A Novel Multi-vision Sensor Dataset for Insect-Inspired Outdoor Autonomous Navigation

Jan K. N. Verheyen(✉), Julien Dupeyroux, and Guido C. H. E. de Croon

Micro Air Vehicle Laboratory, Department of Control and Simulation Faculty of Aerospace Engineering, Delft University of Technology, 2629HS Delft, The Netherlands
jan.verheyen@protonmail.com, g.c.h.e.decroon@tudelft.nl

Abstract. Insects have—over millions of years of evolution—perfected many of the systems that roboticists aim to achieve; they can swiftly and robustly navigate through different environments under various conditions while at the same time being highly energy efficient. To reach this level of performance and efficiency, one might want to look at and take inspiration from how these insects achieve their feats. Currently, no dataset exists that allows bio-inspired navigation models to be evaluated over long >100 m real-life routes. We present a novel dataset containing omnidirectional event vision, frame-based vision, depth frames, inertial measurement (IMU) readings, and centimeter-accurate GNSS positioning over kilometer long stretches in and around the TUDelft campus. The dataset is used to evaluate familiarity-based insect-inspired neural navigation models on their performance over longer sequences. It demonstrates that current scene familiarity models are not suited for long-ranged navigation, at least not in their current form.

Keywords: Long-range navigation · Neuromorphic systems · Event-based camera · RGB Camera · GPS · GNSS

1 Introduction

To date, some insect-inspired aerial robots have been developed [9,19] which mimic the flight capabilities of insects and while basic navigating capabilities have already been shown on board such limited platforms [26], their navigational performance falls short compared to their biological counterparts. Recent neural insect-inspired navigational models [2,3,11,36] show promising results over short distances, but lack the capacity for long-ranged (>100 m) navigation. This could be overcome by chaining local navigation strategies [10], although the difficulty here lies in where to transition from one local navigation strategy into the other. One of the major hurdles still holding back high-performance navigation onboard robots is energy-efficient visual processing. Insects' visual system is event-based, where photosensitive cells react *independently* from each other to changes in light intensity and subsequently generate spikes that propagate

A. Hunt et al. (Eds.): Living Machines 2022, LNAI 13548, pp. 279–291, 2022.
https://doi.org/10.1007/978-3-031-20470-8_28

through the visual system to be processed in their miniature brains. Event cameras are *neuromorphic* vision sensors that mimic that process. Here, pixels take the role of the photosensitive cells and generate events asynchronously. Visual information is thus captured in a stream of events opposed to synchronous frames as taken by traditional cameras. This allows neuromorphic cameras to operate at very high temporal resolution and low latency (in the order of microseconds), very high dynamic range (140 dB compared to 60 dB of standard cameras), high pixel bandwidth (in the order of kHz), and low power consumption (order of mW) [14]. Processing such information requires novel methods to be developed. Biologically plausible methods involve the use of spiking neural networks (SNNs) since they are biologically more similar to networks of neurons found in animal nervous systems than regular artificial neural networks (ANNs). Implemented on neuromorphic processors such as Intel's Loihi and IBM's Truenorth, SNNs can deliver highly powerful computing at a fraction of the power budget of traditional hardware (CPUs, GPUs), making them promising candidates for implementation on robots.

Datasets form an important part of training and evaluating such novel methods. Currently, there are several event-based vision datasets focusing on navigation, covering applications in visual odometry, depth reconstruction, and SLAM, but little focusing on insect-inspired navigation. Images, events, optic flow, 3D camera motion, and scene depth in static scenes using a mobile robotic platform are provided in [4]. A large automotive dataset containing over 250000 human-labeled box annotations of cars and pedestrians is presented by [31]. The dataset provided by [35] includes synchronized stereo event data, augmented with grayscale images, and inertial measurement unit (IMU) readings. Ground truth pose and depth images are provided through a combination of a LiDAR system, IMU, indoor and outdoor motion capture, and GPS. The DDD20 [17] dataset consists of an extensive automotive dataset with events, frames, and vehicle human control data collected from 4000 km of mixed highway and urban driving. However, most insects have compound eyes that cover an almost panoramic FOV and this plays an important role in insects' navigational dexterity [15]. Additionally, insects fuse various sensory inputs from their environment together during navigation [13], making datasets that combine sensors valuable sources for training and evaluating such methods in e.g. multimodal navigation such as [16,30]. None of the datasets above provide event data captured through an omnidirectional lens enhanced with additional sensors over long distances.

This paper presents two main contributions. First, a dataset containing omnidirectional event camera and IMU data, forward-facing high-resolution footage, and centimeter-level accurate GPS data along with a software package to load, process and manipulate the dataset. The dataset, including the software package, will be made available at https://github.com/tudelft/NavDataset. Secondly, an evaluation is presented of three different familiarity-based insect-inspired navigation models from the literature [2,3,36] with respect to their performance in long-ranged navigation.

2 The Biological Principles in Insect Navigation

2.1 Visual Perception

Insects' compound eyes consist of small individual hexagonally-shaped photoreceptive units, called *ommatidia*, which are arranged to form a faceted surface, capable of covering an almost panoramic field of view. Each such ommatidium receives light only from a small angle in the visual field, constricting the visual *acuity* of the insect's visual system. When excited by photons, these photoreceptive cells generate electric signals encoding the amount of light it absorbs, which downstream neurons turn into *spikes* that are passed through to the underlying optic lobes [25]. This low resolution but often almost panoramic vision from insects plays a crucial role in the success of their visual navigation's neural implementations.

2.2 Insect Visual Navigation Models

Insects are adept navigators capable of maneuvering through cluttered environments and memorizing long routes. Cartwright and Collett's [8] seminal snapshot model presented some of the first work that studied and modeled the visual navigation feats of honeybees. It hypothesized that honeybees store a single retinotopic *snapshot* of the place that they later want to navigate back to. Other methods include the Average Landmark Vector (ALV) model [21], image warping [12], and rotationally invariant panoramic methods that utilize Fourier-transformed [29] images or other frequency-domain-based methods [28].

The area surrounding the stored snapshot from which agents can successfully return is categorized as the *catchment area*. The extent of the catchment area changes depending on various factors such as the deployed navigation technique and the complexity and texture of the environment [34]. Insects therefore implement different visual navigation strategies, depending on whether they are on an already familiar route or find themselves in unfamiliar surroundings, respectively switching between route-following and visual homing behaviors [33].

Route-following methods commonly exploit the fact that generally, in natural scenes, the root-mean-square difference between the stored panoramic snapshot and another panoramic snapshot (also called the *image difference*) changes smoothly in correlation with the distance from the stored snapshot, where it terminates in a sharp minimum [34]. Visual route following methods include the Descent in Image Difference (DID) methods, which follows the declining gradient in image difference towards the stored snapshot [23]. Later, more biologically plausible DID models employed an ANN-based approach. In the scene familiarity model [3], a route is learned through training a 2-layer feedforward network to memorize snapshots along a route. An SNN-based scene familiarity model, modeled after the mushroom bodies (MBs) of ants was later formulated by [2]. Other work looked at integrating various cues in a multimodal decentralized neural model, combining input from the Central Complex (CX), the MB and the Path Integration (PI) centers [30].

2.3 Neuromorphic Processing

Event-Based Vision Sensors. Event cameras are vision sensors that take inspiration from the working principle of the biological retina. Each pixel reacts asynchronously to changes in light intensity. The sensor logs the pixel's location, time (in microsecond resolution), and polarity ('ON' or 'OFF'), and sends it over a digital bus in an Address-event Representation (AER) format [14]. Because event cameras only react to small *changes* in light intensity at individual pixels, visual information is more efficiently conveyed compared to frame-based cameras.

Spiking Neural Networks. Analogous to their biological counterparts, artificial neurons in SNNs generate a spike (action potential) if their membrane potential reaches a certain threshold after receiving a series of excitatory spikes from upstream neurons over their synaptic connections. After firing, the neuron lowers its internal voltage to a resting state. For a short time (the so-called refractory period) the neuron will not react to any incoming signals anymore. Various computational models of biological neurons exist to replicate this behavior. The most used neuronal models in artificial spiking neural networks nowadays are the Leaky Integrate-and-Fire (LIF) [27], Spike Response Model (SRM) [20], and the Izhikevich [18] model. The LIF neuron model in Eq. 1 shows how presynaptic spikes $s_j(t)$ arriving from neurons in layer $l-1$ increase (or decrease)—depending on weight matrix $W_{i,j}$, denoting its synaptic connectivity—the neuron's membrane potential $v_i(t)$ (scaled with the time constant λ_v) after which it decays to its resting potential v_{rest} if no more signals arrive. If enough excitatory presynaptic spikes arrive in short succession, the membrane potential will reach a certain internal threshold after which the neuron spikes ($s_i(t)$), resets its membrane potential, and enters a refractory period. Inhibiting presynaptic spikes have the opposite effect and will lower the membrane potential.

$$\lambda_v \frac{dv_i(t)}{dt} = -(v_i(t) - v_{\text{rest}}) + \sum_{j=1}^{n^{l-1}} \left(W_{i,j} s_j^{l-1}(t - \tau_d) \right) \tag{1}$$

Information in SNNs can be encoded in several different manners including position, temporal, rate coding, and subsequent combinations thereof. Learning in SNNs thus takes place in these domains, traditionally with much focus on a mechanism called Spike-Timing-Dependent Plasticity (STDP) [7]. STDP is a (biological) form of Hebbian learning that changes the synaptic strength of neuron connections dependent on their relative spike timing.

3 Dataset Design

The following section presents the utilized sensors and how the dataset was collected. The dataset was collected in both rural and urban environments in and around the TUDelft campus (Delft, The Netherlands). It mainly consists of events and IMU readings captured by a DAVIS240 event camera (as a more

biologically accurate source of visual data) and video from a GoPro Hero 8 Black along with RTK GNSS positioning data. Video was collected in HEVC encoded MP4 and ROS bag files for the rest. The dataset also provides the same raw data in HDF5 containers. The DAVIS240 sensor was fitted with an omnidirectional lens to more accurately represent insects' vision, and its benefits in visual route following. Section 3.1 gives an overview of the dataset collection platform and the acquisition environment. A Python3 package will be made available for converting the bag and HDF5 files to and from various formats, as well as performing the data (pre)processing as elaborated upon in Sect. 3.2.

3.1 Sensors and Data Acquisition

Sensors. Table 1 provides an overview of the sensors with their characteristics. The complete logistical overview of the dataset acquisition platform can be seen in Fig. 1A and B. The dataset junction box forms the housing holding the various sensors as well as the Intel Up board computation platform. The Intel Up board runs ROS Kinetic and is responsible for collecting and time synchronizing the data from the various sensors which are connected over USB2/3 buses. The GNSS antenna was mounted at the back of the bike to minimize interference from the USB3 controllers. A mobile phone with cellular was connected to the Intel Up board by connecting to the phone's wifi hotspot. This allowed for running commands on the Intel Up board over ssh as well as provided the Intel Up board with internet access. This was needed for RTK GNSS positioning; RTCM messages were sent to the ublox ZED-F9P GNSS receiver by connecting to the EUREF-IP network ntrip server allowing for up to centimeter-accurate positioning. The DAVIS240 camera was mounted to an omnidirectional catadioptric lens to achieve omnidirectional vision. The GoPro camera was manually operated, thus a small gap exists between its timing and the rest of the sensors; this has been manually compensated for in the post-processing of the data. An external portable SSD was utilized to offload the collected data after every single run as the internal storage of the Intel Up board was limited. The dataset box was then mounted to a bike (Fig. 1B).

Sequences. The dataset consists of 12 routes traversed by bike from and to the start point (Fig. 2). The runs cover both rural and urban environments in and around the TUDelft campus. The dataset was collected over two days in April in the afternoon with similar partially clouded sunny weather conditions. Each bike run was travelled at about 18 km/h on primarily bike lanes.

3.2 Data Processing

Central to this dataset is the data provided by the DAVIS240 equipped with the omnidirectional catadioptric lens. As can be seen in Fig. 3, the omnidirectional lens projects its light on a circular region on the DAVIS240 sensor. In Fig. 3A the direction of the view with respect to the bike's heading is annotated. The

Table 1. Dataset collected data

Type	Sensor	Characteristics	Container
Visual	DAVIS240	240 × 180 pixel DVS AER	ROS bag/HDF5
Visual	GoPro Hero 8 Black	1920 × 1080 pixel 60 fps	mp4 (HEVC)
Visual	Intel Realsense d435i	720 × 1280 pixel depth 30 fps 16UC1	ROS bag
Position-ing	Ublox ZED-F9P GNSS	NavPVT 5 Hz capture rate Position Accuracy 1.3 cm CEP	ROS bag/HDF5

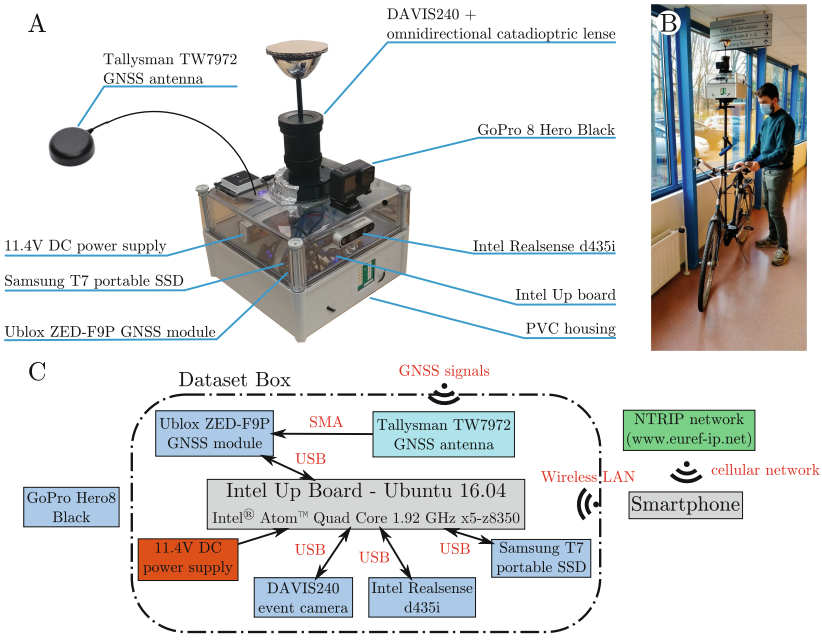

Fig. 1. Dataset acquisition hardware. **A** shows an overview of the various sensors mounted on the dataset box. **B** shows the full setup—the dataset box mounted on a bike to cover the long distances. **C** shows the data flow diagram between the sensors and the central computers.

approximate location of the capture can be seen in Fig. 3B. This circular projection can easily be masked as it stays fixed with respect to the sensor, subsequent unwrapping (in this case unwrapping is simply performed by mapping the polar coordinates of view Fig. 3A directly to image coordinates) results in the view presented by Fig. 3C.

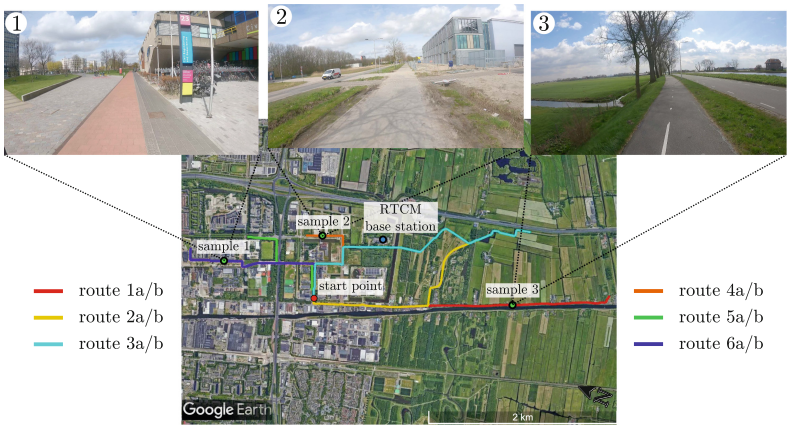

Fig. 2. Map of the routes covered by the dataset. Route 'a' indicates runs away from the start point, 'b' towards the start point. Samples are shown in 1–3.

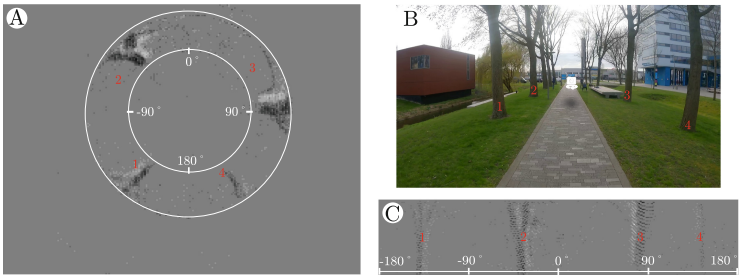

Fig. 3. Example of the omnidirectional events captured by the DAVIS240 and subsequent preprocessing. **A** shows an accumulated event frame as captured by the omnidirectional system, the two concentric circles show the boundaries of the visible field. **B** shows a frame from the GoPro footage, the left snapshot coincides with the camera position in the frame. **C** shows the event stream after masking and unwrapping the events.

4 Experimental Study: Evaluating Familiarity-Based Neural Insect Navigation Models

The following section presents the use of the dataset to investigate a few recent neural insect-inspired familiarity-based navigational models. The experiments compare three neural familiarity-based insect navigation models in terms of their performance for long-ranged navigation, namely [3]'s scene familiarity neural network, [2]'s Mushroom Body (MB) model, and [36]'s MB model. The mushroom bodies are relatively large structures in the insect brain that consist of large parallel arrangements of neurons, called Kenyon Cells (KCs), which are sampled by a relatively small amount of extrinsic neurons, also called Mushroom Body Output Neurons (MBONs). The mushroom bodies' role in visual learning has

been investigated, revealing direct neural connections between the medulla and mushroom bodies [32]. Recent research has shown that MBs are *necessary* for learned visual navigation [6]. The aforementioned scene-familiarity models have been mostly tested in either simulated environments [2,3] or over very short distances in a controlled environment [36]. This dataset provides an interesting testing ground to evaluate these methods in real-life conditions with visual data inspired by the way insects perceive their environment. The inclusion of frame-based video allows for comparison between frame-based [2,3] and event-based [36] methods, but could also be used for multimodal models, utilizing the GNSS data as a virtual compass. We are specifically interested in how these methods hold up over longer distances.

4.1 Neural Familiarity-Based Insect Navigation Models

Baddeley et al.'s Scene Familiarity Model. The (frame-based) scene familiarity model of [3] consists of an input layer with the same dimensions as the number of pixels in the acquired images, which is fully connected by feedforward connections to a novelty layer which consists of tanh activation functions. The information about the input presented by the novelty layer is maximized through weight adaptation following the Information-Maximization (infomax) principle [5]. The infomax principle adjusts the weights of the network in such a way as to maximize the information about the input that the novelty layer presents. This is performed by following the gradient of mutual information [3]. By maximization of information through weight adaptation, the output of the novelty layer units is decorrelated, effectively reducing the network's output for sequences that have already been seen. 'Familiar' frames can be discerned after a single training run.

Ardin et al.'s Mushroom Body Model. The (frame-based) SNN MB model presented by [2] consists of 360 visual projection neurons (vPNs) that are sparsely connected to 20000 KCs which connect to a single MBON. Each such vPN can thus activate only a handful of KCs, representing a sparse encoding of information. Learning is performed by lowering the synaptic weights (Long-Term Depression (LTD)) between the KCs and the MBON through STDP. As a result, after training, the MBON's spike rate is lower for familiar views opposed to novel ones. Due to solely applying LTD, these connections are permanently weakened. This limits the model's capacity to memorize long sequences to the amount of depletable weights and the sparseness of the connections.

Zhu et al.'s Mushroom Body Model. The model presented by [36] (event-based) adapts Ardin et al.'s [2] model based on the finding that 60% of the input synapses of KCs come from other KCs. Instead of performing learning on the weights connecting the KCs to the MBON; when a KC spikes, it inhibits its connection to downstream KCs that spike at a later time based on an STDP rule. The KCs are split up into two groups of 5000 neurons to speed up learning; each

solely acting within its group. Additionally, an anterior paired lateral (APL) [1] neuron is included that inhibits the activity of the KC layer.

4.2 Setup

The aforementioned neural networks were trained on sequences of 8, 16, 24, and 32 s from a 32-s section of route 1a (see Fig. 2). The respective networks were trained to 'memorize' that stretch of the route. Ensuing runs would result in a lower response from the network for already seen sequences, compared to unseen sequences. Ideally having a high (thus comparable to the untrained baseline run in orange) response for unfamiliar views and vice versa for familiar views. During the validation run, the same sequence was injected with unfamiliar views over stretches of 4 s and presented to the networks (gray bars in Fig. 4). The injected parts were sampled from sections of other routes in the dataset. For the event-based networks, the injected sequences were closely matched to the event rate of the original sequence, to maintain similar levels of activity in the network's layers. The frame-based models were presented with 28×8 pixel grayscale histogram equalized frames flattened to a 1D array. Input from the DAVIS240 sensor was max-pooled to 32×7 pixels before passing it to the network.

4.3 Results

The results of the experiment can be seen in Fig. 4. Inspired by [36]'s novelty index, we apply a performance index P

$$P = \frac{\sum s_{\text{unfamiliar}} - \sum s_{\text{familiar}}}{\sum s_{\text{total}}} \tag{2}$$

with s the response (familiarity index/spike rate) of the network, to evaluate the model's performance over increasingly longer test sequences ($P = 1$ is 'perfect' performance). The performance index captures the normalized relative difference in the response of the network between familiar and unfamiliar views. Its results are visualized in Fig. 4D. The Infomax scene familiarity [3] model's performance index decreased overall, while Ardin et al.'s [2] stabilized, but both these frame-based models maintain an adequate performance level to still separate familiar from unfamiliar views. This is in stark contrast with Zhu et al.'s [36] event-based model, which has depleted its 'memory' after about 16 s of learning (so not showing the full 32 s for visual clarity in Fig. 4). This deteriorating performance over longer distances is a result of the limited capacity of the networks, as synaptic connections' weights are depleted during training. This could be improved upon by a number of factors. First, the networks were trained with a constant learning rate, this could be tuned for longer distances, although at the cost of lower performance in general. Secondly, one could lower the number of presented frames 60 Hz to lower rates based on some metric of the input data. Further increases in the network's size by increasing the amount of KCs remains an option as well, although its computational increase would severely limit the

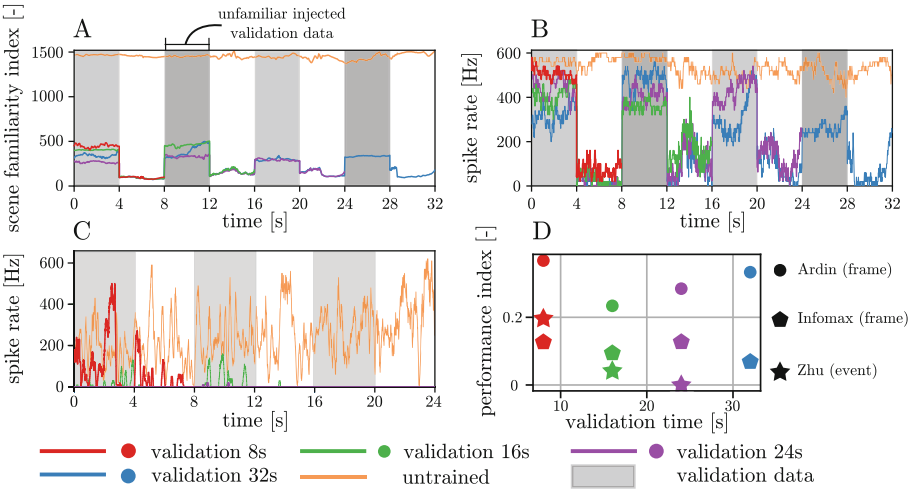

Fig. 4. Insect navigation models response to 8-, 16-, 24-, and 32-second sequences of a section of route 1a. **A** Infomax neural network [3] (frame-based). **B** Ardin et al.'s [2] MB model (frame-based). **C** Zhu et al.'s [36] MB model (event-based). **D** Performance index P (Eq. 2)

number of deployable robotic platforms. The frame-based method's more stable performance could be a consequence of them having more control over their input through well-established techniques such as normalization, which are less developed for event-based vision. Investigating intrinsically modulating mechanisms such as [24]'s adaptive LIF neuron could perhaps provide more fundamental solutions for this. Furthermore, it is known that insects perform a number of preprocessing steps (including elementary motion detection such as optic flow) in their optic lobes [25] as well as have mechanisms present that adjust the learning and forgetting of 'unnecessary' information [13], worth investigating.

5 Conclusion

The aforementioned navigational models have been evaluated on their capability to discriminate between familiar and unfamiliar views, which differs from closed-loop navigation evaluations as presented in e.g. [2,3]. Nonetheless, the experiments in Sect. 4 present some issues with how usable these networks are for memorizing longer sequences. Successful navigational methods could utilize different mechanisms to still obtain dependable navigational e.g. combining attractive and repulsive cues [22] or integrating several navigational cues [16,30]. This work aims to provide a valuable tool with which the further development of such neural insect-inspired long-range navigation methods can be accelerated.

References

1. Amin, H., Apostolopoulou, A.A., Suárez-Grimalt, R., Vrontou, E., Lin, A.C.: Localized inhibition in the Drosophila mushroom body. eLife 9 (2020). https://doi.org/10.7554/eLife.56954, https://elifesciences.org/articles/56954
2. Ardin, P., Peng, F., Mangan, M., Lagogiannis, K., Webb, B.: Using an insect mushroom body circuit to encode route memory in complex natural environments. PLOS Comput. Biol. **12**(2), e1004683 (2016). https://doi.org/10.1371/journal.pcbi.1004683, https://dx.plos.org/10.1371/journal.pcbi.1004683
3. Baddeley, B., Graham, P., Husbands, P., Philippides, A.: A Model of Ant Route Navigation Driven by Scene Familiarity. PLoS Computational Biology 8(1), e1002336 (2012). https://doi.org/10.1371/journal.pcbi.1002336,https://dx.plos.org/10.1371/journal.pcbi.1002336
4. Barranco, F., Fermuller, C., Aloimonos, Y., Delbruck, T.: A dataset for visual navigation with neuromorphic methods. Front. Neurosci. **10**, 49 (2016). https://doi.org/10.3389/fnins.2016.00049, http://journal.frontiersin.org/Article/10.3389/fnins.2016.00049/abstract
5. Bell, A.J., Sejnowski, T.J.: An Information-Maximization Approach to Blind Separation and Blind Deconvolution. Neural Comput. **7**(6), 1129–1159 (1995). https://doi.org/10.1162/neco.1995.7.6.1129,http://www.mitpressjournals.org/doi/10.1162/neco.1995.7.6.1129
6. Buehlmann, C., Wozniak, B., Goulard, R., Webb, B., Graham, P., Niven, J.E.: Mushroom bodies are required for learned visual navigation, but not for innate visual behavior. Ants. Current Biol. **30**(17), 3438-3443.e2 (2020). https://doi.org/10.1016/j.cub.2020.07.013
7. Caporale, N., Dan, Y.: Spike timing-dependent plasticity: a Hebbian learning rule. Ann. Rev. Neurosci. **31**(1), 25–46 (2008). https://doi.org/10.1146/annurev.neuro.31.060407.125639, https://www.annualreviews.org/doi/10.1146/annurev.neuro.31.060407.125639
8. Cartwright, B.A., Collett, T.S.: Landmark learning in bees - Experiments and models. J. Comparat. Physiol. **151**(4), 521–543 (1983). https://doi.org/10.1007/BF00605469, http://link.springer.com/10.1007/BF00605469
9. de Croon, G., de Clercq, K., Ruijsink, R., Remes, B., de Wagter, C.: Design, aerodynamics, and vision-based control of the DelFly. Int. J. Micro Air Vehicles **1**(2), 71–97 (2009). https://doi.org/10.1260/175682909789498288, http://journals.sagepub.com/doi/10.1260/175682909789498288
10. Denuelle, A., Srinivasan, M.V.: A sparse snapshot-based navigation strategy for UAS guidance in natural environments. In: 2016 IEEE International Conference on Robotics and Automation (ICRA). vol. 2016-June, pp. 3455–3462. IEEE, May 2016. https://doi.org/10.1109/ICRA.2016.7487524, http://ieeexplore.ieee.org/document/7487524/
11. Dupeyroux, J., Serres, J.R., Viollet, S.: Antbot: a six-legged walking robot able to home like desert ants in outdoor environments. Sci. Robot. **4**(27), eaau0307 (2019)
12. Franz, M.O., Schölkopf, B., Mallot, H.A., Bülthoff, H.H.: Where did I take that snapshot? Scene-based homing by image matching. Biol. Cybern. **79**(3), 191–202 (1998). https://doi.org/10.1007/s004220050470, http://link.springer.com/10.1007/s004220050470
13. Freas, C.A., Schultheiss, P.: How to Navigate in Different Environments and Situations: Lessons From Ants. Front,. Psychol. **9**, 1–7 (2018). https://doi.org/10.3389/fpsyg.2018.00841, https://www.frontiersin.org/article/10.3389/fpsyg.2018.00841/full

14. Gallego, G., et al.: Event-based vision: a survey. IEEE Trans. Pattern Anal. Mach. Intell. **44**(1), 154–180 (2022). https://doi.org/10.1109/TPAMI.2020.3008413
15. Graham, P., Philippides, A.: Vision for navigation: what can we learn from ants? Arthropod Struct. Dev. **46**(5), 718–722 (2017). https://doi.org/10.1016/j.asd.2017.07.001, https://www.sciencedirect.com/science/article/pii/S1467803917300932
16. Hoinville, T., Wehner, R.: Optimal multiguidance integration in insect navigation. Proc. Natl. Acad. Sci. U.S.A. **115**(11), 2824–2829 (2018). https://doi.org/10.1073/pnas.1721668115
17. Hu, Y., Binas, J., Neil, D., Liu, S.C., Delbruck, T.: DDD20 end-to-end event camera driving dataset: fusing frames and events with deep learning for improved steering prediction. In: 2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC), pp. 1–6. IEEE, September 2020. https://doi.org/10.1109/ITSC45102.2020.9294515, https://ieeexplore.ieee.org/document/9294515/
18. Izhikevich, E.: Simple model of spiking neurons. IEEE Trans. Neural Networks **14**(6), 1569–1572 (2003). https://doi.org/10.1109/TNN.2003.820440, http://ieeexplore.ieee.org/document/1257420/
19. Jafferis, N.T., Helbling, E.F., Karpelson, M., Wood, R.J.: Untethered flight of an insect-sized flapping-wing microscale aerial vehicle. Nature **570**(7762), 491–495 (2019). https://doi.org/10.1038/s41586-019-1322-0, https://www.nature.com/articles/s41586-019-1322-0
20. Kistler, W.M., Gerstner, W., Hemmen, J.L.V.: Reduction of the hodgkin-huxley equations to a single-variable threshold model. Neural Comput. **9**(5), 1015–1045 (1997). https://doi.org/10.1162/neco.1997.9.5.1015, https://doi.org/10.1162/neco.1997.9.5.1015
21. Lambrinos, D., Möller, R., Pfeifer, R., Wehner, R.: Landmark navigation without snapshots: the average landmark vector model. In: Elsner, N., Wehner, R. (eds.) Proceedings of the Neurobiology Conference on Göttingen, p. 30a. Georg Thieme Verlag (1998). www.cs.cmu.edu/~motionplanning/papers/sbp_papers/integrated2/lambrinos_landmark_vector.pdf
22. Le Möel, F., Wystrach, A.: Opponent processes in visual memories: a model of attraction and repulsion in navigating insects' mushroom bodies. PLOS Computational Biology **16**(2), e1007631 (2020). https://doi.org/10.1371/journal.pcbi.1007631, https://dx.plos.org/10.1371/journal.pcbi.1007631
23. Möller, R., Vardy, A.: Local visual homing by matched-filter descent in image distances. Biol. Cybern. **95**(5), 413–430 (2006). https://doi.org/10.1007/s00422-006-0095-3, http://link.springer.com/10.1007/s00422-006-0095-3
24. Paredes-Valles, F., Scheper, K.Y.W., De Croon, G.C.H.E.: Unsupervised learning of a hierarchical spiking neural network for optical flow estimation: from events to global motion perception. IEEE Trans. Pattern Anal. Mach. Intell. **8828**(c), 1–1 (2019). https://doi.org/10.1109/TPAMI.2019.2903179, https://ieeexplore.ieee.org/document/8660483/
25. Sanes, J.R., Zipursky, S.L.: Design Principles of Insect and Vertebrate Visual Systems. Neuron **66**(1), 15–36 (2010). https://doi.org/10.1016/j.neuron.2010.01.018, https://linkinghub.elsevier.com/retrieve/pii/S0896627310000449
26. Scheper, K.Y., Karasek, M., De Wagter, C., Remes, B.D., De Croon, G.C.: First autonomous multi-room exploration with an insect-inspired flapping wing vehicle. In: 2018 IEEE International Conference on Robotics and Automation (ICRA), pp. 5546–5552. IEEE (2018). https://doi.org/10.1109/ICRA.2018.8460702, https://ieeexplore.ieee.org/document/8460702/

27. Stein, R.B.: A theoretical analysis of neuronal variability. Biophys. J. **5**(2), 173–194 (1965). https://doi.org/10.1016/S0006-3495(65)86709-1, https://www.sciencedirect.com/science/article/pii/S0006349565867091
28. Stone, T., Mangan, M., Wystrach, A., Webb, B.: Rotation invariant visual processing for spatial memory in insects. Interface Focus **8**(4), 20180010 (2018). https://doi.org/10.1098/rsfs.2018.0010, https://royalsocietypublishing.org/doi/10.1098/rsfs.2018.0010
29. Stürzl, W., Mallot, H.: Efficient visual homing based on Fourier transformed panoramic images. Robot. Auton. Syst. **54**(4), 300–313 (2006). https://doi.org/10.1016/j.robot.2005.12.001, https://linkinghub.elsevier.com/retrieve/pii/S0921889005002113
30. Sun, X., Yue, S., Mangan, M.: A decentralised neural model explaining optimal integration of navigational strategies in insects. eLife **9** (2020). https://doi.org/10.7554/eLife.54026, https://elifesciences.org/articles/54026
31. de Tournemire, P., Nitti, D., Perot, E., Migliore, D., Sironi, A.: A large scale event-based detection dataset for automotive. arXiv preprint (2020). http://arxiv.org/abs/2001.08499
32. Vogt, K., et al.: Direct neural pathways convey distinct visual information to Drosophila mushroom bodies. eLife **5**(APRIL2016), 1–13 (2016). https://doi.org/10.7554/eLife.14009, https://elifesciences.org/articles/14009
33. Wystrach, A., Beugnon, G., Cheng, K.: Ants might use different view-matching strategies on and off the route. J. Exp. Biol. **215**(1), 44–55 (2012). https://doi.org/10.1242/jeb.059584, https://doi.org/10.1242/jeb.059584
34. Zeil, J., Hofmann, M.I., Chahl, J.S.: Catchment areas of panoramic snapshots in outdoor scenes. J. Opt. Soc. Am. A **20**(3), 450 (2003). https://doi.org/10.1364/JOSAA.20.000450, https://www.osapublishing.org/abstract.cfm?URI=josaa-20-3-450
35. Zhu, A.Z., Thakur, D., Ozaslan, T., Pfrommer, B., Kumar, V., Daniilidis, K.: The multivehicle stereo event camera dataset: an event camera dataset for 3D perception. IEEE Robot. Autom. Lett. **3**(3), 2032–2039 (2018). https://doi.org/10.1109/LRA.2018.2800793, http://ieeexplore.ieee.org/document/8288670/
36. Zhu, L., Mangan, M., Webb, B.: Spatio-temporal memory for navigation in a mushroom body model. In: Living Machines 2020. LNCS (LNAI), vol. 12413, pp. 415–426. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-64313-3_39 https://link.springer.com/10.1007/978-3-030-64313-3_39