Proceedings of the 2018 IEEE
International Conference on Robotics and Biomimetics
December 12-15, 2018, Kuala Lumpur, Malaysia

Performance of a Visual Fixation Model in an Autonomous Micro Robot Inspired by *Drosophila* Physiology

Qinbing Fu[1], Nicola Bellotto[1], Cheng Hu[1,2] and Shigang Yue[1]

***Abstract*— In nature, lightweight and low-powered insects are ideal model systems to study motion perception strategies. Understanding the underlying characteristics and functionality of insects' visual systems is not only attractive to neural system modellers but also critical in providing effective solutions to future robotics. This paper presents a novel modelling of dynamic vision system inspired by *Drosophila* physiology for mimicking fast motion tracking and a closed-loop behavioural response to fixation. The proposed model was realised on the embedded system in an autonomous micro robot which has limited computational resources. A monocular camera was applied as the only motion sensing modality. Systematic experiments including open-loop and closed-loop bio-robotic tests validated the proposed visual fixation model: the robot showed motion tracking and fixation behaviours similarly to insects; the image processing frequency can maintain $25 \sim 45$Hz. Arena tests also demonstrated a successful following behaviour aroused by fixation in navigation.**

I. INTRODUCTION

Motion tracking strategies can vary from traditional computer vision techniques like learning based or prediction schemes to computational biology theories. There have been many learning based methodologies showing good performance on motion detection and tracking, e.g. [1], [2]. Recently, an approach was proposed to learn real-time tracking, which could reach even 100 fps, with deep regression networks [3]. A new object-detection based fast tracking algorithm was presented in [4]. In addition, a monocular vision based solution was implemented to estimate multi-body motion and successfully tested from vehicle-mounted cameras [5]. On the aspect of biologically inspired methods, human-brain inspired attention-based models (e.g. [6]) and learning strategies (e.g. [7]) have been demonstrated robust performance outperforming similar models in the literature.

Balancing the model performance and efficiency still poses a big challenge toward an artificial dynamic vision system. These state-of-the-art methods can achieve significant improvements on motion tracking. They nevertheless are either computationally expensive, or heavily restricted to specialised hardware, like event-driven cameras.

In nature, motion vision is of great importance for animals and humans in every aspects of daily life. Lightweight insects, in particular, have a small number of visual neurons but can smartly navigate through unpredictable and cluttered environments. These energy-efficient neural mechanisms and circuits evolved through thousands of millions of years can provide researchers with many inspirations to explore approaches for motion perception in both a low-powered and reliable mode.

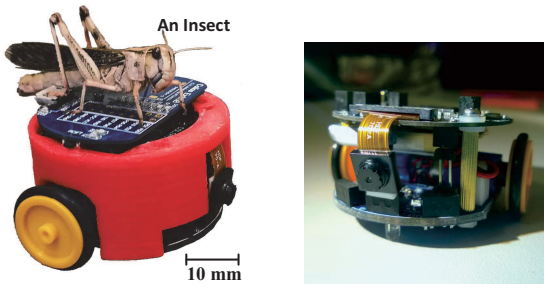


Fig. 1. Prototype of the utilised wheeled micro robot named '*Colias*' [8]: (a) a *Colias* robot with a 3D-printed shell and an insect standing on top to demonstrate its small size, (b) a frontal view of the applied camera.

From insects to robots, a few motion sensitive visual neurons and circuits have been computational modelled and successfully applied in both ground and flying robots. For instance, the looming sensitive neurons in locusts (e.g. [8]–[10]) and the translating sensitive neurons in flies (e.g. [11]) have been found as prominent model systems to study visual motion sensing strategies. Specifically for visual tracker, some bio-inspired models have shown convincing performance. These are inspired by dragonfly visual systems mainly sensing translating small targets [12], [13].

On the behavioural level, there has been significant evidence indicating different motion sensitive neurons or circuits arouses specific behaviours. In *Drosophila*. recent biological studies have demonstrated a correspondence between the preliminary visual systems and the sensorimotor response of 'visual fixation' [14], [15]. Such a visually-guided behaviour depicts insects track a moving object of interest and keep it near the centre of view along with 'turning response'. The visual tracking and fixation is crucial to insects for a variety of activities like foraging and chasing mates. However, it lacks systematic modelling and investigation in potential robotic applications. Here, we propose a novel visual fixation neural model on embedded system in an autonomous wheeled robot, as shown in Fig. 1. We demonstrate its effectiveness and robustness of guiding fast visual fixation via systematic bio-robotic experiments.

In the following sections, the proposed model will be introduced in Section II. The experimental set-up and results will be presented in Sections III and IV. We have further discussion in Section V and conclusion in Section VI.

This work was supported by the grants of EU Horizon 2020 project STEP2DYNA(691154).

[1] Lincoln Centre for Autonomous Systems (L-CAS), University of Lincoln, UK; [2] School of Mechanical and Electrical Engineering, Guangzhou University, China; Corresponding author email: qinbingfu87@gmail.com

978-1-7281-0377-8/18/$31.00 ©

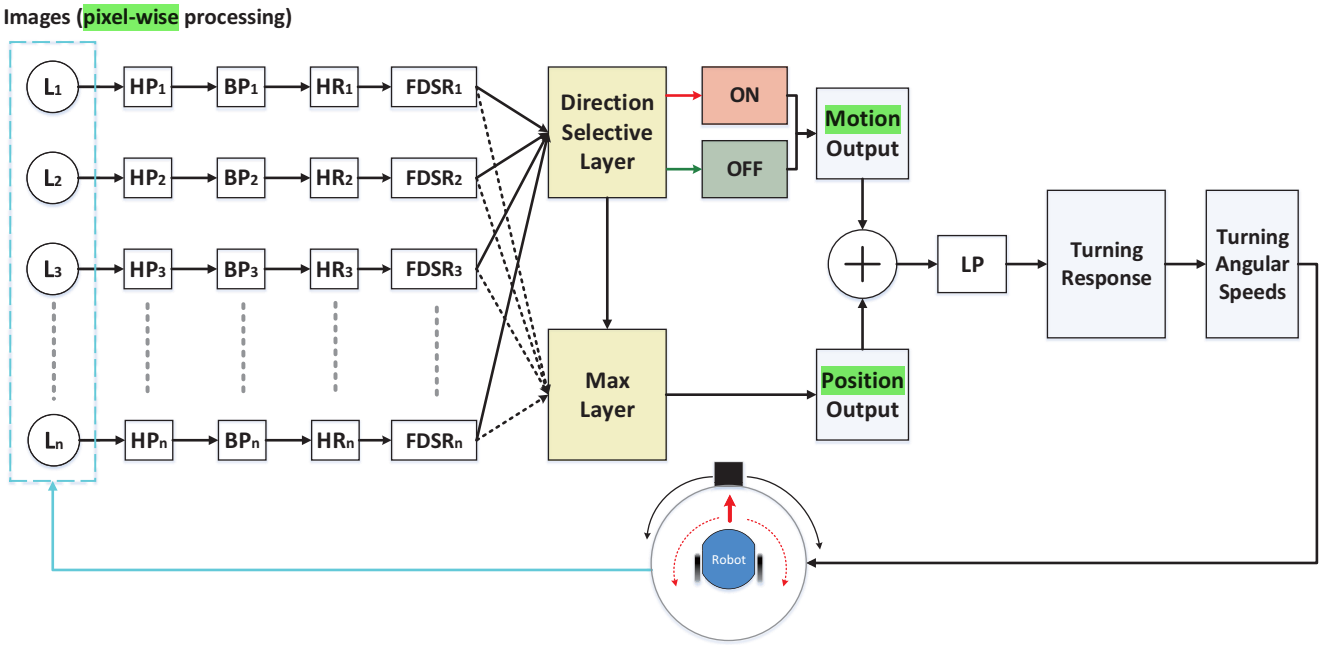


Fig. 2. The proposed embedded vision system consists of ON and OFF and position pathways and a hybrid control for guiding robot fast motion tracking and fixation behaviours. 'n' indicates the total amount of photoreceptors in the first layer; 'L' is the pixel-wise grey-scale luminance; 'HP, BP, LP' denote the high-pass, band-pass and low-pass filters; 'HR' indicates the half-wave rectifier; 'FDSR' denotes an adaptive mechanism for 'fast onset slow offset' responses. **The outputs of two subsystems are linearly combined to form a turning response for robot fixating a translating target in real time.**

II. THE EMBEDDED VISION SYSTEM

In this section, algorithms of the proposed visual fixation model on embedded system is illustrated in detail. This bio-robotic approach was rigorously based on the *Drosophila* physiology research (e.g. [16]), as well as the former modelling studies on fly visual systems [17], [18]. Generally speaking, we highlight the collaboration of three separate pathways or subsystems conducting the fast motion tracking and fixation behaviours: the ON and OFF visual pathways constitute the motion-detecting system, which encode brightness increments and decrements in parallel channels and indicate direction and intensity of wide-field translating stimuli; the position pathway is only sensitive to locational information of max offset response. A schematic diagram is depicted in Fig. 2 and the parameters are given in Table I.

A. Modelling of the ON and OFF Motion Sensing Subsystem

Firstly, we present the modelling of translating sensing pathways. The primary goal of this subsystem is the detection of direction and intensity of foreground translating stimuli against visual clutter. As depicted in Fig. 2, the first computational layer consists of photoreceptors, arranged in a 2-D matrix, capturing grey-scale imagery and retrieving motion information by a high-pass filtering. That is,

$$P(x,y,t) = L(x,y,t) - L(x,y,t-1) + \sum_{i}^{n_p} a_i \cdot P(x,y,t-i), \tag{1}$$

where $P(x,y,t)$ is the brightness change corresponding to each local pixel: x and y are the abscissa and ordinate; t indicates the current frame. The pixel-wise luminance change could last for a short duration of n_p number of frames. We define a coefficient a_i to be calculated by $a_i = (1+e^{u \cdot i})^{-1}$ and $u = 1$, simulating the quick decay of residual visual information.

After that, there is a spatial band-pass filter to achieve the edge selectivity and remove redundant environmental noise. It is represented by an algorithm of 'difference of Gaussians'. In this bio-robotic study, we nevertheless applied linearly distributed weightings to convolve visual signals, so as to save computational power of the embedded system. That is,

$$\begin{aligned} P_e(x,y,t) &= P(x,y,t) \overset{x,y}{*} W_e(x,y), \\ P_i(x,y,t) &= P(x,y,t) \overset{x,y}{*} W_i(x,y), \end{aligned} \tag{2}$$

where $\overset{x,y}{*}$ indicates the convolution at each local cell (x,y) in the visual field. The excitatory kernel $[W_e]$ is as follows:

$$W_e = \begin{pmatrix} 1/16 & 1/8 & 1/16 \\ 1/8 & 1/4 & 1/8 \\ 1/16 & 1/8 & 1/16 \end{pmatrix}, \tag{3}$$

and the inhibitory kernel $[W_i]$ is with twice radius:

$$W_i = \begin{pmatrix} 1/128 & 1/64 & 1/32 & 1/64 & 1/128 \\ 1/64 & 1/32 & 1/16 & 1/32 & 1/64 \\ 1/32 & 1/16 & 1/8 & 1/16 & 1/32 \\ 1/64 & 1/32 & 1/16 & 1/32 & 1/64 \\ 1/128 & 1/64 & 1/32 & 1/64 & 1/128 \end{pmatrix}. \tag{4}$$

After that, the inhibition is subtracted from the excitation:

$$\bar{P}(x,y,t) = P_e(x,y,t) - P_i(x,y,t). \tag{5}$$

Then, there are ON and OFF rectifying transient cells encoding onset and offset responses. Each photoreceptor corresponds to a pairwise ON and OFF detectors:

$$P_{on}(x,y,t) = [\bar{P}(x,y,t)]^{+}, P_{off}(x,y,t) = -[\bar{P}(x,y,t)]^{-}. \tag{6}$$

Here, $[x]^+$ and $[x]^-$ denote $max(0,x)$ and $min(x,0)$, respectively. In this research, we also apply an adaptive mechanism with biological plausibility, that is, the 'fast-depolarisation-slow-repolarisation' (FDSR) to implement an 'adaptation state' with a fast-onset-slow-offset characteristic [12]. Technically speaking, the previous step of spatial filtering can remove environmental motion noise in space, and such a temporal mechanism significantly reduces noise in time. Let $P_{on}(x,y)$, $P_{off}(x,y)$ be abbreviated as P, and the delayed signal $D(x,y)$ as D, such a mechanism can be denoted as:

$$dD(t)/dt = \begin{cases} (P(t)-D(t))/\tau_1, & \text{if } dP(t)/dt \geq 0 \\ (P(t)-D(t))/\tau_2, & \text{if } dP(t)/dt < 0 \end{cases}, \quad (7)$$

where τ_1 and τ_2 are time constants and $\tau_1 < \tau_2$. The delayed signal is then subtracted to the original passed one:

$$F_{on,off}(x,y,t) = P_{on,off}(x,y,t) - D_{on,off}(x,y,t). \quad (8)$$

After that, the same-polarity and neighbouring signals in the 'Direction Selective Layer' (Fig. 2) interact with each other in a non-linear way for parallel computation in the ON and OFF pathways. The computational role is denoted by ensembles of HR detectors with dynamic temporal delays (τ_s) between each combination of ON/OFF local motion detectors. Our former research has demonstrated that such temporal dynamics work effectively to sharpen up the speed response of translating sensitive visual model [17].

In this bio-robotic study, as the ground robot can only move on a 2D surface, we calculate just the horizontal motion signals as follows:

$$\begin{aligned} ON(x,y,t) = \sum_{i=d}^{d\cdot n_c} (&\bar{F}_{on}(x,y,t)\cdot F_{on}(x+i,y,t) \\ &- \bar{F}_{on}(x+i,y,t)\cdot F_{on}(x,y,t)), \end{aligned} \quad (9)$$

where d and n_c are the sampling distance between each pairwise detectors and the number of connected interneurons in the ON pathway. $\bar{F}(x,y)$ denotes a temporal low-pass filtering on $F(x,y)$ similarly to the Eq. 7 yet delayed by a dynamic time parameter τ_s. The computation of interneurons in the OFF pathway corresponds to the Eq. 9. As a result, the output of proposed motion-sensing sub-system is a pooling from all local ON/OFF motion detectors:

$$MO_{on}(t) = \sum_{1}^{C}\sum_{1}^{R} ON(x,y,t), \quad (10)$$

where C and R are the numbers of columns and rows in the visual field and similarity for the integration of OFF channels. Both ON and OFF motion outputs are normalised via a sigmoid transformation. That is,

$$f(x) = \text{sgn}(x)\cdot((1+e^{-|x|\cdot(C\cdot R\cdot K_{sig})^{-1}})^{-1} - \Delta_C), \quad (11)$$

where K_{sig} and Δ_C indicate two coefficients. The output is normalised to $[0, 0.5)$ for the positive input, and $(-0.5, 0]$ for the negative input. The global output of motion-sensing subsystem ($MO(t)$) combines outputs from both ON and OFF pathways which is ranged within $(-1, 1)$.

B. Modelling of the Position Locating Subsystem

With regard to our previous modelling study [18], we present a simplified computational structure of the position locating subsystem. As shown in Fig. 2, it shares some same structures of spatiotemporal processing with the motion sensing subsystem, until the filtered visual streams flow into a different 'Max Layer'. Within this layer, a maximisation operation is proposed to retrieve the horizontal position of strong offset response. Importantly, the motion sensing subsystem provides also local motion information ($LM(x,y)$) to localise a sub-area in the visual field for the following maximisation operation:

$$\begin{aligned} &\hat{LM}(\hat{x},y,t) = MAX_{(x,y)\in\Omega(max_x,y)}\, LM(x,y,t), \\ &\text{where, } LM(x,y,t) = ON(x,y,t) + OFF(x,y,t). \end{aligned} \quad (12)$$

Here $\hat{x}$ indicates the abscissa of the location given by the position pathway, in a neighbouring field $\Omega(max_x, y)$ centred by (max_x, y). And the max_x is the abscissa of the max offset response location. In addition, the radius of this neighbouring field corresponds to the max sampling distance ($d \times n_c$) in the ON and OFF pathways. Therefore, the output of position locating subsystem ($PO(t)$) is activated by an exponential transformation:

$$PO(t) = \text{sgn}(\hat{x}(t) - x_{vc})(1/e^{-(\sigma_1\cdot(\hat{x}(t)-x_{vc})/C)^2} - 1), \quad (13)$$

where x_{vc} is the horizontal location of view centre (VC) in the visual field, and σ_1 is a scale parameter.

C. Modelling of the Hybrid Motion Control System

Finally, we propose a simple motion control strategy for the ground micro robot with a hybrid system integrating the outputs of two subsystems as depicted in Fig. 2. A hybrid turning response (TR) is formed at each frame, and then delayed by a first-order low-pass filtering:

$$\begin{aligned} &TR(t) = \sigma_2\cdot MO(t) + \sigma_3\cdot PO(t), \\ &\text{then, } d\hat{TR}(t)/dt = (TR(t) - \hat{TR}(t))/\tau_3, \end{aligned} \quad (14)$$

where σ_2 and σ_3 are two gain factors for the output of motion and position subsystems, respectively; τ_3 is a time constant.

Importantly, in this research, we map the TR to an angular speed of the differentially driven mobile robot. Given an initial speed v_i, the motor powers of the right (P_R) and left (P_L) wheels can be described as follows:

$$\begin{aligned} P_R(t) &= g_v\cdot v_i(t) - g_w\cdot\hat{TR}(t), \\ P_L(t) &= g_v\cdot v_i(t) + g_w\cdot\hat{TR}(t), \end{aligned} \quad (15)$$

where g_v and g_w are gain values that control motion efficiency. According to our previous research [18], a satisfactory robot fixation behaviour should meet the following requirement:

$$\lim_{t\to t_0} ||\hat{x}(t) - x_{vc}|| \leq \gamma, \quad (16)$$

where γ is a predefined threshold. Notably, this term sets a criterion for identifying the satisfactory fixation response.

TABLE I

PREDEFINED PARAMETERS OF THE EMBEDDED VISION SYSTEM

Parameter	Description	Value
n_p	luminance change persistence	2
τ_1	latency in fast onset response(ms)	1
τ_2	latency in slow offset response(ms)	100
d	sampling distance	$2 \sim 4$
n_c	number of connected interneurons	$2 \sim 4$
τ_s	latency in ON/OFF pathways(ms)	$10 \sim 100$
C, R	columns and rows of the visual field	$99, 72$
K_{sig}	coefficient in sigmoid function	$0.05 \sim 0.3$
Δ_C	scale parameter in sigmoid function	0.5
x_{vc}	abscissa of the view centre	$C/2$
σ_1	scale parameters	$0.5 \sim 2$
$\{\sigma_2, \sigma_3\}$	gain factors	$\{20, 10\}$
τ_3	latency in turning response(ms)	10
$\{g_v, g_w\}$	gain values in motion control	$\{1, 10\}$
γ	threshold for fixation response	10

III. ROBOT AND SYSTEM CONFIGURATION

Within this section, we propose the parameters setting of the embedded vision system, and briefly introduce the robot platform. The proposed model processes visual information with a feed-forward structure. No learning methods were applied in this research. All the parameters in Table I were decided empirically, with considerations of optimisation and realisation in hardware. More precisely, the ON and OFF motion sensing subsystem represents positive and negative responses to preferred (rightward) and non-preferred (leftward) directional translating movements; whilst the position locating subsystem is only sensitive to the relative location of max offset response with respect to the robot view centre.

The monocular vision based micro robot is a low-cost and autonomous ground mobile platform named '*Colias*' [8], [10]. As illustrated in Fig. 1a, it has a small footprint of 40mm in diameter and 30mm in height. Two DC-motors are driven differentially and provide the platform with a maximum speed of roughly 35cm/s. A 3.7V, 320mAh Lithium battery supports the autonomy for $1 \sim 2$ hours. The *Colias* robot has two main boards: the bottom board includes wheels and battery, working as a motion actuator on 2D surfaces; the upper board supports in-chip image processing with an $OV7670$ camera (Fig. 1b). Its main processor is an ARM- Cortex M4 based MCU STM32F427, which runs at 180MHz, with $256K$byte SRAM, 2Mbyte in-chip Flash. The acquired image is set to 99×72 in 8-bit YUV422 format. In addition, the field of view can reach approximately 70 degrees. The only sensor used in this research is the monocular camera. We also used a Bluetooth device, which is connected with the upper board, to retrieve real-time system data including visual model output from the robot. In our bio-robotic tests, the image processing frequency can reach $25 \sim 45$Hz matching well the requirement of most real-time visual tasks.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

Within this section, we present the systematic bio-robotic experiments and analyse the results. To clarify our goals and the significance of this bio-robotic approach, there are two kinds of tests to demonstrate. Firstly, in the open-loop tests, we adopted different categories of basic motion patterns to stimulate a motionless *Colias* robot, aiming at representing the specific response of the two subsystems and the hybrid system in the Fig. 4. We also investigated both the speed and distance response of the proposed visual model in the Fig. 5. Secondly, in the closed-loop tests, we examined the fixating response of a *Colias* robot to other translating robots in the Fig. 8. In addition, an important biological finding has revealed that the *Drosophila* fixation behaviour could be achieved by the position pathway only, while the motion pathway can improve the fixation precision [14]. This theory has also been verified by a recent computational modelling study [18]. For comparison with the ethological study and results shown in the Fig. 3, we also inspected whether our micro robot can show similar behaviour to the *Drosophila*.

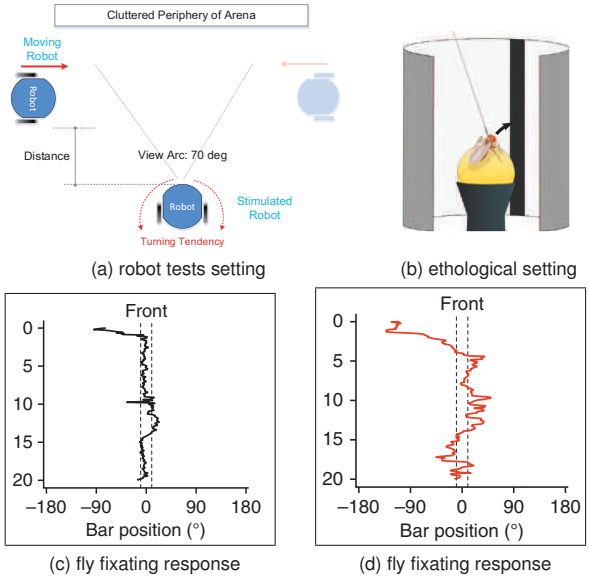


Fig. 3. Experimental settings of (a) the proposed bio-robotic tests and (b) previous ethological study on a motion-blind *Drosophila* with the fly fixating responses to a translating bar by (c) the intact motion and position pathways and (d) only the position pathway, adapted from [14].

A. Open-loop Tests

In the open-loop tests, we used another *Colias* robot as the visual stimuli and collected on-line outputs of the proposed embedded vision system from the stimulated robot (Fig. 3a). All the stimuli can be categorised into three kinds of movements: looming (Fig. 4a), horizontal translating (Fig. 4b and 4c), which are essential and frequent visual challenges to both robots and insects in navigation.

Fig. 4 demonstrates the outputs of both the motion and position subsystems, as well as the tuned hybrid turning response. When challenged by an approaching object, both the motion-sensing subsystem and the robot turning response remain quiet. In the case of translating, the motion subsystem is highly activated generating positive and negative response to rightward and leftward translations, respectively. Notably, the hybrid turning response represents a changing tendency similarly to the outputs of the motion subsystem. Therefore, we can conclude that the motion-sensing subsystem is only sensitive to directional translating features regardless of the location of moving target. It also affects significantly the robot turning response. On the other hand, the outputs of

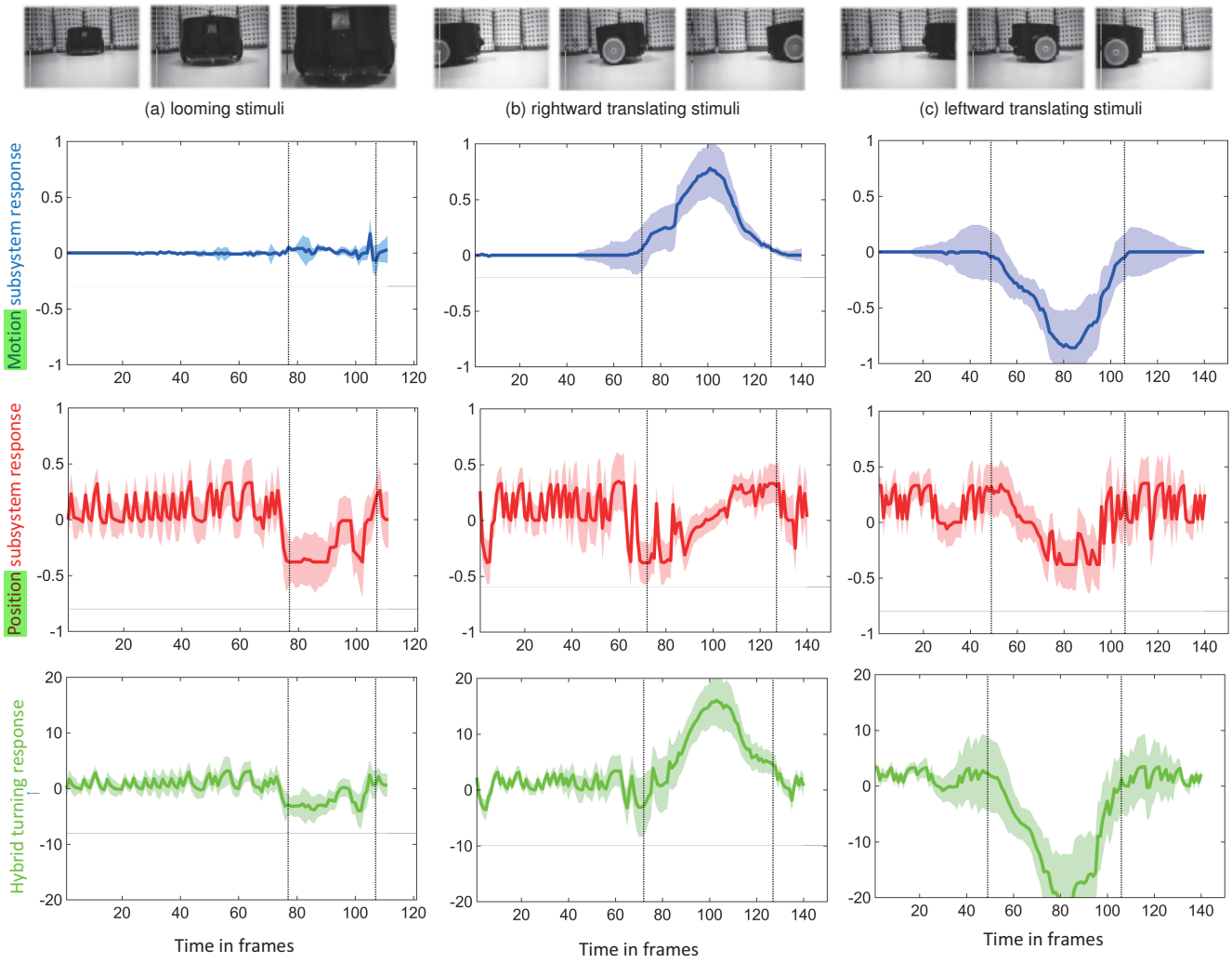


Fig. 4. Neural response from the tested motionless *Colias* robot under open-loop tests, including the outputs of the motion-sensing, the position-locating and the hybrid systems. The example views from the stimulated robot are shown at each top. Each kind of movements was repeated ten times, with colour-shadows indicating the continuous errors. Two vertical dashed lines denote the period that motion features are extracted by the proposed model.

the position system reveal that it is only sensitive to the location instead of the direction of translating features. In practice, it generates a negative response when the object moves within the left-side of visual field, and a positive response to movements within the right-side of receptive field. The two half-fields are separated by the robot view centre. Moreover, it appears that the position system is more likely influenced by background noise, i.e., the response tends to fluctuate within the view field without apparent translating cues extracted by the proposed visual model.

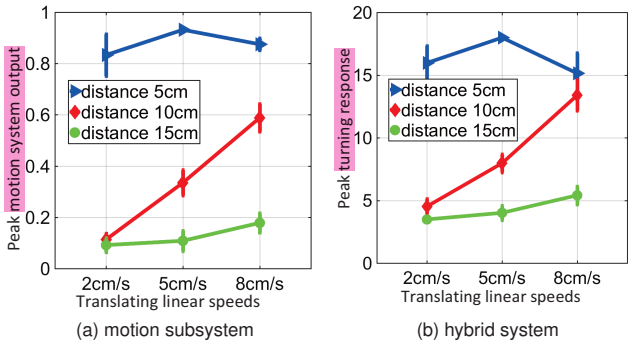
On the aspect of investigations on speed and distance response, the *Colias* robot was tested by rightward translations at three constant linear speeds and from three different distances, separately. The statistics in Fig. 5 demonstrate that the motion subsystem and the hybrid system have similar speed and distance sensitivity. More precisely, the max outputs of both systems are reached by translations from the shortest distance. The speed response is largely weakened by translations far from the view field, since the proposed model is sensitive to wide-field over small-target motion.

B. Closed-loop Tests

In the closed-loop tests, we enabled the motion unit of the tested *Colias* robot to demonstrate its fixating response[1]. More precisely, in the arena tests, if we set the robot initial speed greater than zero, a successful following behaviour can be triggered by the fast motion tracking and fixation, as depicted in the Fig. 6. This could promote research in insect vision-based collective behaviours. On the other hand, the robot only turns to fixate the translating robot by setting $v_i = 0$. The results in Fig. 8 demonstrate that the micro robot can smoothly fixate a translating robot at different speeds and from varied distances which reconciles with the *Drosophila* behaviour [14]. Notably, the hybrid system output delivers a signal that is correct only in presence of a moving target.

Crucially, being consistent with the above mentioned biological findings in [14], our bio-robotic experiments also demonstrated that the visual fixation behaviour could be mediated by the position-locating subsystem only; whilst

[1]This is supported by a supplementary video of the micro robot performance in the closed-loop arena tests.



(a) motion subsystem (b) hybrid system

Fig. 5. Statistical results of peak response of the motion and the hybrid neural systems: the robot was tested by translations at three constant linear speeds and three distances, respectively, each throughout ten times tests. The proposed visual fixation model shows speed and distance response sensitivity to translating movements.

the motion-sensing subsystem can significantly improve the fixation precision, similarly to the results in the Fig. 3c, 3d. To support this, the statistics in Fig. 7 illustrate higher successful fixation rate by the intact visual system, compared with the motion-blocked, i.e. position-only situation, for all tested speeds and distances.

V. DISCUSSION

Through the above systematic experiments, our micro robot has shown similar visual fixation behaviour to *Drosophila*. Compared to learning and/or registration based visual trackers, the proposed model benefited from feed-forward and low-level visual processing, which is energy-saved. This approach has been validated by our bio-robotic tests. The visual fixation can be accomplished from an ordinary camera and under limited storage capability. Compared to recent insect-inspired trackers [12], [13], that are sensitive to only small target movements, the proposed visual model can detect wide-field translating objects. We will further compare the computational complexity of these visual trackers in the near future.

Similarly to other cutting-edge bio-inspired or bio-mimetic dynamic vision systems (e.g. [8]–[10], [13]), performance of the proposed visual fixation model is also restricted by the speed and size of moving features. To the best of our knowledge, a single type of neural system can not handle this challenge well, whereas coordination or competition of multiple neural systems, that are possessing diversity of direction and size selectivity, may provide effective solutions.

The detection of motion is a paramount characteristic in the proposed visual fixation model. This comprises two subsystems and a hybrid control system for generating turning response toward sensorimotor control. In comparison with the position subsystem, the motion subsystem is more stable in extracting translating movement features. Moreover, the position output delivers a signal that is correct only in presence of a moving target. Our experimental results have also verified that this model well matches a biological hypothesis that the insect visual fixation behaviour could be completed by only the position pathway in the preliminary visual system, whilst the motion pathway could help the whole system achieve higher fixation accuracy.

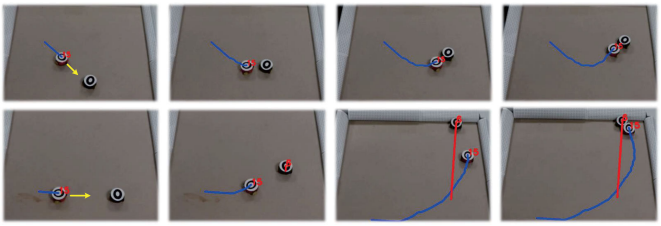
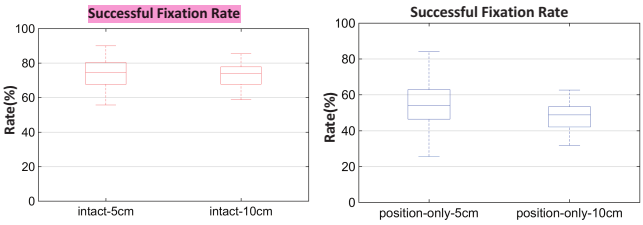


Fig. 6. Performance of the *Colias* robot in arena tests with a cascade of behaviours including motion tracking, fixation and following captured by a top-down camera: yellow arrows indicate the robot initial direction; blue and red lines denote the robot trajectories. In this case, $v_i > 0$.



(a) results of intact system (b) results of motion-blocked system

Fig. 7. Integration of successful fixation rate corresponding to the tests in Fig. 8: **the intact visual system shows higher fixation rates**.

VI. CONCLUDING REMARKS

Inspired by *Drosophila* physiology, we have presented a novel computational model structure mimicking fast motion tracking strategy and a visually-guided closed-loop behaviour to fixation. The proposed visual neural network has been successfully implemented on embedded system in an autonomous micro ground robot which has limited computational resources. Satisfactory experimental results have demonstrated the effectiveness, robustness and efficiency of this bio-robotic approach simulating insect visual fixation behaviour. Most importantly, we have compared the robot test results with biological research: our *Colias* micro robot showed fixating response similarly to insects. In navigation, this visual model can also conduct a following behaviour. The proposed dynamic vision system can be built as neuromorphic sensors for mobile autonomous machines.

REFERENCES

[1] T. B. Moeslund, A. Hilton, and V. Kruger, "A survey of advances in vision-based human motion capture and analysis," *Computer Vision and Image Understanding*, vol. 104, pp. 90–126, 2006.

[2] A. Dosovitskiy, P. Fischer, E. Ilg, P. Hausser, C. Hazirbas, and V. Golkov, "Flownet: Learning optical flow with convolutional networks," in *International Conference on Computer Vision*, 2015, Conference Proceedings, pp. 2758–2766.

[3] D. Held, S. Thrun, and S. Savarese, "Learning to track at 100 fps with deep regression networks," in *European Conference on Computer Vision*, 2016, Conference Proceedings.

[4] J. Redmon and S. Divvala, "You only look once: Unified, real-time object detection," in *Conference on Computer Vision and Pattern Recognition*, 2016, Conference Proceedings, pp. 1–10.

[5] R. Sabzevari and D. Scaramuzza, "Multi-body motion estimation from monocular vehicle-mounted cameras," *IEEE Transactions on Robotics*, vol. 32, no. 3, pp. 638–651, 2016.

[6] V. Mahadevan and N. Vasconcelos, "Biologically inspired object tracking using center-surround saliency mechanisms," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 3, pp. 541–554, 2012.

[7] B. Cai, X. Xu, X. Xing, K. Jia, J. Miao, and D. Tao, "Bit: Biologically inspired tracker," *IEEE Transactions on Image Processing*, vol. 25, no. 3, pp. 1327–1339, 2016.

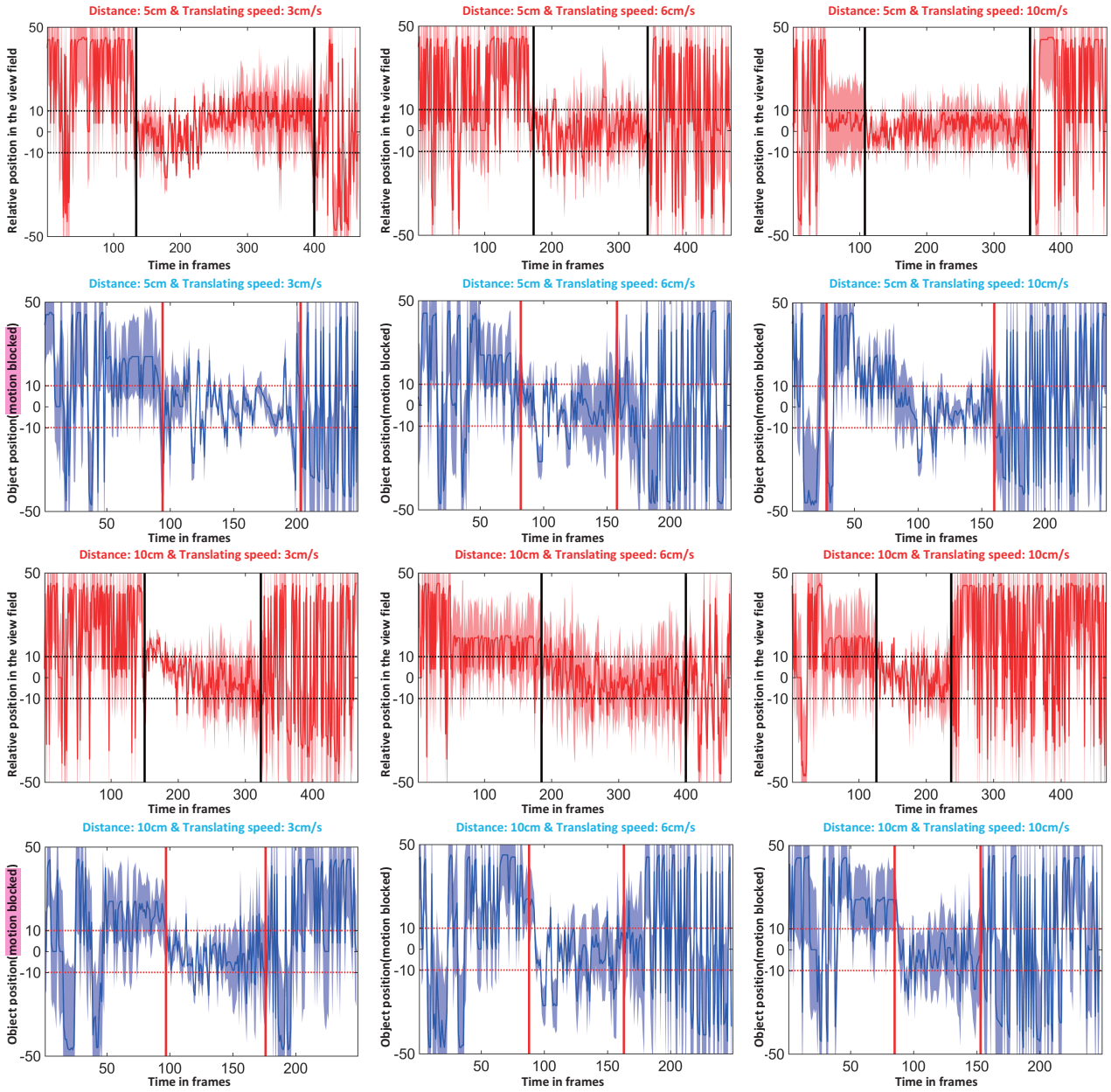


Fig. 8. *Colias* robot fixating response under closed-loop tests: both the intact system (red) and the motion-blocked system (blue) were tested by translations at three speeds and two distances, respectively, each throughout ten repeated tests. Two horizontal dashed lines indicate the range of VC for successful fixation behaviour; two vertical lines denote the period that translating stimuli are extracted by the tested *Colias*.

[8] 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, 2017.

[9] Q. Fu, C. Hu, T. Liu, and S. Yue, "Collision selective lgmds neuron models research benefits from a vision-based autonomous micro robot," in *IROS*, 2017, Conference Proceedings, pp. 3996–4002.

[10] Q. Fu, C. Hu, J. Peng, and S. Yue, "Shaping the collision selectivity in a looming sensitive neuron model with parallel on and off pathways and spike frequency adaptation," *Neural Networks*, vol. 106, pp. 127–143, 2018.

[11] J. R. Serres and F. Ruffier, "Optic flow-based collision-free strategies: From insects to robots," *Arthropod Structure & Development*, vol. 46, no. 5, pp. 703–717, 2017.

[12] Z. M. Bagheri, S. D. Wiederman, B. S. Cazzolato, S. Grainger, and D. C. O'Carroll, "Performance of an insect-inspired target tracker in natural conditions," *Bioinspiration & Biomimetics*, vol. 12, no. 2, p. 025006, 2017.

[13] Z. M. Bagheri, B. S. Cazzolato, S. Grainger, D. C. O'Carroll, and S. D. Wiederman, "An autonomous robot inspired by insect neurophysiology pursues moving features in natural environments," *Journal of Neural Engineering*, vol. 14, no. 4, p. 046030, 2017.

[14] A. Bahl, G. Ammer, T. Schilling, and A. Borst, "Object tracking in motion-blind flies," *Nat Neurosci*, vol. 16, no. 6, pp. 730–738, 2013.

[15] J. W. Aptekar, P. A.Shoemaker, and M. A. Frye, "Figure tracking by flies is supported by parallel visual streams," *Current Biology*, vol. 22, no. 6, pp. 482–487, 2012.

[16] Y. E. Fisher, J. C. Leong, K. Sporar, M. D. Ketkar, D. M. Gohl, T. R. Clandinin, and M. Silies, "A class of visual neurons with wide-field properties is required for local motion detection," *Current Biology*, vol. 25, no. 24, pp. 3178–3189, 2015.

[17] Q. Fu and S. Yue, "Modeling direction selective visual neural network with on and off pathways for extracting motion cues from cluttered background," in *The 2017 IJCNN*, 2017, Conference Proceedings, pp. 831–838.

[18] ——, "Mimicking fly motion tracking and fixation behaviors with a hybrid visual neural network," in *IEEE Int. Conf. on Robotics and Biomimetics*, 2017, Conference Proceedings, pp. 1636–1641.