A Bio-Inspired and Solely Vision-Based Model for Autonomous Navigation

Tingtao Chen[1,2], Xuelong Sun[*,1,2], Qinbing Fu[1,2], Ziyan Qin[1,2] and Jigen Peng[*,1,2]
[1] School of Mathematics and Information Science, Guangzhou University, China
[2] Machine Life and Intelligence Research Centre, Guangzhou University, China

2024 International Joint Conference on Neural Networks (IJCNN) | 979-8-3503-5931-2/24/$31.00 ©2024 IEEE | DOI: 10.1109/IJCNN60899.2024.10651033

***Abstract*—Vision, utilizing eyes or cameras as the predominant sensory input, emerges as the primary and informative perception source for both animals and robots, facilitating crucial functions such as perception, navigation, interaction, and comprehensive understanding of their surroundings. Inspired by insect-like invertebrate animals' remarkable visual processing abilities which are achieved despite their constrained computational resources, this paper delves into the realm of bio-inspired autonomous navigation. The objective is to tackle the substantial challenges of cost efficiency, ensuring a solely vision-based algorithm guides the agent to its destination without collisions. To this end, a vision-guided navigation model is proposed by orchestrating the neural model of ant's visual navigation and crab's looming spatial localization. As the exploratory endeavor in constructing a bio-inspired autonomous visual navigation model, which demands significantly lower computational resources in comparison to prevailing engineering solutions. The effectiveness of this model has been validated through rigorous systematic experiments, lending empirical support to its capabilities. The proposed vision-motion closed-loop framework imparts valuable insights for the development of more efficient and precise autonomous systems, embodying the principle of drawing inspiration from nature's wisdom.**

***Index Terms*—vision-guided, insect navigation, collision detection, autonomous navigation.**

I. Introduction

Vision is of utmost importance for both living creatures and intelligent robots, serving as the primary sense for an agent to understand and interact with the environment [1]–[3]. This capability facilitates essential activities, for instance, motion perception and navigation. Researchers have devoted significant effort to vision-guided navigation methods, providing intelligent robots with collision-free navigation tours through simultaneous localization and mapping (SLAM) [4]–[6]. However, most of these method are computational expensive, thereby limiting their efficacy when applied to micro mobile robots or other agents with limited computation and memory resources [7].

In nature, insect-like animals exhibit remarkable capabilities in capturing and processing visual cues. This enables them to navigate, forage, and avoid predators, showcasing efficient navigation ability despite their small brains and limited computational resources [8], [9]. Inspired by the impressive navigation and visual processing abilities observed in these animals, previous work by Sun and his colleagues [10] has implemented a closed-loop collision-free navigation method based on bee-inspired path integration model and locust-inspired collision perception model. The model takes prior knowledge of speed and heading direction as input for the path integration model, while utilizing the common camera-received image as input for the collision perception model. Yet, achieving a bio-inspired navigation system that relies solely on visual input remains a challenge. Moreover, the model presented in [10] demonstrates sub-optimal performance in complex environments with the presence of various obstacles.

To address the above challenges, in this paper, we introduce a bio-inspired and solely vision-based navigation model that can navigate accurately and safely to a target location under the complex environment. Three key sub-subsystems are combined in the proposed model.

The first sub-subsystem is established as a global control system, responsible for determining a low-cost path to the designated target location. A vision-guided control system is a suitable choice, supported by ample evidence indicating that insects and other animals utilize visual representations from memory to determine target positions [11]. For instance, bees, wasps, ants, and hoverflies all rely on visual cues to locate their nests and food sources [12], [13]. Recent neurological data [14], [15] and neural models [10], [16], [17] also significantly contribute to understanding the neural basis and computational requirements of vision-based navigation behaviors in insects [14]. Visual homing (VH) is one of those navigation behaviors that offers significant opportunities for purely vision-guided autonomous navigation. VH denotes the behavior of insects returning directly to a familiar environment after travelling, and its core mechanism involves the discharge rate of the mushroom bodies (MB, brain region related to memory), showing a gradient that increases monotonically with the distance from the familiar area. This mechanism provides a homing signal by processing visual cues and memory, irrespective of the animal's orientation [16], [17].

The second sub-subsystem builds up a local control system responsible for real-time collision detection and avoidance, enabling the agent to navigate safely [18]. Owing to their remarkable visual information processing ability, insect-like animals exhibit sensitivity to looming stimuli, enabling them to escape from approaching predators or threats [19]–[22]. Neuroscientists have discovered specific visual neurons in the brains of these animals that are specialized in detecting motion cues. Examples include the Lobula Giant Movement Detectors (LGMDs) in locusts [19], [20], [23], the Lobula Plate Tangential Cells (LPTCs) in flies [21], [22], [24], and the Monostratified Lobula Giant type1 (MLG1) neuron in

crabs [25]–[28]. The MLG1 neuron exhibits more precise modulation of spatial position [29] compare to the above mentioned visual neurons. Considering the finding that crabs have a large visual field, a MLG1-inspired collision detection model is employed [30] with panoramic images, providing an parsimony method that balances computational consumption with highly accurate spatial positioning of looming stimuli. Thereby making it an ideal model for the collision detection and avoidance during navigation.

The third sub-subsystem coordinates the global control system VH and the local collision detection system MLG1, ensuring target-directed navigation with real-time collision detection and avoidance abilities. Just as Sun and his colleagues [10] uses global working memory to drive the agent to the target and uses local LGMD detection for collision avoidance, the simulation produces satisfactory results, verifying the integrability of global and local information. From a biological basis, insects adopt different behavioural strategies during target orientation depending on whether the stimulus represents a potential threat or in a goal-oriented approach. If a predator is present, they immediately perform a localised evasive action, then reorientate to the target and correct [31], [32]. To this end, two insect-inspired models, the VH model and the MLG1 model, which implement global goal-oriented navigation and local collision detection respectively, should be appropriately integrated in the proposed navigation control system.

To assess the effectiveness and efficiency of the proposed model, a series of simulated experiments were conducted. The positive outcomes unequivocally demonstrate the efficacy of our autonomous model, emphasizing its primary contribution: **the successful integration of bio-inspired path planning and collision avoidance within a vision-based framework**. In a simulated environment, this model effectively completes navigation tasks, benefiting from the seamless collaboration between global and local visual information.

In summary, our proposed model facilitates navigation tasks through a vision-motor closed-loop approach, paving the way for novel perspectives in the development of future autonomous robots and navigation systems. Serving as a foundational element of a biological navigation model, it exemplifies the principle of learning from nature and sheds light on the intelligent utilization of perceptual cues by animals to make informed decisions [33], [34].

The rest of this paper is structured as follows: Section II describes the developed 3D visual world, the defined navigation task and elaborates the proposed visual navigation model. The experimental process and results are illustrated in Section III. Finally, conclusion and future research directions are discussed in Section IV.

II. Methods

To establish a visual sensory-motor closed loop, we construct a *simulated 3D visual world*. Within this world, the agent is guided by the *proposed navigation model* to explore and successfully accomplish the *defined navigation task*. In the subsequent subsections, we will provide a detailed description of these components.

A. Simulated 3D world and navigation task

The Python programming language, along with packages such as *open3d*, *pyrender*, and *trimesh*, was utilized to construct the 3D world used for simulation. The agent's exploration area is confined to a rectangular region within the simulated world. Visible obstacles are represented as cuboids with dimensions of $[0.08, 0.08, 0.25]$ (see Fig. 1B). The agent is equipped with a panoramic camera that captures reconstructed view images as it moves in the visual world (example images shown in Fig. 1C and Fig. 2A). The navigation task involves starting from an arbitrary position (start point) and navigating to a predefined destination (typically located at $[x = 0, y = 0]$ in our simulations), while avoiding obstacles along the way. It is important to note that prior to commencing navigation, the agent undergoes an artificial training process (see details in II-B for more information).

B. Bio-inspired visual navigation model

As a solely vision-based navigation model, the agent relies solely on visual input for sensing. The reconstructed panoramic image is simultaneously fed into the *visual homing* and *collision detection* subsystems (as shown in Fig. 1A). The specific neural processing involved in these subsystems will be described in detail in the following subsections.

1) Ant visual homing model: In a recent study by Sun et al., the authors demonstrated the effectiveness of using Zernike Moments (ZM) encoding to achieve rotational invariance in the encoding of binary panoramic skyline images [17]. This encoding method transforms the images into frequency components, providing a concise representation that remains consistent across different headings. Zernike moments are mathematically defined as the projections of a function onto an orthogonal basis of polynomials known as Zernike polynomials [35], [36]. The Zernike moment is obtained by mapping the image $L^{x,y}$ onto these orthogonal polynomials $V_{nm}^{\rho,\theta}$, $L^{x,y}(t)$ is the vision input of the agent at time t, the integral can be approximated by a sum:

$$Z_{nm}(\rho, t) = \frac{n+1}{\pi} \sum_x \sum_y L^{x,y}(t) \cdot V_{nm}^{\rho,\theta *}(t), x^2 + y^2 \leq 1 \tag{1}$$

* indicates complex conjugate. For views with different headings at the same position, the amplitudes $A = |Z|$ of the ZM coefficients of these two images satisfy the following relationship:

$$|Z_{nm}^{\theta_r}| = |Z_{nm} e^{-jm\theta_r}| = |Z_{nm}| \; i.e., A_{nm}^{\theta_r} = A_{nm} \tag{2}$$

Therefore, the amplitude of the Zernike Moments coefficients remains invariant even the image is rotated, which is a crucial characteristic for conveying visual information. The amplitude coefficient is extracted and then input into the visual homing model, where it plays a significant role in determining the agent's motion.

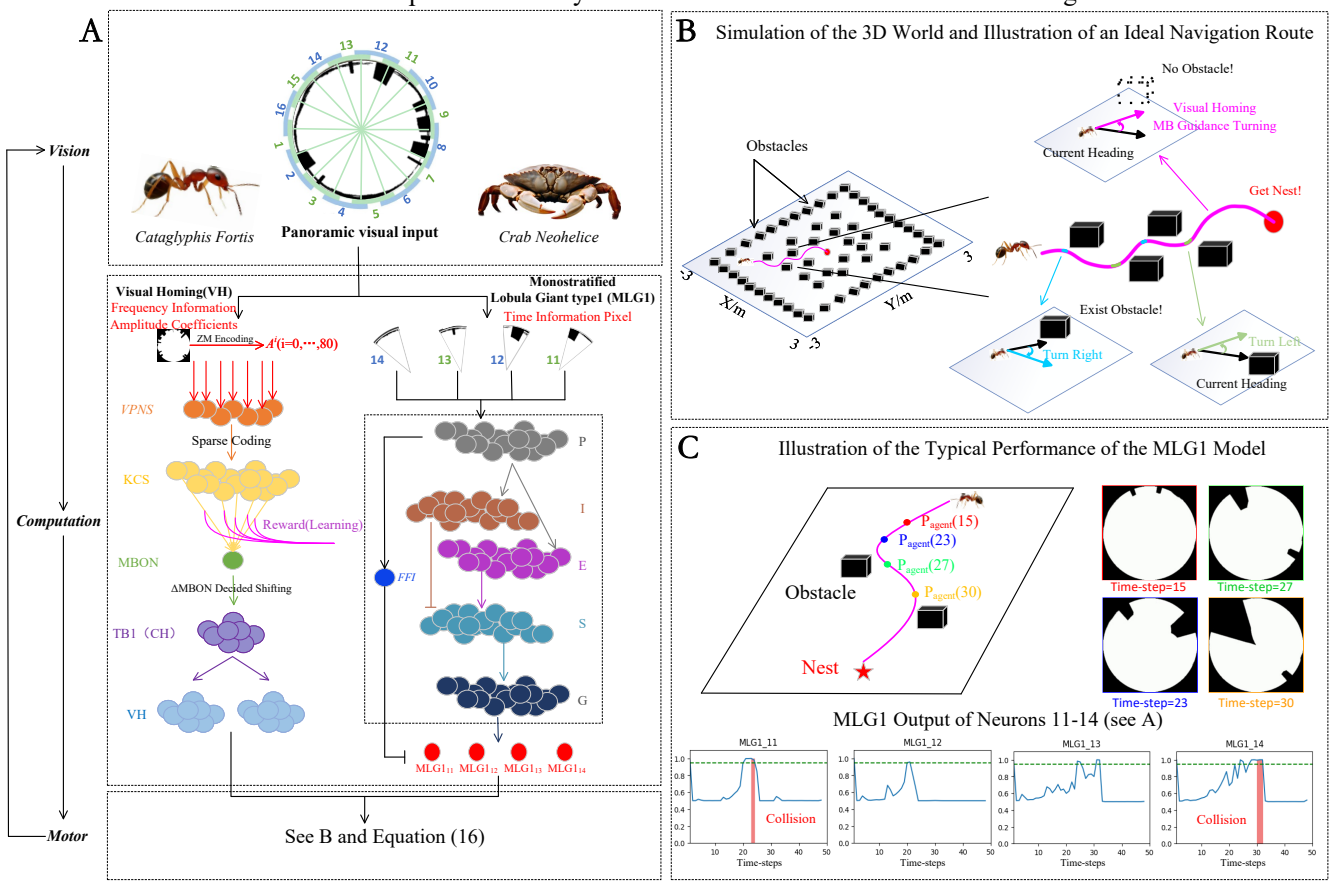

Fig. 1. Overview of the Proposed Model in a Vision-Motor Closed-Loop. (A) Integration of the ant's Visual Homing (VH) model and the crab's Looming Detection model at the motor level. The panoramic image input is simultaneously processed by these two subsystems. The VH model performs frequency encoding of the image and then transmits the resulting amplitude to the neural circuit within the mushroom body (MB), which generates a valence signal based on the current stimuli. The panoramic image is divided into 16 fields of view by the MLG1 neurons, with each neuron having a field of view of $37.5°$. The visual fields of adjacent MLG1 neurons overlap by $15°$, and each MLG1 neuron covers a distinct visual range of $7.5°$. The blue and green arcs represent individual MLG1 neurons. (B) Simulation of the 3D World and Illustration of an Ideal Navigation Route. The left figure presents a 3D representation of the world, demonstrating successful navigation of the agent to the nest without any collisions. On the right figure, different strategies are depicted based on the agent's obstacle detection. The blue segment of the path indicates detection of an obstacle on the left, prompting the agent to make a right turn. The green segment represents obstacle detection on the right, resulting in a left turn by the agent. In the absence of any encountered obstacles, the agent follows guidance inferred from the mushroom body (MB). (C) Illustration of the Typical Performance of the MLG1 Model. The model receives visual inputs from the agent's visual sensor and accumulates activation in each ($MLG1_i$) neuron (i=11, 12, 13, 14) using the sigmoid function. When the accumulation exceeds a certain threshold for consecutive (N_p) time-steps, it signifies that the (i^{th}) segment of the visual field is perceived as looming, triggering a collision response (depicted by the vertical red line).

The visual homing neural network is is designed as an associative network, mirroring the architecture of the insect's mushroom body (MB). In this network, visual projection neurons (vPNs) receive the magnitudes of the Zernike Moments (ZM) coefficients as their firing rates:

$$C_{vpn}^{i,t} = A^{i,t}, i = 0, 1, 2, \ldots, N_{vpn} \tag{3}$$

where N_{vpn} is the number of vPN neurons, which is the same as the total number of ZM magnitudes. vPNs project to Kenyon cells (KCs) through randomly generated binary connections W_{v2k}, causing each KC to receive the activation of 10 randomly selected vPNs' activation:

$$I_{kc}^{j,t} = \sum_{i=0}^{N_{vpn}} W_{v2k}^{ji} C_{vpn}^{i,t} \tag{4}$$

where I_{kc}^{j} represents the total input current received by the j^{th} KC from the vPNs, and KCs are modeled as binary neurons with the same threshold Thr_{kc}. The MBON neurons sum the activation of all Kenyon cells through the plastic connection W_{k2m}:

$$C_{mbon}^{t} = \sum_{i=0}^{N_{kc}} W_{k2m}^{i,t} C_{kc}^{i,t} \tag{5}$$

The plasticity of W_{k2m} is processed using an anti-Hebbian learning rule:

$$W_{k2m}^{i,t} = W_{k2m}^{i,t-1} - \eta_{k2m} \quad if\ C_{kc}^{i,t-1} = 1 \tag{6}$$

The learning process is controlled by the learning rate η_{k2m}, which depends on the activation of the reward signal. The

activation of EN C_{mbon} reflects the familiarity of the current view. The change in C_{mbon} is defined as:

$$\Delta C_{mbon} = C^{t}_{mbon} - C^{t-1}_{mbon} \tag{7}$$

ΔC_{mbon} is used to track the gradient of familiarity and guide the agent to more familiar locations by adjusting the current heading C_{ch} (see supplemental Video 1 or Video 2).

$$C^{i,t}_{vh} = C^{j,t}_{ch}, j = \begin{cases} i + offset & if\ i + offset \leq 7 \\ i + offset - 7 & otherwise \end{cases} \tag{8}$$

where $i = 0, 1, \ldots, 7$. The relationship between ΔC_{mbon} and the offset is shown as follows.

$$offset = \begin{cases} 0 & if\ \Delta C_{mbon} \leq 0 \\ min([k_{vh}\Delta C_{mbon}], 4) & otherwise \end{cases} \tag{9}$$

The angle of rotation is determined by the difference between the sum of the activation of the left and the sum of the activation of the right neurons:

$$\theta^{t}_{vh} = k_{motor}\left[\sum_{i=0}^{7}(C^{i,t}_{vh} - C^{i,t}_{ch}) - \sum_{i=8}^{15}(C^{i,t}_{vh} - C^{i,t}_{ch})\right] \tag{10}$$

2) Crab Neohelice collision detection model: In our neural network, unlike natural MLG1 neurons, we divide the panoramic image into 16 equal sections, with each section covering a field of view of 37.5°. The perception field of 15° is shared with adjacent segments, while the central 7.5° perception field is exclusive to each segment. This image partitioning approach is illustrated in Fig. 1A and can be further visualized in supplemental video 1. Each green and blue arc in Fig. 1(A) represents an MLG1 neuron encoding a specific looming spatial location.

The visual input $L^{x,y}(t)$ is fed into the four-layer structure P, I, E, S as described in [30]. The output is obtained through an additional artificial layer G. In this MLG1-based neural network, feature enhancement is critical to extract collision targets in complex backgrounds. This is achieved by encoding expanded edges through grouping excitation. The membrane potential $MLG1(t)$ of the MLG1 cell is determined by summing the rectified outputs from each pixel in the grouping layer (indicated by circles in Section V:Video 1 or Video 2, with darker colors indicating higher output). The resulting potential is then activated using the sigmoid function, expressed as follows:

$$MLG1(t) = (1 + e^{-C_l \sum_i \sum_j abs(G^{x,y}(t))N^{-1}_{cell}})^{-1} \tag{11}$$

C_l is a balancing factor, N_{cell} is the number of neurons in layer G, which corresponds to the size of the input image. A feed-forward inhibition (FFI) mechanism is incorporated to adapt the threshold in response to sudden changes in the entire visual receptive field. j^{th} represents the segment number of each MLG1 neuron. If the number of $MLG1^{spike}_{j}$ reaches continuous excitation above set time steps, the final collision warning signal C^{mlg1}_{j} will be generated:

$$C_j(t) = \begin{cases} 1 & if\ \sum_{i=t-N_{sp}}^{t} MLG1^{spike}_{j}(t) \geq N_{sp} \\ 0 & otherwise \end{cases} \tag{12}$$

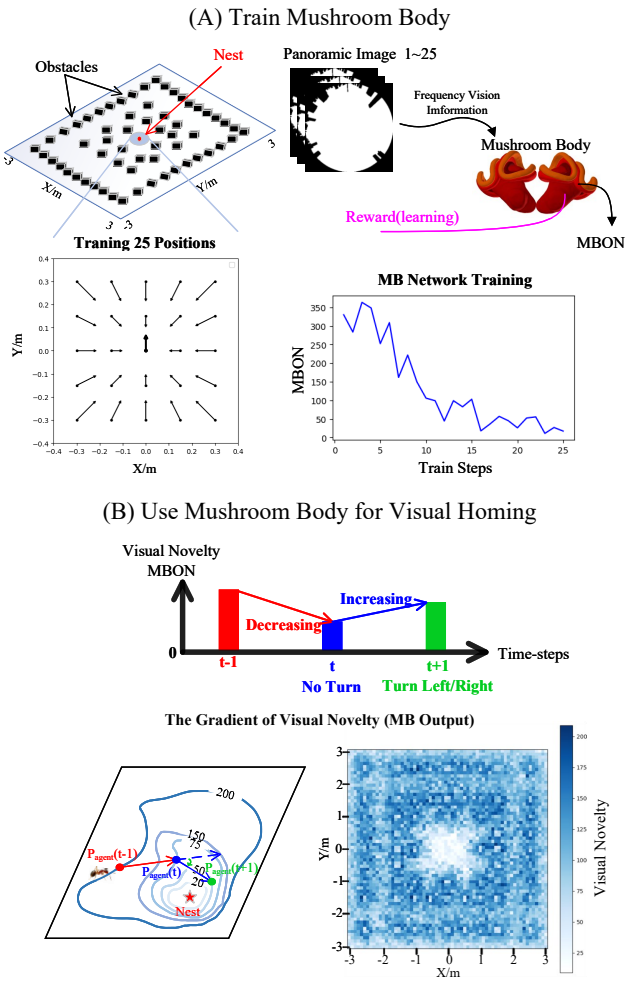

Fig. 2. Training Mushroom Body and Visual Navigation. (A) Training Mushroom Body: Frequency information from panoramic images of the nest and nearby views is fed into the MB reward signal for learning, resulting in decreased MBON activation over training steps. (B) Visual Homing with Mushroom Body: Agent calculates visual novelty based on MB output for its current position. Decreased novelty maintains current heading, while increased novelty initiates turning action. Visual homing model generates a novelty gradient field. Heatmap shows agent's MB activity area output.

where N_{sp} represents the quantity of continuous spikes and time-steps. Note that the collision here refers to the potential risk of collision detected by MLG1, not the actual collision. As shown in Fig. 1C, supplemental Video 1 and Video 2, the blue line represents the output value of $MLG1_i(i = 11, \ldots, 14)$ at time t, and t filled with red represents that the threshold has been exceeded for N_{sp} consecutive moments. The fusion of the Visual Homing (VH) and MLG1-based models represents the integration of global visual memory and local collision detection. The schematic mechanism is illustrated in Fig. 1B.

3) Navigation model: At time t, the state of the navigating agent is denoted as $S(t) = [P_{agent}(t), V_{agent}(t), \theta(t)]$, where $P_{agent}(t) = [x(t), y(t), z(t)]$ and $V_{agent}(t) = [v_x(t), v_y(t), v_z(t)]$. In this study, the agent's movement is confined to the xy-plane, resulting in $z = 0.01$ and $v_z = 0$. Thus, the navigation problem simplifies to determining $V_{agent}(t)$, and the agent's position can be updated accordingly. $V_{agent}(t)$

TABLE I
THE MAIN PARAMETERS OF THE PROPOSED AUTONOMOUS NAVIGATION MODEL.

Parameters	Simulated world
N_{vpn}	80
N_{kc}	4000
Thr_{kc}	0.04
η_{k2m}	0.1
k_{vh}	0.3
k_{motor}	0.25
C_l	0.32
$N_{cell} = n_{row} \times n_{colum}$	$90000 = 300 \times 300$
N_{sp}	3
k	2.25
θ_{MLG1}	$\pi/9$
$\|V(t)\|$	0.05

is configured to move forward at a constant speed in the current heading.

$$P_{agent}(t+1) = P_{agent}(t) + V_{agent}(t+1) \tag{13}$$

$$V_{agent}(t+1) = |V_{agent}(t)|[cos\theta(t+1), sin\theta(t+1)] \tag{14}$$

$\theta(t)$ represents the agent's current heading, which is updated as:

$$\theta(t+1) = \theta(t) + \Delta\theta(t) \tag{15}$$

where $\Delta\theta(t)$ denotes the rotation angle. Collision avoidance is determined based on the MLG1 response (see Fig. 1C); otherwise, visual homing is executed using the global visual memory. In the anti-collision motion control strategy, in the simulation environment, if C_{11}, C_{12}, C_{13}, and C_{14} of the MLG1 network detect a local collision, the agent will initiate angle offsets in the corresponding directions to evade obstacles first. According to the aforementioned model, $\Delta\theta$ is expressed as:

$$\Delta\theta(t) = \begin{cases} k\theta_{mlg1} & if\ C_{11}, C_{12}, C_{13}, C_{14} = 1 \\ \theta_{mlg1} & if\ C_{12}, C_{13} = 1, MLG1_{12} \geq MLG1_{13} \\ -\theta_{mlg1} & if\ C_{12}, C_{13} = 1, MLG1_{12} < MLG1_{13} \\ \theta_{mlg1} & if\ C_{11} + C_{12} > C_{13} + C_{14} \\ -\theta_{mlg1} & if\ C_{11} + C_{12} < C_{13} + C_{14} \\ \theta_{vh}^{t} & otherwise \end{cases} \tag{16}$$

In this study, it is important to highlight that the MLG1 model is capable of detecting potential collision risks before an actual collision occurs. This is achieved by predicting a collision when the MLG1 response reaches its peak, which happens prior to the actual collision time. The key parameters used in the bio-inspired and solely vision-based model for autonomous navigation are summarized in Table I.

III. EXPERIMENTS

A. Evaluation Metrics

In our endeavor to conduct a thorough evaluation of our model's performance, we have identified two crucial metrics. Firstly, we have defined "time-step consumption" as the duration required for the agent to complete the navigation task. Given that the agent moves at a constant speed, distance and time-step consumption demonstrate consistent performance. Let ($T_{consumption}$) represent the time-step consumption.

To assess the effectiveness of collision detection using (MLG1), we employ the metric $N_{collision}$ which represents the number of collisions that occur during navigation. This metric is determined by:

$$N_{collision} = \sum_{t=0}^{T} M_{collision}(t) \tag{17}$$

$$M_{collision}(t) = \begin{cases} 1 & if\ P_{agent}(t) \in \cup_{i=0}^{N_{obs}} \Omega_i \\ 0 & otherwise \end{cases} \tag{18}$$

where $M_{\text{collision}}(t)$ denotes whether a collision occurs between the agent and an obstacle at time t. Ω_i represents the spatial occupancy of the obstacles. These two metrics, $T_{consumption}$ and $N_{collision}$, re utilized to evaluate the performance of completing navigation tasks within the simulated environment. It is important to note that all agents in the simulation are configured to encounter non-physical collisions with obstacles. This implies that the agents have not yet successfully avoided collisions, but the simulation continues without causing any physical damage. Each experiment in the subsequent scenarios is conducted 8 times, with the average of $N_{collision}$ and $T_{consumption}$ recorded.

B. Simulation Results

To thoroughly evaluate the performance of our proposed model, we conducted simulation experiments to validate its effectiveness in completing a defined navigation task. Our model, as illustrated in Fig. 1, successfully navigated the environment, showcasing its superior proficiency compared to the baseline model that employed random walking. With a specific emphasis on key metrics, our evaluation aimed to compare the impact of the proposed visual homing model on $T_{consumption}$ and the effectiveness of the MLG1-based collision detection model in reducing $N_{collision}$. The comparison was conducted under identical initial conditions. The accompanying box plots in Fig. 3 clearly demonstrate that our proposed model significantly decreases both $T_{consumption}$ and $N_{collision}$ during the navigation task.

To investigate the ability of our model to complete the navigation task under various initial headings, we initiated a new experiment on the system. The results revealed that irrespective of the initial heading, the agent demonstrated superior performance in reaching the nest, provided that an adequate number of time-steps were consumed. As depicted in Fig. 4A, the path represented by the red solid line indicates the trajectory with fewer ($N_{collision}$) in the simulation experiment, while the blue solid line represents the path with the best experimental result in terms of reduced ($T_{consumption}$) in the simulation experiment.

In order to further demonstrate the robustness of our model, we considered varying the start positions to illustrate the ability of completing navigation tasks. As demonstrated in Fig. 4B, the proposed system effectively navigates back to

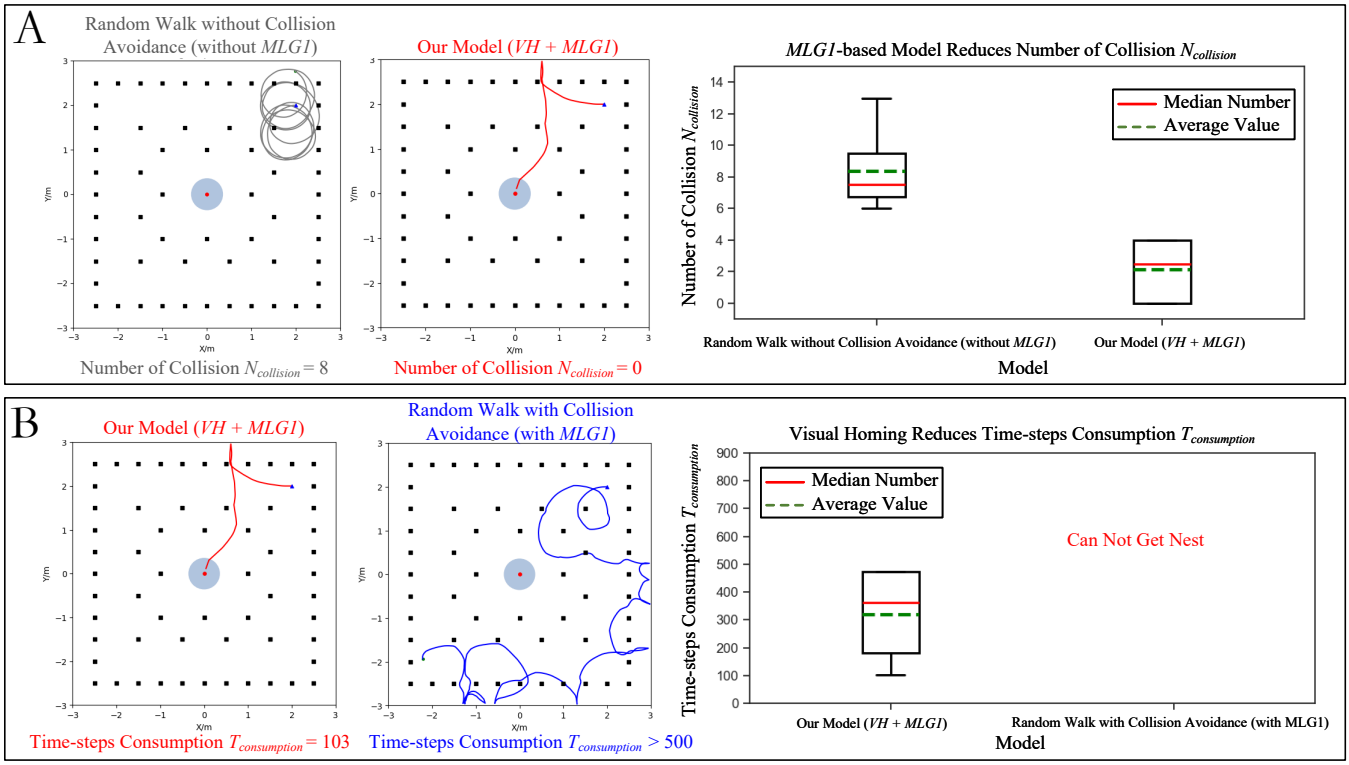

Fig. 3. (A) MLG1-based collision detection model reduces $N_{collision}$. Comparative experiments between our model (VH + MLG1) and a random walk without collision avoidance (without MLG1) with the same initial state of navigation. Box plots show results from multiple experiments, with the solid red line representing the median and the dashed green line representing the mean. (B) Visual homing significantly reduces $T_{consumption}$ for autonomous navigation. Comparative experiments between our model (VH + MLG1) and random walk with collision avoidance (with MLG1) assessed the time-steps consumption required to complete the navigation task.

the nest from different start positions, exhibiting reduced ($T_{consumption}$) and remarkable collision avoidance ability. More detailed experimental results are presented in Table II. Due to the different initial headings and positions, the agent may face towards or away from the nest, resulting in variations in the number of time-steps consumed.

To further examine the adaptability and robustness of our model, we pose an intriguing question: how does the model respond when the density of obstacles within the agent's fixed activity area is altered? By varying the number of obstacles and thus adjusting the obstacle density, we aim to delve into the behavior of our model in the face of changing environmental conditions.

As depicted in Fig. 4C (i-iii), no distinct trend is observed regarding the number of collisions and time consumption as the obstacle density increases. This finding highlights the robustness of the proposed model, as it demonstrates its ability to handle various levels of obstacle density without significant changes in performance. However, the unexpected finding in Fig. 4C (iv) shows that the best performance was achieved in the environment with 76 obstacles. This result suggests that when there are an ample number of visual cues available, the learning efficiency of MB may be enhanced. While this hypothesis warrants further investigation, it falls beyond the scope of the present study.

In summary, these findings highlight the versatility of our model, showcasing its effectiveness and robustness across a range of start positions, initial headings, and obstacle densities.

TABLE II
THE EXPERIENCE RESULTS OF BIO-INSPIRED AND SOLELY VISION-BASED MODEL FOR AUTONOMOUS NAVIGATION.

experiment	obstacle number	start position	initial heading	$N_{collision}$	$T_{consumption}$
Fig.4A	60	$(2, 2)$	$0°$	3.125	334
			$90°$	1.375	195.625
			$180°$	2.125	319.75
			$-90°$	1.875	227.625
Fig.4B	60	$(2, 2)$	$90°$	1.375	195.625
		$(2, -2)$		2.375	289.5
		$(-2, 2)$		2.25	211.625
		$(-2, -2)$		1.5	188.38
Fig.4C	48	$(2, 2)$	$-90°$	1.25	169.875
	60			1.875	227.625
	68			3.25	288.75
	76			1.5	177

IV. Conclusion and Discussions

This paper introduces a bio-inspired and solely vision-based autonomous navigation model that integrates ant-inspired visual homing as a global visual memory and crab-inspired MLG1 model as a local collision cue. The results obtained from simulation experiments demonstrate several key findings: 1) The proposed bio-inspired model enables the accomplishment of autonomous navigation using vision as the sole sensory input. 2) The successful integration of both global visual memory, obtained through visual homing, and local collision detection using the MLG1 model, improves navigation perfor-

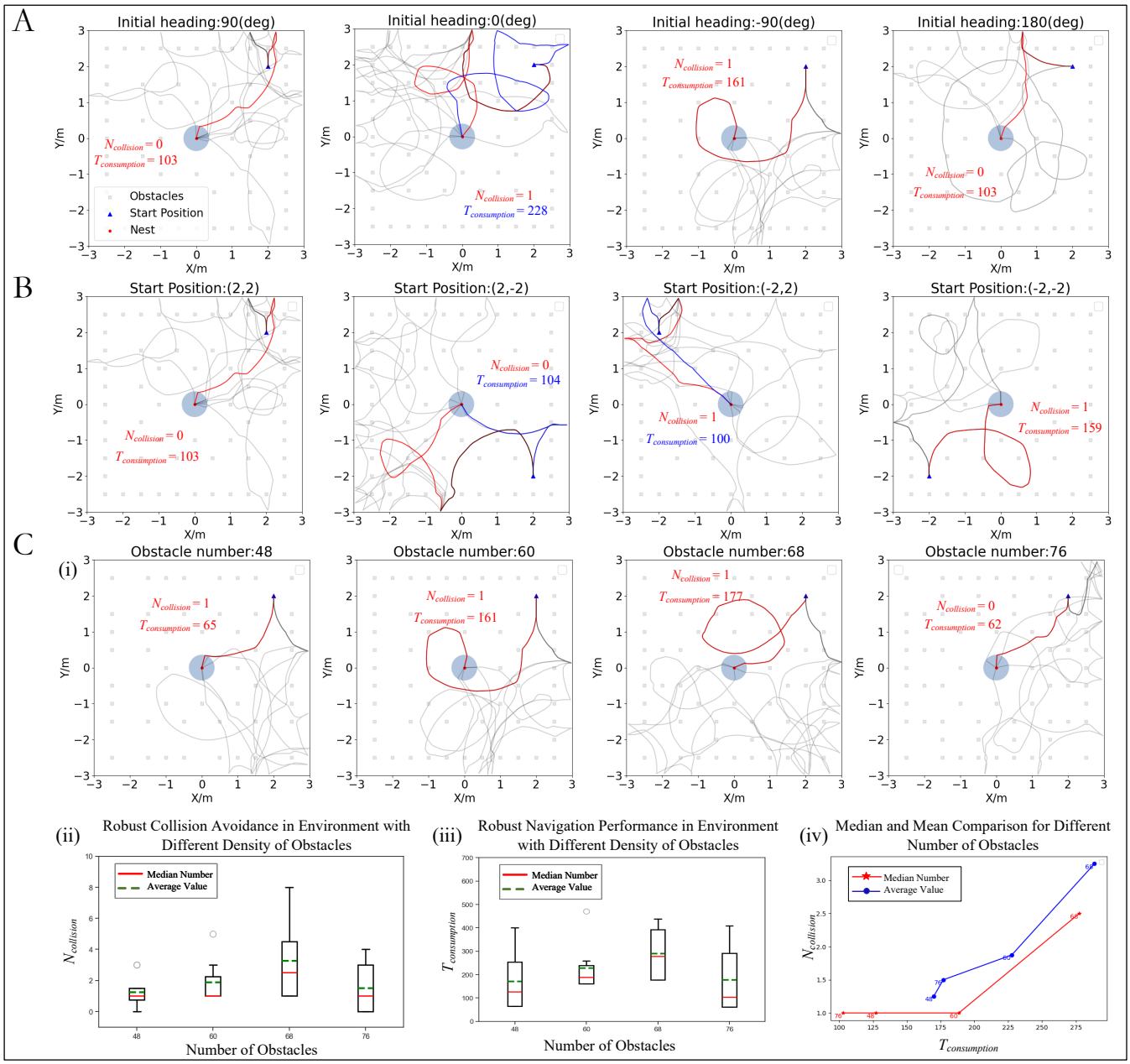

Fig. 4. Collision Avoidance and Navigation Performance in Variable Environments. (A) Navigation experiments ($N = 8$) with the agent starting from the same position but different initial headings (from left to right: 90°, 0°, −90°, 180°). The solid red line represents the optimal path with fewer $N_{collision}$. while the blue solid line represents the optimal path with less $T_{consumption}$. The corresponding $N_{collision}$ and $T_{consumption}$ are labelled. (B) Navigation experiment ($N = 8$) with the agent starting from different positions (from left to right: [2,2], [2,-2], [-2,2], [-2,- 2]) but with the same initial heading. (C) Navigation experiment ($N = 8$) with the agent starting from the same position and heading but in environments with varying obstacle densities. (i) Trajectories of the navigating agents in environments with different obstacle numbers (48, 60, 68, 76). Boxplots illustrate the distribution of $N_{collision}$ (ii) and $T_{consumption}$ (iii) under different obstacle densities. (iv) Correlation between $N_{collision}$ and $T_{consumption}$, represented by the solid red line (median) and the solid blue line (mean). The numbers labelled on the scatter plot indicate the number of obstacles.

mance and effectively reduces the occurrences of collisions. 3) The robustness of the proposed model is validated in a sensory-motor closed loop, offering novel algorithms for robotics applications in environments with dense obstacles.

Indeed, this study faces a challenge that warrants attention and resolution. The global visual memory is learned under a static simulation environment, which raises concerns about its effectiveness in achieving visual homing in dynamic environments. In dynamic environments, the learned views may become invalid as visible obstacles move and alter the visual scene. This poses a significant challenge that requires further investigation to develop robust visual homing strategies in dynamic environments. Additionally, although our model results in relatively fewer occurrences of collisions, achieving 100% collision avoidance remains a challenge. Future work could explore optimization strategies and improvements,

such as incorporating models like Directional Selective Visual Neural Networks (DSN) [37], to enhance collision avoidance capabilities and further refine the proposed navigation model. Addressing these challenges can lead to advancements in the efficiency and effectiveness of the navigation model, allowing for better performance and increased collision avoidance ability.

V. Supporting Information

Video1 - Simulation of navigation starts from [2,2] with initial heading of $-90°$

Video2 - Simulation of navigation starts from [-2,-2] with initial heading of $0°$

Acknowledgement

This research has received funding from the National Natural Science Foundation of China under the Grant No.62206066, No.62376063, No.12031003 and No.12211540710, the China Postdoctoral Science Foundation under the Grant No 2022M710864, the Social Science Fund of the Ministry of Education of China under the Grant No 22YJCZH032 and the Innovation Research for the Postgraduates of Guangzhou University. *Corresponding author: Xuelong Sun (xsun@gzhu.edu.cn)and Jigen Peng (jgpeng@gzhu.edu.cn).

References

[1] E. R. Davies, *Machine vision: theory, algorithms, practicalities*. Elsevier, 2004.

[2] A. Borst, "Drosophila's View on Insect Vision," *Current Biology*, vol. 19, no. 1, pp. R36–R47, Jan. 2009.

[3] N. Franceschini, "Small Brains, Smart Machines: From Fly Vision to Robot Vision and Back Again," *Proceedings of the IEEE*, vol. 102, no. 5, pp. 751–781, May 2014.

[4] H. Durrant-Whyte and T. Bailey, "Simultaneous localization and mapping: part i," *IEEE robotics & automation magazine*, vol. 13, no. 2, pp. 99–110, 2006.

[5] T. Bailey and H. Durrant-Whyte, "Simultaneous localization and mapping (slam): Part ii," *IEEE robotics & automation magazine*, vol. 13, no. 3, pp. 108–117, 2006.

[6] C. Stachniss, J. J. Leonard, and S. Thrun, "Simultaneous localization and mapping," *Springer handbook of robotics*, pp. 1153–1176, 2016.

[7] C. Hu, F. Arvin, C. Xiong, and S. Yue, "Bio-inspired embedded vision system for autonomous micro-robots: The lgmd case," *IEEE transactions on cognitive and developmental systems*, vol. 9, no. 3, pp. 241–254, 2016.

[8] J. Zeil, "Visual homing: an insect perspective," *Current opinion in neurobiology*, vol. 22, no. 2, pp. 285–293, 2012.

[9] C. A. Freas and M. L. Spetch, "Varieties of visual navigation in insects," *Animal cognition*, vol. 26, no. 1, pp. 319–342, 2023.

[10] X. Sun, Q. Fu, J. Peng, and S. Yue, "An insect-inspired model facilitating autonomous navigation by incorporating goal approaching and collision avoidance," *Neural Networks*, vol. 165, pp. 106–118.

[11] J. Zeil, N. Boeddeker, and W. Stürzl, "Visual homing in insects and robots," *Flying insects and robots*, pp. 87–100, 2010.

[12] B. Cartwright and T. S. Collett, "Landmark learning in bees: experiments and models," *Journal of comparative physiology*, vol. 151, pp. 521–543, 1983.

[13] T. S. Collett and M. Land, "Visual spatial memory in a hoverfly," *Journal of comparative physiology*, vol. 100, pp. 59–84, 1975.

[14] B. K. Hulse, H. Haberkern, R. Franconville, D. Turner-Evans, S.-y. Takemura, T. Wolff, M. Noorman, M. Dreher, C. Dan, R. Parekh *et al.*, "A connectome of the drosophila central complex reveals network motifs suitable for flexible navigation and context-dependent action selection," *Elife*, vol. 10, 2021.

[15] M. E. Sayre, R. Templin, J. Chavez, J. Kempenaers, and S. Heinze, "A projectome of the bumblebee central complex," *Elife*, vol. 10, p. e68911, 2021.

[16] P. Ardin, F. Peng, M. Mangan, K. Lagogiannis, and B. Webb, "Using an insect mushroom body circuit to encode route memory in complex natural environments," *PLoS computational biology*, vol. 12, no. 2, p. e1004683, 2016.

[17] X. Sun, S. Yue, and M. Mangan, "A decentralised neural model explaining optimal integration of navigational strategies in insects," *Elife*, vol. 9, p. e54026, 2020.

[18] H. Shin and J. Chae, "A performance review of collision-free path planning algorithms," *Electronics*, vol. 9, no. 2, p. 316, 2020.

[19] F. Gabbiani, H. G. Krapp, and G. Laurent, "Computation of object approach by a wide-field, motion-sensitive neuron," *Journal of Neuroscience*, vol. 19, no. 3, pp. 1122–1141, 1999.

[20] F. C. Rind and P. J. Simmons, "Seeing what is coming: building collision-sensitive neurones," *Trends in neurosciences*, vol. 22, no. 5, pp. 215–220, 1999.

[21] A. Borst and T. Euler, "Seeing things in motion: models, circuits, and mechanisms," *Neuron*, vol. 71, no. 6, pp. 974–994, 2011.

[22] A. Borst and M. Helmstaedter, "Common circuit design in fly and mammalian motion vision," *Nature neuroscience*, vol. 18, no. 8, pp. 1067–1076, 2015.

[23] F. C. Rind, S. Wernitznig, P. Pölt, A. Zankel, D. Gütl, J. Sztarker, and G. Leitinger, "Two identified looming detectors in the locust: ubiquitous lateral connections among their inputs contribute to selective responses to looming objects," *Scientific reports*, vol. 6, no. 1, p. 35525, 2016.

[24] H. Fotowat, A. Fayyazuddin, H. J. Bellen, and F. Gabbiani, "A novel neuronal pathway for visually guided escape in drosophila melanogaster," *Journal of neurophysiology*, vol. 102, no. 2, pp. 875–885, 2009.

[25] M. Berón de Astrada and D. Tomsic, "Physiology and morphology of visual movement detector neurons in a crab (decapoda: Brachyura)," *Journal of Comparative Physiology A*, vol. 188, pp. 539–551, 2002.

[26] J. Sztarker and D. Tomsic, "Binocular visual integration in the crustacean nervous system," *Journal of Comparative Physiology A*, vol. 190, pp. 951–962, 2004.

[27] J. Sztarker, N. J. Strausfeld, and D. Tomsic, "Organization of optic lobes that support motion detection in a semiterrestrial crab," *Journal of Comparative Neurology*, vol. 493, no. 3, pp. 396–411, 2005.

[28] D. Oliva, V. Medan, and D. Tomsic, "Escape behavior and neuronal responses to looming stimuli in the crab chasmagnathus granulatus (decapoda: Grapsidae)," *Journal of Experimental Biology*, vol. 210, no. 5, pp. 865–880, 2007.

[29] D. Oliva and D. Tomsic, "Visuo-motor transformations involved in the escape response to looming stimuli in the crab neohelice (= chasmagnathus) granulata," *Journal of Experimental Biology*, vol. 215, no. 19, pp. 3488–3500, 2012.

[30] H. Luan, Q. Fu, Y. Zhang, M. Hua, S. Chen, and S. Yue, "A looming spatial localization neural network inspired by mlg1 neurons in the crab neohelice," *Frontiers in neuroscience*, vol. 15, p. 787256, 2022.

[31] T. Bockhorst and U. Homberg, "Interaction of compass sensing and object-motion detection in the locust central complex," *Journal of Neurophysiology*, vol. 118, no. 1, pp. 496–506, 2017.

[32] R., Rosner, U., and Homberg, "Widespread sensitivity to looming stimuli and small moving objects in the central complex of an insect brain," *Journal of Neuroscience*, 2013.

[33] R. Pfeifer, M. Lungarella, and F. Iida, "Self-Organization, Embodiment, and Biologically Inspired Robotics," *Science*, vol. 318, no. 5853, pp. 1088–1093, Nov. 2007.

[34] A. Matsiko, "Taking inspiration from nature is a no-brainer," *Science Robotics*, vol. 8, no. 78, p. eadi2720, May 2023.

[35] M. R. Teague, "Image analysis via the general theory of moments," *Josa*, vol. 70, no. 8, pp. 920–930, 1980.

[36] A. Khotanzad and Y. H. Hong, "Invariant image recognition by zernike moments," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 12, no. 5, pp. 489–497, 1990.

[37] S. Yue and F. C. Rind, "Postsynaptic organisations of directional selective visual neural networks for collision detection," *Neurocomputing*, vol. 103, pp. 50–62, 2013.