

COMPUTER VISION

A bug's-eye view

Yiannis Aloimonos* and Cornelia Fermüller

An insect-scale visual sensing system indicates the return of active vision for robotics.

Biorobotics aims to mimic the characteristics of living organisms to enable new robot designs. Insects have been a popular focus of the field because of their amazing sensing and actuation capabilities. However, manufacturing miniaturized actuators at low weight and power is a very challenging task. Solutions to this problem will enable the design of microrobots that can efficiently move and perceive. Writing in this issue of *Science Robotics*, Iyer *et al.* (1) report a low-power rotatable vision system for live insects and insect-scale robots. Their system is lightweight (248 mg), can be controlled wirelessly, and is able to record images to a Bluetooth device. A key innovation is the use of piezo actuators as capacitors with low leakage; this enables the actuators to hold their state without power input. The other components are a custom-designed lens and antenna and off-the-shelf sensor and Bluetooth chips connected via a custom interface. The vision system is coupled with an accelerometer that triggers image capture only when the insect moves. This low-power configuration enabled operational times of up to 6 hours when the system was attached to freely walking beetles (Fig. 1A), allowing the insect's daily routine to be observed, thus potentially demonstrating new opportunities for the field of entomology.

The researchers' vision system was also mounted on a microrobot to record images at 1 frame/s (Fig. 1B), demonstrating that it is now feasible to equip insect-scale robots with what scientists and engineers have named "Active Vision" (2–4). An active observer is one that manipulates the parameters of the sensory apparatus (focal length, field of view, or viewpoint). As a consequence, an active observer is computationally superior to a passive observer with the same resources (2).

Robot vision problems are usually solved using methods from computer vision. The

goal is to recover three-dimensional (3D) models of the scene and recognize objects and actions by assigning labels to the corresponding parts of the image. Computer vision, as a theoretical discipline, has been by definition passive, where the input is assumed to be a given set of images or videos. Biological vision systems, on the other hand, do not work this way. We are not given an image; we take the image. We (the agents) decide which images to collect. Our pupils dilate to adjust to the level of illumination; our eyes move, converge, or diverge, and position themselves to get a better view of the scene. Biological systems with vision, from insects to primates, move their bodies, heads, and eyes. In a similar way, the camera developed in (1) can move independently of the body of the insect, and this enables active perception.

Eye movements serve multiple functions from the perspective of computing visual information. Agents engaged in tasks continuously move their eyes to look at different objects and actions to support visual recognition (5, 6). Eye movements can also facilitate image reconstruction and motion analysis. An active observer can, by manipulating the geometric parameters of its sensory apparatus, introduce additional computational constraints that make the computations easier (2). For example, the problem of estimating 3D rigid motion becomes easier if the observer, while moving, also "tracks" an environmental point (i.e., the camera is moved independently of the body motion to keep a scene point at the image center) (7). Another example is the use of controlled translational movements to estimate distance (8) or solve the motion segmentation problem (finding moving objects as you move). These movements can be induced by vibrating the sensor of visual system developed by Iyer and co-workers. We note,

however, that these tasks require visual motion not available from the low frame rate camera setup of the target paper. One option could be to use low-power, neuromorphic cameras (9), which capture motion at high temporal resolution.

Insect-scale robots are subject to size, weight, area, and power constraints. The classic approach of reconstructing a scene using simultaneous localization and mapping (SLAM) algorithms is not applicable here; the alternative is to design systems that optimally perform a specific set of tasks. This is a multidimensional optimization problem related to hardware software codesign. In the future, insect-scale robots equipped with a steerable cameras may be able to find holes, fly through them find their way back home, avoid obstacles, recognize a variety of objects,

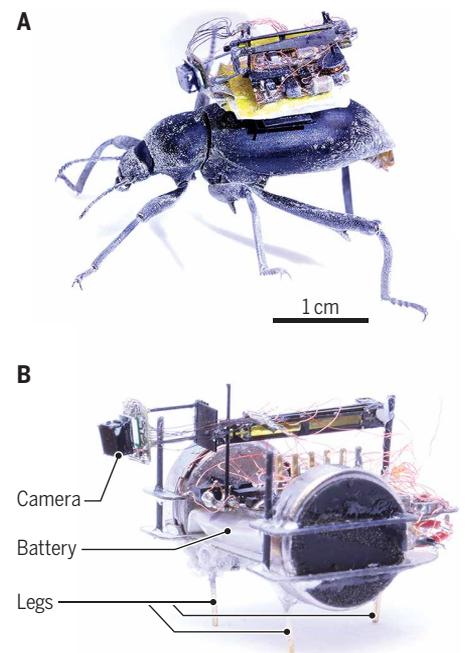


Fig. 1. Wireless steerable vision for insects and insect-scale robots. The vision system of by Iyer and co-workers (1) carried by a live darkling beetle (A) and mounted on an insect-scale robot (B).

Computer Vision Laboratory, Institute for Advanced Computer Studies, Department of Computer Science, Institute for Systems Research, University of Maryland, College Park, MD, USA.

*Corresponding author. Email: jyaloimo@umd.edu

and more, without computing 3D world models. In other words, these new systems will be implementing a SLAM alternative. An agent is an active perceiver (10), if it knows why it wishes to sense; chooses what to perceive; and determines how, when, and where to achieve that perception. The “why” (action expectation), the “what” (scene selection, and fixation), the “where” (viewpoint selection and agent/sensor pose), the “when” (temporal selection), and the “how” (priming, alignment, and proprioception) imply particular task-specific architectures. If the insect robotics community succeeds with this

SLAM alternative, then the results will permeate all robotics.

REFERENCES

1. V. Iyer, A. Najafi, J. James, S. Fuller, S. Gollakota, Wireless steerable vision for live insects and insect-scale robots. *Sci. Robot.* **5**, eabb0839 (2020).
2. J. Aloimonos, I. Weiss, A. Bandyopadhyay, Active vision. *Int. J. Comput. Vis.* **1**, 333–356 (1988).
3. R. Bajcsy, Active perception. *Proc. IEEE* **76**, 966–1005 (1988).
4. D. H. Ballard, Animate vision. *Artif. intell.* **48**, 57–86 (1991).
5. M. Hayhoe, D. Ballard, Eye movements in natural behavior. *Trends Cogn. Sci.* **9**, 188–194 (2005).
6. X. Yu, C. Fermüller, C. L. Teo, Y. Yang, Y. Aloimonos, Active scene recognition with vision and language, in *2011 IEEE International Conference on Computer Vision* (Barcelona, 2011), pp. 810–817.
7. C. Fermüller, Y. Aloimonos, The role of fixation in visual motion analysis. *Int. J. Comput. Vis.* **11**, 165–186 (1993).
8. F. Santini, M. Rucci, Active estimation of distance in a robotic system that replicates human eye movement. *Robot. Auton. Syst.* **55**, 107–121 (2007).
9. P. Lichtsteiner, C. Posch, T. Delbruck, A 128 x128 120 dB 15 μ s latency asynchronous temporal contrast vision sensor. *IEEE J. Solid-State Circuits* **43**, 566–576 (2008).
10. R. Bajcsy, Y. Aloimonos, J. K. Tsotsos, Revisiting active perception. *Auton. Robots* **42**, 177–196 (2018).

10.1126/scirobotics.abd0496

Citation: Y. Aloimonos, C. Fermüller, A bug’s-eye view. *Sci. Robot.* **5**, eabd0496 (2020).