the index in the set of point masses. Figure by Ryan Torn, University at Albany

Finally, to create a representation that is fully invariant to roll, pitch and yaw, spherical harmonics (SH) were used (Kazhdan et al., 2003). This method still uses the summation of basis functions to approximate a function, but on the surface of a sphere rather than a disk. Capturing the shape of the sky in this manner allows sky/ground images to be matched by their coefficient signatures in the amplitude spectrum, but remain completely invariant to any rotation. The combination of skyline and SH are useful for localisation, as global features in the image are compared, making the approach more robust to changes in condition, however, some of the appealing properties from using local descriptors, such as pose invariance, are also retained. Features could also be extracted despite significant motion blur and misalignment of previous views. The ideal use-case for this approach would be on small, fast robots that need to localise while moving across uneven terrain. Results of a robot study of this method are outlined in the attached publication below.

4.2 PUBLICATION: IEEE INTERNATIONAL CONFERENCE ON ROBOTICS AND AUTOMATION 2016

This paper was published at the IEEE International Conference on Robotics and Automation 2016 (Stone et al., 2016a). Personal contributions include:

- Hypothesis development and experimental design. In particular, the concept of using the amplitude in SH to create a signature that could be plugged into a traditional localisation algorithm.
- SeqSLAM localisation algorithm was developed by Michael Milford, ported to Python and modified to take SH coefficients as input by Thomas Stone
- Camera developed in collaboration with Michael Mangan and Paul Ardin. This included the design of a novel UV fisheye sensor.
- Acquisition of suitable robot platform in collaboration with Dario Differt.
- Python wrapper for C++ spherical harmonics library.
- Experiments in collaboration with Dario Differt. Suitable experiments were designed together, and walking based routes were carried out by Thomas Stone and Dario Differt in Edinburgh. Vehicle-based routes were carried out by Dario Differt in Bielefeld.
- All experimental analysis on the collected data and production of plots.

- Primary author on paper content and figures, including the creation of the main body of text. Thomas Stone also coordinated the contributions of other authors during the editing phase.

2016 IEEE International Conference on Robotics and Automation (ICRA)
Stockholm, Sweden, May 16-21, 2016

Skyline-based Localisation for Aggressively Manoeuvring Robots using UV sensors and Spherical Harmonics

Thomas Stone*[1], *Student Member, IEEE*, Dario Differt*[2], Michael Milford[3], *Member, IEEE*, and Barbara Webb[4]

***Abstract*— Place recognition is a key capability for navigating robots. While significant advances have been achieved on large, stable platforms such as robot cars, achieving robust performance on rapidly manoeuvring platforms in outdoor natural conditions remains a challenge, with few systems able to deal with both variable conditions and large tilt variations caused by rough terrain. Taking inspiration from biology, we propose a novel combination of sensory modality and image processing to obtain a significant improvement in the robustness of sequence-based image matching for place recognition. We use a UV-sensitive fisheye lens camera to segment sky from ground, providing illumination invariance, and encode the resulting binary images using spherical harmonics to enable rotation-invariant image matching. In combination, these methods also produce substantial pitch and roll invariance, as the spherical harmonics for the sky shape are minimally affected, providing the sky remains visible. We evaluate the performance of our method against a leading appearance-invariant technique (SeqSLAM) and a leading viewpoint-invariant technique (FAB-MAP 2.0) on three new outdoor datasets encompassing variable robot heading, tilt, and lighting conditions in both forested and urban environments. The system demonstrates improved condition- and tilt-invariance, enabling robust place recognition during aggressive zigzag manoeuvring along bumpy trails and at tilt angles of up to 60 degrees.**

I. INTRODUCTION

Place recognition is an important component of robot navigation [1]. Although impressive results have been demonstrated using view-based approaches (e.g. [2], [3]), many of these approaches only work under the assumption that training and test route are undertaken in similar lighting conditions. Other methods such as SeqSLAM [4] exhibit good condition-invariance but fail when viewpoint changes significantly. Many actual and potential robotic scenarios, from military reconnaissance to mowing the lawn [5], [6], involve small cheap robots moving at high speed over variable terrain, and hence revisiting places at different times with significantly different pose orientations. This paper addresses the challenge of robustly recognising these places despite rotation (yaw) change, tilt (roll, pitch) change, and appearance change.

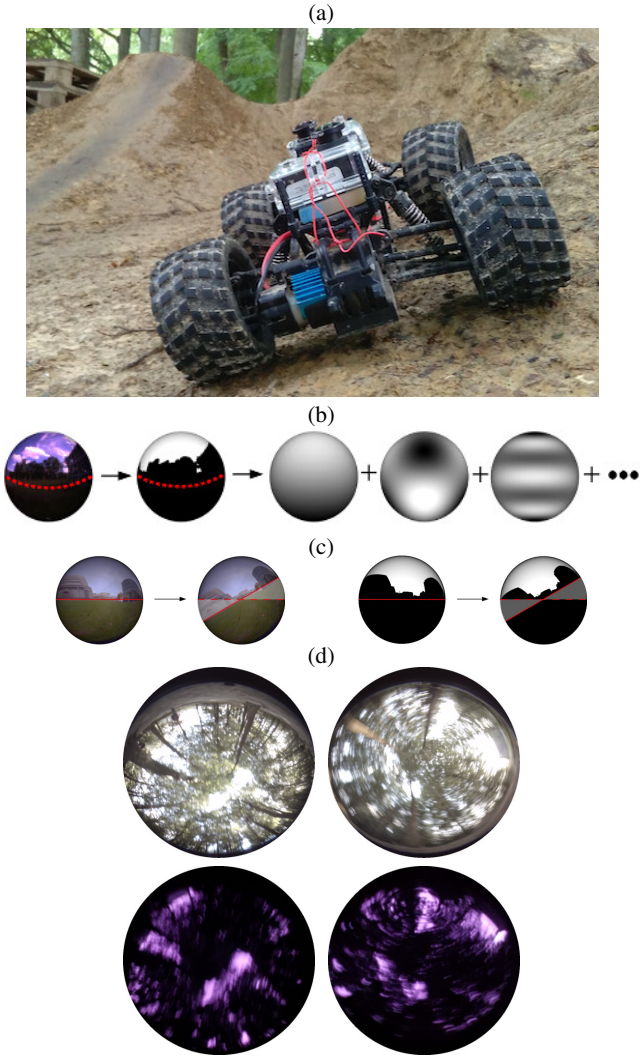


Fig. 1: (a) A high-speed all-terrain platform with UV fisheye lens oriented upwards. (b) The camera captures one hemisphere (above red line); we segment the sky, and convert it to rotation- and tilt-invariant spherical harmonic (SH) amplitudes. (c) Left: RGB images with different tilt have non-overlapping areas that need to be inferred (grey area) to match. Right: in sky-segmented images unknown areas are assumed to be ground (black) and the SH amplitudes will be unaffected by tilt. (d) Four images from the same location across two runs (one run per column, top: RGB, bottom: UV). RGB images are hard to match due to tilt, yaw rotation and motion blur; using SH amplitudes, we can correctly match the UV images.

*First author
[1]Thomas Stone is a PhD student at the School of Informatics, University of Edinburgh, 11 Crichton St, Edinburgh, Midlothian, EH8 9LE, United Kingdom t.j.stone@sms.ed.ac.uk
[2]Dario Differt is with the Computer Engineering Group, Faculty of Technology, Bielefeld University, D-33594 Bielefeld, Germany dario.differt@uni-bielefeld.de
[3]Michael Milford is with the Australian Centre for Robotic Vision at the Queensland University of Technology, Brisbane, Australia michael.milford@qut.edu.au
[4]Barbara Webb is Professor of Biorobotics at the School of Informatics, University of Edinburgh B.Webb@ed.ac.uk

Our approach is inspired by the capacity of ants to perform robust visual navigation despite substantial pitch and tilt variation in their views caused by legged locomotion over uneven terrain [7]. Ants achieve these remarkable feats of navigation by extracting skyline contours [8] exploiting UV sensitivity in the insect ommatidia [9]. We combine a novel sensing modality—a camera equipped with a UV transmitting lens—and a generalisation of the Fourier transform to the sphere known as spherical harmonics (SH) to encode the UV-segmented skyline. As described in section III, this produces descriptors that are invariant to illumination, roll, pitch and yaw, to which a sequence matching approach can be applied to enable robust place recognition along a previously traversed route. We collected novel datasets on routes with substantial condition and tilt variation (section IV), and demonstrate successful performance of our method in section V.

II. BACKGROUND

The low cost, low power consumption, and compactness of cameras has produced much interest in developing robotic navigation systems that operate from camera imagery alone. A standard approach consists of detecting features in the visual field and tracking these in subsequent views so as to reconstruct the translation and rotation of the camera relative to the incrementally expanding feature map [2], [3]. The use of rotation- and scale-invariant features makes these methods robust to different viewpoints [10]. However, they are prone to failure under different lighting conditions [11] and are also challenged in settings that lack reliable, unique features. Recent forays into using features learnt using deep convolutional networks [12] have yielded some success, but their ability to deal with appearance variation is still limited to moderate changes. These methods are also typically computationally intensive and require relatively high-resolution, high-quality imagery; constraints that are not necessarily feasible on small, low-cost robot platforms.

In contrast, methods based on global features or whole images have proven to be highly invariant to substantial lighting changes, even when using low resolution imagery [4]. However, the performance of whole image and global descriptor-based methods typically degrades rapidly when camera viewpoint changes upon multiple revisits to a place [13]. Efforts have been made to develop viewpoint invariant full-image localisation algorithms, for example, by generating multiple views for each location [14]. This solves particular problems such as the translation due to lane position, but it soon becomes intractable to generate views that deal with all types of pose invariance due to roll, tilt and yaw, especially if computational resources are limited. For less agile robot platforms, this problem can be partially addressed by the use of accelerometers that can compensate for tilt on the fly by remapping the view. However, a gravity vector cannot be reliably extracted from an accelerometer on a platform that is regularly accelerating and decelerating [15], and the use of high sampling rate inertial measurement units (IMUs) adds substantial expense.

A potential image-based solution to this problem is to use an omnidirectional camera and a compact image encoding that is pose invariant. In principle, lower order SH can be used in this way by converting an image to greyscale, projecting it onto a hemisphere and filling in the missing areas through interpolation or reflection before calculating the coefficients [16]. However, using SH precludes the ability to incorporate the methods used to ameliorate illumination variation, such as local histogram equalisation on neighbouring pixels (patch normalisation) used in vanilla SeqSLAM [4]. Moreover, unless the full spherical view can be captured, which is difficult for any practical robot sensor configuration, this method can also fail to handle robot pitch and roll. For example, if only the upper hemisphere is captured, then even small camera pose variations can cause substantial differences between two images and their harmonics (fig. 1c). Detecting and correcting for tilt may not suffice because different areas of the view will be missing and would need to be interpolated.

Both limitations can potentially be overcome by preprocessing the images to extract an illumination invariant feature that remains visible in a hemispherical view under tilt and yaw. Such a feature in outdoor scenes is the boundary between sky and ground. The shape of the sky at a particular location can provide a unique signature [17], [18]. Due to low reflectivity of most objects in the ultraviolet (UV) range [19], UV sensors are particularly suitable for segmenting the sky from images with minimal computational cost [20], [21]. We have previously shown that sky extraction using a UV-adapted camera can be successfully combined with SeqSLAM for navigation, albeit on a large and stable platform traversing flat city streets [22]. Here we develop a UV-based tilt-invariant place recognition system by combining this novel sensing modality with the use of SH.

III. APPROACH

In this section we describe the novel method for extracting skylines using a UV-sensitive camera, and encoding a tilt-, rotation- and illumination-invariant snapshot of each place the robot visits based on SH, which can then be fed into a sequence-based place recognition algorithm.

A. Skyline Extraction using UV

Many natural materials show only a low reflectivity to light with short wavelengths, while the light emitted by the sky contains a comparably high portion of UV light [21]. Sensors which are only sensitive to shorter wavelengths will therefore perceive a high contrast between ground objects and the sky [22]. We use a camera with a UV-sensitive lens (see section IV) to extract a skyline by taking an average of all channels in the image and applying a threshold using Otsu's method [23]. The resulting black/white image is invariant to changes of the lighting conditions as long as a classification between the sky and ground is possible. Moreover, when two binary hemispherical images from the same location are tilted differently, we can safely assume that the non-overlapping parts of the image are largely black, due to the

natural position of the sky falling well within the upper hemisphere (fig. 1c). Once a binary sky shape has been extracted, matching it to a previously seen view reduces to a 2D shape recognition problem. Solutions to translation- and rotation-invariant shape recognition problems exist (e.g. [24], [25]). However, as our sky is mapped on a 3D sphere we need to apply a suitable basis that can be used for recognition of shapes on spheres, such as SH [26].

B. Real Spherical Harmonics

The generalisation of the Fourier transform to the sphere is a well-studied topic in the field of Fourier analysis [27], [28]. While the standard approach is to use complex-valued SH, instead we use the real-valued equivalent, real spherical harmonics (RSH), to avoid computations with complex numbers. Since both bases differ only in their representation we will in the following denote them by SH. Fourier analysis states that there exists a base of real-valued orthogonal functions $y_m^l : S^2 \to \mathbb{R}$ on the sphere with

$$y_m^l(\vartheta,\varphi) := \begin{cases} \sqrt{2}N_m^l \cos(m\varphi)P_m^l(\cos\vartheta) & m > 0 \\ \sqrt{2}N_m^l \sin(-m\varphi)P_{-m}^l(\cos\vartheta) & m < 0 \\ N_0^l P_0^l(\cos\vartheta) & m = 0 \end{cases} \quad (1)$$

and $l \in \mathbb{N}, -l \leq m \leq l$, where l denotes the *band* of the corresponding SH, P_m^l are the *associated Legendre polynomials* and $N_m^l = \sqrt{\frac{2l+1}{4\pi}}\sqrt{\frac{(l-|m|)!}{(l+|m|)!}}$ is a normalisation term. Note that we used spherical coordinates (ϑ, φ) in (1). For an efficient implementation the values $y_m^l(\vartheta,\varphi)$ can be calculated by using recursive formulas, e.g. [29]. Any real-valued function $f : S^2 \to \mathbb{R}$ can be expressed in terms of SH by a projection

$$c_m^l := \langle f, y_m^l \rangle = \int_S f(s) y_m^l(s) ds \in \mathbb{R}, \quad (2)$$

such that $f \approx \sum_{l,m} c_m^l y_m^l$ converges quadratically in $L^2(S^2)$ with an increasing value of l. Therefore each function f can be approximated by l^2 coefficients c_m^l by using the first l bands of the SH.

By defining the amplitude spectrum on the SH as

$$a^l := \sqrt{\sum_{m=-l}^{l} (c_m^l)^2}, \quad (3)$$

a rotational-invariant measure can be deduced analogously to the amplitude spectrum in the standard Fourier transform. It contains a total of l entries for a function approximated up to band l. Due to the missing phase information, the total information contained in the amplitude spectrum is less then the information provided by the function f.

C. Place Recognition Algorithms

SeqSLAM is a place recognition algorithm that uses whole images as global features [4] and has been shown to be effective using very low-resolution, blurred images [30]. It works by repeatedly searching previously stored images for locally good matches with the current image. Linear sequences of candidate images are interpreted as a high probability of being in the corresponding location. SeqSLAM can be applied to feature vectors as well as images, and we here apply it to vectors of SH amplitude coefficients, and compare this to vanilla SeqSLAM (i.e., using greyscale images). In the following we will denote the original by *SeqSLAM (vanilla)* and our approach by *SeqSLAM (sky)*.

OpenFABMAP is an implementation of FAB-MAP 2.0 that uses feature detection to build a visual vocabulary [2]. This is then used for a bag-of-words [31] approach to approximate the likelihood of a particular frame matching the current view. Since FAB-MAP is a feature-based method it is invariant to changes in the viewpoint but encounters problems if the feature detection is disturbed (blur, lighting changes, or featureless environments).

IV. EXPERIMENTAL SETUP

In order to evaluate the robustness of the proposed technique to deal with tilt and appearance variation in real environments, we first conducted a diagnostic tilt-controlled experiment, then evaluated the system in two challenging robot scenarios. See attached video for details.

A. Cameras and Rigs

To capture our datasets we used two identical cameras (GoPro Hero 3+ Black) mounted with different lenses. The first camera had a RageCams $1/3.2''$ panoramic fisheye lens with a focal length of $1.19\,\mathrm{mm}$, $185°$ field of view and an infrared (IR) cut filter, which is able to collect RGB images. The second camera is mounted with a similar, but modified, fisheye lens provided by skylinesensors.com. The lens has an additional custom IR cut filter and UV bandpass filter, with a peak response around $350\,\mathrm{nm}$. Raw camera data was recorded with the same settings on both (Video resolution "1080p super view", 24 FPS, wide field of view, automatic exposure time). The only significant cost for making this camera sensitive to the UV spectrum was the filter. Currently, a one-off filter with custom response costs several hundred dollars, but if mass produced would cost far less. Costs could be further reduced by mass producing low pixel count, small sensor arrays sensitive to UV.

The cameras were used in two different experimental rigs (fig. 2). The first rig consisted of each camera mounted (one at a time) on the top of a helmet with an adjustable tilting angle α (fig. 2a) to record a ride on a bicycle. The second rig consisted of an off-the-shelf remote control car (RCC) (Drive&Fly-models TruckFighter) with both cameras mounted next to each other on the top of the RCC, replacing the plastic chassis (fig. 2b).

B. Datasets

The City dataset (fig. 3a, 4a) formed the basis for controlled tilt evaluation. It was collected using the bike helmet rig on a circular track of $520\,\mathrm{m}$ along urban streets in Bielefeld at around 3pm. The route was traversed eight times with varying tilting angles $\alpha \in (-30, -20, \ldots, 30)$—where $\alpha = 0$ has been recorded twice—and with each camera

(a) Variable tilt (b) RCC

Fig. 2: (a) Side-view of the bicycle helmet setup with a single camera mounted at each run. The additional tilting is denoted by the angle α, where $\alpha = 0$ with the camera approximately level. (b) Top-view of the RCC platform mounted with the RGB and UV cameras simultaneously, both pointing upwards.

(a) City (variable tilt) (b) BMX Track (RCC) (c) Disposal Site (RCC)

Fig. 3: Aerial imagery of the three different testing environments, with the red arrow line indicating the route, traversed at least twice for each experiment. Map data: Google, Geobasis-DE/BKG.

(a) City (variable tilt) (b) BMX Track (RCC) (c) Disposal Site (RCC)

Fig. 4: Photos of the three testing environments taken with a regular digital camera, showing the wide range of urban, forested and industrial areas encountered.

(RGB, UV), totalling 16 runs. Over the course of all the runs, substantial scene dynamics were encountered including those due to varying lighting conditions and traffic.

The other two data sets were gathered using the RCC rig, on two different days, at different times of day, to increase the visual dynamics (e.g. obstacles, illumination) of the observed scenes. The BMX Track data (fig. 3b, 4b) was collected in an area close to Bielefeld university and contains a high diversity of visual cues. The $600\,\mathrm{m}$ route starts on a bicycle motocross (BMX) track in a forest, containing sloping terrain and a heavy forest canopy with minimal unique landmarks. The route then leaves the forest and continues along a border between a forest (right side) and the gardens of adjoining houses (left side). Finally, the route ends along a suburban street. The Disposal Site dataset (fig. 3c, 4c) was recorded on a highly repetitive track in Borgholzhausen, presenting potential aliasing issues. The $510\,\mathrm{m}$ track consists mainly of a road through a forest marked by a single road junction, a disposal site and a small building.

To ground truth the datasets, key corresponding frames in all environment traverses were visually identified. Image pairs between these key frames were matched through linear interpolation. When calculating precision, images were deemed a correct match when the estimated frame correspondence fell within 5 frames from the ground truth.

C. Parameter Values

OpenSeqSLAM was run using all default parameters: $ds = 10$, $v_{min} = 0.8$, $v_{max} = 1.2$, $r_{window} = 10$. For OpenFABMAP, we trialled a range of different detectors used at varying resolutions. For best performance we used 480×270 pixel images with Speeded Up Robust Features (SURF) [10] detectors and extractors to build a visual vocabulary, using training data supplied from a separate video dataset recorded in similar environments for both the BMX Track and Disposal Site datasets. All other parameters were kept as provided in the latest sample script from the OpenFABMAP public repository (accessed Aug. 2015).

For SeqSLAM (sky) we used 2000 pixels to sample the upper hemisphere from the UV images and Otsu's method for the binarisation which does not depend on any parameters. A total of $l = 120$ bands were used for the creation of the feature vector containing a total of 14400 entries per input image (equivalent to the data contained in a square 120×120 pixel monochromatic image).

V. RESULTS

A. City Tilt Variation

As can be seen in figure 5b, place recognition performance using SeqSLAM (vanilla) and SeqSLAM (sky) is broadly comparable at low tilt angles, but diverges rapidly as the tilt angle increases. At a tilt difference of $60°$, a recall of nearly $20\,\%$ can still be achieved with a precision of $100\,\%$ using the UV images. The vanilla method is unable to correctly match any images at a similar precision level, even at $50°$. Figure 6 shows an example of two frames that were successfully matched using the sky segmented imagery.

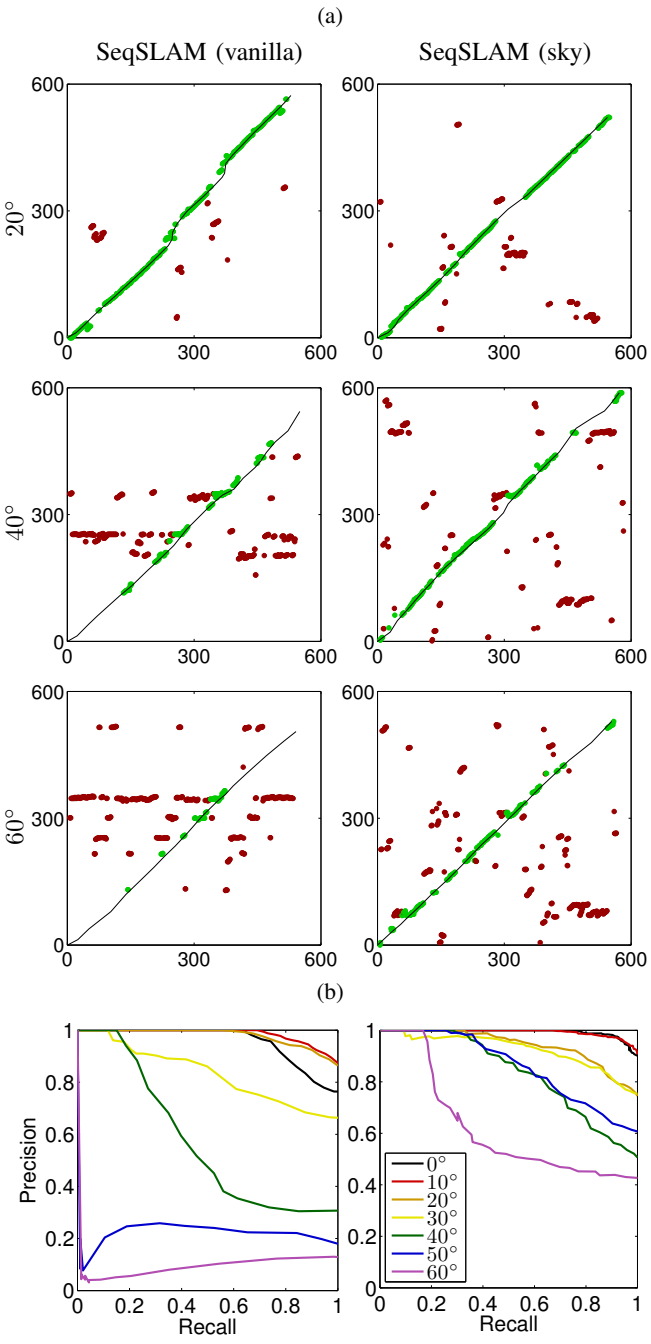


Fig. 5: (a) Frame correspondences (place matches between the test and training environment traverses) on the City tilt set at $20°$, $40°$, and $60°$ tilt differentials. Left: vanilla SeqSLAM. Right: improved performance achieved with sky SeqSLAM, both in terms of total number of correct matches and matching coverage throughout the dataset. Solid line: ground truth; Green dots: near ground truth; Red dots: far from ground truth; Axis units: frame indices. (b) Precision versus recall at various degrees of tilt. Runs were captured at values of α between $-30°$ and $30°$ and compared by pairing $\alpha_1 \approx -\alpha_2$. As the difference in tilt between runs increases, performance degrades more slowly for sky SeqSLAM than for vanilla SeqSLAM.

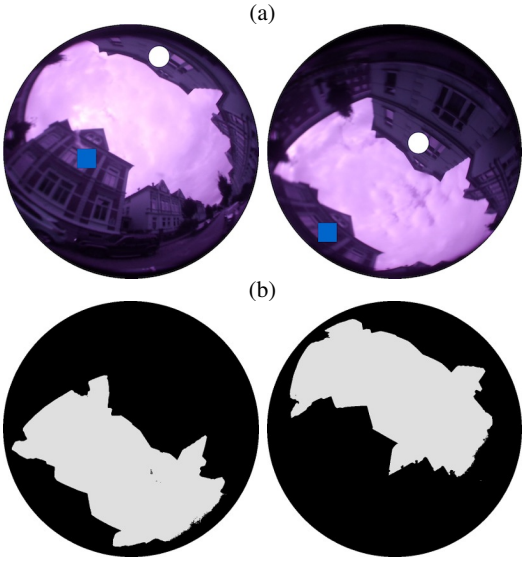


Fig. 6: (a) Two hemispherical UV images taken at the same location with a 60° tilt offset from each other. The white circle and blue square show corresponding parts of the images. Distortion and misalignment results in poor place recognition performance with vanilla SeqSLAM. (b) Extracting the shape of the sky gives an excellent match using SH amplitudes.

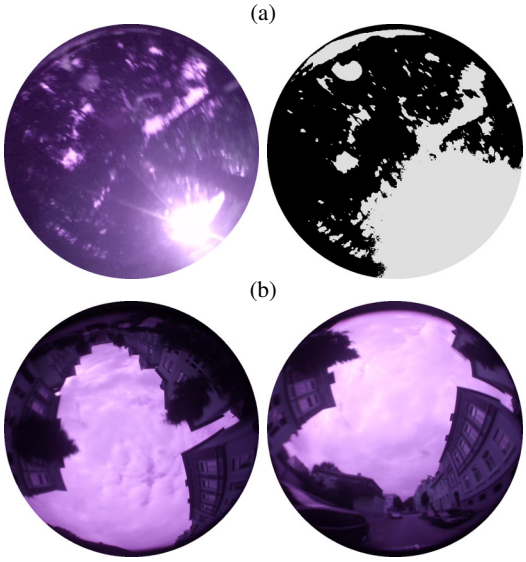


Fig. 7: Failure cases: (a) (left) An image with bright direct sunlight in a dark area causing high amounts of lens flare, leading to (right) incorrect segmentation of sky and ground. (b) Strongly tilted views recorded at the same position, which can not be matched due to cropping of the UV segmented skyline image.

The difference in matching performance can also be examined in the frame correspondence plots shown in figure 5a. At tilt angles of 20°, 40° and 60°, the UV-based method finds more true positive place matches than the vanilla SeqSLAM method. Place recognition coverage throughout the route is also improved.

B. BMX Track

As described in section IV and as shown on the map in figure 3b, the BMX Track dataset consists of three sections which strongly differ in their visual appearance. This particularly affects the correspondences found by FAB-MAP (fig. 8a), which can mostly distinguish between the three different areas on the track but struggles to find the RCC's exact location, with precision no better than 50 % for most recall levels (fig. 8d). In comparison, both SeqSLAM methods show good performance on the complete BMX track at 100 % precision, with recall rates of 33 % and 37 % for the vanilla and UV segmented skyline approach, respectively. Vanilla SeqSLAM tends to fail on parts of the track that include misaligned images due to either tilt or rotation. The precision levels achieved while maintaining 100 % recall (the complete dataset) are 68 % for vanilla and 83 % for sky SeqSLAM respectively, showing superior localisation with the UV segmented skyline. One problem we encountered on this dataset is lens flare, leading to errors during the skyline segmentation as can be seen in figure 7a. However, we found that the restricted view of the sky through foliage (fig. 1d) did not impact the performance of SeqSLAM using UV segmented skyline images, although no clear skyline can be seen. Instead, the spots of sky visible through the canopy were sufficient on our dataset to localise the RCC.

C. Disposal Site

On the Disposal Site all algorithms performed worse, probably due to the highly repetitive nature of the track (fig. 9d). FAB-MAP failed catastrophically using the best performing parameters we could find (fig. 9a). Vanilla SeqSLAM failed due to a combination of zigzagging (rotational differences) and the strongly repetitive visuals, achieving poor frame correspondence towards the end of the route (fig. 9b) and a precision level of 33 % at 100 % recall. Using the sky does not completely fix this problem as there is no viewpoint invariance for translations across the x-y plane. Despite this our algorithm performed well (fig. 9c) with 64 % precision at 100 % recall.

D. Tilt Invariance versus Sequence Length

As indicated in figures 8d and 9d the precision rates can be improved by increasing the sequence length used in sky SeqSLAM. With a sequence length of 20, we obtain a precision level of 89 % for the BMX track and 70 % for the Disposal Site at 100 % recall. A comparison using single images (rather than sequences) shows a surprisingly high precision level of 64 % and 49 % at 100 % recall over the whole dataset, despite the fact that the amplitude spectrum generally provides ambiguous information for images.

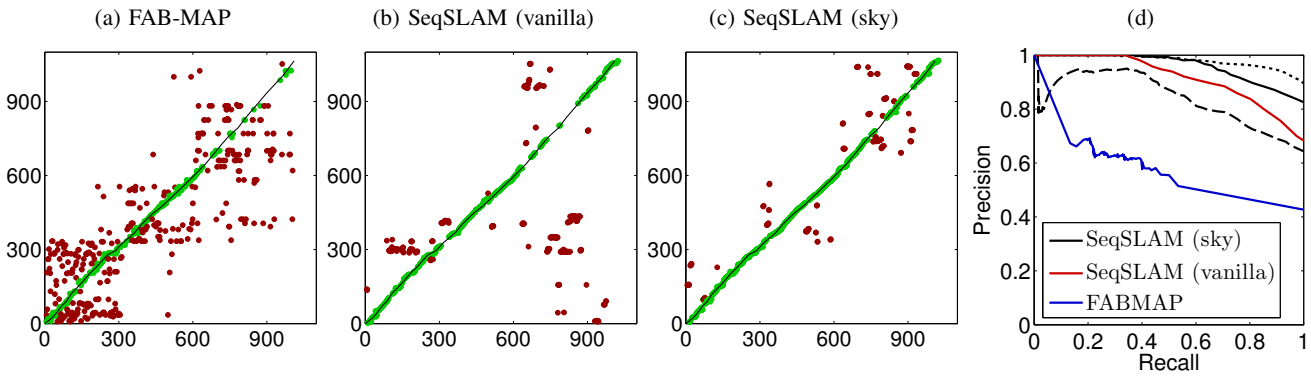


Fig. 8: (a-c) Frame correspondences on the BMX Track dataset. Using UV sky segmentation improves the matching performance over both FAB-MAP and vanilla SeqSLAM. (d) Precision vs. recall. SeqSLAM (sky) performs best, and can be improved by increasing the sequence length from the default of $w = 10$, which equates to 2 seconds of travel, to $w = 20$ (dotted line). Dashed line is ($w = 1$).

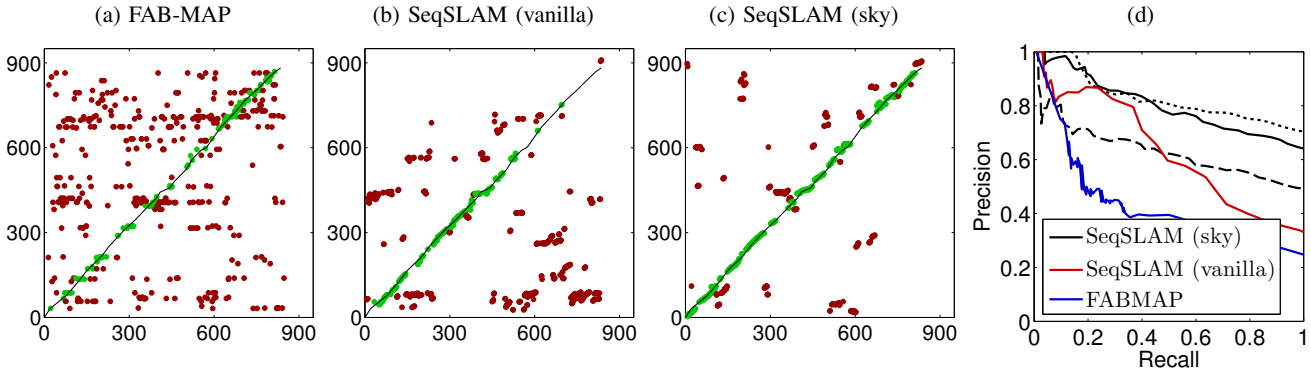


Fig. 9: (a-c) Frame correspondences on the Disposal Site dataset. As well as improving overall matching performance, the UV sky segmented implementation particularly improved performance at the end of the dataset where aggressive zig-zag manoeuvring was performed. (d) Precision vs. recall. Performance of adjusted sequence length using UV is shown by dashed ($w = 20$) and dotted ($w = 1$) lines, as in fig. 8d.

E. Computational Efficiency

In our implementation, the computation of the Fourier transform is $O(mn^2)$ and the amplitude spectrum $O(n^2)$, where m is the number of sample points and n the number of bands. For the presented results we used $n = 120$ bands and $m = 2000$ visual points equally distributed over the sphere. The run-time for both the Fourier transform and the calculation of the amplitude spectrum is around $40\,\mathrm{ms}$ on a Lenovo T-530 laptop. An additional speed-up could be gained by reducing the number of used bands and/or visual points.

VI. DISCUSSION AND FUTURE WORK

In this paper we have presented a novel method for place recognition that combines UV-sensitive camera imaging with a compact tilt- and rotation-invariant skyline representation using SH amplitudes. The proposed system is particularly relevant for low-cost, high-speed robotic platforms operating over challenging terrain and across changing environmental conditions. The UV-modification can be cheap and provides appearance robustness and the SH representation provides invariance to the major viewpoint change challenges faced by these platforms. This can be exploited for use in a sequence-based matching algorithm, which would not work on tilt-sensitive imagery. The controlled tilt and robot platform experiments presented here demonstrate that the approach is feasible over a wide range of environments and platforms, and performs significantly better than current state of the art condition-invariant (SeqSLAM) and viewpoint-invariant (FAB-MAP) approaches. While SH are a natural choice for the stated problem, other shape recognition techniques could be adapted to work on a sphere. Here we discuss some of the shortcomings of this approach and promising areas of future work.

In large flat areas such as a desert, the use of a skyline feature is less beneficial, both due to homogeneity of the skyline and the introduction of skyline distortion at a smaller tilt angle than in wooded or urban environments where the skyline is high. These shortcomings could be somewhat mitigated if a greater than $180°$ sensor was used, such as the Ricoh M15, a full $360° \times 180°$ dual fisheye camera.

Like all camera-based methods, our method is sensitive to

extreme lens flare, such as that caused by direct sunlight very early or late in the day. In such situations, ground pixels can be labelled as sky during segmentation, leading to degraded place recognition performance. This problem could potentially be overcome in a number of ways; firstly by semantic recognition of similar situations and a consequent downgrading of the place recognition confidence, or secondly by attempting to match snapshots using a mask that excludes or adapts areas of maximum brightness in the image, similar to the approach in [20].

Although we have applied this system on high-speed ground-based platforms, it could also have utility on aerial platforms which can rapidly change their attitude when moving. The method would have particular advantages in complex environments such as cities and forests.

The use of UV and skyline for navigation was inspired by research on the behaviour and visual system of ants. While it is not explicitly known how ants represent or process this information, a basis encoding such as SH is a possibility that would be consistent with current hypotheses for visual processing in other animals [32]. Investigating the underlying neural mechanisms in the insect brain could lead to further insight into compact and low cost computation for navigation.

ACKNOWLEDGMENT

This work was supported by a Scottish Informatics and Computer Science Alliance Distinguished Visting Fellow grant, Research Council Future Fellowship FT140101229 and Microsoft Research Faculty Fellowship to Michael Milford. We thank the Deutsche Forschungsgemeinschaft (DFG) for financial support (grant numbers MO 1037/8-1 and MO 1037/10-1), and also acknowledge grants EP/F500385/1 and BB/F529254/1 for the University of Edinburgh School of Informatics Doctoral Training Centre in Neuroinformatics and Computational Neuroscience (www.anc.ac.uk/dtc).

REFERENCES

[1] I. Ulrich and I. Nourbakhsh, "Appearance-based place recognition for topological localization," in *IEEE International Conference on Robotics and Automation (ICRA)*, vol. 2. IEEE, 2000, pp. 1023–1029.

[2] M. Cummins and P. Newman, "Highly scalable appearance-only SLAM - FAB-MAP 2.0." in *Robotics: Science and Systems*, vol. 5. Seattle, USA, 2009.

[3] A. J. Davison, I. D. Reid, N. D. Molton, and O. Stasse, "MonoSLAM: Real-time single camera SLAM," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, pp. 1052–1067, 2007.

[4] M. J. Milford and G. F. Wyeth, "SeqSLAM: Visual route-based navigation for sunny summer days and stormy winter nights," in *International Conference on Robotics and Automation (ICRA)*. IEEE, 2012, pp. 1643–1649.

[5] L. Matthies, Y. Xiong, R. Hogg, D. Zhu, A. Rankin, B. Kennedy, M. Hebert, R. Maclachlan, C. Won, T. Frost, *et al.*, "A portable, autonomous, urban reconnaissance robot," *Robotics and Autonomous Systems*, vol. 40, no. 2, pp. 163–172, 2002.

[6] J. Yang, S.-J. Chung, S. Hutchinson, D. Johnson, and M. Kise, "Omnidirectional-vision-based estimation for containment detection of a robotic mower," in *International Conference on Robotics and Automation (ICRA)*. IEEE, 2015.

[7] P. Ardin, M. Mangan, A. Wystrach, and B. Webb, "How variation in head pitch could affect image matching algorithms for ant navigation," *Journal of Comparative Physiology A*, vol. 201, no. 6, pp. 585–597, 2015.

[8] P. Graham and K. Cheng, "Ants use the panoramic skyline as a visual cue during navigation," *Current Biology*, vol. 19, no. 20, pp. R935–R937, 2009.

[9] R. Möller, "Insects could exploit UV–green contrast for landmark navigation," *Journal of Theoretical Biology*, vol. 214, no. 4, pp. 619–631, 2002.

[10] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (SURF)," *Computer vision and image understanding*, vol. 110, no. 3, pp. 346–359, 2008.

[11] C. Valgren and A. J. Lilienthal, "SIFT, SURF & seasons: Appearance-based long-term localization in outdoor environments," *Robotics and Autonomous Systems*, vol. 58, no. 2, pp. 149–156, 2010.

[12] N. Sünderhauf, S. Shirazi, A. Jacobson, F. Dayoub, E. Pepperell, B. Upcroft, and M. Milford, "Place recognition with convnet landmarks: Viewpoint-robust, condition-robust, training-free," *Proceedings of Robotics: Science and Systems XII*, 2015.

[13] N. Sünderhauf, P. Neubert, and P. Protzel, "Are we there yet? Challenging SeqSLAM on a 3000 km journey across all four seasons," in *Proc. of Workshop on Long-Term Autonomy, International Conference on Robotics and Automation (ICRA)*. IEEE, 2013, p. 2013.

[14] M. Milford, C. Shen, S. Lowry, N. Sünderhauf, S. Shirazi, G. Lin, F. Liu, E. Pepperell, C. Lerma, B. Upcroft, and I. Reid, "Sequence searching with deep-learnt depth for condition- and viewpoint-invariant route-based place recognition," in *Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*. IEEE, 2015.

[15] S. Thurrowgood, D. Soccol, R. J. Moore, D. Bland, and M. V. Srinivasan, "A vision based system for attitude estimation of UAVs," in *International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2009, pp. 5725–5730.

[16] H. Friedrich, D. Dederscheck, M. Mutz, and R. Mester, "View-based robot localization using illumination-invarant spherical harmonics descriptors," in *Proceedings of the International Joint Conference on Computer Vision and Computer Graphics Theory and Applications (VISAP)*. INSTICC, 2008.

[17] F. Stein and G. Medioni, "Map-based localization using the panoramic horizon," in *International Conference on Robotics and Automation*. IEEE, 1992, pp. 2631–2637.

[18] J.-i. Meguro, T. Murata, Y. Amano, T. Hasizume, and J.-i. Takiguchi, "Development of a positioning technique for an urban area using omnidirectional infrared camera and aerial survey data," *Advanced Robotics*, vol. 22, no. 6-7, pp. 731–747, 2008.

[19] T. Kollmeier, F. Röben, W. Schenck, and R. Möller, "Spectral contrasts for landmark navigation," *Journal of the Optical Society of America A*, vol. 24, no. 1, pp. 1–10, 2007.

[20] M. H. Tehrani, M. Garratt, and S. Anavatti, "Horizon-based attitude estimation from a panoramic vision sensor," in *Embedded Guidance, Navigation and Control in Aerospace*, vol. 1, no. 1, 2012, pp. 185–188.

[21] D. Differt and R. Möller, "Insect models of illumination-invariant skyline extraction from UV and green channels," *Journal of theoretical biology*, vol. 380, pp. 444–462, 2015.

[22] T. Stone, M. Mangan, P. Ardin, and B. Webb, "Sky segmentation with ultraviolet images can be used for navigation," in *Proceedings Robotics: Science and Systems*, 2014.

[23] N. Otsu, "A threshold selection method from gray-level histograms," *Automatica*, vol. 11, no. 285-296, pp. 23–27, 1975.

[24] M.-K. Hu, "Visual pattern recognition by moment invariants," *Transactions on Information Theory*, vol. 8, no. 2, pp. 179–187, 1962.

[25] A. Khotanzad and Y. H. Hong, "Invariant image recognition by Zernike moments," *Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, no. 5, pp. 489–497, 1990.

[26] M. Kazhdan, T. Funkhouser, and S. Rusinkiewicz, "Rotation invariant spherical harmonic representation of 3D shape descriptors," in *Symposium on geometry processing*, vol. 6, 2003, pp. 156–164.

[27] W. E. Byerly, *An elementary treatise on Fourier's series and spherical, cylindrical, and ellipsoidal harmonics, with applications to problems in mathematical physics*, 1st ed. London, England: Ginn & Company, 1893.

[28] G. B. Folland, *Fourier analysis and its applications*, 1st ed. Providence, Rhode Island: American Mathematical Society, 1992.

[29] W. Press, S. Teukolsky, W. Vetterling, and B. Flannery, *Numerical recipes in C: The art of scientific computing*, 2nd ed. Cambridge, England: Cambridge University Press, 1992.

[30] M. Milford, "Vision-based place recognition: How low can you go?" *The International Journal of Robotics Research*, vol. 32, no. 7, pp. 766–789, 2013.

[31] H. M. Wallach, "Topic modeling: Beyond bag-of-words," in *Proceedings of the 23rd international conference on Machine learning*. ACM, 2006, pp. 977–984.

[32] B. A. Olshausen *et al.*, "Emergence of simple-cell receptive field properties by learning a sparse code for natural images," *Nature*, vol. 381, no. 6583, pp. 607–609, 1996.

4.3 ROTATION INVARIANT HOMING IN A SIMULATED ANT WORLD

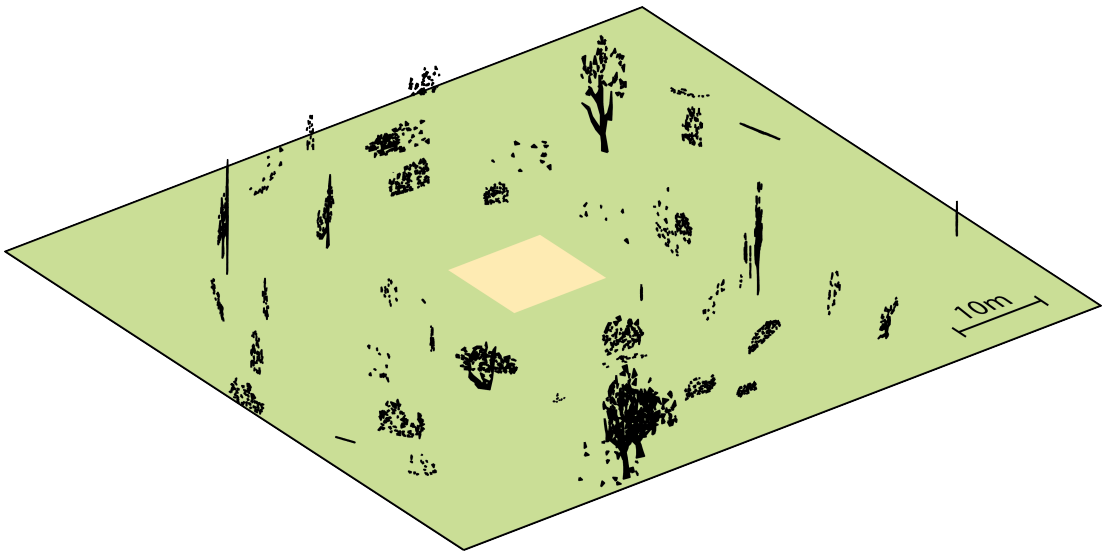
The previous section described a useful application of a frequency representation to robot navigation. An interesting follow-up question is whether a similar mechanism could occur to help insects to localise in a rotation invariant manner. This section describes a similar approach to invariant place recognition, applied to an insect behavioural simulation as an attempt to explain certain behaviours. The work was presented as a poster at the 2016 international congress of neuroethology (ICN) (Appendix 7.2).

Many aspects of visual navigation in ants can be explained through retinotopic matching of the current scene with a visual memory. Models have exploited this property to successfully recreate both route following and visual homing behaviours observed in ants (Zeil et al., 2003; Baddeley et al., 2012). Crucially, visual familiarity peaks when the animal is facing the same direction as when the memory was stored but drops as the animal rotates. Yet, recent behavioural data suggest that ants can also measure visual familiarity independent of their viewing direction (Ardin et al., 2016b).

A rotationally independent representation of the current location can be achieved by Fourier transform, which has previously been shown as useful for insect models of navigation by visual homing. Stürzl and Mallot (2006) extracted a horizontal strip of greyscale values from across the middle of images from a panoramic camera mounted on a robot. The array was converted to a frequency representation and successfully used to home, with catchment areas being more pronounced and narrow for higher frequencies and wider for lower frequencies. The proposed method of homing in this section works similarly to the localisation algorithm outlined earlier in this chapter, by ignoring phase information and descending the gradient of image familiarity, defined by the similarity of frequency amplitudes, towards the nest in an semi-exploratory manner. In the approach outlined in the following section, the entire sky/ground image is transformed to a frequency representation rather than simply a 1D signal.

4.3.1 *Methods*

A virtual ant world previously made by Baddeley et al. (2012) was used to create panoramic greyscale views (Fig. 4.3a). Views were generated in a 10×10 m box at 2 cm intervals (Fig. 4.3b). Twenty evenly spaced locations were used as reference positions, to which all other locations within the box were compared. Two methods of comparison were used: retinotopic and using amplitude coefficients of ZMs of the panoramic images. The Euclidean distance between moments at each location was



(a) 3D Computer model based on ant environment.

(b) Panoramic view in ant world

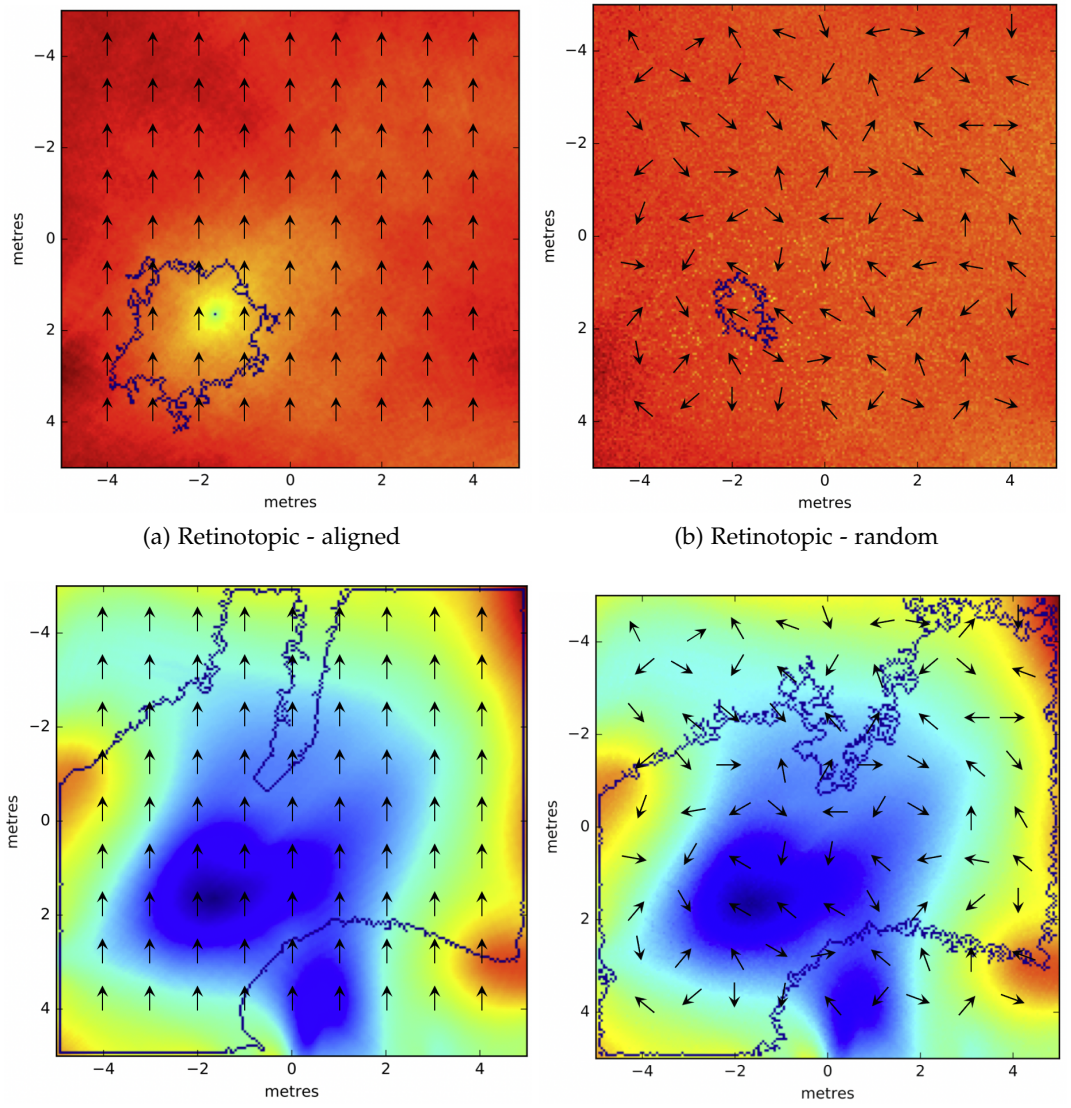
(c) Wrapped panorama.

Figure 4.3: **Simulated ant world. (a)** A simulated world was populated with random vegetation at different densities creating an ant-like environment in which 360° panoramic views were generated every 2 cm in the test area (light orange) to test homing models. **(b)** Panoramic view from world. **(c)** Images were wrapped before converting to a Zernike moments (ZM) representation.

plotted on a 2D surface and gradient descent was used to determine which points would fall within the basin of attraction by iteratively moving a small step in the direction of steepest descent, where the local gradient was inferred using bicubic interpolation. If the starting location eventually led back to the nest it was deemed to be within the catchment area (Fig. 4.4).

4.3.2 *Results*

The catchment area for a frequency representation was significantly larger than using retinotopic matching (Fig. 4.4 and 4.5). Encoding images in the frequency domain can result in algorithms like visual homing working in simulated ant environments, without the need for mental rotations of image stored. The results can be explained by visualising two images that have been displaced in a virtual environment, while



(a) Retinotopic - aligned

(b) Retinotopic - random

(c) Frequency - aligned

(d) Frequency - random

Figure 4.4: **Frequency based homing, catchment areas.** All plots show similarity to a panoramic image from the same reference location measured at 2 cm intervals in the virtual environment. In *(a)* and *(c)* the images are taken facing north, in *(b)* and *(d)* with random headings. In *(a)* and *(b)* comparison is pixelwise, in *(c)* and *(d)* by taking the Euclidean distance between vectors made up by Zernike moments (ZM) amplitude coefficients. Blue contour lines indicate the border of the catchment area.

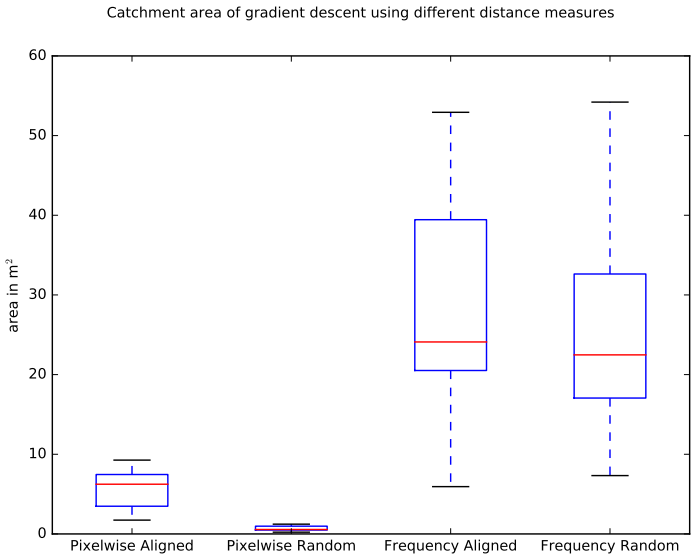


Figure 4.5: **Homing performance.** $N = 20$. Retinotopy provides guidance but only when aligned with memory. Frequency encoding performs well, regardless of orientation.

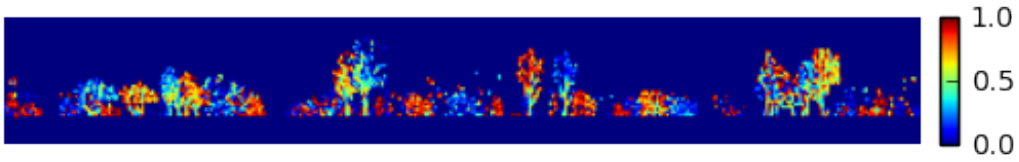
still allocentrically aligned (Fig. 4.6). Parallax errors cause landmarks across the horizon to be misaligned (Fig. 4.6c), however the frequency approach, especially at low frequencies is more robust (Fig. 4.6d). Seemingly poor performance of the catchment area on allocentrically aligned images is most likely due to the higher spatial resolution used on this study, showing that a simple gradient descent using those features is hard, even in uncluttered environments.

This could provide a first explanation to some observed behaviours. A biologically plausible implementation to extract the amplitude at various frequencies would be through randomly aligned sinusoidal receptive fields, which is plausible given the ants known optics and early visual processing, such as bar-sensitive, orientation-biased cells in insect lobula (O'Carroll, 1993) or lateral triangle (Seelig and Jayaraman, 2013). Scene recognition in this manner would resemble the use of gist descriptors used in computer vision, where the image is convolved with gabor filters at different frequencies and orientations (Oliva and Torralba, 2001).

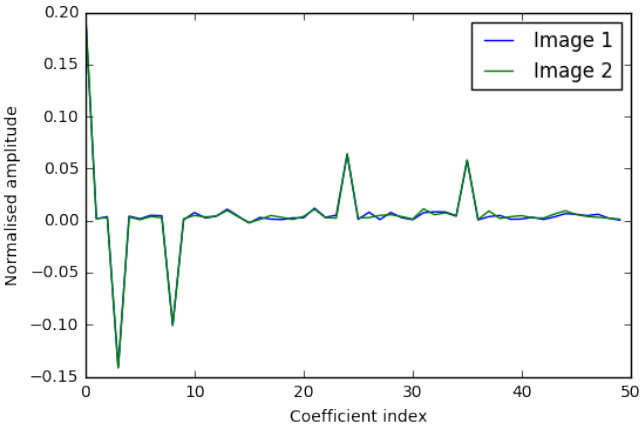
If ants are able to use rotation invariant features, why would they exhibit scanning behaviour (Wystrach et al., 2014)? The most obvious explanation is that some phase information is retained, meaning the alignment is still useful. Although highly speculative, a different interpretation of scanning behaviour in this context could be the re-sampling of the visual environment to obtain a higher fidelity estimate of the amplitude signature. This could also partially address the fact that in desert ants and many other insects there is a 60° blind spot that is not covered by their visual field. A

(a) North-aligned image 1

(b) North-aligned image 2 (displaced by 3.6 metres)



(c) Difference between images (red indicates the maximum possible difference)



(d) First 50 coefficients of Zernike transform amplitude of both images

Figure 4.6: **Explanation of poor performance of retinotopic matches between allocentrically aligned images.** *(a,b)* Two panoramic views taken 3.6 metres apart in the virtual ant environment shown in figure 4.3a. *(c)* Absolute image difference between the two images. Even a relatively small translational displacement in an open area with only distant landmarks can cause the majority of matching skyline features to misalign. *(d)* Conversely, normalised Zernike amplitude coefficients of each image remain relatively similar.

similar role could be suggested for the dung beetle dance, a scanning-like behaviour performed before rolling the ball (El Jundi et al., 2016).

4.4 CONCLUSION

Findings described in this chapter build upon work in chapter 3, showing that besides being a stable, compact feature, the skyline provides additional benefits for localisation. A frequency representation of the skyline works well for localising on bumpy terrain, retaining advantages of whole image matching such as condition invariance while being somewhat invariant to pose. Some future applications of this work are outlined below.

4.4.1 *Navigation under foliage*

Surprisingly, extracted sky features worked well for localising in forests by matching patches of light in the canopy. This could perhaps have implications for understanding how insects manage to visually navigate in highly foliated areas, such as the wood ant *Formica rufa* (Graham and Collett, 2002).

4.4.2 *Learnt features*

The trend for many computer vision problems is towards using convolutional neural networks (CNNs), where features are learnt rather than hand-designed (Sec. 2.1.1). A major advantage of these features is their invariance to many types of distortion including pose and lighting conditions (LeCun et al., 2004). For now, most localisation implementations through CNNs have focused on pre-trained networks from object recognition datasets (Sünderhauf et al., 2015). Localisation algorithms need to be invariant to specific types of change, such as new development of buildings, foliage from seasonal changes and weather. It should become clear in the near future whether training on visual localisation datasets can overcome many of the problems outlined in the previous two chapters by specifically learning good features that are invariant to the type of changes we encounter when localising in the real world. Early results on this problem can be seen in Arandjelović et al. (2015), where images depicting the same place but taken at different times are used to learn localisation-specific invariant features.

An interesting outcome of this would be to visualise how features learned by the network correspond to hand designed ones. At lower levels we certainly see the Ga-

bor filter style tuning of units, which indicate some type of frequency approximation (Krizhevsky et al., 2012).It would be interesting to investigate whether any higher level features approximated skyline contours as a salient feature.

4.4.3 *Sky/ground separation algorithm*

In this chapter Otsu's algorithm was used to separate sky and ground (Otsu, 1975) in UV images, rather than the watershed method Stone et al. (2014) described in chapter 3. In the previous chapter a third peak was often prominent in the intensity histograms due to sun pixels, which prevented accurate segmentation using conventional methods. As performance between the algorithms was comparable on the datasets described in this chapter, the more conventional segmentation algorithm was opted for. A smaller and less sensitive sensor was used on the robot in this chapter than on the human trials in the previous one, which could have contributed to this discrepancy. It is also plausible that the weather conditions during the data collection could have reduced the advantages of the watershed algorithm and performance gains from using it could still be made in other conditions, especially in sunny weather.

4.4.4 *Improving lookup*

The current SeqSLAM approach uses a grid search to retrieve matching images by searching for diagonal lines in an image similarity matrix (Milford and Wyeth, 2012). For the approach outlined in this chapter, all images are stored by a vector representing the amplitude of different frequencies. A better way to match a sequence of images might be developed by taking inspiration from sequence detection in audio and video files, which is a common operation for tasks such as piracy detection (Zobel and Hoad, 2006) or even song recognition. An example is the song detection algorithm used by Shazam, where peaks at each frequency are extracted and the constellation formed by local peaks on a spectrogram is used as a key to match against stored sequences in a large database (Wang et al., 2004). Applying a similar approach to the skyline frequency representation could help to create a lookup system for sequence based localisation algorithms that would scale to much larger datasets. Challenges include making this representation invariant to changes in speed that would distort the constellations.

Part II

BEES

5

A NEURAL SUBSTRATE FOR PATH INTEGRATION IN THE BEE BRAIN

5.1 INTRODUCTION

In this part, another attempt is made to learn more about the navigational abilities of insects by examining part of their visual system. However, the approach described here uses a neural model, based on observed individual cellular connections in the sweat bee *Megalopta genalis*, to understand the underlying algorithms of path integration. This described model is the first plausible explanation of path integration that is fully grounded in known cell-to-cell connections.

5.1.1 *Justification for modelling the central complex*

It is known that the central complex (CX) provides the substrate for a celestial compass where cues, including the solar azimuth and polarisation patterns in the sky, are combined to determine an allocentric heading (Heinze and Homberg, 2007). A recently discovered group of cells in the noduli of the CX of *Megalopta genalis*, and described in Stone et al. (2016b), display activity proportionate to front-to-back optic flow, essentially acting as a visual odometer. As this species also makes use of the celestial compass, both the components required to determine velocity are now identified at the cellular level which, if combined with a form of memory, are necessary and sufficient to carry out path integration. We know that many insects path integrate using visual input (Collett and Collett, 2000b) and that both speed and heading information converge in the CX. From this it follows that the circuitry in the immediate vicinity of those cells provides a strong candidate for a possible neural implementation of path integration. After tracing connectivity between the speed and compass neurons in the CX, and also including cells immediately downstream, the resulting connectivity pattern resembled a hypothetical model for path integration by Wittmann and Schwegler (1995); Haferlach et al. (2007). These findings justify the investigation of this neural circuitry described in this chapter.

5.1.2 *Modelling approach*

In this chapter, around one hundred cells in the CX were modelled using an average firing rate approach (Dayan and Abbott, 2001), that was chosen for a number of reasons: First, fewer assumptions needed to be made about the firing properties of the cells being modelled. Secondly, the time saved due to both low computational power and ease of network configuration allowed more time to be invested into analysing the circuitry in detail. Finally, using rate-based neurons provided the flexibility to develop a noise-free theoretical model alongside the noisy neural model. Swapping individual modules between models facilitated troubleshooting and improved the overall understanding of how the system operates.

5.2 PUBLICATION: NATURE 2016 (UNDER REVIEW)

This paper was submitted as an article to Nature (Stone et al., 2016b). Personal contributions include:

- Development of the idea of TB1 neurons as a ring attractor due to mutual inhibition.
- Discovery of topological ring structure in TB1, CPU4 and CL1 cells.
- Partial development of the concept of CX as a path integrator with Barbara Webb and Stanley Heinze.
- Expansion of mathematical theory describing why the CX can function as a path integrator.
- Concept of the CX as a holonomic path integrator by using two separate speed inputs. In particular the development of the idea of the TB1 acting as a sinusoid centred on an external point of reference to map motion from one basis (egocentric) to another (allocentric).
- Development of the idea that CPU4 cells function as a compound memory.
- Suggestion of a new role for Pontine cells that normalise memory output.
- Supervision of Andrea Adden and Nicolai Ben Weddig in development of early CX model prototypes and experiments to test robustness of the circuit.
- Design of computational model and implementation of almost all code in Python.
- Designing experiments for the model to generate all the data used in the manuscript.

- Significant contributions to the manuscript and content, including figures and supplementary material. Creation of initial content for methods and supplementary material describing the model.

A neural substrate for path integration in the bee brain

Thomas Stone, Barbara Webb, Andrea Adden, Nicolai Ben Weddig, Anna Honkanen, Rachel Templin, William Wcislo, Eric Warrant, Stanley Heinze

Abstract

Path integration is one of the most widespread navigational strategies among animals. Path integrating animals continuously combine information about directional changes and distance covered while foraging, enabling them to return home in a straight line. The neural mechanisms underlying this integration are unknown in any species, but behavioural data has shown that bees use a visual compass and an optic flow based odometer for this purpose. Using intracellular electrophysiology we show that polarized-light-sensitive compass neurons and neurons sensitive to optic flow converge on specific identified neurons in the central complex of the bee brain. We demonstrate that a circuit model based on the mapped connectivity patterns of these cells functions as a holonomic path integrator, able to steer an animal directly home after an arbitrary outward journey. In a broader context this network is relevant to all situations in which current and desired headings are compared to initiate behaviour and thus suggests a general neural mechanism for navigational decisions.

Introduction

Animals with a fixed nest use many different sensory cues to ensure that they can return home after extended foraging trips. One widespread yet sophisticated strategy for carrying out this task quickly and efficiently is path integration[1]: the animal continuously integrates the distance cov-

ered in each direction over the duration of the outbound foraging trip to generate a homeward-pointing vector that allows it to return to the nest in a straight line. While in most animals, a combination of different self-generated motion cues are used for path integration[2], in bees, visual information alone is key to this task, making them uniquely accessible for identifying the neural components involved. Celestial skylight cues, such as polarized light, are used as a consistent external compass reference[3], while movement-generated optic flow patterns across the retina are used for estimating speed[1,4]. Across the planet, bee species perform this behaviour in a large variety of sensory environments, suggesting that the underlying neural mechanisms are both extremely robust and highly conserved. However, despite these rich behavioural data, nothing is known about the neural basis of path integration in bees or insects in general. In mammals, there is evidence that neural activity in the hippocampus, the retrosplenial cortex, and parahippocampal cortex are involved in human path integration[5], and head direction[6] and grid cells[7] are crucial elements for path integration in rodents. Nevertheless, the exact nature of the neural circuits involved has remained unresolved across all animals[1], although a number of hypothetical circuits have been proposed (reviewed and compared in[8]). One region of the insect brain thought to play a prominent role in all orientation behaviours is the central complex (CX), a conglomerate of intricately structured brain compartments highly conserved across species[9]. In migratory locusts and fruit flies this region houses an ordered array of visually based compass neurons[10,11]. The activity of individual CX-neurons can also reliably predict intended movement directions in freely walking cockroaches, indicating direct involvement in steering[12]. These and similar findings place the CX on the boundary between higher-order sensory processing and behavioural planning. We here provide converging evidence that one role of the CX is to serve as the neural substrate for path integration. Uniquely, we are able to explain the high level brain processing

that supports a complex behaviour, from integrating multiple sensory streams to producing motor steering commands, with a complete, anatomically grounded circuit model at the level of identified neurons.

Visual compass sensing in the bee CX

In locusts and butterflies polarized-light information is relayed to the CX via a highly conserved neural pathway[13,14], which, in locusts, has been shown to produce an ordered array of compass neurons, encoding heading in a global reference frame[10]. In flies, homologous cells encode head direction based on visual landmark cues, suggesting that mapping of directional space in the CX is a shared feature across insects[11,15]. As bees use polarized light to obtain compass information during foraging flights[3] and possess specialized eye regions for perceiving polarized light[16,17], we first ask whether compass neurons also exist in the bee CX (Fig. 1). For physiological recordings we focused on the CX of the tropical nocturnal bee *Megalopta genalis*[18] (Fig. 1a). These bees forage at times of the day when polarized skylight provides the single most reliable directional cue in their rainforest habitat and they possess all optical specializations typical for polarized-light perception[17]. Bees were captured with light traps directly from their natural habitat in Panama during foraging flights and tested within two weeks after capture. Using a newly developed artificial sky, we successfully recorded from 160 *Megalopta* bees to test responses of CX-neurons to linearly polarized light (Fig. 1b,c), continuously rotated above the animal either in clockwise or counter-clockwise directions (Fig. 1d,e). We found strong sinusoidal modulations of firing frequency in response to this stimulus, i.e. polarized-light tuning, in 10 neurons.

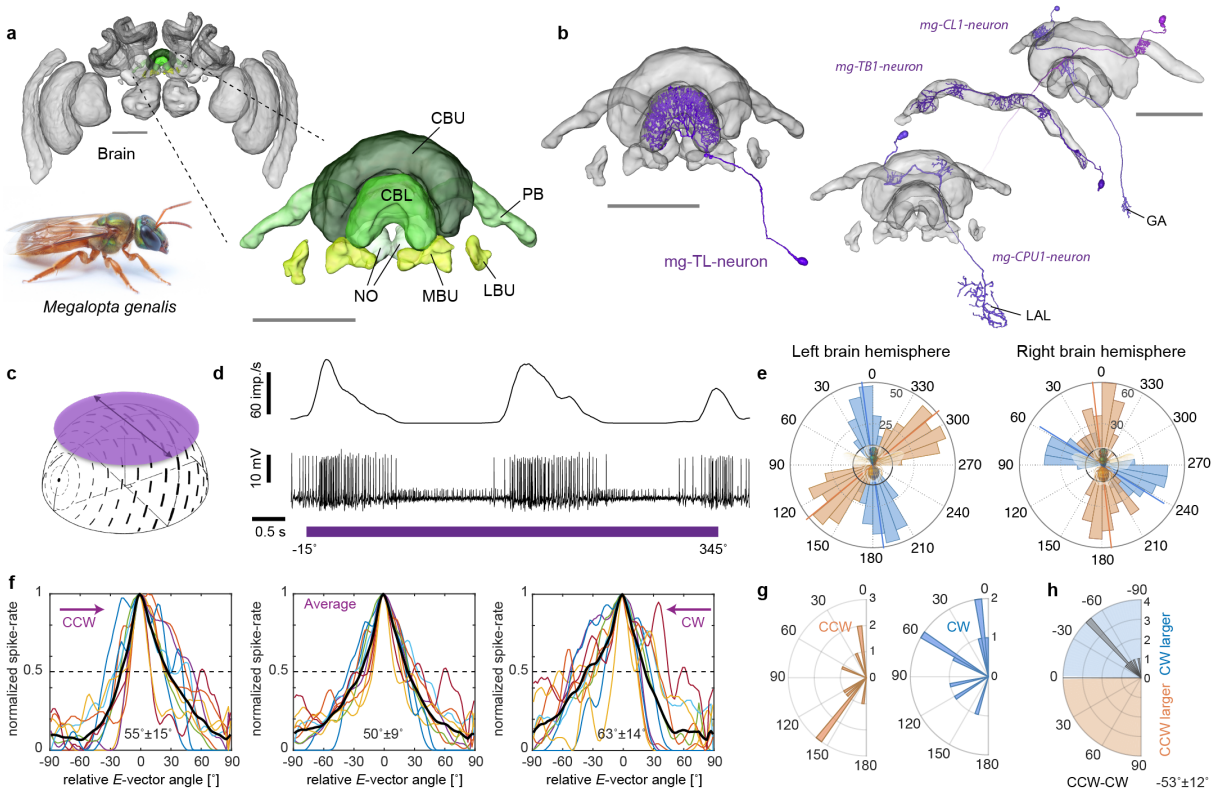


Figure 1 | Compass neurons of the bee central complex. a, 3D reconstruction of the brain of the solitary sweat bee *Megalopta genalis* (pictured below; body-length:1-2cm; photo credit Ajay Narendra). Highlight and inset: compass regions (central complex and bulbs). CBL, lower division of the central body; CBU, upper division of the central body; PB, protocerebral bridge; NO, noduli; MBU, medial bulb; LBU, lateral bulb. **b,** 3D anatomy of polarized-light-sensitive compass neurons based on intracellular injections, which together form a compass network highly similar to other insect species. Proposed information flow: TL-CL1-TB1-CPU1. LAL, lateral accessory lobes; GA, gall. **c,** Schematic stimulus: rotating, large polarizer, illuminated with an array of UV LEDs provided an artificial sky. **d,** Intracellular recording trace from individual TL-neuron during full rotation of the polarizer (bottom: membrane voltage; top: mean activity based on sliding average). **e,** Tuning of two TL neurons to polarized-light angle for clockwise (blue; 360˚-0˚) and counter-clockwise (orange; 0˚-360˚) filter rotations (binned mean activity during individual filter rotations). Black circles: background firing rate. **f**, Tuning curves of all recorded compass neurons (n = 10), normalized to peak activity with minimum activity set to zero. Peak activity was shifted to 0˚ for all cells and tuning width was measured at half-maximal excitation. Black curves: mean of population. Mean width ± SD is indicated for each graph. Left: counter-clockwise rotations (CCW); Right: clockwise rotations (CW); Middle: Average of CW and CCW. **g**, Distribution of tuning angles for CW and CCW filter rotations. **h**, Distribution of difference angles between CW and CCW rotations (mean ± SD is indicated numerically). Scale bars: 100 µm; brain in (**a**), 200µm.

The neurons showed an average tuning width of 50˚ and a difference in tuning between clockwise and counter-clockwise rotations of 53˚ on average, with tuning optima anticipating rotation direction (Fig. 1f-h), a phenomena found consistently in compass neurons across species that may aid correct compass encoding during fast body rotations[19]. Seven neurons were analyzed anatomically. All arborized in the lower division of the central body (ellipsoid body in flies; Fig. 1b), a part of the CX tightly associated with compass encoding in migratory insects and a key component of the *Drosophila* head direction network[11,13,14]. Indeed, both identified neuron types with compass-like activity in *Megalopta* (TL, and CL1-neurons) are either homologous to the GABAergic (inhibitory) ring-neurons or to the wedge-neurons that comprise the head direction system of the CX in *Drosophila*. These cells make up an estimated 5-10% of all CX-neurons and identical neurons have been described in detail in locusts[20,21], monarch butterflies[13] and dung beetles[22] with physiological responses to polarized light that are highly similar to those in *Megalopta*. Additionally, we identified bee-counterparts of all remaining locust compass neurons, occupying the protocerebral bridge (PB) (TB1 neurons) and the upper division of the central body (CBU, fan-shaped body in flies) (CPU1 neurons) (Fig. 1b). Together these findings strongly suggest that the CX serves as an internal compass in bees as well.

Optic flow sensing in the bee CX

The second sensory stream required to compute the bee's homing vector is translational information from optic flow[4] caused by image motion across the retina when an animal moves through its environment. Recent evidence from *Drosophila* and cockroaches shows that the CX houses neurons sensitive to large-field motion cues[23-25].

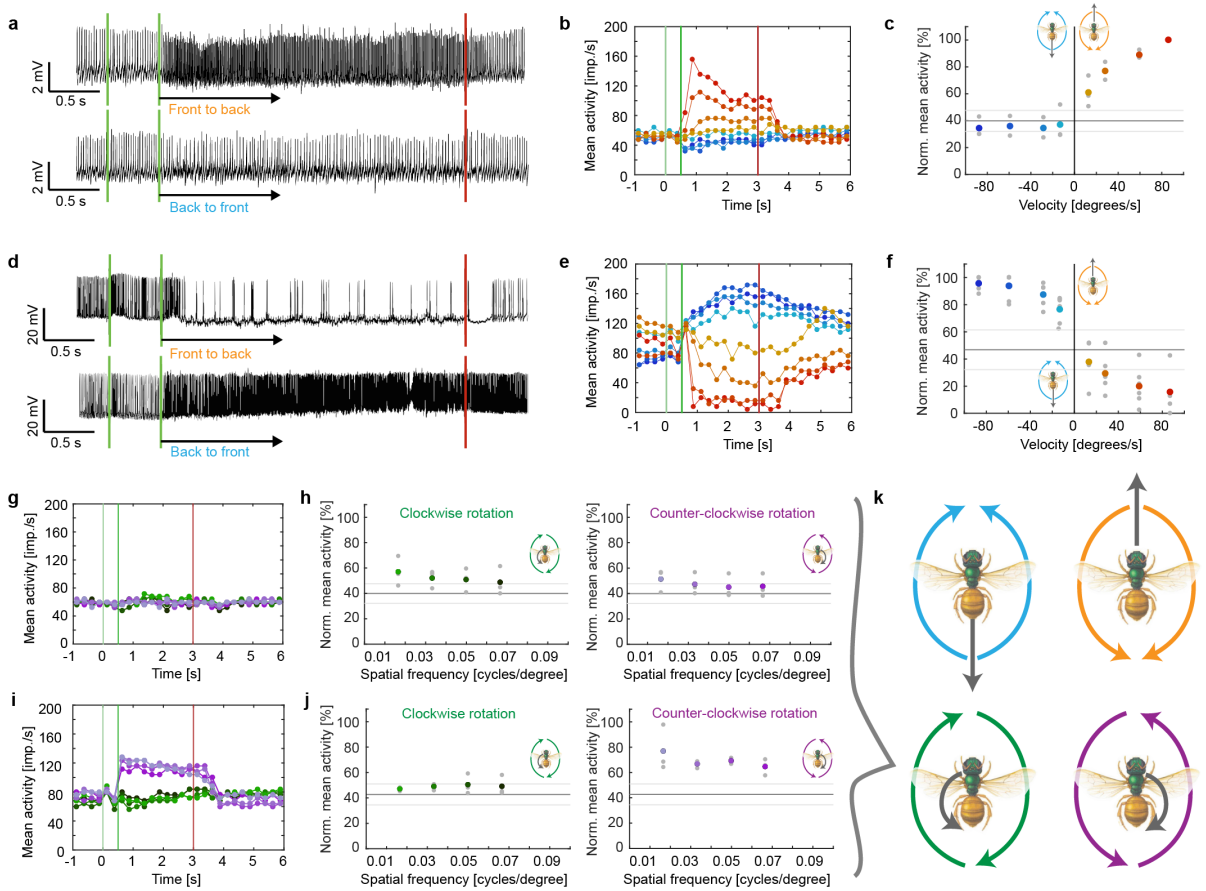


Figure 2 | Neurons of the bee central complex are sensitive to translational optic flow. a, Unfiltered voltage traces from one intracellular recording to front-to-back (top) and back-to-front translational optic flow. First green line: grating presented; second green line: motion onset; red line: motion stop. **b,** Mean activities of the same neuron to different stimulus velocities (–90˚/s to 90˚/s (blue to red)). Vertical lines as in (**a**). **c,** Mean tuning curve of all neurons with responses of the type shown in (**a**) and (**b**). Plotted is the normalized mean activity during the final 2s of movement for each stimulus bout; coloured circles: mean; grey circles: individual data-points; solid line: background activity (± SD, grey lines); n=3. This neuron type acts as direct speed indicator. **d-f,** As in (**a-c**), but for neurons with opposite response characteristics; n=5 in (**f**). Note that the responses in (**e**) consistently outlast the stimulus. **g/h,** Responses to rotational optic flow at different spatial frequencies of the same neuron as shown in (**a**) and (**b**). **i/j,** Responses to rotational optic flow of the same neuron shown in (**d**) and (**e**). Rotational optic flow was tested only in four out of five cells displayed in (**f**) and is shown for cells located in the right brain hemisphere as cells on the left side had inverted preferences to clockwise and counter-clockwise rotations; n = 3. **k,** Types of movement simulated by the different presented optic flow stimuli.

During intracellular recordings from CX-neurons we presented large-field optic flow stimuli (high-contrast sinewave gratings moving at different speeds) to bees located in the centre of a 360˚ LED arena (Fig. 2). Two types of CX-neurons responded strongly to translational optic flow. The first was excited by simulated forward flight (Fig. 2a-c). When exposed to different grating velocities, the neuron increased its firing approximately linearly with stimulus velocity, up to a sustained peak activity of approximately 150 impulses/s at a stimulus velocity of 90˚/s. This grating velocity matches the preferred optic flow of *Megalopta* in flight tunnels and their ground speed of 20 cm/s [26]. This cell type could thus serve as a neural speed indicator similar to speed-cells in the rat entorhinal cortex[27]. The second cell-type was inhibited by simulated forward flight but strongly excited by simulated backwards flight (peak activities of up to 200 impulses/s; Fig. 2d-f). In approximately half the neurons of that type the neural response outlasted the stimulus and decayed slowly until overridden by the next stimulus three seconds later (Fig. 2e). This characteristic is highly unusual for optic-flow encoding neurons (e.g. [28]) and might constitute a low pass filter that provides the bee's odometer with a smooth, continuous optic-flow based speed signal in a discontinuous, complex environment of fast changing spatial density (e.g. forests).

Both cell types responded weakly or not at all to rotational optic flow (Fig. 2g-j) despite identical grating parameters (contrast, velocity, spatial frequency), suggesting the response must be based on the structure of the neuron's receptive fields. By mapping receptive fields with a narrow, bright bar moving around the bee, we indeed found remarkably complex receptive field structures. Unlike neurons providing output from the optic-lobe motion-vision systems (Extended Data Fig 1; [28-30]), the CX-neurons received information from the entire panorama.

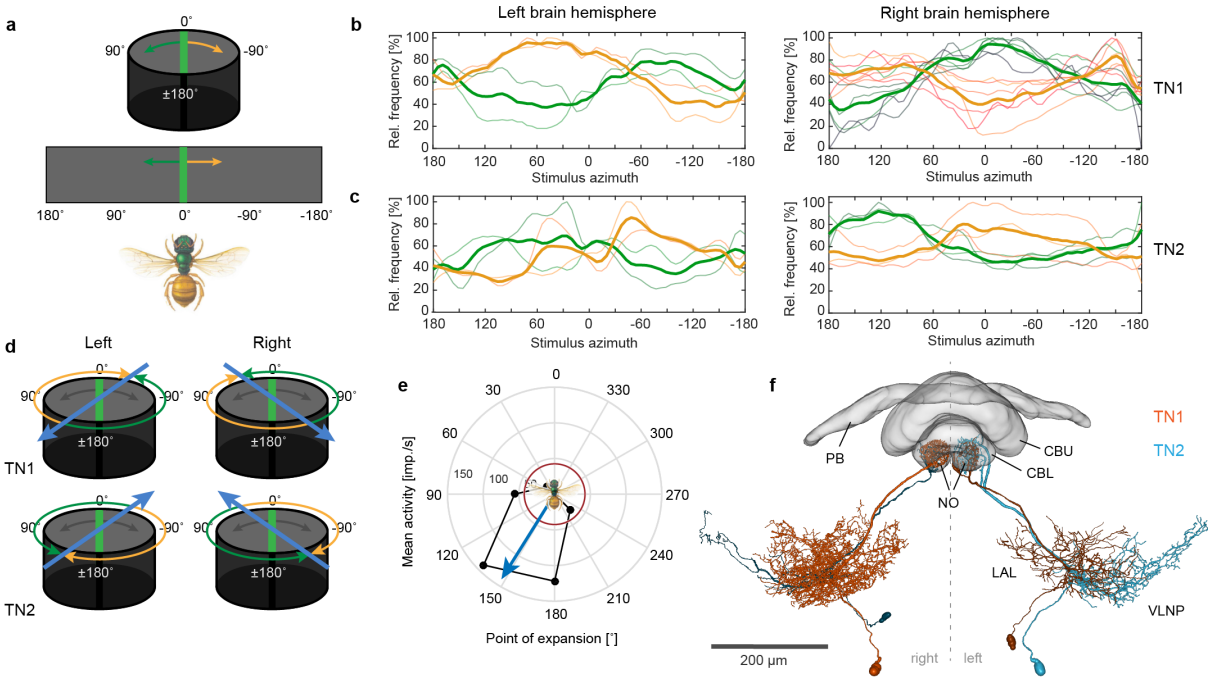


Figure 3 | Receptive fields of optic-flow-sensitive neurons define encoded movement space. a, To map receptive fields, a green bar was moved around the bee at constant speed. Bottom: schematic display of flattened arena used for graphs in (**b**) and (**c**). **b,** Mean normalized activity (thick curve) during receptive field mapping of TN1-neurons (inhibited by front to back optic flow) shown on top of individual response curves. Yellow traces: clockwise movement (CW); green traces: counter-clockwise movement (CCW). Neurons respond to motion in the whole panorama, but CW or CCW movement responses dominate specific areas of the visual field, with crossover occurring at an azimuth ca. 60˚ left or right of the body axis. Responses of neurons in right (n = 6) and left brain hemispheres (n = 2) (see **f**) are mirror-symmetric, but directionally inverted. **c,** As in (**b**), but for TN2 neurons (activated by front to back optic flow). Responses are inverted compared to TN1 neurons (left: n = 2; right: n = 3). **d,** Based on directional preference crossover within the receptive field, the preferred point of expansion of translational optic flow can be predicted (blue arrows). **e,** Translational optic flow was presented from different origins to a neuron with TN1-like physiology in the left brain-hemisphere. The resulting activity matches the predictions from (**d**). Blue arrow: maximally effective point of expansion; red circle: background activity. **f,** Morphology of TN1 (orange) and TN2 neurons (blue), shown as 3D reconstructions of intracellularly injected neurons registered into a common reference brain. For clarity, only the largest branches are shown for TN1-left and TN2-right, while the full extent of arborisations is shown for TN1-right and TN2-left (see Extended Data Fig. 2 for additional morphologies). PB, protocerebral bridge; CBU, upper unit of the central body; CBL, lower unit of the central body; NO, noduli; LAL, lateral accessory lobes; VLNP, ventrolateral neuropils.

Like many optic lobe cells, the neurons possess preferred and anti-preferred motion directions within their receptive fields, and several of these regions with opposite responses tiled the visual space around the animal (Fig. 3b,c). This combination of spatial and directional opponency allowed us to predict the maximally effective optic-flow stimuli for each type of neuron. Based on the receptive field data, four different neurons could be distinguished, each of which would be maximally excited by translational optic flow with a distinct point of expansion, offset from the body axis by approximately 45˚ (Fig. 3d,e). This set of cells thus encodes four cardinal directions of movement that equally divide the bee's movement space and together provide an accurate, robust representation of the bee's movement, and therefore are a basis for encoding holonomic movements in the bee's flightpath (bee-typical movements during which the body axis is not aligned with the movement direction, e.g. side slip).

A subset of the recorded optic-flow processing neurons was also characterized anatomically. The morphological identity of the cells correlated with the four physiologically identified types. All neurons selectively provided input to two CX compartments called the noduli and were named noduli tangential neurons (TN-neurons). While all TN-cells received input only in the ipsilateral brain hemisphere, those that were inhibited by simulated forward flight (TN1) had dendrites in different regions of the brain compared to those that were excited by the same stimulus (TN2), thus also representing two anatomically distinct cell types (Fig. 3f; Extended Data Fig. 2; Supplementary Video 1). Additionally, the preference in movement direction towards either the right or the left side of the body axis was reflected in the hemisphere in which each cell received its input. We not only identified and functionally characterized these neurons in the tropical *Megalopta* but, partly at least, also in the bumblebee (Extended Data Fig. 3). In this species, block-face electron microscopy revealed that indeed only two of these cells exist per nodu-

lus and that each of them possesses many hundreds of synaptic outputs (active zones) associated with large pools of synaptic vesicles (Extended Data Fig. 4j,k). The identical morphology and very similar response features in two distantly related species suggest that these cells exist across all bees and are part of the core network of the bee CX.

Integrating compass and speed

Only very few identified neurons with arborisations in the noduli have been functionally characterized in any species. The only data available derives from locust neurons (CPU4-neurons[21]) that connect the noduli and the PB and which respond to polarized-light stimuli in a context-dependent way. These CPU4 neurons are thus highly suited to combine direction information from the ordered array of direction cells in the PB with speed information from the noduli. We identified CPU4-like neurons in *Megalopta* and found they were frequently co-stained with dye-injected TN-neurons, revealing a close proximity of their neural branches (Fig. 4f). By neural tracing after performing block-face electron microscopy of the bumblebee noduli we confirmed that TN cells are presynaptic to numerous columnar fibres most likely corresponding to CPU4 cells (Extended Data Fig. 4).

CPU4 cells also possess output fibres in yet another part of the CX (the fan shaped body or CBU), which houses the dendrites of the major CX-output neurons (columnar CPU1 neurons)[22,31-34] and those of pontine cells, which interconnect specific columns between CBU hemispheres (Extended Data Fig. 5e,f). Recordings from *Megalopta* CPU1 cells revealed neuronal responses to compass stimuli (Extended Data Fig. 5a-d), supporting a role of these cells in navigation, whereas pontine cells potentially provide additional links between CPU4 and CPU1 cells.

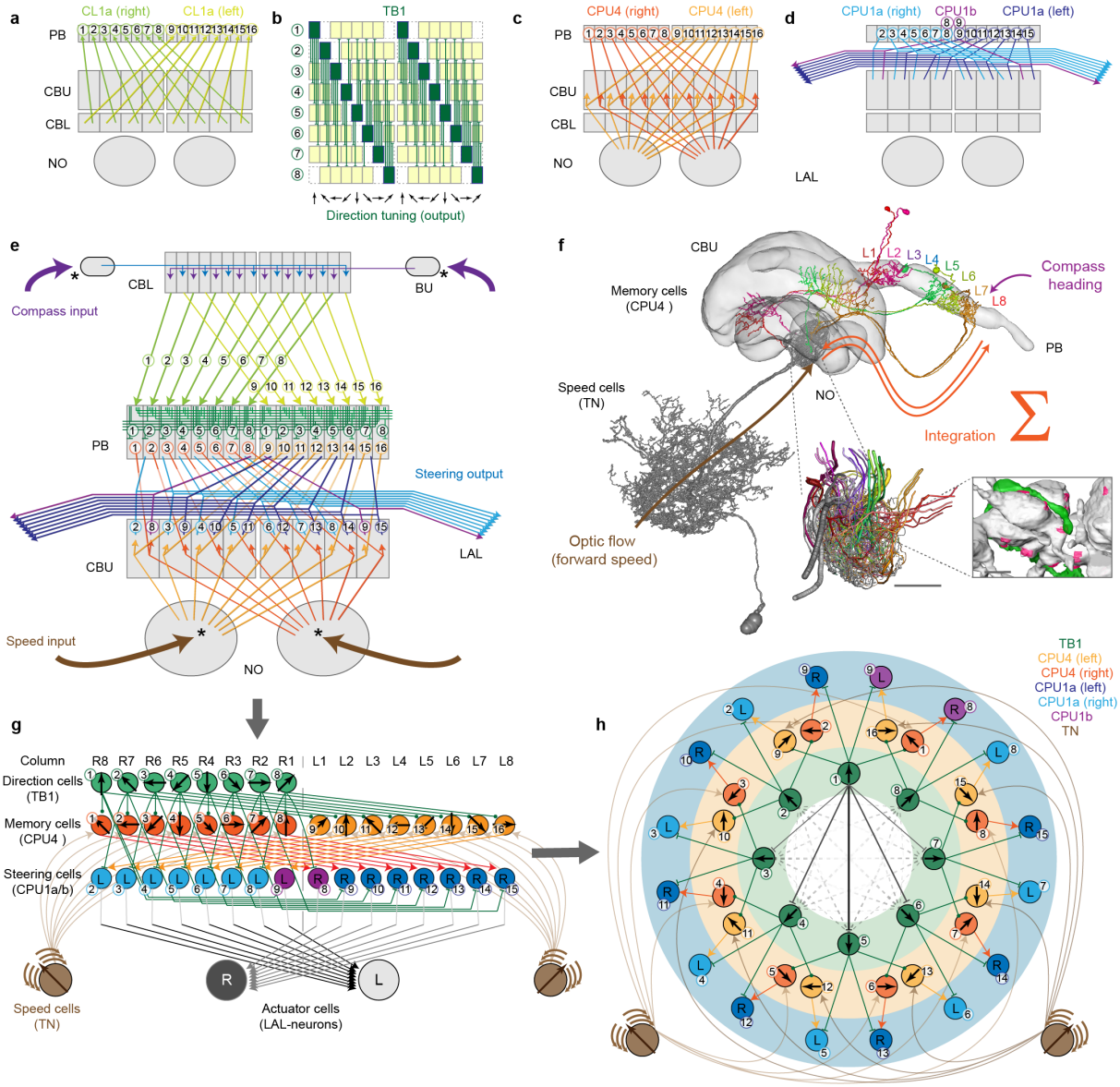


Figure 4 | Wiring diagram of anatomically plausible connections in the central complex and the resulting circuit model. a-d, Schematic connectivity diagrams of CL1a neurons (**a**), TB1 neurons (**b**), CPU4 neurons (**c**) and CPU1 neurons (**d**), based on single cell injections across various insect species. In (**b**) each line represents one neuron with columnar input regions (yellow) and output regions (green). CBU, upper division of the central body; PB, protocerebral bridge; NO, noduli; LAL, lateral accessory lobes; BU, bulbs. **e,** Anatomically constrained wiring based on overlapping dendritic and axonal projections of all cell types from (**a-d**), and the speed-encoding TN-neurons. Asterisks: Verified synaptic connections (see Supplementary Discussion for evidence for remaining connections). For clarity, connections between individual TB1 neurons are omitted and CBL is shown on top. **f,** The proposed integrator circuit. CPU4 neurons in individual columns of the CBU accumulate optic flow speed input from the noduli proportionally to their simultaneous inhibition by polarization compass directional input from the PB. They thus store an integrated memory of the flight path. All neurons shown are 3D reconstructions from one *M. genalis* preparation; five out of eight cells of one hemisphere have been labelled. The enlargement shows tracing of TN and CPU4 cells based on block-face electron microscopical data. Inset: EM-based 3D reconstruction of presynaptic terminals from TN cell (grey with active zones in magenta) and one postsynaptic CPU4 cell. **g,** Abstraction of (**e**) with neurons arranged in conceptual layers: direction cell layer, memory cell layer, and steering cell layer. The

arrows indicate the directional tuning (TB neurons) or the integrated direction preference (CPU4 cells). All steering cells from one hemisphere converge on one hypothetical actuator cell, which initiates right or left turns. **h,** Topological structure of (**g**) obtained using a force-directed graph: circuit forms a ring. Connections between TB1 neurons create a ring attractor network (solid lines: connections of one cell; dashed lines: remaining cells; line thickness: strength of connection). CPU1 neurons receive inputs from CPU4 neurons with a rightwards (orange) or leftwards (yellow) shift relative to their TB1 input. For morphology and connectivity of pontine cells, which maintain balance of the memories in both brain hemispheres, see Extended Data Fig. 5.

The axons of CPU1 output cells converge in neighbouring brain regions (lateral accessory lobes, LAL) implicated in steering control in moths (e.g. [35]). Given the high degree of conservation of both the CPU1 cells and the general LAL layout, these regions are a likely source of premotor control signals for supervising turning behaviour across insects.

We hence propose that the bee CX contains a neuronal network that combines polarized-light based compass information with translational optic flow information and links both to control of turning maneuvers. In the following, we show that this predicted circuit is indeed sufficient to explain not only how a home vector is generated during path integration, but also how it is used to control the return path and the search around the home location.

A model for path integration

Taking anatomical overlap of arborisation trees as a predictor of synaptic connections, we extracted a detailed connectivity pattern of all cells involved (Fig. 4; for a comprehensive list of all model assumptions see Supplementary Discussion). A force-directed graph layout revealed an underlying logic to this anatomically suggested connectivity (Fig. 4h) that strikingly resembles previous hypothetical circuits proposed for path integration[36,37]: In these models, direction cells divide the azimuthal space and are activated by the current heading of the animal. Corresponding memory cells accumulate the distance moved in each direction. Homing is controlled by comparing the memory cell activity to the current direction cell activity to determine if the animal

should steer left or right. This mechanism generates a distributed representation of the home vector and has been shown in principle to produce successful direct homing after convoluted outward journeys[36,37]. Each component can now be directly mapped to identified neurons in the CX.

The direction neurons are a set of eight neurons of the PB (TB1), which are predicted to create a ring-attractor network through systematic, mutually inhibitory connections with one another (Fig. 4b, Extended Data Fig. 6). Consistent with recent data from homologous neurons in *Drosophila*[38], TB1 neurons are anatomically well suited to perform this computation and have indeed been suggested to be inhibitory (Supplementary Discussion)[6]. They receive input from columnar cells of the ellipsoid body/lower division of the central body (CL1, Fig. 4a), characterized as compass neurons in flies[11], locusts[21], butterflies[13], beetles[22] and bees (this work). In the model, each TB1-neuron is tuned to one specific azimuth direction and together they tile the azimuth around the animal with a resolution of 45˚. This 360˚ representation of the azimuth within the PB-columns of each brain hemisphere closely reflects data from flies [11], which have revealed a mapping of body orientations to CL1 neurons targeting the PB.

The substrate for memory is then predicted to be the columnar noduli neurons (CPU4). Above we showed through EM studies that they receive input from the newly found speed-sensing neurons (TN) via branches in the CX-noduli (Fig. 4c; Extended Data Fig. 4), and it is highly plausible that they receive input from the PB direction cells[21]. Using block-face electron microscopy, we observed a set of 19 CPU4-like neurons per CX-column, i.e., for each of the eight represented compass directions, repeated for each hemisphere (Extended Data Fig. 4f). Thus these cells comprise a pool of neurons that could, in each column, integrate the speed information from optic flow in negative proportion to the inhibitory input from direction cells, e.g. by forming local microcircuits or recurrent connections between the PB and the noduli. The ex-

istence of such microcircuits is strongly supported by the presence of both synaptic inputs and outputs in noduli branches of individual CPU4 neurons, as well as synaptic connections between different CPU4 cells. By combining excitatory TN-speed input with inhibitory TB1-direction signals, memory would increase proportionate to the bee's speed in those CPU4-columns that face *opposite* to the current flight heading. The combined activity would thus represent the home vector in distributed form. Interestingly (and not foreseen in the earlier models[36,37]), as the CPU4 cells in each brain hemisphere can obtain input only from the TN cells on the corresponding side, the experimentally determined 90˚ offset in the optimal centre of optic-flow expansion between right and left TN-cells produces a home vector encoding that is robust to a significant degree of holonomic movement of the animal (i.e. a ground velocity that is not precisely aligned with the body orientation).

The final element required for homing, the steering cells, are then predicted to be the aforementioned CPU1 cells, which are suited to receive information from the direction cells and from the memory (Fig. 4d). Crucially, this identification utilizes a specific columnar offset observed between CPU1- and CPU4 cell innervation patterns that is conserved across all examined insects (Fig. 4g). This offset would produce an activity level in the left vs. right steering cells that corresponds to whether a left or right rotation would produce a better match between the current heading direction and the memory (Extended Data Fig. 7) with the amplitude of the output from the two hemispheres balanced by the cross-connecting pontine neurons.

We implemented a computational model of this anatomically defined circuit using rate-based neurons, demonstrated its functionality mathematically by adapting a known algorithmic framework[37] (Supplementary Notes; Extended Data Fig. 7), and tested it in an agent simulation (Fig. 5; Supplementary Video 2; Extended Data Fig. 8,9).

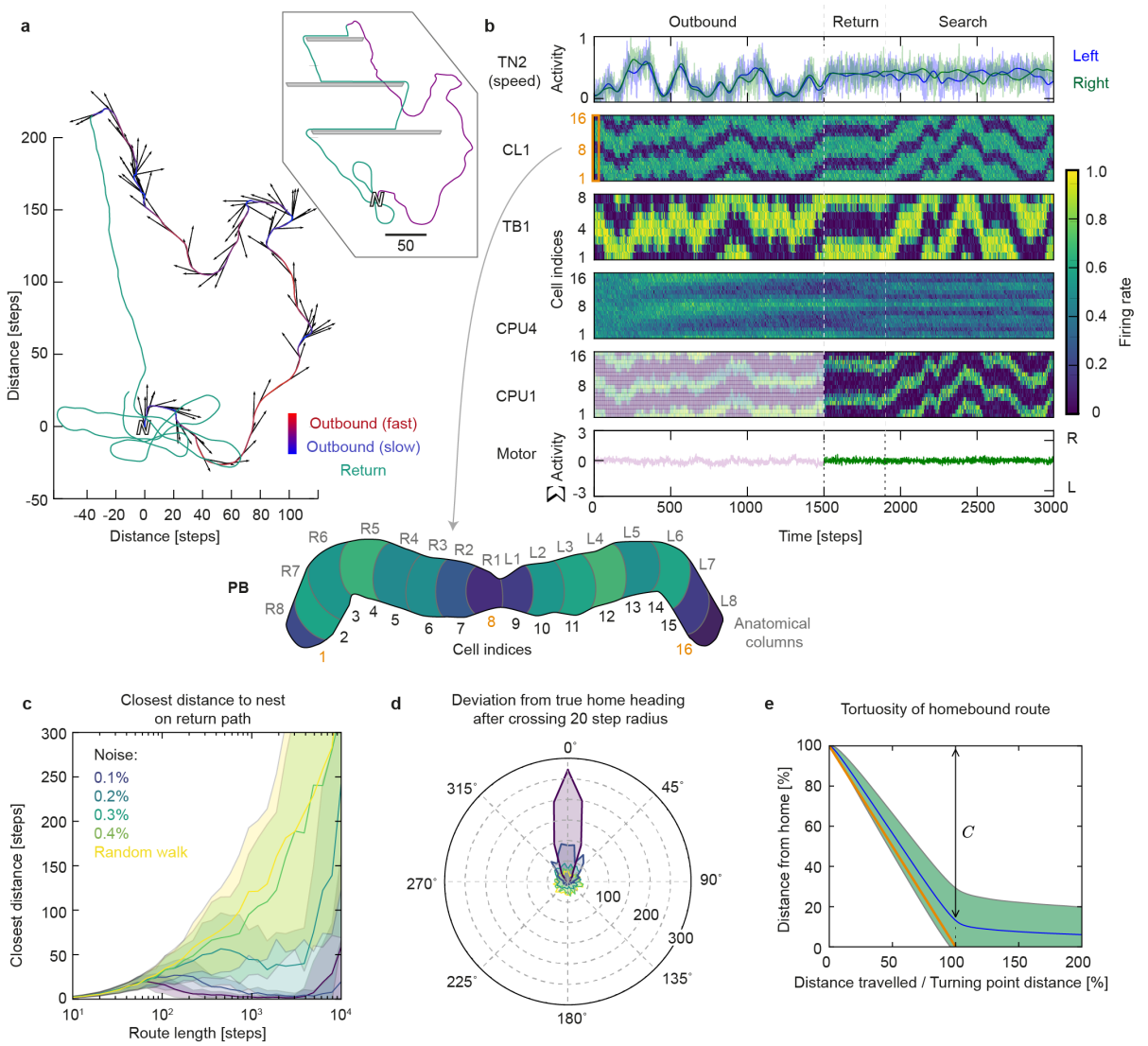


Figure 5 | The biologically constrained circuit successfully performs path integration. a, Example trace: The agent is driven on a random outbound journey with variable speed for 1500 steps, during which the circuit accumulates memory by integrating speed in each direction. For return, the steering cells become functionally connected to the motor output and guide the agent's movements back to the point of origin. Close to the nest, the agent automatically initiates searching behaviour, similar to path-integrating insects [43]. Inset: Same outbound route, but obstacles were present during homing. The circuit continues to integrate movement during the deviation to provide the correct guidance. **b,** Activity of all neurons in the circuit over time during the trial in (**a**). Neural noise level: 10%; cell indices indicate the identity of the individual neuron in each set; firing rate is normalized to peak rate within each cell-type. **c,** Performance of the path integration circuit at increasing lengths of the outbound journey (10-10000 steps) with different levels of neural noise (10 to 40%). Closest distance to the nest during homing (mean ± SD from 1000 runs) is plotted against the length of the outbound journey. **d**, Heading of the agent after steering has been initiated at the beginning of homing (after crossing 20 step radius): Distribution of angular deviations from true home (10˚ bins, 1000 trials) for different noise levels (colours as in **f**). **e**, Deviation from the best possible route (straight line) during homing (at 10% noise). *C* is distance actually covered towards the nest after L steps, where L is the straight line distance to the nest; L/C is a measure of path tortuosity, which is on average very low (=1.150).

The agent can be moved along an arbitrary outbound path, during which its heading determines the activation of the direction neurons (TB1), which is subtracted from input from the corresponding speed neuron (TN2) and accumulated in the memory neurons (CPU4) with a constant leak. When instructed to return home, the steering neurons (CPU1) take control of the agent and it turns according to the relative left and right activity. Tested with random outbound routes of up to 10000 steps, the agent homes directly to within a few steps of its starting location (Fig. 5a,b; Supplementary Video 2; Extended Data Fig. 8). As the memory continuously updates, the agent can deal with forced deviations on its homeward path (Fig. 5a inset) and also (in the absence of a stopping criteria) produces an emergent 'search' pattern around the home location, deviating in a random direction until the activity in the memory neurons becomes unbalanced enough to drive the steering neurons for another return path. The homing ability is not affected by biologically relevant amounts of holonomic movement, typically produced during bee flight (±45˚). In repeated runs, the network is robust to added noise of 10-20%, to non-linearity of the speed input, and to turning-direction-dependent shift in the preferred response of compass neurons, generating highly accurate homing until the outbound route extends beyond the capacity of the memory cells (Fig. 5c-e; Supplementary Discussion; Extended Data Fig. 9). It is worth nothing that even though the central complex in our model stores the home vector using polar coordinates, it retains all advantages of a geocentric Cartesian model, which is believed to be the most robust type of path integration network for biological systems[8].

Discussion

We have presented a complete neural circuit for path integration that combines a previously well-described compass network with speed-sensing neurons reported here for the first time, and

which makes concrete predictions about synaptic connectivity in line with all functional and anatomical data across insect species.

Several highly conserved structural features of the insect CX gain relevance through our model. The distinct sensory pathway transmitting translational optic flow information to the CX reveals a first functional role for the CX-noduli, a region present in all flying insects. The complex pattern of dendritic and axonal fibres in the multiglomerular TB1 neurons are the core of the ring attractor circuit. The characteristic connectivity schemes of different types of columnar neurons (CPU4, CPU1) are fully accounted for as subserving the functions necessary for memory and steering. Finally, the functionally undescribed CX-pontine cells, the last cell-class of the CX that is also present in all species examined, are crucial in maintaining stability of the proposed path integration memory.

Our model also produces a range of concrete predictions suited to open up multiple new lines of structural, physiological and behavioural research: First, the CPU4 columnar cells are the proposed physical substrate for path integration memory, suggesting memory capacity is linked to the number of parallel CPU4 neurons per CX-column. Thus, path-integrating insects should in principle have more CPU4 cells compared to species with different navigational strategies. Second, as the homing vector is hypothesized to be represented by ongoing recurrent neural activity, the path integration memory should be less stable compared to memory laid down by synaptic remodelling (e.g. in the mushroom body) and thus be more prone to experimental disturbance, e.g. via brain cooling at the beginning of a homing flight. Third, physiologically, our network predicts that during path integration the activity of CPU4 cells as well as pontine cells increases over time and that specific patterns of activity across the CX result from specific outbound flights. Although it would not be trivial to generate the right motivational and behavioural condi-

tions, these patterns as well as the accumulation of activity could potentially be observed by electrophysiology or functional imaging. Finally, our model can also be tested in behavioural experiments. Manipulating optic flow to introduce controlled biases during outbound routes (e.g. generating the impression of continuous backwards flight) as well as direct manipulations of neuronal activity can be used to mimic conditions that break our model in the simulations in predictable ways (Extended Data Fig. 10).

In a broader context, the proposed circuit compares the animal's current heading with its desired heading and initiates compensatory steering commands in case of a mismatch, a process common to a range of navigational strategies in animals. In its simplest form a circuit containing only few recurrent CPU4 neurons per CX-column could store an ongoing direction of movement, allowing steering back to the desired direction in case of disturbance[39,40]. At the opposite end of the scale, CPU4 neurons could permanently encode the migratory heading of long distance navigators, such as the monarch butterfly.

We suggest CPU1 neurons perform the comparison between intended and actual heading. The dendritic trees of these cells are extensive and reside in a brain region (the fan shaped body, CBU) that receives input from vast areas of the insect brain[34,41,42]. It is thus conceivable that neurons other than CPU4 also deliver information about the animal's desired heading to these cells in different sensory/motivational contexts, a possibility supported by pronounced state dependency of visual responses in the fan shaped body of flies[24]. This simple expansion of our model could thus also explain behaviours such as steering towards landmarks, choosing between targets, and following routes.

Finally, the underlying principles of this circuit potentially bear relevance for navigation behaviours in mammals as well. Mammals are known to use externally centred (allocentric) spa-

tial representations, e.g. via grid cells[6], but it is not clear how the brain converts egocentric motion signals from the sensory system into the allocentric grid cell coordinate system. We propose that in the insect CX, such a remapping of reference frames is achieved by a neuroarchitecture that effectively converts the angle and length of a movement vector into the phase and amplitude of a sine wave relative to external reference cues; and by first decomposing velocity into a basis at two offset angles, supports integration of holonomic motion. This solution suggests possible circuit motifs that could be relevant to decoding navigational abilities in more complex brains.

References

1. Srinivasan, M. V. Where paths meet and cross: navigation by path integration in the desert ant and the honeybee. *J. Comp. Physiol. A* **201,** 533–546 (2015).
2. Mittelstaedt, M. L. & Mittelstaedt, H. Homing by path integration in a mammal. *Naturwissenschaften* **67,** 566–567 (1980).
3. Evangelista, C., Kraft, P., Dacke, M., Labhart, T. & Srinivasan, M. V. Honeybee navigation: critically examining the role of the polarization compass. *Philos. Trans. R. Soc. B* **369,** 20130037–20130037 (2014).
4. Srinivasan, M. V. Going with the flow: a brief history of the study of the honeybee's navigational 'odometer'. *J. Comp. Physiol. A* **200,** 563–573 (2014).
5. Chrastil, E. R., Sherrill, K. R., Hasselmo, M. E. & Stern, C. E. There and back again: Hippocampus and retrosplenial cortex track homing distance during human path integration. *J. Neurosci.* **35,** 15442–15452 (2015).
6. Sargolini, F. *et al.* Conjunctive representation of position, direction, and velocity in entorhinal cortex. *Science* **312,** 758–762 (2006).

7. Allen, K. *et al.* Impaired path integration and grid cell spatial periodicity in mice lacking GluA1-containing AMPA receptors. *J. Neurosci.* **34,** 6245–6259 (2014).

8. Cheung, A. & Vickerstaff, R. Finding the way with a noisy brain. *PLoS Comput. Biol.* **6,** e1000992 (2010).

9. Pfeiffer, K. & Homberg, U. Organization and functional roles of the central complex in the insect brain. *Annu. Rev. Entomol.* **59,** 165–184 (2014).

10. Heinze, S. & Homberg, U. Maplike representation of celestial E-vector orientations in the brain of an insect. *Science* **315,** 995–997 (2007).

11. Seelig, J. D. & Jayaraman, V. Neural dynamics for landmark orientation and angular path integration. *Nature* **521,** 186–191 (2015).

12. Martin, J., Guo, P., Mu, L., Harley, C. M. & Ritzmann, R. E. Central-complex control of movement in the freely walking cockroach. *Curr. Biol.* **25,** 2795–2803 (2015).

13. Heinze, S. & Reppert, S. M. Sun compass integration of skylight cues in migratory monarch butterflies. *Neuron* **69,** 345–358 (2011).

14. Homberg, U., Heinze, S., Pfeiffer, K., Kinoshita, M. & Jundi, el, B. Central neural coding of sky polarization in insects. *Phil. Trans. R. Soc. B* **366,** 680–687 (2011).

15. Seelig, J. D. & Jayaraman, V. Feature detection and orientation tuning in the *Drosophila* central complex. *Nature* **503,** 262–266 (2013).

16. Labhart, T. Specialized Photoreceptors at the dorsal rim of the honeybee's compound Eye: polarizational and angular sensitivity. *J. Comp. Physiol. A* **141,** 19–30 (1980).

17. Greiner, B., Cronin, T. W., Ribi, W. A., Wcislo, W. T. & Warrant, E. J. Anatomical and physiological evidence for polarisation vision in the nocturnal bee *Megalopta genalis*. *J. Comp. Physiol. A* **193,** 591–600 (2007).

18. Warrant, E. J. *et al.* Nocturnal vision and landmark orientation in a tropical halictid bee. *Curr. Biol.* **14,** 1309–1318 (2004).

19. Träger, U. & Homberg, U. Polarization-sensitive descending neurons in the locust: connecting the brain to thoracic ganglia. *J. Neurosci.* **31,** 2238–2247 (2011).

20. Vitzthum, H., Müller, M. & Homberg, U. Neurons of the central complex of the locust *Schistocerca gregaria* are sensitive to polarized light. *J. Neurosci.* **22,** 1114–1125 (2002).

21. Heinze, S. & Homberg, U. Linking the input to the output: new sets of neurons complement the polarization vision network in the locust central complex. *J. Neurosci.* **29,** 4911–4921 (2009).

22. Jundi, el, B. *et al.* Neural coding underlying the cue preference for celestial orientation. *Proc. Natl. Acad. Sci. USA* **112,** 11395–11400 (2015).

23. Bausenwein, B., Müller, N. R. & Heisenberg, M. Behavior-dependent activity labeling in the central complex of *Drosophila* during controlled visual stimulation. *J. Comp. Neurol.* **340,** 255–268 (1994).

24. Weir, P. T. & Dickinson, M. H. Functional divisions for visual processing in the central brain of flying *Drosophila*. *Proc. Natl. Acad. Sci. USA* **112,** E5523–32 (2015).

25. Kathman, N. D., Kesavan, M. & Ritzmann, R. E. Encoding wide-field motion and direction in the central complex of the cockroach *Blaberus discoidalis*. *J. Exp. Biol.* **217,** 4079–4090 (2014).

26. Baird, E., Kreiss, E., Wcislo, W. T., Warrant, E. J. & Dacke, M. Nocturnal insects use optic flow for flight control. *Biology Letters* (2011).

27. Kropff, E., Carmichael, J. E., Moser, M.-B. & Moser, E. I. Speed cells in the medial entorhinal cortex. *Nature* **523,** 419–424 (2015).

28. Borst, A., Haag, J. & Reiff, D. F. Fly motion vision. *Annu. Rev. Neurosci.* **33,** 49–70 (2010).

29. Krapp, H. G. & Hengstenberg, R. Estimation of self-motion by optic flow processing in single visual interneurons. *Nature* **384,** 463–466 (1996).

30. Barnett, P. D., Nordström, K. & O'Carroll, D. C. Retinotopic organization of small-field-target-detecting neurons in the insect visual system. *Curr. Biol.* **17,** 569–578 (2007).

31. Heinze, S. & Homberg, U. Neuroarchitecture of the central complex of the desert locust: Intrinsic and columnar neurons. *J. Comp. Neurol.* **511,** 454–478 (2008).

32. Wolff, T., Iyer, N. A. & Rubin, G. M. Neuroarchitecture and neuroanatomy of the *Drosophila* central complex: A GAL4-based dissection of protocerebral bridge neurons and circuits. *J. Comp. Neurol.* **523,** 997–1037 (2015).

33. Lin, C.-Y. *et al.* A comprehensive wiring diagram of the protocerebral bridge for visual information processing in the *Drosophila* brain. *Cell Rep.* **3,** 1739–1753 (2013).

34. Heinze, S., Florman, J., Asokaraj, S., Jundi, el, B. & Reppert, S. M. Anatomical basis of sun compass navigation II: the neuronal composition of the central complex of the monarch butterfly. *J. Comp. Neurol.* **521,** 267–298 (2013).

35. Namiki, S., Iwabuchi, S., Pansopha Kono, P. & Kanzaki, R. Information flow through neural circuits for pheromone orientation. *Nat. Comm.* **5,** 5919 (2014).

36. Haferlach, T., Wessnitzer, J., Mangan, M. & Webb, B. Evolving a neural model of insect path integration. *Adaptive Behavior* **15,** 273–287 (2007).

37. Wittmann, T. & Schwegler, H. Path integration - a network model. *Biological Cybernetics* **73,** 569–575 (1995).

38. Kakaria, K. S. & de Bivort, B. L. Ring attractor dynamics emerge from a spiking model of

the entire protocerebral bridge. Preprint at http://www.biorxiv.org/content/early/2016/10/16/081240 (2016).

39. Neuser, K., Triphan, T., Mronz, M., Poeck, B. & Strauss, R. Analysis of a spatial orientation memory in Drosophila. *Nature* **453,** 1244–1247 (2008).

40. Jundi, el, B. *et al.* A Snapshot-Based Mechanism for Celestial Orientation. *Curr. Biol.* (2016).

41. Hanesch, U., Fischbach, K. F. & Heisenberg, M. Neuronal architecture of the central complex in *Drosophila melanogaster*. *Cell Tissue Res.* **257,** 343–366 (1989).

42. Young, J. M. & Armstrong, J. D. Structure of the adult central complex in *Drosophila*: organization of distinct neuronal subsets. *J. Comp. Neurol.* **518,** 1500–1524 (2010).

Online Content Methods, additional Extended Data display items, Supplementary Notes, Supplementary Discussion, and Supplementary Videos are available in the online version of the paper at www.nature.com/nature; references unique to these sections appear only in the online paper.

Acknowledgements

This work was supported by a junior project grant from the Swedish Research Council (to S.H.; VR 621-2012- 2213), a Marie-Curie Intraeuropean Fellowship (to S.H. 327901), the Wenner-Gren Foundation (to A.H.), the Australian Research Foundation and the Air Force Office for Scientific Research (to Justin Marshall, supporting R.T.), the Swedish Research Council, the Air Force Office for Scientific Research and the Knut and Alice Wallenberg Foundation (to E.W.),

as well as in part by grants EP/F500385/1 and BB/F529254/1 for the University of Edinburgh School of Informatics Doctoral Training Centre in Neuroinformatics and Computational Neuroscience (www.anc.ac.uk/dtc) from the UK Engineering and Physical Sciences Research Council (EPSRC), UK Biotechnology and Biological Sciences Research Council (BBSRC), and the UK Medical Research Council (MRC). The work has made use of resources provided by the Edinburgh Compute and Data Facility (ECDF; www.ecdf.ed.ac.uk), which has support from the eDIKT initiative (www.edikt.org.uk). We are additionally grateful for logistical support from the Smithsonian Tropical Research Institute, in particular to Dr. Julia Schuckel for help with catching bees and construction of the recording setup. For neuronal tracings we would like to thank Johanna Chavez, Henrik Martenson, David Szakal, and Olle Claesson. Finally, the authors would like to acknowledge the facilities, and the scientific and technical assistance, of the Australian Microscopy & Microanalysis Research Facility at the Centre for Microscopy and Microanalysis, The University of Queensland, especially Robyn Chapman for technical assistance.

Author Contributions: S.H. designed the study and T.S., S.H. and B.W. developed the concept of the manuscript; S.H. and A.H. performed electrophysiological and anatomical work; S.H. analyzed the physiological data; T.S., N.W., B.W. and A.A. carried out computational circuit modelling; R.T. performed electron microscopy and block-face imaging; S.H., T.S. and B.W. wrote the manuscript and designed the figures; W.W., E.W., A.A., A.H., N.W., B.W, R.T., T.S., and S.H. critically read the manuscript and significantly contributed to its final format and content.

Author Information: Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interest. Correspondence and requests for materials should be addressed to S.H. (stanley.heinze@biol.lu.se)

Methods

Animals

Bees of the genus *Megalopta* (species *M. genalis* and *M. centralis*) were caught from the wild using light traps (white sheets illuminated by a bright light source containing UV wavelengths). Traps were placed ca. 2 m above ground within small canopy openings of the tropical forest on Barro Colorado Island (field station of the Smithsonian Tropical Research Institute), located in the Panama Canal, Panama. Trappings were carried out during the activity phase of the bees between 4:30 am and 5:30 am in the morning, i.e. during early morning twilight. Caught bees were kept individually in 50 ml plastic vials, equipped with two cotton balls, one soaked in honey solution as well as one soaked in water. Vials were kept at room temperature in a dark secondary container (small amounts of natural light were allowed to reach the bees to ensure continuous circadian entrainment). Bees were used for experiments within two weeks after capture. With few exceptions used bee were large to medium sized females.

Bumblebees (*Bombus terrestris*) were obtained from a commercial supplier (Koppert, Berkel en Rodenrijs, The Netherlands) and kept at room temperature in a room sized flight cage at Lund University, Sweden. Feeders with honey solution were available at all times.

Visual Stimulation

For visual stimulation we developed a new virtual reality environment combining a 360˚ panoramic LED arena (570 nm LEDs; IO-Rodeo, Pasadena, USA) with an artificial sky. The arena had an angular resolution of 1.5˚ and covered 55˚ of vertical space (equal parts above and below the horizon). The artificial sky consisted of a planar array of UV LEDs (365 nm) illuminating a large polarizer (BVO-UV, Boldervision; 88˚ of dorsal space) mounted on a custom built rotation

stage (Prototypverkstaden, Lund, Sweden) driven by a Micos DT-50 rotation stage (controlled via MoCo controller; Micos). All LED panels as well as the rotation stage were controlled via an integrated, custom designed MATLAB-based software.

Polarized-light stimuli were applied by switching the LED illumination of the sky on and rotating the polarizer through 360˚ at constant speed (60˚/s or 30˚/s) clockwise and counter-clockwise (as seen from the animal's point of view).

Optic-flow stimuli were shown in stereotypical series of individual stimulus bouts separated by darkness. Each bout consisted of 0.5 s stationary display of the stimulus pattern followed by 3 s of movement at constant velocity, followed by 3 s of darkness. As patterns we used sinus-gratings of different spatial frequencies (ranging from 0.017 cycles/˚ to 0.067 cycles/˚) at maximal contrast, moving at velocities between 10˚/s to 160˚/s. For translational optic flow, the stimulus moved clockwise in one hemisphere and counter-clockwise in the other hemisphere, while the entire panorama moved in one coherent direction during rotational optic flow.

Receptive fields were mapped using a narrow vertical stripe (width: 7.5˚) that moved around the entire panorama at constant speed (60˚/s) either clockwise or counter-clockwise. Each stimulus consisted of two clockwise rotations followed by two counter-clockwise rotations. The bar was introduced into the arena behind the bee and remained stationary for 0.5 s before movement commenced. Control voltages were recorded for all stimuli, indicating the timing of displayed frames in the virtual reality arena and the angular position of the rotation stage controlling the polarizer.

Electrophysiology

Intracellular recordings were carried out with sharp-tipped electrodes (resistance 50-150 M) drawn from borosilicate glass capillaries as described in detail elsewhere (e.g. [21]). In short, bees were cooled on ice (bumblebees in freezer) until immobile and waxed to a plastic holder. Legs and wings were removed for increased stability of the preparation. The head capsule was opened frontally between the antennal base and the ocelli, and air-sacks and fat tissue were pushed aside or removed if necessary. The brain surface was shortly exposed to Pronase (crystals applied directly), after which the neural sheath was removed with tweezers. A silver wire was placed in the ventral part of the head (near mandibles) as reference electrode. After placing the preparation in the center of our virtual reality arena (vertical orientation), the recording electrode (tip filled with 4% neurobiotin (Vector Laboratories) in 1M potassium chloride, backed up with 1M potassium chloride) was frontally inserted into the brain using the antennal lobes and the vertical lobes of the mushroom body as landmarks (Sensapex micromanipulator, stepping mode). Target areas were the noduli and the central body of the central complex. Once cells were impaled and the stimulation protocol was successfully tested, a depolarizing current (1-3 nA) was applied to iontophoretically inject neurobiotin into the recorded neuron in most experiments.

Recordings were performed throughout the day on 160 female worker bees (with the exception of very few male *Megalopta*). With few exceptions, only one cell was recorded and analyzed per bee to ensure a clear correspondence between anatomy and physiology. Whenever the recording was lost before dye injection, the physiology of more than one cell was recorded, but never were two cells of the exact same type encountered within the same bee in these cases. Therefore, for physiological data the number of cells always equals the number of bees used for the displayed cell population. Data were amplified with a BA-03X amplifier (NPI) (Panama set-

up) or a SEC05-LX amplifier (NPI) (Sweden setup), digitized using CED-1401 micro (Cambridge Electronics Design), and recorded with Spike2 software (Cambridge Electronics Design).

Data analysis

Action potentials were extracted from the recorded voltage traces using threshold based event detection in MATLAB. Only recordings with stable baseline were evaluated. Timing of events was then correlated to the recorded stimulus traces by custom designed analysis scripts.

For optic-flow stimuli each stimulus bout was analyzed independently. Spikes were counted in bins of 0.25 s during stimulation intervals and the resulting mean frequencies were plotted for display of individual stimulus responses. To calculate tuning curves, the last 2 s of each stimulus interval was used to compute the mean response frequency for the analyzed condition and plotted against either stimulus velocity or spatial frequency. Background activity of a neuron was calculated as the mean activity during 2 s before onset of the first stimulus. For averaging tuning curves, we normalized each individual tuning curve to its maximal value and then calculated the mean and standard deviation of each stimulus condition across all recorded neurons of the same type. Mean normalized background activity was calculated accordingly.

Receptive field mapping was analyzed by finding the number of events during each frame of the display of the moving bar (192 frames per rotations; 6 s per rotation) and averaging the resulting frequencies over three neighbouring frames (resulting bin duration: ca. 0.1 s). The mean across the two stimulations with identical movement direction was calculated for each bin and the result was low pass filtered (window-size: 3 bins) to display receptive field responses of individual cells. Filtered data were normalized to peak frequency and averaged across neurons of the same type. Mean and individual receptive fields were displayed either as heat maps projected onto a

32x192 matrix (dimension of LED arena; colour represents mean frequency) or as line plots (mean ±SD).

Translational optic flow with different expansion points was analyzed as for tuning curve analysis of other optic flow stimuli. The resulting mean frequencies during each stimulation interval were then displayed in a circular plot against the azimuth of expansion of the optic flow. The maximally effective stimulus was calculated by converting the mean spike frequencies of each response to a distribution of azimuth angles (spike frequency = frequency of corresponding azimuth angle in the distribution). The circular mean of this distribution was calculated with the circular statistics package for MATLAB [44].

Polarized-light stimuli were analyzed by assigning angles (position of the polarizer) to each spike-event. This list of angles was used to display data in circular plots (bin size 10˚, converted from spike count to frequency by dividing by bin duration) and to perform statistical analysis. A cell was rated polarization sensitive if the distribution of action potentials during a rotation of the polarizer was significantly different from a uniform distribution (Rayleigh test for axial data; alpha level 0.05). Clockwise and counter-clockwise rotations of the polarizer were analyzed separately. Tuning curves were calculated by normalizing each response to its peak firing rate, adjusting the minimum firing rate during stimulation to zero, and shifting the curve, so that peak activity coincided with 0˚ E-vector angle. The width of the tuning curve at half-maximal excitation was taken as the tuning width.

Histology for light microscopy

Neurobiotin injected brains were processed as described elsewhere (e.g. [21]). In short, injected brains were removed from the head capsule and fixed in neurobiotin fixative (4% paraformalde-

hyde, 2% saturated picric acid, 0.25% glutaraldehyde) over night at 4˚C. Brains were transferred to 0.1M PBS until further processing. After rinsing the brains for 4x15min in PBS, they were incubated with Streptavidin conjugated to Cy3 (1:1000, in 0.1M PBT (PBS plus 0.3% TritonX-100)) for 3 days. The brains were then washed (4x20 min PBT, 2x 20 min PBS) and dehydrated in an increasing ethanol series. Finally, they were cleared in Methyl salicylate and mounted in Permount (between two coverslips, separated by spacers).

Immunohistochemical stainings were performed to visualize the boundaries of brain regions, in particular the central complex. Antibodies against the synaptic protein Synapsin (monoclonal, mouse antibodies; obtained from Drs. Erich Buchner and Christian Wegener) were used for this purpose [45]. Brains were dissected and fixed in Zinc-Formaldehyde fixative [46] (overnight at 4˚C). After washing in Hepes-buffered saline (HBS, 8x30 min; [46]), the brains were subjected to a permeabilisation step (85 min incubation with a fresh mixture of DMSO and Methanol (20:80)) and washed 3x 10 min in Tris-HCL buffer. After pre-incubation in normal goat serum (NGS, 5% in PBT), the brains were incubated with the primary antibody solution (anti-Synapsin, 1:25, in PBT with 1% NGS) for 5-6 days at 4˚C. The brains were washed in PBT (8x 30 min) and incubated with the secondary antibody solution (goat-anti-mouse antibody conjugated to Cy5, 1:300, in PBT with 1% NGS) for 4-5 days at 4˚C. After washing (4x 30 min PBT, 2x 30 min PBS) the brains were dehydrated, cleared and mounted as described above.

Block-face electron microscopy

Bumblebee brains were dissected in fixative (4% paraformaldehyde, 2% glutaraldehyde, in Cacodylate-buffer) over night at 4˚C. After washing 4x 15 min in 0.1M PBS, the brains were embedded in albumin/gelatin and postfixed over night in 4% paraformaldehyde. To cut out the re-

gion of interest (central-complex noduli), a single thick section (200 µm) was cut from each brain using a vibrating blade microtome. These sections were stored in PBS until further processing.

Samples were then rinsed in PBS buffer for 5 min, before osmification (in 2% osmium tetroxide and 1.5% potassium ferricyanide, aqueous solution) for 1 h at room temperature. Then, samples were washed in pure water (3 x 5 min) and incubated for 20 min in thiocarbohydrazide solution (1% aqueous), washed in pure water (3 x 5 min) and incubated for another 30 min in osmium tetroxide (2% aqueous; room temperature). After more rinsing (3 x 5 min) in pure water, samples were transferred into 1% uranyl acetate for overnight incubation at 4˚C. Finally, the samples were incubated with lead aspartate (after 3 x 5 min washing in pure water) for 60 min at 60˚C (oven). After more washing (3 x 5 min pure water), the samples were dehydrated in an increasing ethanol series (20%, 50%, 70% 90%, 2x 100%; 5 min each). Next, the samples were slowly infiltrated with a mixture of Durcupan and ethanol (25%, 50%, 75%, 2 h each), before transferring them to 100% Durcupan overnight. After transferring to fresh Durcupan for 2 h, samples were embedded in fresh resin for 48 h at 60˚C. Samples were embedded in a thin layer of resin (just enough to cover the sample). They were trimmed and mounted onto aluminium stubs using 2 part conductive silver epoxy.

Imaging and image processing

Electron-microscopical imaging was carried out using a Zeiss Sigma VP scanning electron microscope equipped with a Gatan 3view. Low resolution scans were obtained with a voxel size of 100 x 100 nm (field of view: 400 x 400 µm) in 100 nm steps at 2 kV, medium resolution scans were done with a voxel size of 23 x 23 nm (field of view: 95 x 95 µm) in 50 nm

steps and at an energy of 2 kV, while high-resolution scans were carried out with a voxel-size of 11.5 x 11.5 nm at an energy of 2 kV (field of view: 46 x 46 µm) in 50nm steps. Acquired images of each stack were aligned and several consecutive imaging runs were merged into one continuous stack using Amira 5.3 software. After contrast optimization, the image stack was down-sampled to 8 bit image depth. This enabled use of the Skeletonize plugin for Amira [47] to perform neuron tracing, as well as to perform image segmentation (Amira segmentation editor).

Neuron profiles were traced manually by adding short, straight skeleton segments linked by branchpoints. The midline of large branches was automatically fitted to match the image information. The same tracing was carried out in both the high-, medium- and low-resolution stacks. For low-resolution stacks only the main neurite and the largest branches of cells were traceable, albeit over longer distances. Neuropil boundaries of the noduli were generated after down-sampling the image stack to 1x1x1 µm voxel size. We used the segmentation editor of Amira to manually segment key cross sections of the noduli in all three spatial planes. This provided a 3D scaffold of the noduli that was then used as the basis for generating a surface model of this brain region using the Amira wrapping algorithm. Volumetric reconstruction of selected neuronal fibres was also carried out with the Amira segmentation editor. For this purpose, cross sections of neuronal profiles were traced in the highest resolution image-plane (x-y plane) for each section until the 3D shape of the neurite of interest became apparent. The resulting label field was used to generate a surface model of the neurite. Active zones were identified by aggregations of synaptic vesicles adjacent to an electron-dense part of the neuronal membrane. Profiles directly opposite of the active zone were identified as postsynaptic.

Confocal imaging was carried out with a Zeiss LSM 510 equipped with a 10x objective (Plan Neofluar 0.45, water immersion, Zeiss) for imaging large neuropil structures and whole brains, as well as a 25x long distance objective (LD LCI Plan-Apochromat 25×/0.8 Imm Corr DIC, Zeiss) for obtaining single-neuron morphologies and high-resolution neuropil data. Neurons were imaged at a voxel size of 0.3x0.3x0.88 µm. All image stacks together covering the full extent of one cell were aligned to a common reference frame (using Amira) and used as input to the skeletonize plugin[47]. Neurons were traced manually and the resulting skeletons were finalized by automatic midline fitting and diameter adjustment (using local brightness information of the image data). Neuropils were reconstructed by manual image segmentation of key cross-sections of each structure combined with automatic surface wrapping in Amira. This was carried out for neuropils innervated by each neuron (based on background staining) as well as from anti-Synapsin antibody labelled preparations.

Reconstructions based on antibody labelling were used to generate a representative reference brain, which then served as a shared frame of reference for individual neuron data (neuron atlas). Individual neuron data were registered into this reference brain via manual affine registration (nine degrees of freedom). Registration parameters were obtained by registration of individual neuropil models to their counterparts in the reference brain. For TN-neurons we used the compartments of the central complex, as well as the lateral antennal lobe tract as reference structures.

Simulation Methods

The proposed path integration network was implemented in Python 2.7, using a simple firing rate model for each neuron[48], in which the output firing rate r is a sigmoid function of the input I:

$$r = \frac{1}{(1+e^{-(aI-b)})}$$

where parameters a and b control the slope and offset of the sigmoid (see Extended Data Fig. 6 for parameter values and curve shapes). Optional Gaussian noise $\epsilon_r \sim \mathcal{N}(0, \sigma_r^2)$ can be added to the output, which is then clipped to fall between 0 and 1. The input I is given by the weighted sum of activity of neurons that synapse onto neuron j:

$$I_j = \sum_i w_{ij} r_i$$

In the current simulation these weights only take values of 0 (no connection), 1 (excitation) or -1 (inhibition) with optional added Gaussian noise $\epsilon_w \sim \mathcal{N}(0, \sigma_w^2)$. In the case of added noise, the sign of weights is preserved by clipping any values that fall outside this range, i.e. excitatory connections cannot become inhibitory and vice versa. Rather than tuning the weights, we tune the sigmoid function parameters for each neuron type to balance the number and scale of the inputs to each layer (see Extended Data Fig. 6a). Tuning was carried out visually, by attempting to ensure that each layer would cover a full range of firing rates during a batch of typical runs. Due to the robust nature of this network, many combinations of parameters work and there was no need to carry out extensive parameter tuning, e.g. through a grid search.

The model consists of five layers of neurons, some of which have additional properties to those above (described in more detail below): the TN and TL (input) layers receive direct sensory input from the agent; the TB1 (compass) layer has self-connection weights with values that can fall between -1 and 0; the CPU4 (proposed memory) layer has additional synaptic accumulation; and the CPU1 (output) layer connects to the agent's motor system. In the following description of the individual layers of our model, we use θ for allocentric and φ for egocentric angles. A superscript in parentheses is used to represent values at a particular time step, and subscripts to differentiate parameters by layer and cell index.

Speed Layer 1 - TN neurons: In our simulation the speed estimate, in terms of forward-to-backward optic flow originating from the diagonally offset preference angles of TN cells on each hemisphere, is calculated by

$$I_{\mathrm{TN_L}} = \begin{bmatrix} \sin(\theta_\mathrm{h} - \phi_\mathrm{TN}) & \cos(\theta_\mathrm{h} - \phi_\mathrm{TN}) \end{bmatrix} \mathbf{v}$$

$$I_{\mathrm{TN_R}} = \begin{bmatrix} \sin(\theta_\mathrm{h} + \phi_\mathrm{TN}) & \cos(\theta_\mathrm{h} + \phi_\mathrm{TN}) \end{bmatrix} \mathbf{v}$$

where v is the velocity of the agent in Cartesian coordinates, $\theta_\mathrm{h} \in [0,\ 2\pi)$ is the current heading of the agent and φ_{TN} is the preference angle of a TN neuron, i.e. the point of expansion of optic flow that evokes the biggest response. For our model, a default preference angle of $\phi_\mathrm{TN} = \frac{\pi}{4}$ was used. TN2 neurons act as a rectified linear function, meaning they respond in a positive linearly proportional manner to I_{TN}, but have no response to negative flow (backwards motion) (Extended Data Fig. 6a).

$$r_{\mathrm{TN2}} = \max(0, 2I_\mathrm{TN} - 1)$$

Optional Gaussian noise ε_r can be added to the output, after which activity is clipped to fall between 0 and 1, as above.

Heading Layer 1 - TL neurons: The first direction-related layer consists of 16 inhibitory TL neurons, which have been shown to be polarization sensitive across a range of insect species [13,20,22] and to encode visual landmarks used to compute heading direction in flies (ring neurons [11]). Each TL neuron has a preferred direction θ_{TL}, with the 16 neurons representing 8 cardinal directions $\theta_{TL} \in \{0, 45, 90, 135, 180, 225, 270, 315\}$ twice over. Collectively they encode the heading of the agent at every time step, by each receiving input activation corresponding to the cosine of the angular difference between the current and their preferred heading:

$$I_\mathrm{TL} = \cos(\theta_\mathrm{TL} - \theta_\mathrm{h})$$

Heading Layer 2 - CL1 neurons: The 16 CL1 neurons have a response as described in [49] i.e., they are inhibited by TL neuron activity, effectively inverting the polarization response. This is included for completeness but makes no functional difference in our current model.

Heading Layer 3 – TB1 neurons: The 8 TB1 neurons receive excitatory input from each pair of CL1 neurons that share same directional preference, θ_{TB1}. The TB1 layer also contains mutually inhibitory connections (see Extended Data Fig. 6b), with a weighting that reflects stronger inhibition for greater difference in their preferred directions [9,10,14]:

$$w_{ij} = \frac{\cos(\theta_{\mathrm{TB1},i} - \theta_{\mathrm{TB1},j}) - 1}{2}$$

Where $\theta_{TB1,i}$ and $\theta_{TB1,j}$ are the preferred directions of their respective CL1 inputs. The total input for each TB1 neuron is:

$$I^{(t)}_{\mathrm{TB1},j} = (1-c) r^{(t)}_{\mathrm{CL1},j} - c \sum_{i=1}^{8} w_{ij} r^{(t-1)}_{\mathrm{TB1},i}$$

where c = 0.33 is a scaling factor for the relative effect of lateral TB1 inhibition compared to the direct CL1 excitation. This layer thus acts as a ring attractor[50], which creates a stable sinusoidal encoding of the heading direction, reducing noise from the previous layers, and forming the underpinning for the accurate memory and steering functions in subsequent layers.

Layer 4 - CPU4 neurons: The CPU4 layer consists of 16 neurons. The input for these neurons is an accumulation of heading $\theta_h^{(t)}$ of the agent represented by the sinusoidal TB1 response, modulating the speed signal from the TN2 neurons in the noduli, as reported in the current paper. In addition, there is a constant memory decay to all CPU4 cells:

$$I^{(t)}_{\mathrm{CPU4}} = I^{(t-1)}_{\mathrm{CPU4}} + h(r^{(t)}_{\mathrm{TN2}} - r^{(t)}_{\mathrm{TB1}} - k)$$

where h = 0.0025 determines the rate of memory accumulation and k = 0.1 the uniform rate of memory loss. All memory cells are initialized with a charge of $I^{(0)} = 0.5$ and as they accumulate are clipped on each time step to fall between 0 and 1. The eight TB1 neurons each provide input to two CPU4 neurons, each of which also receives input from a single TN2 cell, coming from either the left or right hemisphere. As these neurons integrate the velocity (speed and direction) of the agent, activity across this layer at any point in time provides a population encoding of the home vector.

Layer 5 - Pontine neurons: 16 pontine neurons project contralaterally and connect two CBU columns eight columns apart from one another [31,34] (see Extended Data Fig. 5e,f and Extended Data Fig. 6b). Each cell receives input from one CPU4 column:

$$I_{\mathrm{Pontin}}^{(t)} = r_{\mathrm{CPU4}}^{(t)}$$

Layer 6 - CPU1 neurons: The CPU1 layer has 16 neurons. It consists of two subtypes of neurons, CPU1a and CPU1b that exhibit distinct projection patterns between the PB and the CBU and are conserved across insect species[31,34]. Each TB1 neuron provides inhibitory inputs (weight = -1) to two CPU1 neurons, in the same pattern as TB1-CPU4 connections. Additionally, each CPU4 neuron provides input to a CPU1 neuron, but with the offset connectivity pattern shown in figure 4g & h, which produces the connectivity matrix shown in Extended Data Fig. 6b. As TB1 input is inhibitory and CPU4 input excitatory, the effective input to CPU1 is the difference of the activity in these units, representing the difference between the integrated path and the current heading direction. Finally, CPU1 cells also receive inhibitory input from contralateral pontine neurons so their total input is:

$$I_{\mathrm{CPU1}}^{(t)} = r_{\mathrm{CPU4}}^{(t)} - r_{\mathrm{Pontin}}^{(t)} - r_{\mathrm{TB1}}^{(t)}$$

The CPU1 neurons form two sets, connecting to either the right or left motor units (postulated to be located in the lateral accessory lobes, the anatomical convergence site of CPU1 neurons). The activation of each set is summed, and the difference determines the turning direction and angle of the agent. Currently this is done by multiplying the difference in summed activity by a constant m=0.5, which is used to change the heading of the agent by that amount of radians:

$$\theta_{\mathrm{h}}^{(t)} = \theta_{\mathrm{h}}^{(t-1)} + m\left(\sum_{i=1}^{8} r_{\mathrm{CPU1L},i}^{(t)} - r_{\mathrm{CPU1R},i}^{(t)}\right)$$

All connection weight matrices and other model parameters can be seen in Extended Data Fig. 6.

In our model, the unit of distance is arbitrary, so we describe everything here in terms of steps in x and y and time steps t, which provides a meaningful measure of accuracy on a homing task by examining the tortuosity of a homing route, the angular errors, and errors relative to the distance of the outbound path. Outbound routes were generated by a filtered noise process, approximating a second order stochastic differential equation (SDE):

$$\omega^{(t)} = \lambda\omega^{(t-1)} + \epsilon_{\omega}$$

$$\theta_{\mathrm{h}}^{(t)} = \theta_{\mathrm{h}}^{(t-1)} + \omega^{(t)}$$

where for each time step the change in angular velocity ε_{ω} was generated by drawing from a von Mises distribution with zero mean:

$$\epsilon_{\omega} \sim VonMises(\epsilon_{\omega}|\mu = 0, \kappa)$$

where μ measures the location and κ is the concentration. For our simulations $\kappa = 100$, with smaller values increasing the tortuosity of the outbound route. We used $\lambda = 0.4$, to minimize excessive spiralling motion. Acceleration a for outbound routes is generated by drawing $\frac{T}{50}$ evenly spaced values from a uniform distribution:

$$a \sim U(a_{\min}, a_{\max})$$

and setting the acceleration between those points using third order spline interpolation, causing the agent to speed up and slow down in a smooth manner, thus imitating natural flight behaviour. Velocity of the agent is determined at each time step by a linear drag model:

$$\mathbf{v}^{(t)} = \mathbf{v}^{(t-1)} + \begin{bmatrix} \sin(\theta_{\mathrm{h}}^{(t)}) \\ \cos(\theta_{\mathrm{h}}^{(t)}) \end{bmatrix} \cdot a^{(t)}(1 - F_D)$$

where F_D=0.15 is the default drag. For regular trials a_{min} = 0 and a_{max} = 0.15 were tuned to cause **v** to mostly fall below 0 and 1, allowing the TN cells to capture all speeds without their activity saturating, whereas for inbound paths a constant a = 0.1 was used. The agent's starting position on each simulation is $x^{(0)} = 0$, $y^{(0)} = 0$. These are updated iteratively depending on the velocity.

$$x^{(t)} = x^{(t-1)} + v_0^{(t)}$$

$$y^{(t)} = y^{(t-1)} + v_1^{(t)}$$

All code used for the described simulation is available at https://github.com/InsectRobotics/path-integration.

Testing of the model in an agent simulation

Each trial was evaluated at five values of noise: $\sigma_r^2 \in \{0, 0.1, 0.2, 0.3, 0.4\}$. The effect of these noise levels on the sigmoid I/O functions for all neurons is illustrated in Extended Data Fig. 6a. As a control, attempting to find home using a random walk with similar parameters to the outbound route generation mechanism was also included. N=100 trials were run at twenty distances ranging between 10 and 10,000 steps. The trials distances were equally spaced on a logarithmic scale:

$$d_x = 10^{1+\frac{3x}{20}}$$

where x ∈{0,1, …, 20}. In each trial a unique route was generated up to a turning point. For each route the agent was given the same number of return steps as the outbound steps. The closest distance (in steps) that the agent obtained from the nest during this length-limited return path was used to quantify the homing success (Fig. 5f). Accuracy can also be evaluated by whether the agent was heading the right way towards the nest shortly after it began homing, a commonly used technique when conducting field experiments with homing insects [51]. Homing direction was determined by measuring the angle when the agent exited a perimeter with 20 step radius around the turning point. The homing angle and correct angle between the nest and the turning point was compared (Fig. 6g). Performance is good with a strong bias towards the nest angle that breaks down gradually with increased noise. Location estimate was decoded directly from the built up memory representation of CPU4 neurons (before the addition of output noise) by summing the two CPU4 subpopulations as shown in Extended Data Fig. 7. Finally, to show that the agent was moving towards the nest in a relatively straight manner we measured the tortuosity of the return route, using the formula:

$$\tau = \frac{L}{C}$$

where *L* is the straight distance to the nest, and C is the distance towards the nest that the agent has actually covered after L steps, i.e., $C = L - d_L$, where d_L is the distance of the agent from the nest after L steps (Fig. 5h).

Additional references

44. Berens, P. CircStat: A MATLAB toolbox for circular statistics. *Journal of Statistical Software* **31,** (2009).

45. Klagges, B. *et al.* Invertebrate synapsins: A single gene codes for several isoforms in *Drosophila*. *J Neurosci* **16,** 3154–3165 (1996).

46. Ott, S. R. Confocal microscopy in large insect brains: zinc-formaldehyde fixation improves synapsin immunostaining and preservation of morphology in whole-mounts. *J Neurosci Methods* **172,** 220–230 (2008).

47. Schmitt, S., Evers, J. F., Duch, C., Scholz, M. & Obermayer, K. New methods for the computer-assisted 3-D reconstruction of neurons from confocal image stacks. *Neuroimage* **23,** 1283–1298 (2004).

48. Hertz, J., Krogh, A. & Palmer, R. G. Introduction to the theory of neural computation. (Basic Books, 1991).

49. Bockhorst, T. & Homberg, U. Amplitude and dynamics of polarization-plane signaling in the central complex of the locust brain. *J Neurophysiol* **113,** 3291–3311 (2015).

50. Zhang, K. Representation of spatial orientation by the intrinsic dynamics of the head-direction cell ensemble: a theory. *J Neurosci* **16,** 2112–2126 (1996).

51. Collett, M., Collett, T. S., Bisch, S. & Wehner, R. Local and global vectors in desert ant navigation. *Nature* **394,** 269–272 (1998).

Supplementary Information

1) SUPPLEMENTARY NOTES

Mathematical framework

An intuitive way to understand how our central complex model achieves path integration is to think in terms of the addition of sinusoids at the fundamental frequency, i.e. with a period of 2π, similar to a theoretical solution to path integration proposed by [37]. Due to the lateral inhibition of the TB1 (compass) cells, the activity bump across these neurons in the protocerebral bridge is roughly sinusoidal and can be used to encode an allocentric heading. We propose that the amplitude of this activity bump is scaled in a speed dependent manner by TN (speed) neurons and is repeatedly added at the CPU4 (memory) neurons where these two signals converge. A theoretical outline of why this works is given below, before elaborating on how this fits in with the architecture of the central complex.

Any point in R^2 (real coordinate space in two dimensions) can be expressed as a sinusoid representing polar coordinates, where the amplitude captures the radial coordinate and the phase shift captures the angular coordinate. A population of cells using a sinusoid to encode a location could be described as:

$$A\cos(\theta+\alpha)$$

where A represents the distance travelled, α represents the heading and each cell has its own preference angle θ. A property of this representation is that any linear combination of two sinusoids with the same period results in another sinusoid with same period, but different amplitude and phase shift:

$$\begin{aligned} A\cos(\theta+\alpha)+B\cos(\theta+\beta) &= A\left[\cos(\theta)\cos(\alpha)-\sin(\theta)\sin(\alpha)\right] \\ &\quad + B\left[\cos(\theta)\cos(\beta)-\sin(\theta)\sin(\beta)\right] \\ &= \underbrace{\left[A\cos(\alpha)+B\cos(\beta)\right]}_{x}\cos(\theta) - \underbrace{\left[A\sin(\alpha)+B\sin(\beta)\right]}_{y}\sin(\theta) \\ &= x\cos(\theta)-y\sin(\theta) \end{aligned}$$

For any x,y ∈ R we can write $x = R\cos(\psi)$ and $y = R\sin(\psi)$ with $\psi \in [0, 2\pi)$ and R>0.

$$\begin{aligned} A\cos(\theta+\alpha)+B\cos(\theta+\beta) &= R[\cos(\psi)\cos(\theta)-\sin(\psi)\sin(\theta)] \\ &= R\cos(\theta+\psi) \end{aligned}$$

To find R and ψ we can use the Pythagorean identity:

$$\begin{aligned} R^2 &= R^2\cos^2(\phi)+R^2\sin^2(\phi) \\ &= \left[A\cos(\alpha)+B\cos(\beta)\right]^2+\left[A\sin(\alpha)+B\sin(\beta)\right]^2 \\ R &= \sqrt{\Big[A\cos(\alpha)+B\cos(\beta)\Big]^2+\Big[A\sin(\alpha)+B\sin(\beta)\Big]^2} \end{aligned}$$

$$\begin{aligned}\psi &= \operatorname{atan2}(R\sin(\psi), R\cos(\psi)) \\ &= \operatorname{atan2}(A\sin(\alpha) + B\sin(\beta), A\cos(\alpha) + B\cos(\beta))\end{aligned}$$

$$\begin{aligned}A\cos(\theta+\alpha) + B\cos(\theta+\beta) = &\sqrt{\Big[A\cos(\alpha) + B\cos(\beta)\Big]^2 + \Big[A\sin(\alpha) + B\sin(\beta)\Big]^2} \\ &\cdot \cos\Big(\theta + \operatorname{atan2}\left[A\sin(\alpha) + B\sin(\beta), A\cos(\alpha) + B\cos(\beta)\right]\Big)\end{aligned}$$

After summing two sinusoids representing consecutive movement vectors the resulting amplitude and phase shift correspond to a sinusoid matching a vector that is the sum of both (Extended Data Fig. 7c). Therefore, by summing the corresponding activity of both cell groups we retain a population code that remains accurately pointing to the new location.

These properties prove to be useful in several ways for the central complex. A single sinusoid (TB1) is repeatedly subtracted (due to inhibition) to form a consistent memory representing a home vector (CPU4). Different speeds can also be captured by modulating the amplitude of the sinusoid (TN).

Motion on a plane can be tracked by measuring the velocity component (speed) along vectors at two offset angles ($\theta_h \pm \varphi_{TN}$), forming a basis. In general, the motion may be holonomic, that is the head direction (origin of egocentric coordinate system) and direction of motion may not be aligned. The two velocity components are stored in two separate memories, the two duplicate populations of CPU4, and could be summed and decoded by re-aligning the sinusoids correctly. In our case, we propose the modulation of the sinusoid happens by inhibition, meaning that the shape of the compass (TB) cannot be inverted (multiplied by a negative constant). Because of this only partial holonomic motion can be accurately integrated, with a direction of motion between $\phi_{TN} - \frac{\pi}{4}$ and $\phi_{TN} + \frac{\pi}{4}$ as defined in the Methods section when describing the Speed Layer 1. Another way to think about this circuit is as a re-projection from an egocentric basis defined with respect to the heading angle at $\theta_h \pm \varphi_{TN}$ to an allocentric one defined with respect to the angle of the sun at $\theta_{TB1} \pm \varphi_{TN}$, where all CPU4 cells with input from the same TB1 cell store motion projected on to their basis vector, at $\frac{\pi}{4}$ intervals (8 columns) creating a redundant memory.

By element-wise multiplying two discrete sinusoids and summing the output, you can determine how much their phase is aligned. The offset between current heading (TB1) and desired heading (CPU4) is estimated through inhibition of the CPU4 activity from connecting TB1 cells. As the animal rotates, the TB1 sinusoid is shifted with respect to the memory. A low response indicates high overlap as all highly active cells are suppressed by the TB1 inhibition, meaning current and desired heading are similar. However, due to the CPU4-TB1-CPU1 connectivity pattern, when comparing CPU4 to the TB1 signal, the two populations that make up the composite memory are each shifted by one column, increasing the existing offset of both memories caused by the TN cell preferences to approximately $\pm\frac{\pi}{2}$. Therefore, a strong alignment with one CPU4 sub-population means the desired heading is approximately $\frac{\pi}{2}$ to the left, and with the other $\frac{\pi}{2}$ to the right. These two values are combined to drive steering as can be seen in Extended Data Fig 7d-

f. In the case of $\phi_{\mathrm{TN}} = \frac{\pi}{4}$ we can prove that the circuit steers correctly. The compound memory can be expressed as a function f(θ), the sum of two memory sinusoids in the same coordinate space (Extended Data Fig. 7e), where the peak indicates the desired heading:

$$f(\theta) = A\cos(\theta + \alpha) + B\cos(\theta + \beta)$$

To find the stationary points which implicitly defines the θ corresponding to the desired heading:

$$f'(\theta) = -A\sin(\theta + \alpha) - B\sin(\theta + \beta) = 0$$

$$-A\sin(\theta + \alpha) = B\sin(\theta + \beta)$$

$$A\cos\left(\theta + \alpha + \frac{\pi}{2}\right) = B\cos\left(\theta + \beta - \frac{\pi}{2}\right)$$

or

$$A\cos\left(\theta + \alpha - \frac{\pi}{2}\right) = B\cos\left(\theta + \beta + \frac{\pi}{2}\right)$$

The correct heading is found when both sinusoids are offset by $\pm\frac{\pi}{2}$ and are equal. This is the case due to the combined $\phi_{\mathrm{TN}} = \frac{\pi}{4}$ and the additional $\frac{\pi}{4}$ columnar offset. However, the circuit still operates robustly with different φ_{TN} values (See results below). At a cellular scale, the steering mechanism can also be understood by observing the activity of a single TB1 compass cell, the corresponding pairs of CPU1 steering cells, and both CPU4 memory cells that connect to them, (Extended Data Fig. 7b). When $\mathrm{TB1}_i = \theta_{\mathrm{h}} - 180°$ it is most inactive, so the CPU4 synapsing on to the corresponding CPU1 will be least suppressed. As the two CPU4 are ±90˚ offset from the TB1 cell, the memory cells that capture the portion of the home vector pointing in θ_h±90˚ will be weighted the most when steering. If the CPU4 with a preference angle of θ_h-90˚ is higher than that of the with preference θ_h+90˚, the agent will turn to the left to head home and vice versa. The pontine cells projecting from CPU4 to CPU1 act as an inhibitory stabilizer. In effect this normalizes both memory sinusoids, ensuring they have the same mean activity, meaning one will not dominate due to unbalanced charging by different TN neurons.

A useful side-effect of this representation is that it becomes trivial to extract a polar coordinate from the population code, e.g. for a waggle dance, by finding the cells with peak activity, while retaining many of the attractive properties of a Cartesian coordinate system [8]. For species that migrate or just want to move in a fixed direction for a while, a permanent or semi-permanent bump in the CPU4 cells would cause them to move consistently in that direction. If retrieving previously visited locations and putting them to CPU4 memory is possible, then this could be done as either a home bound or outward vector, depending on whether the stored memory is added or subtracted from a zero state memory. In addition, a sequence of known paths that link together known landmarks could be sequentially added to memory, resulting in a total home vector, similarly to the addition during normal path integration, meaning the animal could perform shortcuts over previously unvisited terrain.

2) SUPPLEMENTARY DISCUSSION

Responses to noise, parameter changes at different levels

The first experiment to test limitations of the network was to increase the speed of the agent, with the intention of saturating the activity of TN neurons. Acceleration was increased in 5 equal increments up to three times the amount in our default settings, which was initially set to provide a full range of neural activity. As activity was increased, at first barely any drop in path integration performance was noticeable. This is because only a small proportion of motion occurs at the highest speeds. As speed increased beyond 1.5× we see a gradual degradation of performance and the agent undershoots during homing.

In addition to varying the neural activity noise and the route distance (Fig. 5c,d), we tested how robust the model is to small perturbations by running a series of trials. A set of values were chosen for each parameter and N=2000 homing tasks were completed for each value: 1000 trials at default neural activity noise variance σ_r^2=0.1 and 1000 at no noise. Small changes to weight matrices between cells were tested by modifying $\sigma_w^2 \in \{0, 0.01, 0.02, 0.05, 0.1, 0.2\}$. Noise on the weights is less likely to balance out over time than additive noise on each time step, leaving intrinsic bias and making it more detrimental to the system. Regardless, with noise variance up to σ_w^2= 0.05 the network was still able to achieve good homing behaviour.

Perturbations to the angular offset of the speed neurons were made by changing the preference angle between 0 and $\frac{\pi}{2}$ so that $\phi_{\mathrm{TN}} = \frac{x\pi}{16}$ where x ∈ {0,...,8}. Due to the predominant direction of motion being forwards, on trials with neural noise the agent homed with similar good performance on all preference angles where $\phi_{\mathrm{TN}} < \frac{\pi}{4}$, as the direction allowed the full response of the neuron to be taken into account, making the noise proportionately smaller. On noise-free experiments the best angle was around 45˚ as expected. On both tests as angles exceeded 45˚ homing ability decreased gracefully.

Experiments were carried out with a rotation-dependent offset to the preference angle of the compass input neurons (TL and CL). Peak activity would pre-empt the preference angle, similar to the tuning found in real recordings as shown in Fig. 1f. A gradually increasing angle of perturbation was used:

$$\theta_{\mathrm{CL}_{\mathrm{offset}}} = \theta_{\mathrm{CL}} + \mathrm{sgn}(\theta_{\mathrm{h}}^{(t)} - \theta_{\mathrm{h}}^{(t-1)})\frac{x\pi}{32}$$

where x ∈ {0,...,8}. In these tests we did indeed get improved performance from a small bump shift. This could be due to counteracting lag from the signal propagating through the model, especially due to the dynamics of the TB1 ring attractor. Again, after exceeding the ideal angle homing ability decreased in a graceful manner.

Finally, TN input was smoothed to mimic delayed responses seen in Fig. 2b,e. An exponential moving average was used so that:

$$I_{\mathrm{TN}_{\mathrm{smooth}}}^{(t)} = nI_{\mathrm{TN}}^{(t)} + (1-n)I_{\mathrm{TN}_{\mathrm{smooth}}}^{(t-1)}$$

where $n \in \left\{\frac{1}{1}, \frac{1}{2}, \frac{1}{5}, \frac{1}{10}, \frac{1}{20}, \frac{1}{50}\right\}$.

The main conclusion from these tests is that the properties of the model are quite resilient. Failure can be caused by exceeding the limits of the network both in terms of noise tolerance, memory limitations and in terms of changing cell tuning curves from their ideal/observed preference angles. However, no small changes cause catastrophic failure of the model. It works well for reasonable path lengths with a realistic amount of noise and will fail gradually.

List of model assumptions and circuit predictions

- TB1 neurons form a ring attractor network by mutually inhibitory connections onto one another, whose synaptic weights increase with distance of each TB-neuron pair.

Evidence: TB1 cells have no input fibres in PB-columns that neighbour their output fibres, i.e. neighbouring TB1 cells cannot connect to one another. Second, density of input fibres is highest in columns most distant from output fibres. Third, in locusts, responses of CPU1 cells (likely postsynaptic to TB1) are maximally out of phase with TB1 cells, indicating that TB1 cells are likely inhibitory. Also, locust TB1 cells contain the neuropeptide Allatostatin[10], that has been shown to possess inhibitory receptors.

- CPU4 cells synapse onto CPU1 cells in a way that is reflected in their arborisation patterns.

Evidence: In the CBU the polarity of CPU1 cells is postsynaptic, while CPU4 anatomy clearly indicates presynaptic terminals[31,33,34]. Polarity predicted by anatomical means (clear varicose versus smooth terminals) in locust and monarch butterfly CX has proven correct across species in all cases by mapping of presynaptic marker proteins in *Drosophila* (compare e.g. [31] and [33]). This makes the CPU1 cells (together with the highly similar CPU2 cells) the only potential target cells for CPU4 outputs that exist across all species[31,33,34]. The only cell type that is equally likely to receive input from the CPU4 cells are pontine cells, which provide direct connections between ipsi- and contralateral columns within the CBU[31,34,52]. As these cells likely also target the CPU1 neurons, there is either a direct connection between CPU4 and CPU1 cells, an indirect connection between them, or a combination. In fact, our model becomes most robust, if both of these connections are used, which also represents the most parsimonious interpretation of the anatomical data: All overlapping arborisations with the correct polarity are equally likely to correspond to synaptic connections.

- TB1 cells synapse onto CPU1 and CPU4 cells in regions of overlapping arborisations.

Evidence: Both TB1 and CPU1 cells contribute to the mapping of polarized light tunings in locusts, with CPU1 cells out of phase from TB1 cell output columns[10]. Additionally, anatomical polarities of CPU1 and TB1 cells are very pronounced and CPU1 cells additionally show clear postsynaptic potentials when recorded from the PB[53]. Moreover, the TB1 output fibres are the most prominent columnar output arborisations of the PB conserved across insects[22,31-34]. This leaves these cells as the most likely presynaptic cells to all columnar cells with input in the PB, including CPU4 cells.

In locusts, CPU4 cells are conditionally polarization sensitive, with the PB being the most likely source of this information[21]. The only other cell type with major outputs in the PB (likely presynaptic to TB1) are CL1a neurons[21,32,34]. While a network involving CL1a to CPU1 and CPU4 connections with similar overall properties is thinkable, it would be more complex and require more subsequent assumptions.

- CPU4 columns have a mechanism to integrate activity, and thus motion.

Evidence: Neuronal terminals of *Megalopta* CPU4 cells in the PB and the noduli appear of mixed type, indicating both input and output. In the noduli, our EM-work in bumblebees shows that arborisation patterns of CPU4 cells are highly overlapping, in particular between the 19 cells from the same fibre bundle (corresponding to one PB-column), generating a sufficiently large population of cells for memory formation and the spatial proximity needed for column-intrinsic circuits. Furthermore, synaptic connections exist between small EM-profiles within the noduli, most likely corresponding to CPU4 cells, while individual CPU4 cells contain both input and output synapses in the noduli. Finally, preliminary recordings in *Megalopta* have shown responses to translational optic-flow in CPU4 cells (data not shown), while in locusts these cells show context dependent responses to compass stimuli[21].

- Activity is summed for CPU1 on each hemisphere to drive steering.

Evidence: Anatomically, CPU1 neurons converge in highly overlapping regions of the LAL in all species examined[31-34]. Furthermore, in Monarch butterflies and locusts, neurons potentially postsynaptic to CPU1 cells occupy identical regions of the LAL and connect to posterior regions of the brain that provide input to descending pathways[13,21]. Additionally, polarized-light tuning has been found in locust neurons interconnecting the LALs of the right and left brain hemisphere [21]. Highly similar cells are involved in generating steering in response to pheromone pulses in the moth *Bombyx mori* [35].

- CPU4 cells have no influence on steering when the animal is not in a homing state.

Evidence: The decoupling of memory and steering is necessary during behavioural episodes where memory does not drive behaviour. This occurs either at the output or input of CPU1 neurons, thus disabling memory to either reach CPU1 cells or the motor circuit of the LAL. Even though there is no direct evidence for this strong state-dependent switch in active synaptic transmission, there is strong evidence for state-dependent change in neural signalling in the CX in general[21,24], combined with a rich variety of neuromodulator expression in this region[54,55].

- Pontine cells of the CX use inhibition to balance the memory in the CPU4 populations, causing the overall activity on the right brain hemisphere and the left brain hemisphere to remain the same.

Evidence: Pontine cells are highly conserved and are found across all species examined[31,34,41,52]. They have a pattern of midline crossing connectivity uniquely suited to bridge output regions of CPU4 cells and contralateral input regions of CPU1 cells. These cells have been shown to contain

neuromodulators in locusts (CCAP (crustacean cardioactive peptide)[56] and nitric oxide[52,57]) and thus could act through inhibitory metabotropic receptors. Finally, preliminary intracellular recordings from *Megalopta* pontine cells have revealed pronounced, yet non-stereotypical responses to translational optic flow (data not shown).

Expansion into fully holonomic model

The described model can successfully capture holonomic movements in a range of ±45° deviation of body axis and movement direction. This is in line with the observation that honey-bees maintain a steady head direction aligned with their average flight direction during behavioural experiments [58] and likely captures all movements of bees during long-distance flight segments. However, bees also perform hovering flights when close to food sources as well as orientation flights when leaving the nest. During those flight segments, no significant distances are covered. Even though these flight segments are not perfectly captured, our model performs with only minor errors when including them in the outbound trajectory. Nevertheless, given that bees perform fully holonomic flight maneuvers, we deemed it desirable to explore the possibility of expanding our model to enable integration of those flight segments. Surprisingly, only a single additional assumption has to be made to transform our partial holonomic model into a fully holonomic model, without contradicting current anatomical and physiological data. To achieve this the TB1 activity bump indicating compass direction must be able to be stored in either an additive or subtractive manner to CPU4 memory using the input of all four TN neurons identified in this study (i.e. speed neurons and inverted speed neurons; Figs. 2,3). This could be done by the activation from TB1 cells effectively acting as a gating mechanism to either positive or negative effects on the memory from opposing TN cells.

Model predictions

Structural Experiments: First, the number of CPU4 cells, as well as the type of microcircuits formed by parallel CPU4 cells should differ between path-integrating and non-path-integrating insects, reflecting the need for memory in those cells during path integration. Second, block-face EM studies should reveal the proposed synaptic connections between TB1, CPU1, CPU4 cells in the PB and CBU. Third, combined block-face EM in the noduli and the PB should uncover microcircuits formed within CPU4 populations that reveal the concrete memory mechanism for path integration. Fourth, in migratory insects, synaptic remodelling correlated with particular columns of the CBU should be found during the process of seasonal reversal of migratory heading (in the CPU4-CPU1 synapse or the TN-CPU4 synapse). This is because we propose that the migratory heading could be encoded in the form of biased synaptic weights between memory (CPU4) and steering (CPU1) cells. Constant speed input into CPU4 cells then would generate strong output only from those CPU4 cells that correspond to the migratory heading and our circuit would steer the animal towards that heading whenever the animal is in flight. As synaptic weights are plastic, the migratory heading can be changed over time and therefore account for seasonal reversals in migration directions.

Physiological Experiments: First, during path integration the activity of CPU4 cells increases over time and specific patterns of activity across the CX result from specific outbound flights. These patterns as well as the direct accumulation of activity could be observed by electrophysiology or functional imaging. For instance, after training a population of bees on specific outbound routes, these individuals could be captured before returning to the nest. In those full-vector bees, background activity in CPU4 neurons should predictably vary in a column-dependent manner according to their outbound route. Across a population of bees with identical homing vectors these patterns can be established over time. Second, if bees can be maintained in the correct behavioural state (i.e. path integrating while in the recording setup), recordings should reveal increasing activity of CPU4 cells (as well as pontine cells) during path integration. Likely, extracellular recordings in tethered flight will be needed to achieve this. Third, such recordings could also be used to compare summed activity in the right and left LAL during active steering movements with dual extracellular recordings. Finally, it has been shown to be possible to inject current into specific parts of the central complex of cockroaches and thereby affecting behaviour. Similar experiments, complemented possibly by pharmacological inhibition or excitation, are also conceivable in bees. We mimicked such experiments by temporarily increasing the value of the input I (as described in Simulation Methods) to certain neurons, causing a higher firing rate. In our model this was done by clamping the activity of a cell to their maximum firing rate for a fixed period (200 time steps). Injection occurs either when the animal is held still but being exposed to optic flow representing an outbound journey, or during free movement but in a homing state. Injection during inbound journeys can impact all the navigational abilities of the agent. On injection during outbound journeys the agent is unable to steer, however false memories can be induced, allowing for predictable behaviour on the inbound part. Current injections into particular cells cause predictable behaviour (Extended Data Fig. 10): Injecting TB1 cells during either inbound or outbound trips causes the agent to predictably alter their course, due to the compass being offset to the actual direction of motion. The magnitude of the effect depends on the current heading. Injected cells that are misaligned to the current compass bump will cause the biggest adjustments. Injecting individual TN cells causes the agent to adjust course in a predictable manner as well. As one TN cell is induced to fire at a higher rate, the agent will store memory as if severe holonomic motion had occurred, biased towards the direction of higher activity. Injecting CPU1 cells directly affects the steering, causing tight turns towards one side that result in spiralling behaviour. As more cells are injected, the spiralling becomes tighter.

Behavioural experiments: Exposing a bee in a foraging state to optic flow associated with a particular combination of direction of motion and heading should lead to predictable homing errors. Predictions made here depend on the assumptions that a bee's position estimate is primarily based on optic flow integration and is not averaging between other conflicting senses, e.g. smell, wind, proprioception. Our model predicts that when exposed to optic flow beyond a certain speed the TN cells will begin to saturate, meaning the bee will undershoot when homing. Where this threshold lies will depend on the bee species and its natural flight speed and optic flow preference.

For honeybees and bumblebees this value will be higher (in the range of 200˚/s) than in *Megalopta* (ca. 100˚/s).

A number of experiments were also carried out on our model assuming either non-holonomic, partially holonomic and fully holonomic path integration ability (Extended Data Fig. 10). If exposed to optic flow at ±45° relative to heading, an agent that is unable to integrate holonomic motion should also return home at a ±45° offset. For offsets greater than 90° and spurts of randomly angled flow, the agent will not home at all. In our model, the shape of TB activity is modulated to drive memory buildup, however cannot be inverted without making an additional assumption about the nature of memory accumulation (see above). A partially holonomic model such as this causes homing to deteriorate when exposed to flow beyond ±45° and no homing occurs beyond ±135°. For random optic flow the agent should home in the correct direction, but roughly half the distance, whereas fully holonomic integration results in correct homing for all trials.

In addition to the flight paths, by decoding the CPU4 memory population directly we can see the model position estimate. Honeybees, different from *Megalopta* or bumblebees, are known to communicate food sources using polar coordinates through a waggle dance [59]. Exposing the bee to optic flow similarly to above could lead to predictable dances that reflect the positional error made by our neural circuit. It would be very interesting to compare our decoded memory to animals that waggle dance.

Supplementary References

52. Siegl, T., Schachtner, J., Holstein, G. R. & Homberg, U. NO/cGMP signalling: L: -citrulline and cGMP immunostaining in the central complex of the desert locust Schistocerca gregaria. *Cell Tissue Res* **337,** 327–340 (2009).

53. Heinze, S., Gotthardt, S. & Homberg, U. Transformation of polarized light information in the central complex of the locust. *J Neurosci* **29,** 11783–11793 (2009).

54. Kahsai, L. & Winther, A. M. E. Chemical neuroanatomy of the Drosophila central complex: distribution of multiple neuropeptides in relation to neurotransmitters. *The Journal of Comparative Neurology* **519,** 290–315 (2011).

56. Nässel, D. R. & Homberg, U. Neuropeptides in interneurons of the insect brain. *Cell Tissue Res* **326,** 1–24 (2006).

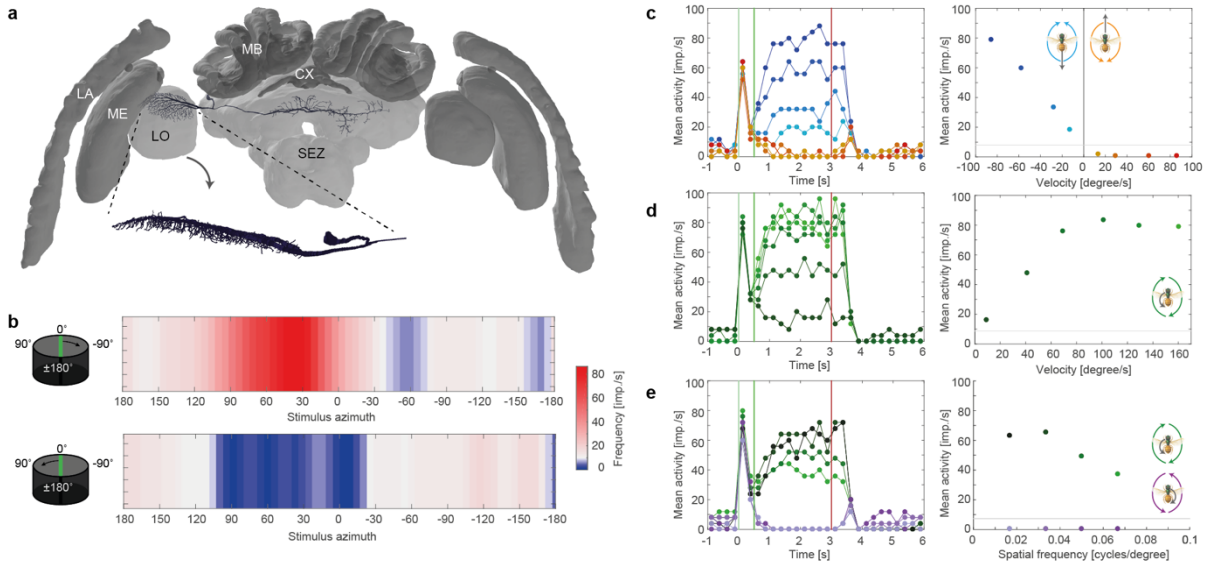
57. Dircksen, H. & Homberg, U. Crustacean cardioactive peptide-immunoreactive neurons innervating brain neuropils, retrocerebral complex and stomatogastric nervous system of the locust, Locusta migratoria. *Cell Tissue Res* **279,** 495–515 (1995).

58. Kunst, M., Pförtner, R., Aschenbrenner, K. & Heinrich, R. Neurochemical architecture of the central complex related to its function in the control of grasshopper acoustic communication. *PLoS ONE* **6,** e25613 (2011).

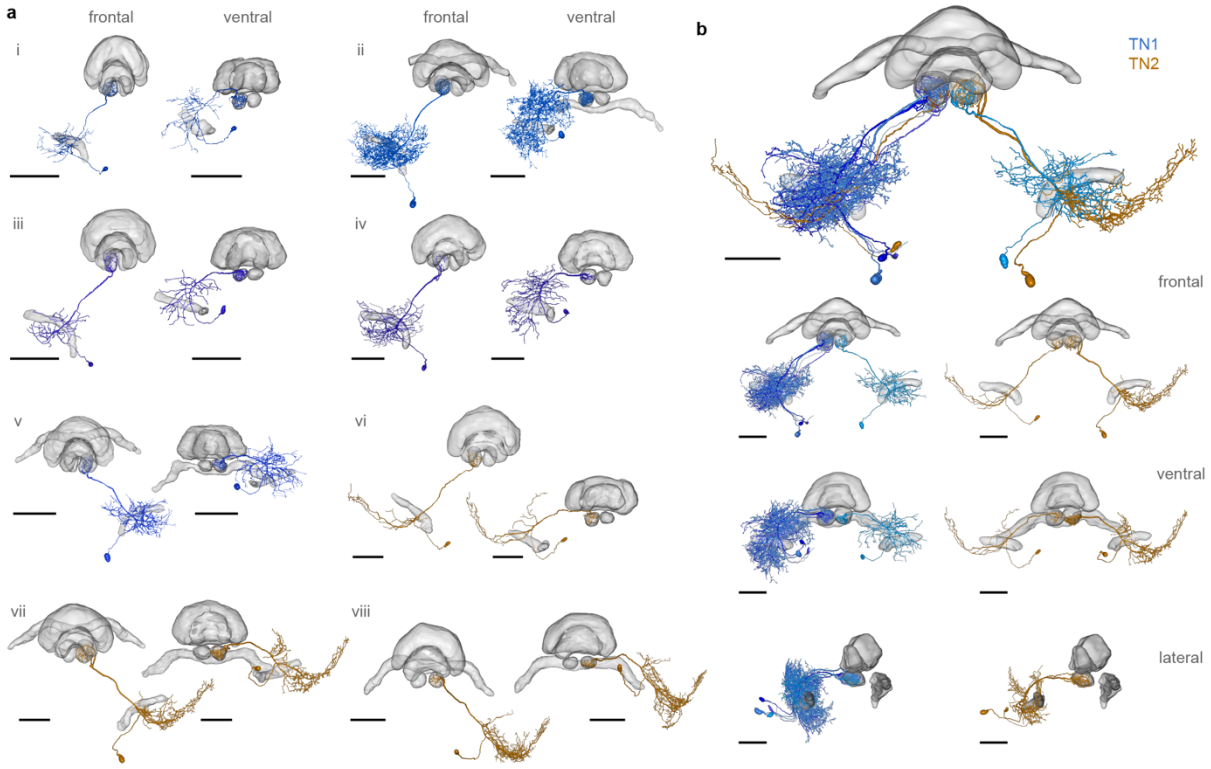
59. Baird, E. Visual Flight Control in the Honeybee. PhD thesis 1–122 (2007).

60. Frisch, von, K. Die Polarisation des Himmelslichtes als orientierender Faktor bei den Tänzen der Bienen. *Experientia* **5,** 142–148 (1949).

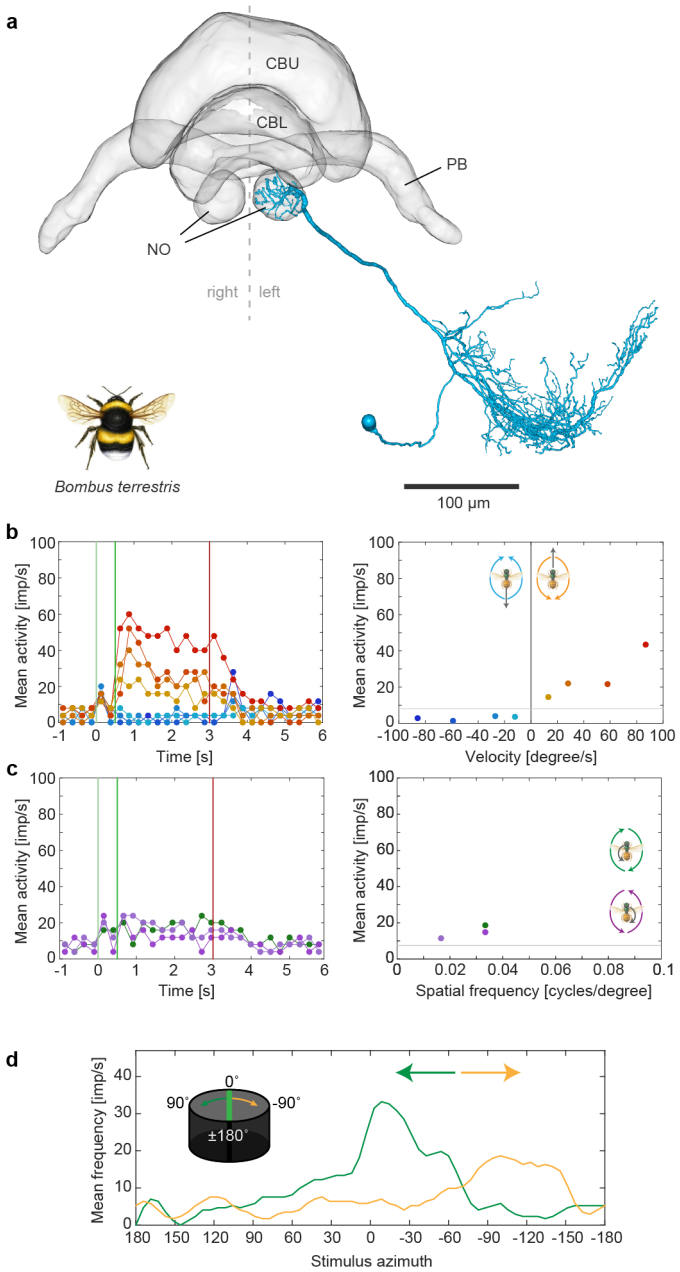
Extended Data Figures



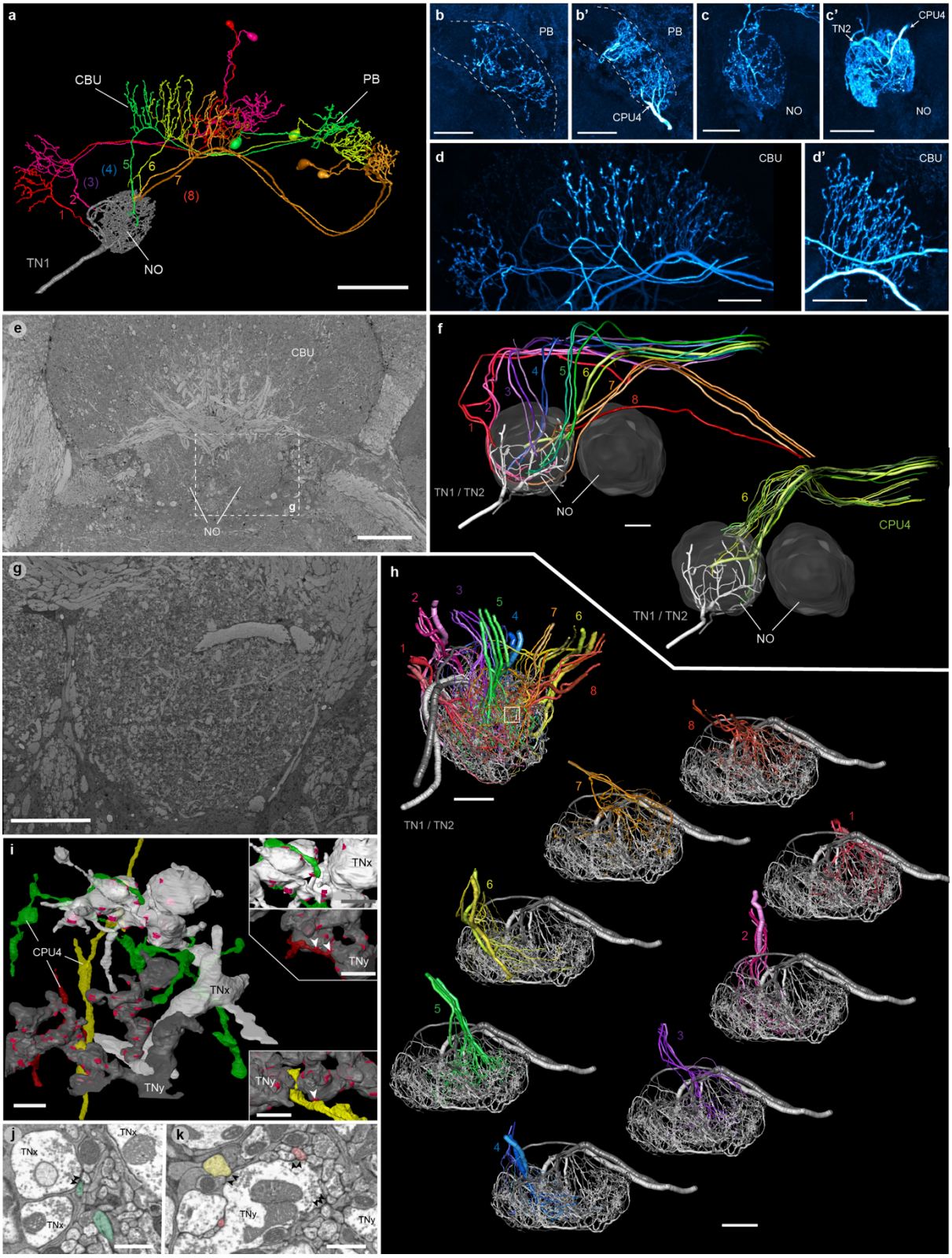
Extended Data Figure 1 | Anatomy and physiology of a lobula output neuron of *Megalopta genalis*. a, 3D-reconstruction of a lobula output neuron projecting from a single layer of the dorsal lobula to bilateral regions of the posterior protocerebrum. The cell is shown in the context of the neuropils from a representative reference brain. Enlargement: input fibres in the lobula are confined to one layer (dorsal view). LA, lamina; ME, medulla; LO, lobula; CX, central complex; MB, mushroom body; SEZ, subesophageal zone. **b,** Receptive field of the same cell type. Top: Response to a green bar moving around the bee clockwise. Bottom: counter-clockwise movement. The responses are opposite in sign depending on movement direction and confined to the frontal, left visual field. **c,** Responses to different velocities of translational optic flow (blue: back to front; orange: front to back). Left: mean frequency during stimulation bouts (0.25 s bins); right: mean frequency during the last 2 s of each stimulation plotted against movement speed. **d,** As in (**c**), but for clockwise rotational optic flow at different speeds. **e,** As in (**c**), but data for different spatial frequencies of clockwise (green) and counter-clockwise (purple) rotational optic flow. Note that the small receptive field of the cell does not enable the neuron to distinguish between the presented rotational and translational optic flow stimuli.



Extended Data Figure 2 | Morphology of TN1 and TN2 neurons. a, 3D-reconstructions of TN1 neurons (i-v) and TN2 neurons (vi-viii) from *Megalopta* bees (i-vii) and bumblebees (viii) shown together with reconstruction of the central-complex neuropils and the lateral antennal-lobe tract (lALT) for better orientation. Left panels: frontal views; Right panels: ventral views. **b,** All *Megalopta* neurons from (**a**) affinely registered into a common reference brain, TN1 cells in blue, TN2 cells in orange. Bottom rows: TN1 and TN2 cells shown separately from frontal, ventral and lateral views to illustrate the different shapes of arborisations between the two cell types, as well as their highly consistent arborisations within each cell type. Note that both cell types overlap in one small region ventral of the lALT. Scale bars: 100 µm.

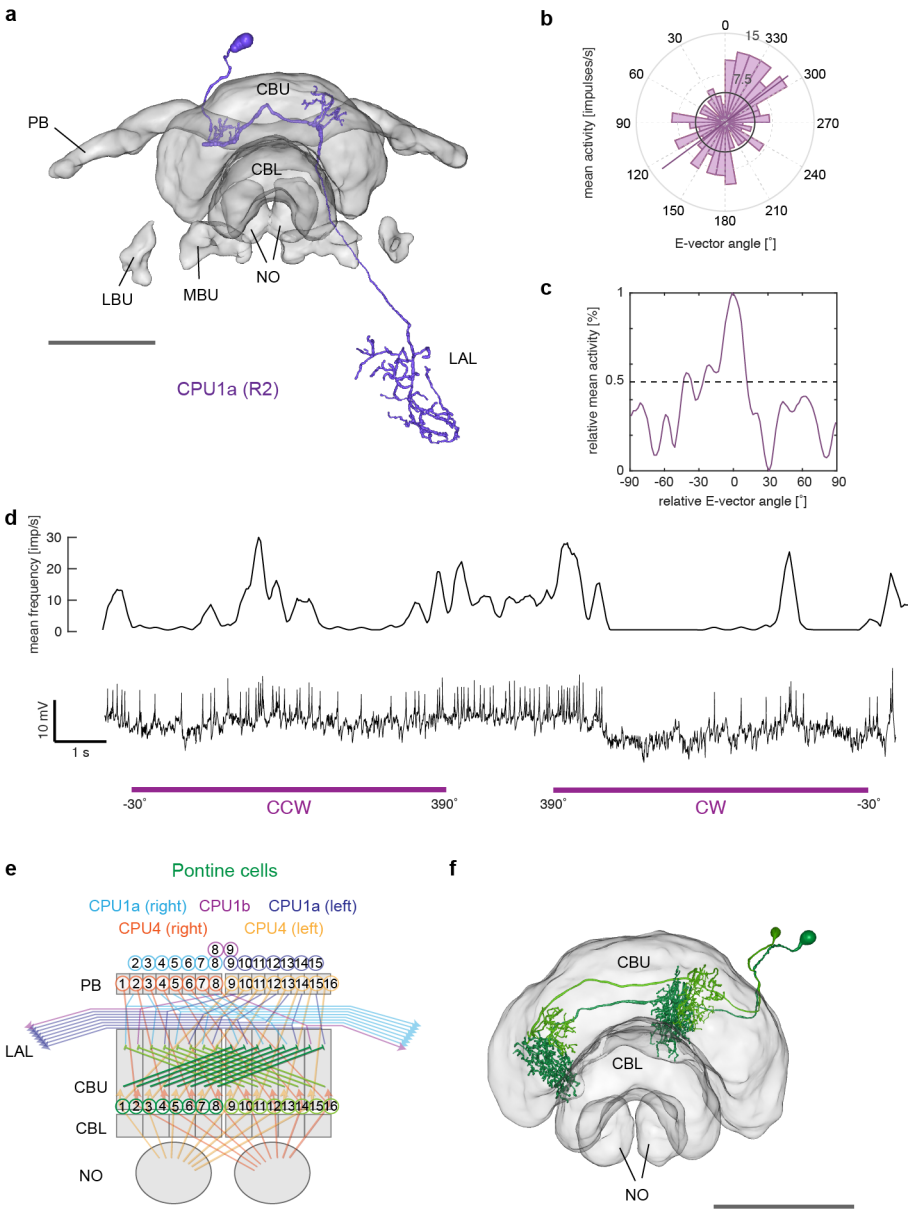


Extended Data Figure 3 | Physiology and anatomy of a bumblebee speed neuron. a, 3D reconstruction of a bumblebee TN2 cell (homologous to *Megalopta* TN2 cell from Fig. 3). CBU, upper division of the central body; CBL, lower division of the central body; PB, protocerebral bridge; NO, noduli. **b,** Responses to different velocities of translational optic flow (blue: back to front; orange: front to back). Left: mean frequency during stimulation bouts (0.25 s bins); right: mean frequency during the last 2 s of each stimulation plotted against movement speed. **c,** As in (**b**), but data for different spatial frequencies of clockwise (green) and counter-clockwise (purple) rotational optic flow. Note that only three data points could be obtained during this stimulus presentation. **d,** Receptive field of the same cell type. Yellow trace: Response to a green bar moving around the bee clockwise. Green trace: counter-clockwise movement. In summary, all stimulus responses are highly similar to the ones from *Megalopta* presented in Figs. 2&3 and indicate conservation of both structure and function of the presented neural circuit.

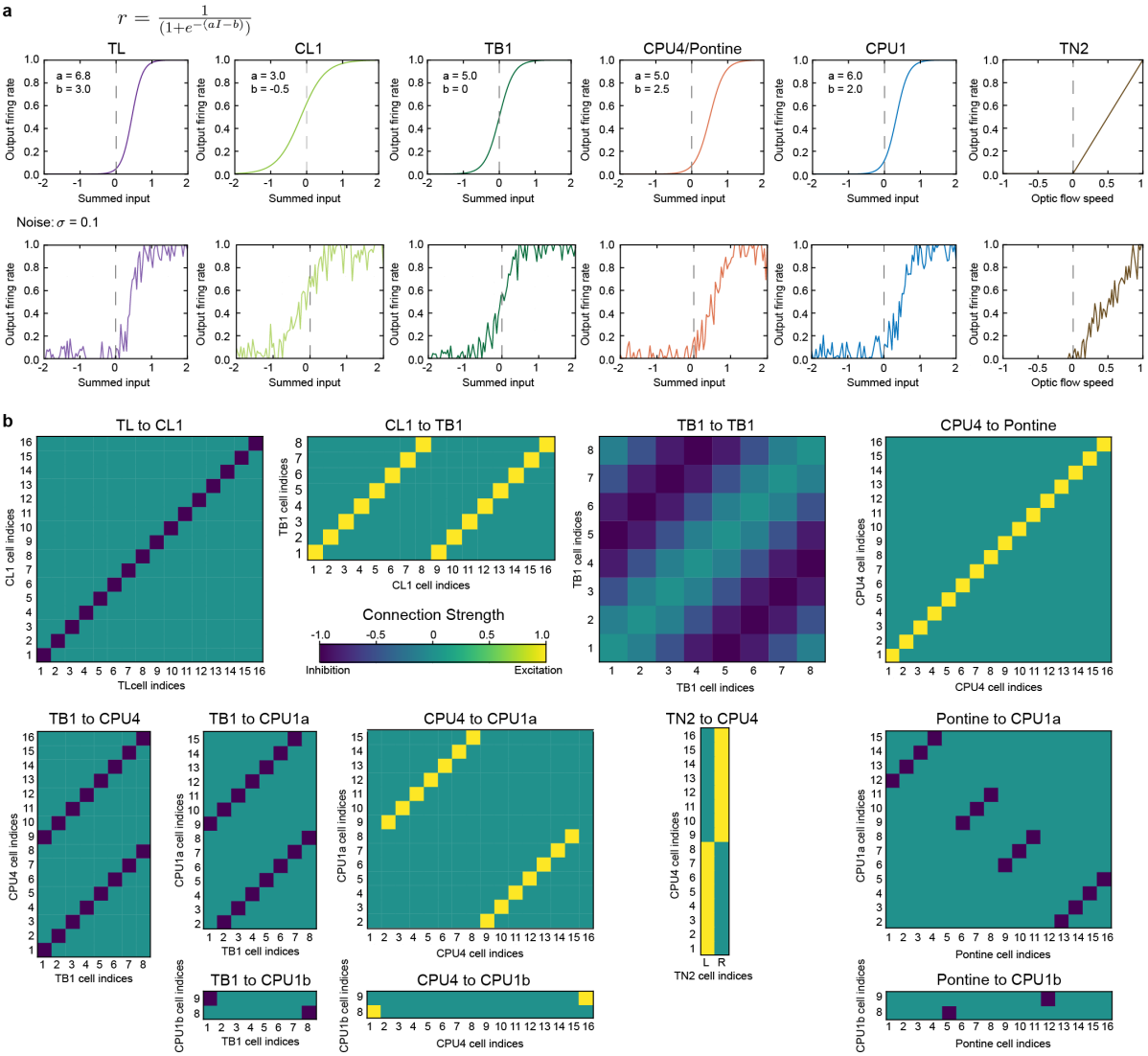


Extended Data Figure 4 | Spatial relation between speed neurons and proposed integrator cells. a, 3D-reconstruction of six CPU4 cells (coloured) simultaneously stained with a single TN1 neurons (grey) from *Megalopta genalis*. The identity of the fibre bundles corresponding to the eight central-complex columns are displayed in identical colours as CPU4 neurons. Bundles not containing any filled neuron are shown in brackets. CBU, upper division of the central body; PB, protocerebral bridge; NO, nodulus. **b-d,** Maximal intensity projections of intracellularly filled CPU4 neuron arborisations in the PB (**b, b'**), noduli (**c,c'**), and the CBU (**d,d'**); two examples shown for each neuropil. Polarity of cells is clearly postsynaptic in CBU, but inconclusive in PB and noduli, suggesting mixed terminals. **e,** Single section from low-resolution (voxel size: 100x100x100

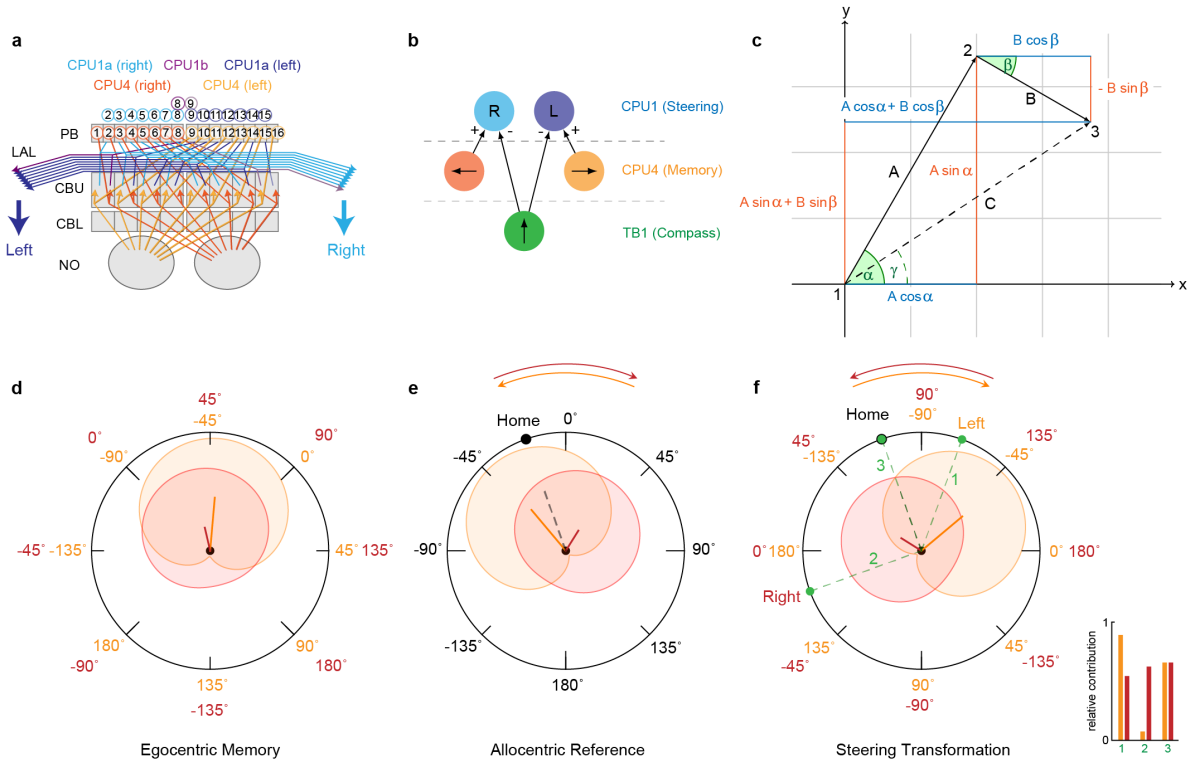
nm) block-face electron-microscopical image-stack of the bumblebee noduli. **f,** 3D tracing of all readily discernible large fibres innervating the right nodulus (based on image stack in **e**). Colours correspond to confocal data in (**a**). Fibre trajectories match the confocal data and allow identification of reconstructed fibres as likely CPU4 cells. Inset: Comprehensive reconstruction of all easily identifiable fibres in bundle six (19 identified cells). **g,** Single section from medium-resolution (voxel size: 23x23x50 nm) data stack from the bumblebee. **h,** Top: 3D tracing of three to five representative CPU4 neurons from seven bundles as well as the two TN neurons have strongly overlapping projection fields (based on image stack in **g**)**.** Bottom: Selective display of reconstructions of CPU4 neurons from six single bundles (same as in **h**) shows that all neurons belonging to the same bundle innervate specific subdomains within one nodulus and thus are well-suited to form local microcircuits that could underlie memory formation. **i,** Surface reconstruction of one terminal branch of each TN neuron (grey; TNx and TNy, as identity cannot be established as TN1 or TN2 based on noduli fibres alone). Clearly identifiable active zones are highlighted in magenta. Three columnar neurons postsynaptic to TN-cells (colours according to bundle identity) are reconstructed as well and were traced to their bundle of origin. Insets: Reconstruction of contact points of CPU4 and TN cells (arrowhead) shown in (**j,k**). **j,k,** Single sections from high-resolution image stack (voxel-size: 11.5x11.5x50 nm) revealing synaptic vesicles associated with active zones (arrowheads) within TN-cells directly opposite of CPU4 cells (coloured as in (**i**)). Scale bars: **a/e,** 50µm; **b-d,f-h,** 20µm; **e,** 50µm; **i,** 2µm; **i(inset),j,k,** 1µm.



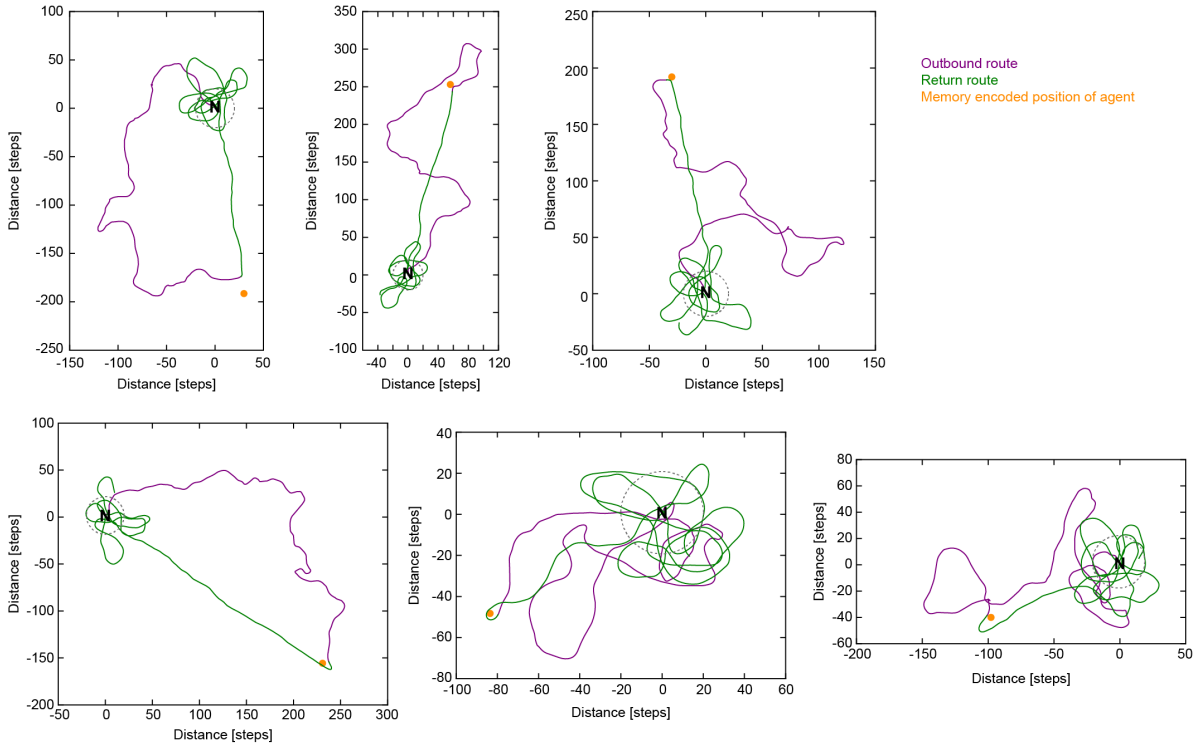
Extended Data Figure 5 | Anatomy and Physiology of *Megalopta* CPU1 neurons. a, 3D reconstruction of CPU1 neuron shown with 3D surface reconstruction of surrounding neuropils. **b,** Average response of the cell in (**a**) to a rotating polarizer (2 clockwise (CW) and 2 counter clockwise (CCW) rotations). Black circle: background activity; Purple line: mean tuning angle. **c,** Tuning width of the same cell (as in Fig. 1f). **d,** Raw spike train of the same cell. Top trace: gliding average mean activity; bottom trace: spike train. Consistent with data from locusts, monarch butterflies, and dung beetles these responses are much noisier compared to input-stage compass neurons (Fig.1) , which is, in fact, predicted by our model for our experimental situation, in which the state of the second main input, the CPU4 cells, is not experimentally controlled. **e,** Connectivity scheme of pontine neurons required to balance memory between right and left brain hemispheres. **f,** Two examples of pontine cells from *Megalopta*, shown as 3D reconstructions together with the structures of the central complex (grey). Abbreviations: PB, protocerebral bridge; CBU, upper division of the central body; CBL, lower division of the central body; NO, noduli; LAL, lateral accessory lobe; LBU lateral bulb; MBU, medial bulb. Scale bars: 100 µm.



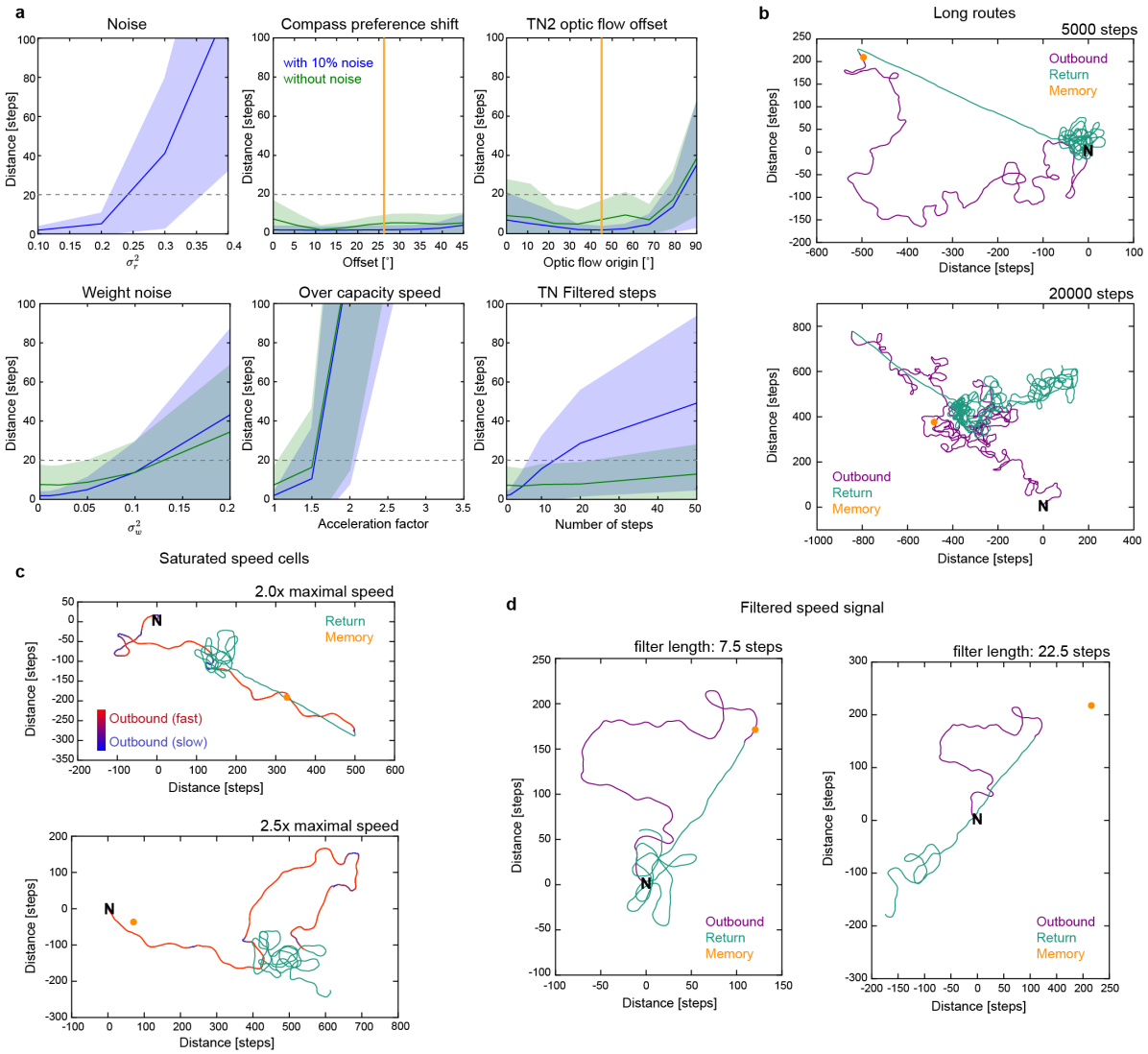
Extended Data Figure 6 | Connectivity and model parameters of the components of the path integration circuit. a, Top: equation used to compute output from summed input of neurons. Parameters a and b were tuned to generate different neuron types. Resulting shapes of sigmoids for each neuron type without added noise (top row) and with 10% added noise (bottom row); parameters a and b are shown. Pontine cells and CPU4 cells have identical sigmoid parameters. **b,** Connectivity matrices between all cell types used in the circuit. Cells were either fully inhibitory or fully excitatory (synaptic strengths were not tuned), except mutually inhibitory interconnections between TB1 neurons, which formed a sinusoidal weight matrix crucial for the resulting ring-attractor network.



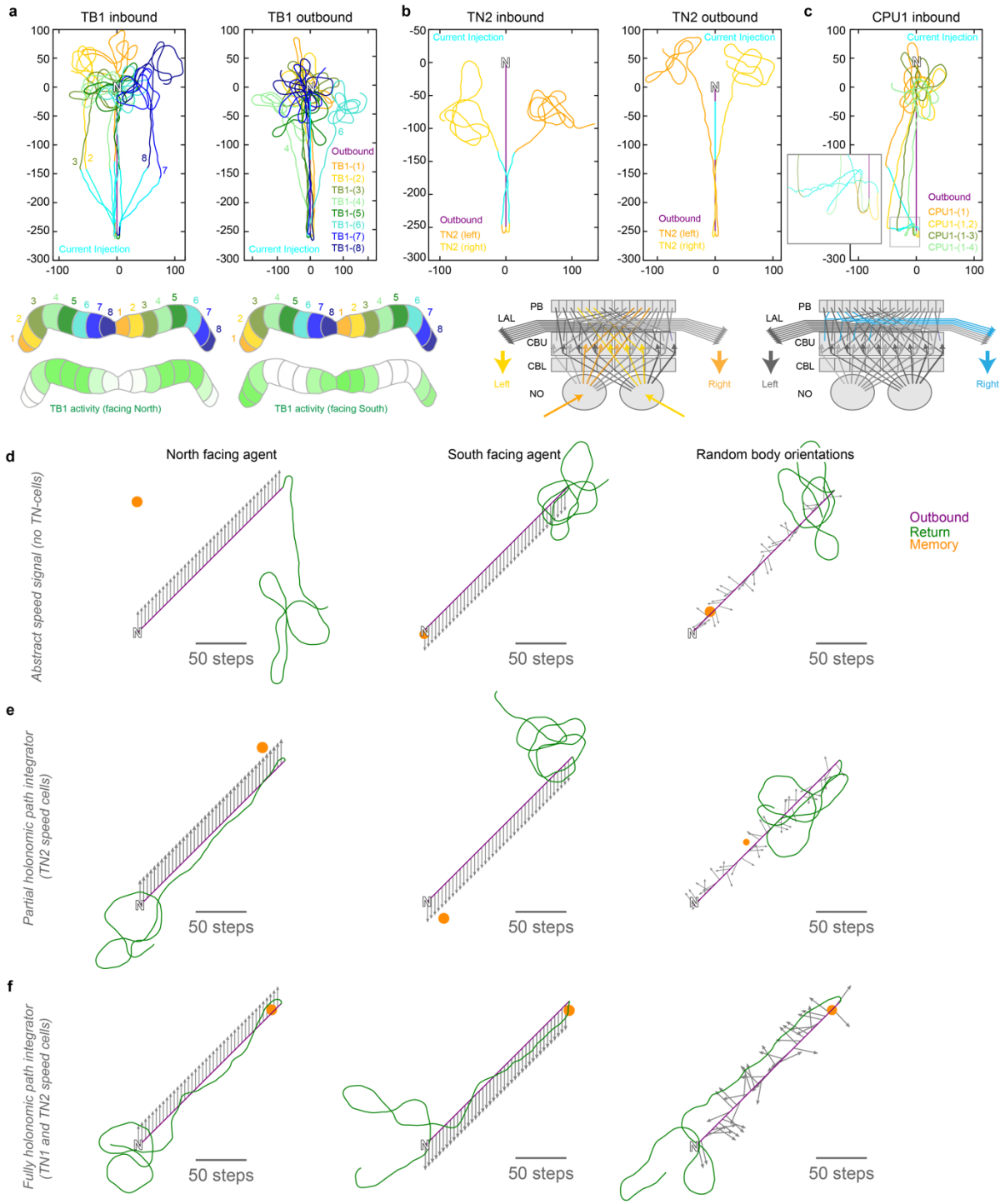
Extended Data Figure 7 | Mathematical explanation of memory and steering components of model. a, Schematic connectivity between CPU4 (memory) and CPU1 (steering) neurons. Memory is charged relative to TN excitation proportionate to translatory motion along their preferred angle of expansion for optic flow, so each CPU4 column is effectively storing motion at a ±45° offset from their TB1 (compass) input preference angle. An additional columnar offset between CPU1 and CPU4 means that each steering cell receives input from TB1 and CPU4 at ±90° offsets. **b,** A segment of the circuit (Fig. 4h) illustrates how it gives rise to homing. When facing downwards the TB1 neuron (green) is minimally excited. The CPU4 columns representing home vector to the left and the right of the downwards direction (orange) are effectively compared. The memory that evokes the highest output will drive steering in that direction by exciting CPU1 cells (blue). **c,** Motion can be described using polar coordinates, which in turn can be expressed as a sinusoid. Successive motion from points 1 to 2 and 2 to 3 could be represented by two sinusoids with phase α and β and amplitude A and B. Summing them results in another sinusoid representing the motion from 1 to 3 with phase γ and amplitude C. **d-f,** Homing can be described by comparison of two populations of CPU4, forming a compound memory to store a sinusoidal home vector. **d,** When aligning the activity bumps by actual CPU4 columns, each population stores motion at a ±45° offset. **e,** For visualization purposes, when rotated the sinusoids sum to give the home vector (dashed grey line). **f,** The connectivity columnar offset to CPU1 shifts by an additional ±45°, allowing populations to be compared to determine left or right turning. Facing towards direction 1 (green line) the agent turns left as activity from the light orange population dominates. When facing towards direction 2 the agent turns right. Finally, when facing to direction 3 the agent will keep going straight as both population activities balance. If facing directly away from home, the noise in the system breaks the symmetry, causing the agent to turn in either direction. Inset: Activation ratio of right and left memory for the three directions.



Extended Data Figure 8 | Example routes. Random outbound routes of 1500 step lengths. The dashed circle indicates the 20 step home-range around the nest used to illustrate successful homing.



Extended Data Figure 9 | The limits of the proposed path integration circuit. a, Responses to disturbances reveal the robustness of the circuit. Plotted is the closest distance to home during homing after a 1500 step random outbound route against different types of disturbances (mean ± SD). Dashed grey line: 20 step home range used to illustrate successful homing. Top left: stochastic noise in the sigmoids of each neuron. The 0.1 level was also used for all remaining trials (blue curves; green curves are without sigmoid noise). Top middle: Effect of shifting clockwise and counter-clockwise preferences angles of compass signal. Orange line: Value found for *Megalopta* compass neurons (Fig.1). Top right: Effect of deviation of translational optic-flow preference axis from the bee's body axis in TN2 neurons. Orange line: Value obtained in *Megalopta* TN-neurons (Fig. 3). Bottom left: Effect of randomly varying synaptic weights. Bottom middle: Effect of actual speed exceeding capacity of speed neuron (TN2) coding range. Bottom right: Effect of low-pass filtering of the speed signal (as found for TN1 neurons). **b,** Effect of very long routes on homing ability. After memory saturates beyond 5000 steps outbound routes, the agent homes into the correct direction, but undershoots. **c,** A similar effect is observed when speed neurons saturate during outbound route. **d,** Example of how a smoothened speed signal affects the encoded position memory.



Extended Data Figure 10 | Predictions of the model for physiological and behavioural experiments. a-c, Results of virtual current injections into components of the path integration circuit predict specific effects on homing routes in real bees. Injections were either performed during outbound or inbound trips (highlighted as cyan traces). **a,** Injections into TB1 neurons during distinct directional movements cause directional shift in homing vector. Which injections are most effective depends on the location of the injected neuron relative to the peak of the compass activity bump in the PB (bottom schematic traces). As the circuit is not actively steering during the outbound trip, the injections generate false memories that manifest during homing. **b,** Injections into the speed neurons lead to partial unbalancing of right and left memory and to directional deviations during homing. Effects are opposite for southwards and northwards facing agent. Schematic below traces illustrates, in principle, which neurons are affected. **c,** Effects of directly injecting the steering neurons (single or several CPU1 neurons) during the homing phase, when the circuit is actively steering. The agent deviates from its homing direction in a predictable direction by an angle correlated to the number of injected neurons. The inset highlights characteristic spiralling, increasing upon injection of more than one CPU1 cell. **d-f,** Results of

outbound routes with specific, optic-flow induced mismatches of body angle (grey arrows) and flight direction for a model without speed neurons, i.e. containing an abstract speed signal (**d**), the partly holonomic model used for the results in this paper (**e**), and a fully holonomic model, which incorporates all identified speed neurons, but requires several additional assumptions (**f**). Right panels: optic flow direction changes randomly every 25 steps and leads to different predictions for each model.

5.3 CONCLUSION

The computational model of the CX presented in this chapter is a valuable contribution to the field of animal navigation. It is the first published model that implements a functioning path integrator based on identified neural connections at the level of individual cells. It also further develops the mathematical perspective on path integration outlined by Wittmann and Schwegler (1995), where distance is captured by summing sinusoid arrays. This representation combines several desirable traits of Cartesian coordinates such as vector additivity, but with a more direct mapping to behaviours associated with polar coordinates, such as waggle dances (Von Frisch, 1967). By having two speed inputs, this system is able to function as a holonomic path integrator, a crucial factor in explaining how to fully integrate motion by flying insects. Finally, in this chapter, a small but natural extension to the model was described by the addition of external activity to the memory circuit. This would allow the CX to both potentially retrieve memories associated with certain locations and combine ego- and geocentric cues, both unsolved problems of insect navigation (Wehner et al., 1996). The use of vector retrieval through landmarks and the ability to fly on shortcuts between two known feeding sites is shown previously by Menzel and Greggers (2015).

The link between work in this chapter on visual path integration and the previous two chapters on visual place recognition is likely to become stronger in the near future. In robotics applications, it is clear that place recognition and dead reckoning need to be combined to achieve the best location estimates (Sec. 1.1.2). Although insects are not performing simultaneous localisation and mapping (SLAM), it is likely that path integration and navigational behaviour based on place recognition interact with each other at some stage. A good candidate for this is the CX, where the circuitry facilitates the combination of both ego- and geocentric motion cues.

5.3.1 *Future work*

Immediate follow-up work should focus on whether CPU4 columns actually implement a form of distributed memory as predicted by the model. Suggested electrophysiological and behavioural experiments are described in the attached paper, so this section will focus on computational modelling aspects of this study.

The level of abstraction for the CX model was chosen to learn about the identified circuitry in an intuitive manner. When exploring this circuit, it was important to easily change properties of the network and run a large number of experiments. For this reason, assumptions about neural dynamics were intentionally kept to a minimum and averaging firing rate neurons were determined to be the most suitable. However,

to explore plausible memory implementations, the correct approach would be to use more realistic neurons, providing a more stringent test on the capacity and stability of the proposed network.

The next clear step that needs to be taken is to add real units in time and space. This would allow tests to be done that more closely approximate an actual route, allowing replication of behavioural experiments while testing memory capacity with real firing rates. After units are added to the simulation and model, several approaches could be taken to implement a form of memory, acting as a proof of concept. We know from the block-face electron microscopy (EM) described above that each column has approximately twenty CPU4 type neurons, meaning that there is potential for an activity-based memory, implemented through recurrent connections. Under the assumption that memory is purely created by recurrent activity, it may be worth implementing it using software where populations are dealt with in an abstract manner and encoding is taken care of automatically, e.g. Nengo (Bekolay et al., 2013). An alternative approach would be to manually implement a feedback loop using some form of spiking neurons, like those proposed by Izhikevich et al. (2003). An advantage of this approach is that neural dynamics of cells already observed electrophysiological experiments, e.g. Heinze and Homberg (2009); Heinze et al. (2009), could be incorporated to make the model match the CX even more closely.

6

GENERAL DISCUSSION

The motivation for this thesis was to learn more about the abilities of some of the best navigators on the planet: insects. In the previous three chapters, several approaches were taken to learn more about the insect brain and, in particular, how insects use aspects of their visual input to localise. A number of these solutions were also successfully mimicked and applied to robotics problems, beating state of the art solutions for visual localisation on fast-moving robots across uneven terrain.

6.1 KEY CONTRIBUTIONS

CHAPTER 3: By photographing a wide range of environments with an ultraviolet (UV) camera, it was shown that UV sensors produce images that are particularly suited to sky and ground separation. The plausibility of using the skyline as a feature for localisation was demonstrated by applying it to a well-known robotics localisation algorithm, SeqSLAM (Milford and Wyeth, 2012). Binary sky and ground images were shown to perform comparably to greyscale imagery in a localisation task along a recorded urban route. In addition, experiments showed that extracting the skyline by using UV is more robust to changes in lighting.

CHAPTER 4: Inspired by insects' need to navigate on uneven terrain, it was shown that sky, when represented in the frequency domain, provides a feature for localisation that is invariant to tilt and yaw rotation. This addresses a common problem in visual localisation, where it is hard to develop representations that are both pose and condition invariant. It also proved particularly useful on fast, lightweight robots that are unable to extract point features due to motion blur and computational constraints, yet also unable to use whole images due to misalignment. A demonstration of localisation using these features was carried out on a rapidly manoeuvring robot. From this test, it became clear that segmented sky can remain a strong cue for localisation in highly foliated areas, providing evidence that ants and other insects could potentially also be using sky features to localise in these areas. Finally, a follow-up pilot study demonstrated the ability of an agent to visually home in a simulated ant environment, using only the amplitude spectrum of a panoramic view.

CHAPTER 5: A novel model of path integration in the insect brain was developed, inspired by the discovery of speed sensitive TN neurons arborising in the central complex (CX) of the bee. The model was based on identified individual cellular connections ascertained from overlapping dendritic arborisation and activity corresponded well to observed electrophysiological data. New neural roles were suggested based on this model, including CPU4 columns acting as a compound memory, consisting of two subpopulations. The model is the first of its kind that is able to path integrate, despite holonomic motion, due to charging of these memory subpopulations by the TN cells of separate hemispheres. Pontin cells, which previously had no known role, were crucial to normalise the memory populations, allowing the model to function in a theoretically favourable manner by summing sinusoids.

6.2 DISCUSSION AND FUTURE WORK

The bodies of work described in the previous chapters link together in a number of ways. Both the ant and bee studies used a biorobotics approach to learn about the insect brain, where a proposed algorithm was implemented and behaviour of either an agent or robot was observed to test algorithmic plausibility. The biological subjects were also similar as both ants and bees are social insects with brains that bear strong structural resemblances (Wolff and Strausfeld, 2015). They also both display similar navigational behaviour, such as path integration, visual homing and systematic search (Collett, 1996). It is plausible that the neural basis for place recognition and path integration is at least partially conserved between the two species. For this reason, it is conceivable that modellers may eventually integrate the two described systems at some point, to better understand how visual place recognition and path integration systems interact inside the insect brain. A clue may lie in the proposal made in section 5.3 regarding the loading of pre-stored patterns into the CPU4 memory, allowing recognised views to be associated with vectors. Another plausible interaction could occur through differing behavioural choices made by place recognition and path integration. If both cues come into conflict, a plausible brain area for this to be resolved could be the CX, possibly by adding additional egocentric activity to the memory circuit to perform a form of optimal cue integration (Wystrach et al., 2015) before the steering phase.

The studies in part i both completed the biorobotics loop described in section 1.1.1, in the sense that both provided insight into insect navigation while also resulting in a bioinspired system that proved to be useful in the field of robotics. Although the immediate robotic use cases for the path integration system developed in part ii are

less apparent, a follow-up Masters project has already implemented the network on a wheeled robot. It may be foreseeable that lightweight unmanned aerial vehicles (UAVs) would benefit from a similar approach, especially due to their holonomic movement, however even the smallest robots can now plausibly run algorithms like an extended Kalman filter (EKF) on board. Despite this, an aerial robot implementation would be a useful follow-up step, allowing a physical proof of concept to be developed to test the robustness of this integrator and visual odometer in an environment that is not visually uniform and has to deal with more realistic challenges, e.g. z-axis motion.

6.2.1 *General insect navigation abilities*

It is perhaps questionable whether all knowledge gained from navigational studies between different insects is transferable, as many insects have evolved to deal with environment-specific problems. Can we learn things about the navigational toolkit of one insect and apply it to another? The structural consistency of the CX across insects and its clear role in many navigational tasks could provide a clue. Chapter 5 suggests that the CX could be more broadly seen as a way to convert between ego- and allocentric coordinate systems and also determine how to turn from a current to a desired heading. Although having a CX does not necessarily mean an insect can path integrate, it perhaps does suggest that tools necessary for various navigational strategies, e.g. egocentric and allocentric coordinate systems with the ability to integrate desired headings in both coordinate systems in a seamless manner, are broadly available. Future work should focus on exploring the differences and similarities of the CX across different species of insects and, in particular, try to determine how these anatomic differences translate to different functional abilities. This could provide insight about shared navigational strategies and to what extent findings from studies across species can be shared.

6.2.2 *Deep learning*

Chapter 4 briefly mentions the increased use of learnt features through convolutional neural networks (CNNs). If this becomes the dominant way to solve computer vision problems, can it still be used to help our understanding of insects? Neural networks are often compared to brains, however the way they learn, through gradient descent (Rumelhart et al., 1988), may actually more closely resemble the evolution of genes than brains. After being exposed to a set of similar problems over many generations the network eventually converges on a solution. If this analogy holds, then perhaps training networks on problems that insects are exposed to could provide

some indication of the different ways these problems are solvable. Insects normally live for a relatively short period of time, meaning many behaviours that deal with problems they encounter in their environment are probably hard-wired rather than learnt. This makes insects particularly suitable as a subject to look for similar motifs in the networks of their brains compared to converged weights in artificial neural networks (ANNs) that perform similar tasks.

6.3 CLOSING REMARKS

Like many neuroscientists, I began my studies inspired by the unbelievable abilities of the brain and wanting to understand how it works. After becoming only vaguely familiar with the field, it soon became apparent that learning how the brain works is actually an enormously complicated problem. The only way to attack this puzzle is by having individuals focus on a small subproblem and slowly working on their part piece by piece. A number of researchers seem to have concluded that the easiest way to make progress was to simplify the problem by scaling down to smaller brains. Some study rats, fish or even simpler organisms like larvae, with a trade off between interestingness through richness of behaviour and structural simplicity. There is a good argument to be made for the study of bees, ants and other insects for precisely these reasons. Insects have tiny brains, yet still manage to perform complex tasks, and nothing makes this more apparent than their impressive feats of navigation. It is incredibly humbling to realise that a small ant can perform similar navigational tasks to a car covered in expensive sensors, with a trunk full of computers and using a global positioning system (GPS). Insects are also an attractive subject because they have short lifespans, making them ethically more accessible than mammals. They can be forced to carry out strenuous and monotonous tasks, e.g. performing navigational tasks with large parts of their bodies removed to enable recordings. Another big advantage of their short life is that many lines of mutants can easily be bred, allowing individual neurons to be genetically targeted, meaning they can be turned off or even made to fluoresce when firing. During the time spent on this thesis, neuroscientists have been able to tinker with and peer inside the brain at an increasingly high resolution, even during behaviour. A remarkable result of this is that insect navigation is no longer a black box problem and is beginning to seem solvable. This transition has made working in the field of neuroscience especially exciting in recent years, with a strong sense of progress felt within the scientific community. I hope that the chapters in this thesis will contribute in a small way to solving a tiny part of the puzzle and look forward to the exciting developments in both insect and robotic navigation that will be made over the next few years.

7

APPENDIX

The next two pages include conference posters that were used to present some of the work described in this thesis.

7.1 POSTER FROM CENTRAL COMPLEX IV: A NEW HOPE TO UNDERSTAND A MULTIFACETED BRAIN REGION, JANELIA FARM, 2016[1]

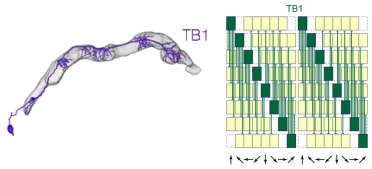
A plausible neural substrate for path integration in the bee brain

Thomas Stone, Nicolai Ben Wedding, Anna Honkanen, Andrea Adden, Rachel, William Wcislo, Eric Warrant, Barbara Webb and Stanley Heinze

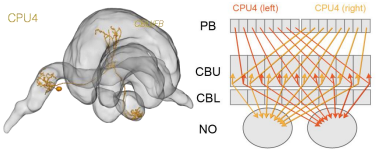
Abstract

Many insects are able to return directly home after convoluted excursions. They are assumed to maintain a home vector through integration of velocity, i.e., to accumulate the distance travelled in each direction. The polarisation compass in protocerebral bridge of the central complex could provide the directional information, and recently discovered optic flow responsive cells in the noduli could provide speed information. Anatomical analysis suggests both inputs converge on CPU4 cells, and that these in turn connect to CPU1 cells which provide output to the lateral accessory lobes. We show that the apparently complex mapping between these layers corresponds in principle to a previously hypothesised neural circuit for path integration, with CPU4 cells forming a distributed memory of the outbound path, and CPU1 cells using a comparison of current and stored directions to produce a steering signal . We have implemented a computational model of the circuit based on the mapped anatomy and show it can reliably return an agent to its starting location.

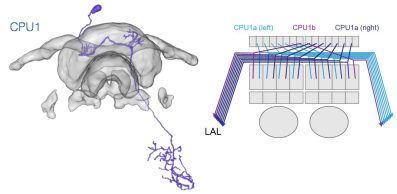
Layer 1: TB1 neurons form a ring attractor for compass direction



Layer 2: CPU4 neurons integrate the speed signal from noduli, modulated by TB1 activity

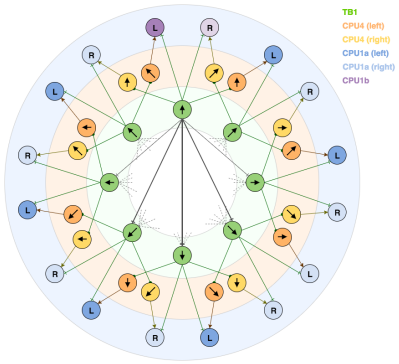


Layer 3: CPU1 neurons are activated by TB1 and CPU4 neurons with different preferred directions

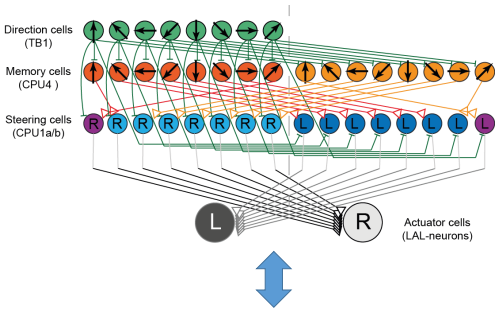


Layer 4: Left and right CPU1 neuron outputs are summed, and produce turning to the stronger side

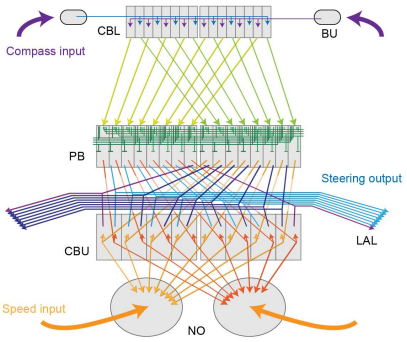
The circuit unravelled using force graph



Logic of circuit mapped

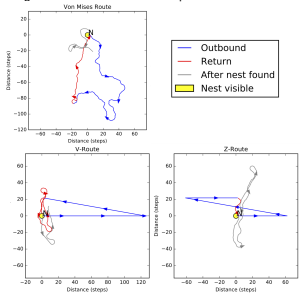


Connections mapped by dendritic arborisation

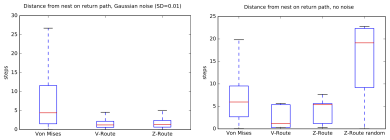


Result: the circuit steers

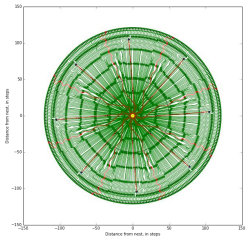
We implemented this circuit and used it to control a simulated agent. We test it with different outward routes (here for 250 steps), with varying directions and speeds. Homing is controlled by 'switching on' the steering cells and allowing the agent to run for another 250 steps.



For repeated tests (N=200) we calculate the nearest approach to the starting point. Homing is reliable, with and without noise.



Using the same outward route in multiple directions we see consistent bias in homing (without noise) when two memory cells are equally charged



Conclusion

A model based on identified neurons in the CX and their hypothesised connectivity (based on dendritic arborisation) is sufficent to produce homing by path integration in an artificial agent. We note that the TB1-CPU4-CPU1 circuit provides a general mechanism for steering through comparison of a stored and a current direction.

Future work

- Currently the model uses a simplified firing rate representation of neural activity. It is not very robust to noise so may be difficult to generalise to a more realistic representation.
- We assume linear and permanent accumulation of activity in the memory cell layer. We need to explore whether there is biophysically plausible mechanism by which this could occur.
- The outward distance limit vs. return accuracy depends on assumptions about the accumulator and how it is transformed into CPU1 activation.

Vetenskapsrådet
MARIE CURIE ACTIONS
BBSRC
EPSRC
Engineering and Physical Sciences Research Council

1 won best poster prize.

Insect Navigation:
From retinotopy to rotation invariant views


Tom Stone


Paul Ardin


Mike Mangan


Antoine Wystrach


Barbara Webb

Summary

- Many aspects of visual navigation in ants can be explained through retinotopic matching of the current scene with a visual memory. For example, homing can be performed by moving up the gradient of 'familiarity' produced by simple pixel-wise comparison after the images are aligned.
- Recent behavioural data (panel B) suggests ants' homing ability is relatively independent of their viewing direction.
- We demonstrate that if ants encoded visual memories in the frequency domain (panel A) they could successfully home without retinotopic alignment of the view (panel C).

A) From Images to Frequency Histograms

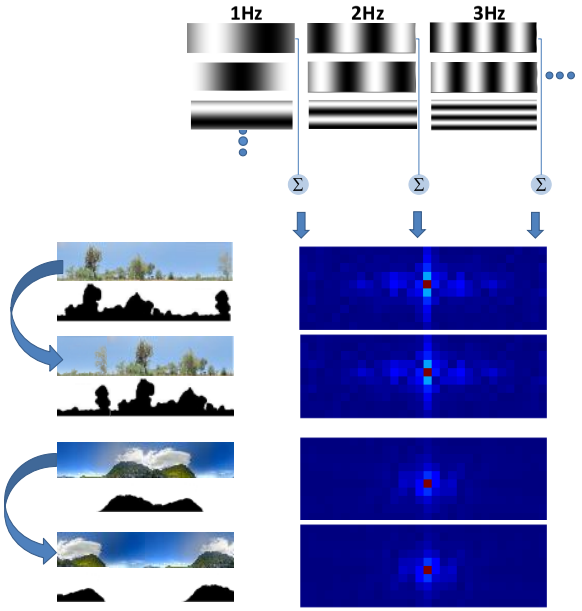


Fig 2. Frequency histograms provide a spatially varying but rotationally invariant encoding of panoramic views. *Images can be decomposed into their frequency components. By summing the total contribution of each frequency, and neglecting the phase relationship, views can be encoded (and compared) in a rationally invariant way.*

B) Ants can home backwards

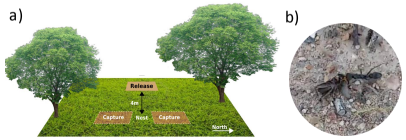
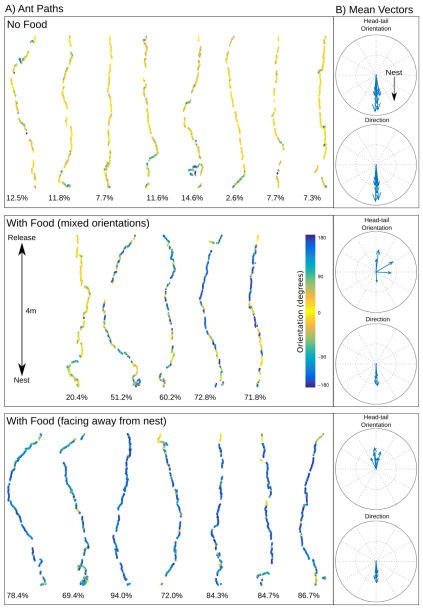


Fig 2. Experimental paradigm to test the ability of ants to home backward. *(a) Ardin et al. (in press) released ants (Myrmecia croslandi) captured 1m north or south of their nest from a location 4m to the west. Ants either travelled with no food, or were carrying or dragging a large prey item (b). Video analysis was used to determine the position and head-tail orientation of each ant along its path, from release to reaching the nest.*

Fig. 3 (below) *All ants returned directly to the nest, despite the fact that ants dragging food spent a substantial portion of their path facing backward. This suggests that homing information used may be independent of viewing direction and retinotopic alignment..*



C) Using Frequency Coding for Visual Homing

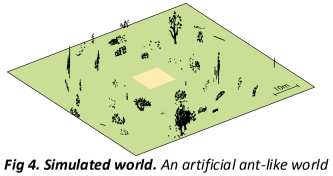


Fig 4. Simulated world. *An artificial ant-like world was created in which 360 degree panoramic views could be generated every 2cm in the test area (central square) to test homing models.*

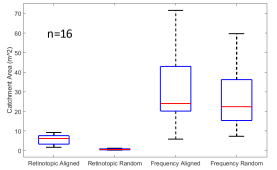


Fig 5. Catchment areas. *Frequency matching performs well, whereas retinotopic matching fails completely for random orientations.*

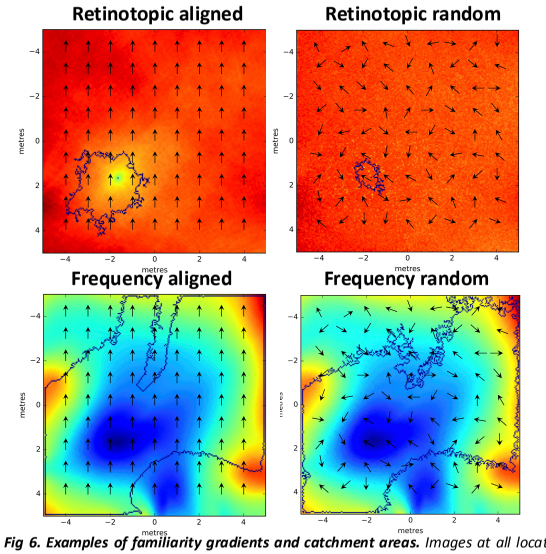


Fig 6. Examples of familiarity gradients and catchment areas. *Images at all locations are compared to a reference image, using retinopic matching (top) or frequency based matching (bottom), and with image orientation aligned (left) or randomised (right). Gradient descent was used to compute the catchment area (the distance from which an agent could home) and used as performance metric to compare the methods (fig 5)..*

Acknowledgments

Thanks to Jochen Zeil for hosting Paul Ardin at ANU for the experimental study. This project was funded by the EPSRC DTC in Neuroinformatics and BBSRC grant 'Invisible Cues'.

BIBLIOGRAPHY

Arandjelović, R., Gronat, P., Torii, A., Pajdla, T., and Sivic, J. (2015). NetVLAD: CNN architecture for weakly supervised place recognition. *arXiv preprint arXiv:1511.07247*.

Ardin, P., Mangan, M., Wystrach, A., and Webb, B. (2015). How variation in head pitch could affect image matching algorithms for ant navigation. *Journal of Comparative Physiology A*, 201(6):585–597.

Ardin, P., Peng, F., Mangan, M., Lagogiannis, K., and Webb, B. (2016a). Using an insect mushroom body circuit to encode route memory in complex natural environments. *PLoS Comput Biol*, 12(2):e1004683.

Ardin, P. B., Mangan, M., and Webb, B. (2016b). Ant homing ability is not diminished when traveling backwards. *Frontiers in behavioral neuroscience*, 10:69.

Arena, P., Patané, L., and Termini, P. S. (2010). An insect brain computational model inspired by drosophila melanogaster: Simulation results. In *The 2010 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE.

Baddeley, B., Graham, P., Husbands, P., and Philippides, A. (2012). A model of ant route navigation driven by scene familiarity. *PLoS Comput Biol*, 8(1):e1002336.

Bay, H., Ess, A., Tuytelaars, T., and Van Gool, L. (2008). Speeded-up robust features (SURF). *Computer vision and image understanding*, 110(3):346–359.

Bekolay, T., Bergstra, J., Hunsberger, E., DeWolf, T., Stewart, T. C., Rasmussen, D., Choo, X., Voelker, A. R., and Eliasmith, C. (2013). Nengo: a python tool for building large-scale functional brain models. *Frontiers in neuroinformatics*, 7:48.

Ben-Yishai, R., Bar-Or, R. L., and Sompolinsky, H. (1995). Theory of orientation tuning in visual cortex. *Proceedings of the National Academy of Sciences*, 92(9):3844–3848.

Bender, J. A., Pollack, A. J., and Ritzmann, R. E. (2010). Neural activity in the central complex of the insect brain is linked to locomotor changes. *Current biology*, 20(10):921–926.

Benhamou, S., Sauvé, J.-P., and Bovet, P. (1990). Spatial memory in large scale movements: efficiency and limitation of the egocentric coding process. *Journal of Theoretical Biology*, 145(1):1–12.

Berry, R. P., Stange, G., and Warrant, E. J. (2007). Form vision in the insect dorsal ocelli: an anatomical and optical analysis of the dragonfly median ocellus. *Vision research*, 47(10):1394–1409.

Biswas, J. and Veloso, M. (2010). WiFi localization and navigation for autonomous indoor mobile robots. In *Robotics and Automation (ICRA), 2010 IEEE International Conference on*, pages 4379–4384. IEEE.

Bosse, M. and Zlot, R. (2009). Keypoint design and evaluation for place recognition in 2D lidar maps. *Robotics and Autonomous Systems*, 57(12):1211–1224.

Bradski, G. et al. (2000). The opencv library. *Doctor Dobbs Journal*, 25(11):120–126.

Briscoe, A. D. and Chittka, L. (2001). The evolution of color vision in insects. *Annual review of entomology*, 46(1):471–510.

Bullmore, E. and Sporns, O. (2009). Complex brain networks: graph theoretical analysis of structural and functional systems. *Nature Reviews Neuroscience*, 10(3):186–198.

Burak, Y. and Fiete, I. R. (2009). Accurate path integration in continuous attractor network models of grid cells. *PLoS Comput Biol*, 5(2):e1000291.

Byrne, M., Dacke, M., Nordström, P., Scholtz, C., and Warrant, E. (2003). Visual cues used by ball-rolling dung beetles for orientation. *Journal of Comparative Physiology A*, 189(6):411–418.

Calbo, J., González, J.-A., et al. (2005). Empirical studies of cloud effects on UV radiation: A review. *Reviews of Geophysics*, 43(2).

Cartwright, B. and Collett, T. S. (1983). Landmark learning in bees. *Journal of comparative physiology*, 151(4):521–543.

Cerdá, X., Retana, J., and Cros, S. (1998). Critical thermal limits in mediterranean ant species: trade-off between mortality risk and foraging performance. *Functional Ecology*, 12(1):45–55.

Chittka, L., Shmida, A., Troje, N., and Menzel, R. (1994). Ultraviolet as a component of flower reflections, and the colour perception of Hymenoptera. *Vision research*, 34(11):1489–1508.

Chli, M. (2010). *Applying information theory to efficient SLAM*. Imperial College, London.

Collett, M. and Collett, T. S. (2000a). How do insects use path integration for their navigation? *Biological cybernetics*, 83(3):245–259.

Collett, M. and Collett, T. S. (2006). Insect navigation: no map at the end of the trail? *Current biology*, 16(2):R48–R51.

Collett, T. (1996). Insect navigation en route to the goal: multiple strategies for the use of landmarks. *Journal of Experimental Biology*, 199(1):227–235.

Collett, T. S. and Collett, M. (2000b). Path integration in insects. *Current opinion in neurobiology*, 10(6):757–762.

Coulson, K. L. (1988). *Polarization and Intensity of Light in the Atmosphere*. A Deepak Pub.

Cruse, H. and Wehner, R. (2011). No need for a cognitive map: decentralized memory for insect navigation. *PLoS Comput Biol*, 7(3):e1002009.

Cummins, M. and Newman, P. (2008). FAB-MAP: Probabilistic localization and mapping in the space of appearance. *The International Journal of Robotics Research*, 27(6):647–665.

Dacke, M., Baird, E., Byrne, M., Scholtz, C. H., and Warrant, E. J. (2013). Dung beetles use the Milky Way for orientation. *Current Biology*, 23(4):298–300.

Dacke, M., el Jundi, B., Smolka, J., Byrne, M., and Baird, E. (2014). The role of the sun in the celestial compass of dung beetles. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 369(1636):20130036.

Dacke, M., Nilsson, D.-E., Scholtz, C. H., Byrne, M., and Warrant, E. J. (2003). Animal behaviour: insect orientation to polarized moonlight. *Nature*, 424(6944):33–33.

Darwin, C. (1873). Origin of certain instincts. *Nature*, 7:417–418.

Davison, A. J., Reid, I. D., Molton, N. D., and Stasse, O. (2007). MonoSLAM: Real-time single camera SLAM. *IEEE transactions on pattern analysis and machine intelligence*, 29(6):1052–1067.

Dayan, P. and Abbott, L. F. (2001). *Theoretical neuroscience*, volume 10. Cambridge, MA: MIT Press.

Differt, D. and Möller, R. (2015). Insect models of illumination-invariant skyline extraction from UV and green channels. *Journal of theoretical biology*, 380:444–462.

Differt, D. and Möller, R. (2016). Spectral skyline separation: Extended landmark databases and panoramic imaging. *Sensors*, 16(10):1614.

Dorigo, M., Maniezzo, V., and Colorni, A. (1996). Ant system: optimization by a colony of cooperating agents. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 26(1):29–41.

Drumheller, M. (1987). Mobile robot localization using sonar. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 9(2):325–332.

Durrant-Whyte, H. and Bailey, T. (2006). Simultaneous localization and mapping: part I. *IEEE robotics & automation magazine*, 13(2):99–110.

El Jundi, B., Foster, J. J., Khaldy, L., Byrne, M. J., Dacke, M., and Baird, E. (2016). A snapshot-based mechanism for celestial orientation. *Current Biology*, 26(11):1456–1462.

el Jundi, B., Warrant, E. J., Byrne, M. J., Khaldy, L., Baird, E., Smolka, J., and Dacke, M. (2015). Neural coding underlying the cue preference for celestial orientation. *Proceedings of the National Academy of Sciences*, 112(36):11395–11400.

Etienne, A. S. and Jeffery, K. J. (2004). Path integration in mammals. *Hippocampus*, 14(2):180–192.

Fallon, M. F., Johannsson, H., and Leonard, J. J. (2012). Efficient scene simulation for robust Monte Carlo localization using an RGB-D camera. In *Robotics and Automation (ICRA), 2012 IEEE International Conference on*, pages 1663–1670. IEEE.

Franz, M. O. and Mallot, H. A. (2000). Biomimetic robot navigation. *Robotics and autonomous Systems*, 30(1):133–153.

Franz, M. O., Schölkopf, B., Mallot, H. A., and Bülthoff, H. H. (1998). Where did I take that snapshot? scene-based homing by image matching. *Biological Cybernetics*, 79(3):191–202.

Fukushi, T. (2001). Homing in wood ants, Formica japonica: use of the skyline panorama. *Journal of Experimental Biology*, 204(12):2063–2072.

Fuller, S. B., Karpelson, M., Censi, A., Ma, K. Y., and Wood, R. J. (2014). Controlling free flight of a robotic fly using an onboard vision sensor inspired by insect ocelli. *Journal of The Royal Society Interface*, 11(97):20140281.

Furgale, P. and Barfoot, T. D. (2010). Visual teach and repeat for long-range rover autonomy. *Journal of Field Robotics*, 27(5):534–560.

Gomez-Ojeda, R., Lopez-Antequera, M., Petkov, N., and Gonzalez-Jimenez, J. (2015). Training a convolutional neural network for appearance-invariant place recognition. *arXiv preprint arXiv:1505.07428*.

Graham, P. and Cheng, K. (2009a). Ants use the panoramic skyline as a visual cue during navigation. *Current Biology*, 19(20):R935–R937.

Graham, P. and Cheng, K. (2009b). Which portion of the natural panorama is used for view-based navigation in the australian desert ant? *Journal of Comparative Physiology A*, 195(7):681–689.

Graham, P. and Collett, T. S. (2002). View-based navigation in insects: how wood ants (Formica rufa L.) look at and are guided by extended landmarks. *Journal of experimental Biology*, 205(16):2499–2509.

Greiner, B., Ribi, W. A., and Warrant, E. J. (2004). Retinal and optical adaptations for nocturnal vision in the halictid bee megalopta genalis. *Cell and tissue research*, 316(3):377–390.

Guizzo, E. (2011). How Google's self-driving car works. *IEEE Spectrum Online, October*, 18.

Gustafsson, F., Gunnarsson, F., Bergman, N., Forssell, U., Jansson, J., Karlsson, R., and Nordlund, P.-J. (2002). Particle filters for positioning, navigation, and tracking. *IEEE Transactions on signal processing*, 50(2):425–437.

Haberkern, H. and Jayaraman, V. (2016). Studying small brains to understand the building blocks of cognition. *Current opinion in neurobiology*, 37:59–65.

Haferlach, T., Wessnitzer, J., Mangan, M., and Webb, B. (2007). Evolving a neural model of insect path integration. *Adaptive Behavior*, 15(3):273–287.

Hafting, T., Fyhn, M., Molden, S., Moser, M.-B., and Moser, E. I. (2005). Microstructure of a spatial map in the entorhinal cortex. *Nature*, 436(7052):801–806.

Harley, C. and Ritzmann, R. E. (2010). Electrolytic lesions within central complex neuropils of the cockroach brain affect negotiation of barriers. *Journal of Experimental Biology*, 213(16):2851–2864.

Hartmann, G. and Wehner, R. (1995). The ant's path integration system: a neural architecture. *Biological Cybernetics*, 73(6):483–497.

Heinze, S. (2015). Insect brain database. http://www.insectbraindb.org. Accessed: 2016-12-13.

Heinze, S., Gotthardt, S., and Homberg, U. (2009). Transformation of polarized light information in the central complex of the locust. *The Journal of neuroscience*, 29(38):11783–11793.

Heinze, S. and Homberg, U. (2007). Maplike representation of celestial e-vector orientations in the brain of an insect. *Science*, 315(5814):995–997.

Heinze, S. and Homberg, U. (2008). Neuroarchitecture of the central complex of the desert locust: intrinsic and columnar neurons. *Journal of Comparative Neurology*, 511(4):454–478.

Heinze, S. and Homberg, U. (2009). Linking the input to the output: new sets of neurons complement the polarization vision network in the locust central complex. *The Journal of neuroscience*, 29(15):4911–4921.

Herz, A. V., Gollisch, T., Machens, C. K., and Jaeger, D. (2006). Modeling single-neuron dynamics and computations: a balance of detail and abstraction. *science*, 314(5796):80–85.

Homberg, U., Heinze, S., Pfeiffer, K., Kinoshita, M., and El Jundi, B. (2011). Central neural coding of sky polarization in insects. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 366(1565):680–687.

Homberg, U., Hofer, S., Pfeiffer, K., and Gebhardt, S. (2003). Organization and neural connections of the anterior optic tubercle in the brain of the locust, Schistocerca gregaria. *Journal of Comparative Neurology*, 462(4):415–430.

Hu, M.-K. (1962). Visual pattern recognition by moment invariants. *IRE transactions on information theory*, 8(2):179–187.

Hubel, D. H. and Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *The Journal of physiology*, 195(1):215–243.

Ioalè, P., Nozzolini, M., and Papi, F. (1990). Homing pigeons do extract directional information from olfactory stimuli. *Behavioral Ecology and Sociobiology*, 26(5):301–305.

Izhikevich, E. M. et al. (2003). Simple model of spiking neurons. *IEEE Transactions on neural networks*, 14(6):1569–1572.

Julier, S. J. and Uhlmann, J. K. (2004). Unscented filtering and nonlinear estimation. *Proceedings of the IEEE*, 92(3):401–422.

Kakaria, K. S. and de Bivort, B. L. (2016). Ring attractor dynamics emerge from a spiking model of the entire protocerebral bridge. *bioRxiv*, page 081240.

Kaneko, T. and Okudaira, M. (1985). Encoding of arbitrary curves based on the chain code representation. *IEEE Transactions on Communications*, 33(7):697–707.

Kazhdan, M., Funkhouser, T., and Rusinkiewicz, S. (2003). Rotation invariant spherical harmonic representation of 3D shape descriptors. In *Symposium on geometry processing*, volume 6, pages 156–164.

Khotanzad, A. and Hong, Y. H. (1990). Invariant image recognition by Zernike moments. *IEEE Transactions on pattern analysis and machine intelligence*, 12(5):489–497.

Kim, D. and Lee, J. (2011). Path integration mechanism with coarse coding of neurons. *Neural processing letters*, 34(3):277–291.

Kimchi, T., Etienne, A. S., and Terkel, J. (2004). A subterranean mammal uses the magnetic compass for path integration. *Proceedings of the National Academy of Sciences of the United States of America*, 101(4):1105–1109.

Kirchner, W. H. and Braun, U. (1994). Dancing honey bees indicate the location of food sources using path integration rather than cognitive maps. *Animal behaviour*, 48(6):1437–1441.

Kohler, M. and Wehner, R. (2005). Idiosyncratic route-based memories in desert ants, Melophorus bagoti: how do they interact with path-integration vectors? *Neurobiology of learning and memory*, 83(1):1–12.

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105.

Kühn-Bühlmann, S. and Wehner, R. (2006). Age-dependent and task-related volume changes in the mushroom bodies of visually guided desert ants, cataglyphis bicolor. *Journal of neurobiology*, 66(6):511–521.

Labhart, T. (1986). The electrophysiology of photoreceptors in different eye regions of the desert ant, Cataglyphis bicolor. *Journal of Comparative Physiology A: Neuroethology, Sensory, Neural, and Behavioral Physiology*, 158(1):1–7.

Labrosse, F. (2007). Short and long-range visual navigation using warped panoramic images. *Robotics and Autonomous Systems*, 55(9):675–684.

Labrosse, F. et al. (2006). The visual compass: Performance and limitations of an appearance-based method. *Journal of Field Robotics*, 23(10):913.

Lambrinos, D., Kobayashi, H., Pfeifer, R., Maris, M., Labhart, T., and Wehner, R. (1997). An autonomous agent navigating with a polarized light compass. *Adaptive behavior*, 6(1):131–161.

Lambrinos, D., Möller, R., Labhart, T., Pfeifer, R., and Wehner, R. (2000). A mobile robot employing insect strategies for navigation. *Robotics and Autonomous systems*, 30(1):39–64.

Latecki, L. J., Lakamper, R., and Eckhardt, T. (2000). Shape descriptors for non-rigid shapes with a single closed contour. In *Computer Vision and Pattern Recognition, 2000. Proceedings. IEEE Conference on*, volume 1, pages 424–429. IEEE.

LeCun, Y., Bottou, L., Bengio, Y., and Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324.

LeCun, Y., Huang, F. J., and Bottou, L. (2004). Learning methods for generic object recognition with invariance to pose and lighting. In *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, volume 2, pages 97–104. IEEE.

Lemaire, T., Berger, C., Jung, I.-K., and Lacroix, S. (2007). Vision-based SLAM: Stereo and monocular approaches. *International Journal of Computer Vision*, 74(3):343–364.

Li, B., Gallagher, T., Dempster, A. G., and Rizos, C. (2012). How feasible is the use of magnetic field alone for indoor positioning? In *Indoor Positioning and Indoor Navigation (IPIN), 2012 International Conference on*, pages 1–9. IEEE.

Linsker, R. (1988). Self-organization in a perceptual network. *Computer*, 21(3):105–117.

Loewenstein, Y. and Sompolinsky, H. (2003). Temporal integration by calcium dynamics in a model neuron. *Nature neuroscience*, 6(9):961–967.

Lowe, D. G. (1999). Object recognition from local scale-invariant features. In *Computer vision, 1999. The proceedings of the seventh IEEE international conference on*, volume 2, pages 1150–1157. IEEE.

Lowry, S., Sünderhauf, N., Newman, P., Leonard, J. J., Cox, D., Corke, P., and Milford, M. J. (2016). Visual place recognition: A survey. *IEEE Transactions on Robotics*, 32(1):1–19.

Mangan, M. and Webb, B. (2012). Spontaneous formation of multiple routes in individual desert ants (Cataglyphis velox). *Behavioral Ecology*, 23(5):944–954.

Markram, H. (2012). The human brain project. *Scientific American*, 306(6):50–55.

Maurer, R. and Séguinot, V. (1995). What is modelling for? a critical review of the models of path integration. *Journal of Theoretical Biology*, 175(4):457–475.

McManus, C., Upcroft, B., and Newmann, P. (2014). Scene signatures: Localised and point-less features for localisation. In *Proceedings of Robotics: Science and Systems*, Berkeley, USA.

McNaughton, B. L., Battaglia, F. P., Jensen, O., Moser, E. I., and Moser, M.-B. (2006). Path integration and the neural basis of the 'cognitive map'. *Nature Reviews Neuroscience*, 7(8):663–678.

Menzel, R. and Greggers, U. (2015). The memory structure of navigation in honeybees. *Journal of Comparative Physiology A*, 201(6):547–561.

Menzel, R., Ventura, D. F., Hertel, H., De Souza, J., and Greggers, U. (1986). Spectral sensitivity of photoreceptors in insect compound eyes: comparison of species and methods. *Journal of Comparative Physiology A*, 158(2):165–177.

Menzel, R. and Wehner, R. (1970). Augenstrukturen bei verschieden großen arbeiterinnen von Cataglyphis bicolor fabr. (Formicidae, Hymenoptera). *Zeitschrift für vergleichende Physiologie*, 68(4):446–449.

Merkle, T., Knaden, M., and Wehner, R. (2006a). Uncertainty about nest position influences systematic search strategies in desert ants. *Journal of Experimental Biology*, 209(18):3545–3549.

Merkle, T., Rost, M., and Alt, W. (2006b). Egocentric path integration models and their application to desert arthropods. *Journal of Theoretical Biology*, 240(3):385–399.

Merlin, C., Heinze, S., and Reppert, S. M. (2012). Unraveling navigational strategies in migratory insects. *Current opinion in neurobiology*, 22(2):353–361.

Milford, M. J. and Wyeth, G. F. (2012). SeqSLAM: Visual route-based navigation for sunny summer days and stormy winter nights. In *Robotics and Automation (ICRA), 2012 IEEE International Conference on*, pages 1643–1649. IEEE.

Milford, M. J., Wyeth, G. F., and Prasser, D. (2004). RatSLAM: a hippocampal model for simultaneous localization and mapping. In *Robotics and Automation, 2004. Proceedings. ICRA'04. 2004 IEEE International Conference on*, volume 1, pages 403–408. IEEE.

Mittelstaedt, H. (1962). Control systems of orientation in insects. *Annual review of entomology*, 7(1):177–198.

Mittelstaedt, M.-L. and Mittelstaedt, H. (1980). Homing by path integration in a mammal. *Naturwissenschaften*, 67(11):566–567.

Mokhtarian, F., Abbasi, S., Kittler, J., et al. (1997). Efficient and robust retrieval by shape content through curvature scale space. *Series on Software Engineering and Knowledge Engineering*, 8:51–58.

Möller, R. (2002). Insects could exploit UV–green contrast for landmark navigation. *Journal of theoretical biology*, 214(4):619–631.

Möller, R. (2009). Local visual homing by warping of two-dimensional images. *Robotics and Autonomous Systems*, 57(1):87–101.

Möller, R. (2012). A model of ant navigation based on visual prediction. *Journal of theoretical biology*, 305:118–130.

Moser, E. I., Kropff, E., and Moser, M.-B. (2008). Place cells, grid cells, and the brain's spatial representation system. *Neuroscience*, 31(1):69.

Moser, E. I., Roudi, Y., Witter, M. P., Kentros, C., Bonhoeffer, T., and Moser, M.-B. (2014). Grid cells and cortical representation. *Nature Reviews Neuroscience*, 15(7):466–481.

Mote, M. I. and Wehner, R. (1980). Functional characteristics of photoreceptors in the compound eye and ocellus of the desert ant, Cataglyphis bicolor. *Journal of comparative physiology*, 137(1):63–71.

Müller, M. and Wehner, R. (1988). Path integration in desert ants, Cataglyphis fortis. *Proceedings of the National Academy of Sciences*, 85(14):5287–5290.

Müller, M. and Wehner, R. (1994). The hidden spiral: systematic search and path integration in desert ants, Cataglyphis fortis. *Journal of Comparative Physiology A*, 175(5):525–530.

Müller, M. and Wehner, R. (2010). Path integration provides a scaffold for landmark learning in desert ants. *Current Biology*, 20(15):1368–1371.

Murillo, A. C. and Kosecka, J. (2009). Experiments in place recognition using gist panoramas. In *Computer Vision Workshops (ICCV Workshops), 2009 IEEE 12th International Conference on*, pages 2196–2203. IEEE.

Nern, A., Pfeiffer, B. D., and Rubin, G. M. (2015). Optimized tools for multicolor stochastic labeling reveal diverse stereotyped cell arrangements in the fly visual system. *Proceedings of the National Academy of Sciences*, 112(22):E2967–E2976.

Neuser, K., Triphan, T., Mronz, M., Poeck, B., and Strauss, R. (2008). Analysis of a spatial orientation memory in Drosophila. *Nature*, 453(7199):1244–1247.

Nistér, D., Naroditsky, O., and Bergen, J. (2004). Visual odometry. In *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, volume 1, pages 652–659. IEEE.

Nityananda, V., Tarawneh, G., Rosner, R., Nicolas, J., Crichton, S., and Read, J. (2016). Insect stereopsis demonstrated using a 3D insect cinema. *Scientific reports*, 6:18718–18718.

O'Carroll, D. (1993). Feature-detecting neurons in dragonflies. *Nature*, 362:541–543.

Ofstad, T. A., Zuker, C. S., and Reiser, M. B. (2011). Visual place learning in drosophila melanogaster. *Nature*, 474(7350):204–207.

Ogawa, Y., Falkowski, M., Narendra, A., Zeil, J., and Hemmi, J. M. (2015). Three spectrally distinct photoreceptors in diurnal and nocturnal Australian ants. *Proc. R. Soc. B*, 282(1808):20150673.

O'Keefe, J. and Dostrovsky, J. (1971). The hippocampus as a spatial map. preliminary evidence from unit activity in the freely-moving rat. *Brain research*, 34(1):171–175.

Oliva, A. and Torralba, A. (2001). Modeling the shape of the scene: A holistic representation of the spatial envelope. *International journal of computer vision*, 42(3):145–175.

Otsu, N. (1975). A threshold selection method from gray-level histograms. *Automatica*, 11(285-296):23–27.

Persoon, E. and Fu, K.-S. (1977). Shape discrimination using Fourier descriptors. *IEEE Transactions on systems, man, and cybernetics*, 7(3):170–179.

Pfeiffer, K. and Homberg, U. (2014). Organization and functional roles of the central complex in the insect brain. *Annual review of entomology*, 59:165–184.

Philippides, A., Baddeley, B., Cheng, K., and Graham, P. (2011). How might ants use panoramic views for route navigation? *Journal of Experimental Biology*, 214(3):445–451.

Pieron, H. (1904). Du role du sens musculaire dans l'orientation de quelques espèces de fourmis. *Bull Inst Gen Psychol*, 4:168–186.

Poggio, T. and Serre, T. (2013). Models of visual cortex. *Scholarpedia*, 8(4):3516. revision #149958.

Ronacher, B. and Wehner, R. (1995). Desert ants Cataglyphis fortis use self-induced optic flow to measure distances travelled. *Journal of Comparative Physiology A*, 177(1):21–27.

Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1988). Learning representations by back-propagating errors. *Cognitive modeling*, 5(3):213–220.

Samsonovich, A. and McNaughton, B. L. (1997). Path integration and cognitive mapping in a continuous attractor neural network model. *The Journal of neuroscience*, 17(15):5900–5920.

Santschi, F. (1911). *Observations et remarques critiques sur le mécanisme de l'orientation chez les fourmis*. Albert Kündig.

Schöne, H. (1996). Optokinetic speed control and estimation of travel distance in walking honeybees. *Journal of Comparative Physiology A*, 179(4):587–592.

Schoonderwoerd, R., Holland, O. E., Bruten, J. L., and Rothkrantz, L. J. (1997). Ant-based load balancing in telecommunications networks. *Adaptive behavior*, 5(2):169–207.

Schroff, F., Kalenichenko, D., and Philbin, J. (2015). Facenet: A unified embedding for face recognition and clustering. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 815–823.

Schultheiss, N. and Redish, A. (2015). The compass within. *Nature neuroscience*, 18(4):482–483.

Schultheiss, P., Wystrach, A., Schwarz, S., Tack, A., Delor, J., Nooten, S. S., Bibost, A.-L., Freas, C. A., and Cheng, K. (2016). Crucial role of ultraviolet light for desert ants in determining direction from the terrestrial panorama. *Animal Behaviour*, 115:19–28.

Schwarz, S., Narendra, A., and Zeil, J. (2011). The properties of the visual system in the Australian desert ant Melophorus bagoti. *Arthropod Structure & Development*, 40(2):128–134.

Seelig, J. D., Chiappe, M. E., Lott, G. K., Dutta, A., Osborne, J. E., Reiser, M. B., and Jayaraman, V. (2010). Two-photon calcium imaging from head-fixed Drosophila during optomotor walking behavior. *Nature methods*, 7(7):535–540.

Seelig, J. D. and Jayaraman, V. (2013). Feature detection and orientation tuning in the Drosophila central complex. *Nature*, 503(7475):262–266.

Seelig, J. D. and Jayaraman, V. (2015). Neural dynamics for landmark orientation and angular path integration. *Nature*, 521(7551):186–191.

Serre, T., Kouh, M., Cadieu, C., Knoblich, U., Kreiman, G., and Poggio, T. (2005). A theory of object recognition: computations and circuits in the feedforward path of the ventral stream in primate visual cortex. Technical report, DTIC Document.

Seung, H. S. (1996). How the brain keeps the eyes still. *Proceedings of the National Academy of Sciences*, 93(23):13339–13344.

Sivic, J. and Zisserman, A. (2003). Video Google: A text retrieval approach to object matching in videos. In *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*, pages 1470–1477. IEEE.

Srinivasan, M., Zhang, S., Lehrer, M., and Collett, T. (1996). Honeybee navigation en route to the goal: visual flight control and odometry. *Journal of Experimental Biology*,

199(1):237–244.

Srinivasan, M. V. (2014). Going with the flow: a brief history of the study of the honeybee's navigational 'odometer'. *Journal of Comparative Physiology A*, 200(6):563–573.

Srinivasan, M. V., Chahl, J. S., Weber, K., Venkatesh, S., Nagle, M. G., and Zhang, S.-W. (1999). Robot navigation inspired by principles of insect vision. *Robotics and Autonomous Systems*, 26(2):203–216.

Srinivasan, M. V. and Gregory, R. (1992). How bees exploit optic flow: behavioural experiments and neural models [and discussion]. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 337(1281):253–259.

Stange, G., Berry, R., Van Kleef, J., et al. (2007). Design concepts for a novel attitude sensor for micro air vehicles, based on dragonfly ocellar vision. In *European Micro Air Vehicle Conference and Flight Competition*, volume 2007, page 7.

Stange, G. and Howard, J. (1979). An ocellar dorsal light response in a dragonfly. *J. exp. Biol*, 83:351–355.

Stieb, S. M., Muenz, T. S., Wehner, R., and Rössler, W. (2010). Visual experience and age affect synaptic organization in the mushroom bodies of the desert ant Cataglyphis fortis. *Developmental neurobiology*, 70(6):408–423.

Stone, T., Differt, D., Milford, M., and Webb, B. (2016a). Skyline-based localisation for aggressively manoeuvring robots using UV sensors and spherical harmonics. In *2016 IEEE International Conference on Robotics and Automation (ICRA)*, pages 5615–5622. IEEE.

Stone, T., Mangan, M., Ardin, P., and Webb, B. (2014). Sky segmentation with ultraviolet images can be used for navigation. In *Proceedings of Robotics: Science and Systems*, Berkeley, USA.

Stone, T., Weddig, N. B., Adden, A., Honkanen, A., Templin, R., Wcislo, W., Warrant, E., Webb, B., and Heinze, S. (2016b). A neural substrate for path integration in the bee brain. *Nature (UNDER REVIEW)*.

Strauss, R. and Berg, C. (2010). The central control of oriented locomotion in insects-towards a neurobiological model. In *The 2010 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE.

Stürzl, W. and Mallot, H. A. (2006). Efficient visual homing based on Fourier transformed panoramic images. *Robotics and Autonomous Systems*, 54(4):300–313.

Sünderhauf, N., Shirazi, S., Dayoub, F., Upcroft, B., and Milford, M. (2015). On the performance of ConvNet features for place recognition. In *Intelligent Robots and Systems (IROS), 2015 IEEE/RSJ International Conference on*, pages 4297–4304. IEEE.

Tarr, M. J. and Pinker, S. (1989). Mental rotation and orientation-dependence in shape recognition. *Cognitive psychology*, 21(2):233–282.

Taube, J. S., Muller, R. U., and Ranck, J. B. (1990). Head-direction cells recorded from the postsubiculum in freely moving rats. I. description and quantitative analysis. *The Journal of neuroscience*, 10(2):420–435.

Taylor, C. P. (1981). Contribution of compound eyes and ocelli to steering of locusts in flight: I. behavioural analysis. *Journal of Experimental Biology*, 93(1):1–18.

Thrun, S., Fox, D., Burgard, W., and Dellaert, F. (2001). Robust Monte Carlo localization for mobile robots. *Artificial intelligence*, 128(1):99–141.

Towne, W. F. and Moscrip, H. (2008). The connection between landscapes and the solar ephemeris in honeybees. *Journal of Experimental Biology*, 211(23):3729–3736.

Ulanovsky, N. and Moss, C. F. (2007). Hippocampal cellular and network activity in freely moving echolocating bats. *Nature neuroscience*, 10(2):224–233.

Ulrich, I. and Nourbakhsh, I. (2000). Appearance-based place recognition for topological localization. In *Robotics and Automation, 2000. Proceedings. ICRA'00. IEEE International Conference on*, volume 2, pages 1023–1029. IEEE.

Van Oudenhove, L., Boulay, R., Lenoir, A., Bernstein, C., and Cerda, X. (2012). Substrate temperature constrains recruitment and trail following behavior in ants. *Journal of chemical ecology*, 38(6):802–809.

Varga, A. G. and Ritzmann, R. E. (2016). Cellular basis of head direction and contextual cues in the insect brain. *Current Biology*, 26(14):1816–1828.

Vickerstaff, R. and Di Paolo, E. (2005). Evolving neural models of path integration. *Journal of Experimental Biology*, 208(17):3349–3366.

Vickerstaff, R. J. and Cheung, A. (2010). Which coordinate system for modelling path integration? *Journal of Theoretical Biology*, 263(2):242–261.

Vincent, J. F., Bogatyreva, O. A., Bogatyrev, N. R., Bowyer, A., and Pahl, A.-K. (2006). Biomimetics: its practice and theory. *Journal of the Royal Society Interface*, 3(9):471–482.

Von Frisch, K. (1967). *The dance language and orientation of bees.* Harvard University Press.

Wang, A. et al. (2004). An industrial strength audio search algorithm. In *SPIE proceedings series*, pages 582–588. Society of Photo-Optical Instrumentation Engineers.

Warrant, E. and Dacke, M. (2011). Vision and visual navigation in nocturnal insects. *Annual review of entomology*, 56:239–254.

Webb, B. (2001). Can robots make good models of biological behaviour? *Behavioral and brain sciences*, 24(06):1033–1050.

Webb, B. and Wystrach, A. (2016). Neural mechanisms of insect navigation. *Current Opinion in Insect Science*, 15:27–39.

Wehner, R. (1976). Polarized-light navigation by insects. *Scientific American*, 235(1):106–115.

Wehner, R. (1997). The ant's celestial compass system: spectral and polarization channels. In *Orientation and communication in arthropods*, pages 145–185. Springer.

Wehner, R. (2003). Desert ant navigation: how miniature brains solve complex tasks. *Journal of Comparative Physiology A*, 189(8):579–588.

Wehner, R. and Menzel, R. (1990). Do insects have cognitive maps? *Annual review of neuroscience*, 13(1):403–414.

Wehner, R., Michel, B., and Antonsen, P. (1996). Visual navigation in insects: coupling of egocentric and geocentric information. *Journal of Experimental Biology*, 199(1):129–140.

Wehner, R. and Srinivasan, M. V. (1981). Searching behaviour of desert ants, genus Cataglyphis (Formicidae, Hymenoptera). *Journal of comparative physiology*,

142(3):315–338.

Weyand, T., Kostrikov, I., and Philbin, J. (2016). PlaNet-photo geolocation with convolutional neural networks. *arXiv preprint arXiv:1602.05314*.

Willshaw, D. J., Buneman, O. P., and Longuet-Higgins, H. C. (1969). Non-holographic associative memory. *Nature*, 222(5197):960–962.

Wittlinger, M., Wehner, R., and Wolf, H. (2006). The ant odometer: stepping on stilts and stumps. *Science*, 312(5782):1965–1967.

Wittmann, T. and Schwegler, H. (1995). Path integration – a network model. *Biological Cybernetics*, 73(6):569–575.

Wolff, G. H. and Strausfeld, N. J. (2015). Genealogical correspondence of mushroom bodies across invertebrate phyla. *Current Biology*, 25(1):38–44.

Wolff, T., Iyer, N. A., and Rubin, G. M. (2015). Neuroarchitecture and neuroanatomy of the Drosophila central complex: A GAL4-based dissection of protocerebral bridge neurons and circuits. *Journal of Comparative Neurology*, 523(7):997–1037.

Wystrach, A., Beugnon, G., and Cheng, K. (2012). Ants might use different view-matching strategies on and off the route. *Journal of Experimental Biology*, 215(1):44–55.

Wystrach, A. and Graham, P. (2012). What can we learn from studies of insect navigation? *Animal Behaviour*, 84(1):13–20.

Wystrach, A., Mangan, M., Philippides, A., and Graham, P. (2013). Snapshots in ants? new interpretations of paradigmatic experiments. *Journal of Experimental Biology*, 216(10):1766–1770.

Wystrach, A., Mangan, M., and Webb, B. (2015). Optimal cue integration in ants. *Proceedings of the Royal Society of London B: Biological Sciences*, 282(1816):20151484.

Wystrach, A., Philippides, A., Aurejac, A., Cheng, K., and Graham, P. (2014). Visual scanning behaviours and their role in the navigation of the Australian desert ant Melophorus bagoti. *Journal of Comparative Physiology A*, 200(7):615–626.

Yosinski, J., Clune, J., Nguyen, A., Fuchs, T., and Lipson, H. (2015). Understanding neural networks through deep visualization. *arXiv preprint arXiv:1506.06579*.

Zeil, J. (2012). Visual homing: an insect perspective. *Current opinion in neurobiology*, 22(2):285–293.

Zeil, J., Hofmann, M. I., and Chahl, J. S. (2003). Catchment areas of panoramic snapshots in outdoor scenes. *JOSA A*, 20(3):450–469.

Zeiler, M. D. and Fergus, R. (2014). Visualizing and understanding convolutional networks. In *European Conference on Computer Vision*, pages 818–833. Springer.

Zhang, K. (1996). Representation of spatial orientation by the intrinsic dynamics of the head-direction cell ensemble: a theory. *The journal of neuroscience*, 16(6):2112–2126.

Zheng, Y.-T., Zhao, M., Song, Y., Adam, H., Buddemeier, U., Bissacco, A., Brucher, F., Chua, T.-S., and Neven, H. (2009). Tour the world: building a web-scale landmark recognition engine. In *Computer vision and pattern recognition, 2009. CVPR 2009. IEEE conference on*, pages 1085–1092. IEEE.

Ziegler, P. and Wehner, R. (1997). Time-courses of memory decay in vector-based and landmark-based systems of navigation in desert ants, Cataglyphis fortis. *Journal of*

Comparative Physiology A, 181(1):13–20.

Zobel, J. and Hoad, T. C. (2006). Detection of video sequences using compact signatures. *ACM Transactions on Information Systems (TOIS)*, 24(1):1–50.

Zollikofer, C., Wehner, R., and Fukushi, T. (1995). Optical scaling in conspecific Cataglyphis ants. *Journal of Experimental Biology*, 198(8):1637–1646.

Bibliographies for included papers can be found on pages 42, 57, 87, 110 and 119.