Integrative and Comparative Biology
Integrative and Comparative Biology, volume 61, number 3, pp. 917–925
https://doi.org/10.1093/icb/icab126
Society for Integrative and Comparative Biology

SYMPOSIUM

# Review of Methods for Animal Videography Using Camera Systems that Automatically Move to Follow the Animal

Andrew D. Straw[1]

Institute of Biology I and Bernstein Center Freiburg, Faculty of Biology, Albert-Ludwigs-Universität Freiburg, Freiburg 79104, Germany

From the symposium "Spatial-Temporal Dynamics in Animal Communication" presented at the annual meeting of the Society for Integrative and Comparative Biology virtual annual meeting, January 3–February 28, 2021.

[1]E-mail: straw@bio.uni-freiburg.de

**Synopsis** Digital photography and videography provide rich data for the study of animal behavior and are consequently widely used techniques. For fixed, unmoving cameras there is a resolution versus field-of-view tradeoff and motion blur smears the subject on the sensor during exposure. While these fundamental tradeoffs with stationary cameras can be sidestepped by employing multiple cameras and providing additional illumination, this may not always be desirable. An alternative that overcomes these issues of stationary cameras is to direct a high-magnification camera at an animal continually as it moves. Here, we review systems in which automatic tracking is used to maintain an animal in the working volume of a moving optical path. Such methods provide an opportunity to escape the tradeoff between resolution and field of view and also to reduce motion blur while still enabling automated image acquisition. We argue that further development will be useful and outline potential innovations that may improve the technology and lead to more widespread use.

## Introduction

### Animal behavior occurs on multiple scales

Animal behavior is a multiscale phenomenon in which details of the movement of body parts, often very intricate and fast, can be decisive while the body as a whole makes large-scale movement through the environment. For example, the head angle of a bumblebee is critical for the visual cues learned that enable the bee to return to the nest (Doussot et al. 2021). Unfortunately, acquiring video footage that simultaneously captures animal movement at both fine and large scales is technically challenging. This is particularly challenging for small animals such as insects that are capable of traveling kilometers or more but whose appendages may be millimeters or smaller in dimension. In this review, we cover technologies that solve this challenge by automatically tracking a moving animal with a moving camera.

There are several ways in which using a moving camera to automatically record images of an animal could be useful. First, there may be some behaviors in which the moment of small body parts is important specifically in relation to the larger scale. The visual navigation of bumblebees mentioned above is one such example in which the alignment of the head to distant visual landmarks is thought to be important for navigating in this species. Second, while it may be possible to establish a high-resolution video recording setup covering a small region of space and waiting for an animal to perform a behavior of interest in this region by chance, it may be tedious and inefficient to do so. In some cases, it might be possible to modify conditions such that this becomes more likely, but in other cases this may not be possible. Third, by actively tracking the animal in real-time, a new ability to interact with an animal becomes possible. The experimenter might create interactive sensory

Advance Access publication June 12, 2021

feedback via custom computer programs that reacts to the details of the animal's behavior. This could be used to create virtual reality by automatically updating sensory input to the animal. As a special case, we could consider collective behavior, which by definition depend on the interaction between animals. In this case, live tracking of a focal animal may open the possibility of creating hybrid biological-artificial systems in which the rules of dynamic, real-time interaction can be altered experimentally and in which these rules depend on subtle cues of fine scale movement.

### Muybridge's "Horse in Motion" as inspiration for automatic tracking during movement

Since the early days of photography, animal behavior and biomechanics have long been subjects of and even drivers of technical innovation. Eadweard Muybridge, for example, was commissioned by Leland Stanford to employ a novel photographic technique to address the question of whether the horse lifts all feet off the ground simultaneously during trotting and galloping. In a now famous example of early high-speed photography, he rigged multiple cameras arranged along a racetrack to acquire images as the horse passed in front of them and pulled tripwires. The resulting image series approximates a high framerate camera moving alongside the horse while maintaining the animal in its field of view (Muybridge 1878) and is an ancestor of the "bullet time" special effect made famous by *The Matrix* movie (Wachowski and Wachowski 1999). In the case of the trotting and galloping horses, this new method established unequivocally that they lift all four legs off the ground simultaneously. Thus, this is an ideal historical example to set the stage for this review in which we outline photographic systems that automatically maintain a moving animal in the field of view of a camera with the goal of addressing biological questions.

### Scope of this review

Here, we review systems that keep an animal in the field of view of a camera whose optical position or angle is automatically moved via commands from an automatic tracking system. Before doing so, we cover two basic points about the geometry and physics of photography and explain how these lead to a tradeoff between resolution versus field of view and potential problems with motion blur when using stationary cameras. We focus mainly on small animals such as insects. We conclude with a discussion of common problems and ideas about the future of these technologies.

### Problem 1: Geometry creates a tradeoff between resolution and field of view

As a consequence of a fixed number of pixels of a given camera sensor, lenses with low magnification will have a large field of view but a given object will cover relatively few pixels. Fine detail will be lost due to the object image being recorded by few pixels. With higher magnification in the absence of other limiting factors, more object detail will be recorded in the available pixels, but this detail comes at a cost of a reduced field of view. This tradeoff between resolution and field of view is thus a fundamental consequence of sensor pixel count and basic geometry.

### Problem 2: Motion blur is a physical limit to image sharpness of moving animals

Fundamentally, photon counts and sensor noise limit the performance of photography (and animal vision), ultimately leading to *motion blur* in photographs and videos. This is the appearance of blurred or streaked objects that were moving relative to the camera during exposure. Motion blur can be used for aesthetic effects. For example, photographers often intentionally track a fast moving subject with the camera to create a blurred background that conveys a feeling that the subject is moving quickly. However, motion blur is often a practical problem with stationary cameras in which visible detail of moving animals is limited by motion blur.

The reasons for motion blur are intrinsic to the physics of light and light detection. Even in bright sunlight, photon counts are limited. Short exposures are necessary for high framerate video or for "stopping motion" to eliminate motion blur. Short exposures, however, limit the amount of light that can be collected. Due to the quantum nature of light and properties of detectors, this creates a fundamental physical problem. Concretely, in semiconductor-based photosensors—the category to which modern digital cameras and photodiodes belong—incoming photons are detected by triggering electron displacement and thus generating electrical current. High-end sensors are designed to maximize the conversion from photons into displaced electrons (measured as *quantum efficiency*), and to minimize spontaneous current unrelated to photons that would occur even in complete darkness (*dark noise*). Saturation of the readout can also be a problem. CCD and CMOS image sensors have an array of pixels in which photons are converted to electrical charge. This is accumulated for a given integration time, also called exposure time, and then read it out. Photodiodes are typically amplified via "transimpedance amplifier" that converts electrical current to voltage. Regardless of the specific type, readout technology, and performance of a particular sensor, these limits of quantum efficiency and dark noise establish what is possible with any given sensor (Hobbs 2009; Lambert and Waters 2014). Real cameras are limited by these factors and consequently a robust market exists for high-performance sensors. In some instances, a high-performance sensor can reduce motion blur to the point where it not problematic, but the limit is established by the physics of light.

### A proposed solution: automatically tracking the animal with a high-resolution camera.

The basic problems outlined above have been understood by photographers and cinematographers since the creation of these technologies. So too, has one solution: keep the subject fixed at the focus of a camera that is moved to maintain this relation. The camera can be translated linearly, rotated in angle, or both together. The magnification of a zoom lens can also be adjusted. In this way, the field of view of moving camera can be used to dynamically cover a large volume while still obtaining high detail and minimizing motion blur. The earliest historical precedent we know for this concept is the system developed by Muybridge (1878) described above. In that case, since it was known in advance that the horse would follow a particular trajectory, the track could be prepared with tripwires that would serve as the technological ancestor of an automatic tracking system. Today, this general approach, using a moving camera to obtain visual information, falls under the term *active vision* in the computer vision and robotic community. Note that even when overall relative motion between an animal's body and the camera is minimized, rapid limb motion may still cause motion blur due to the residual relative motion between these body parts and the camera. Indeed, close inspection of Muybridge's photos reveals motion blur of the airborne legs.

While we focus here on tracking of animals, parallel systems are also used for other purposes. For example, related optical tracking systems for military use have been developed for decades. Many companies in the security industry sell products advertising "auto tracking" in their pan-tilt-zoom (PTZ) camera systems that typically support tracking humans. Auto tracking is now also available in drones for consumer and professional use. Even compact disc and Blu-Ray players use similar principles to keep the optical reader aligned with data tracks on the disc.

Before we mention developments that enable automatic tracking of animals, we note that a sufficiently skilled camera operator can bypass the need for such automation. Professional camera operators are often capable of tracking quickly moving animals, athletes, or objects. The optical systems can be specialized to support the camera operator in tracking and they can be additionally instrumented to record the tracking movements made by the camera operator. For example, a custom rotating telescope was built such that a skillful operator can track a flying bird with its stereo-view to track the bird in 3D by recording azimuth, elevation and convergence angles (Margerie et al. 2015). This system was used to show that swifts (*Apus apus*), despite their great maneuverability, spend only a surprisingly small amount of time (25%) flapping during foraging flights (Hedrick et al. 2018). Haalck and colleagues developed a related concept in which an operator moves handheld or drone-mounted camera to follow animal in their natural habitat and, using techniques from computer vision, reconstructs both the animal path and the camera path (Haalck et al. 2020). So far, they have applied this method to beetles, ants, woodlice and wild dogs in a variety of challenging imaging conditions. Similarly, Brady (2021) built a system that reconstructs animal trajectories and environmental approximations. Here, we discuss systems that can automatically keep an animal in the field of view of a moving camera without relying on the skills of a human operator.

Sufficient speed is a prerequisite for automatic tracking to keep a camera pointed at a moving animal. Clearly, the optical system must react to animal movement with sufficiently short delay to keep the animal in the field of view of the camera. For high magnification and fast maneuvering animals, this puts a premium on low-latency operation and the systems described here all have dealt with this latency issue. The sensors involved used to generate signals must be fast, any processing that runs on these signals must complete quickly, and to move the optical system quickly, powerful motors or low-mass parts (or both) must be used. For a mathematical treatment of the required system speed given optical and organism motion parameters, the reader is referred to Ogawa et al. (2005).

In many of the systems described below, the signal used to update the camera position is from the camera sensor itself or shares part of the optical system but uses a different sensor such as a quadrant photo-diode. In these *feedback*-driven *closed-loop* schemes, sensor signals are used to generate movement that in turn changes subsequent sensor signals (Fig. 1). Using a closed-loop system is accompanied by the challenge that a single momentary loss of tracking can result in an inability to resume tracking later and thus a fast and robust system is required. *Open-loop* schemes, in which the camera is directed by an external sensor system, come with the requirement of bringing the external sensor system into calibration with the motor system.

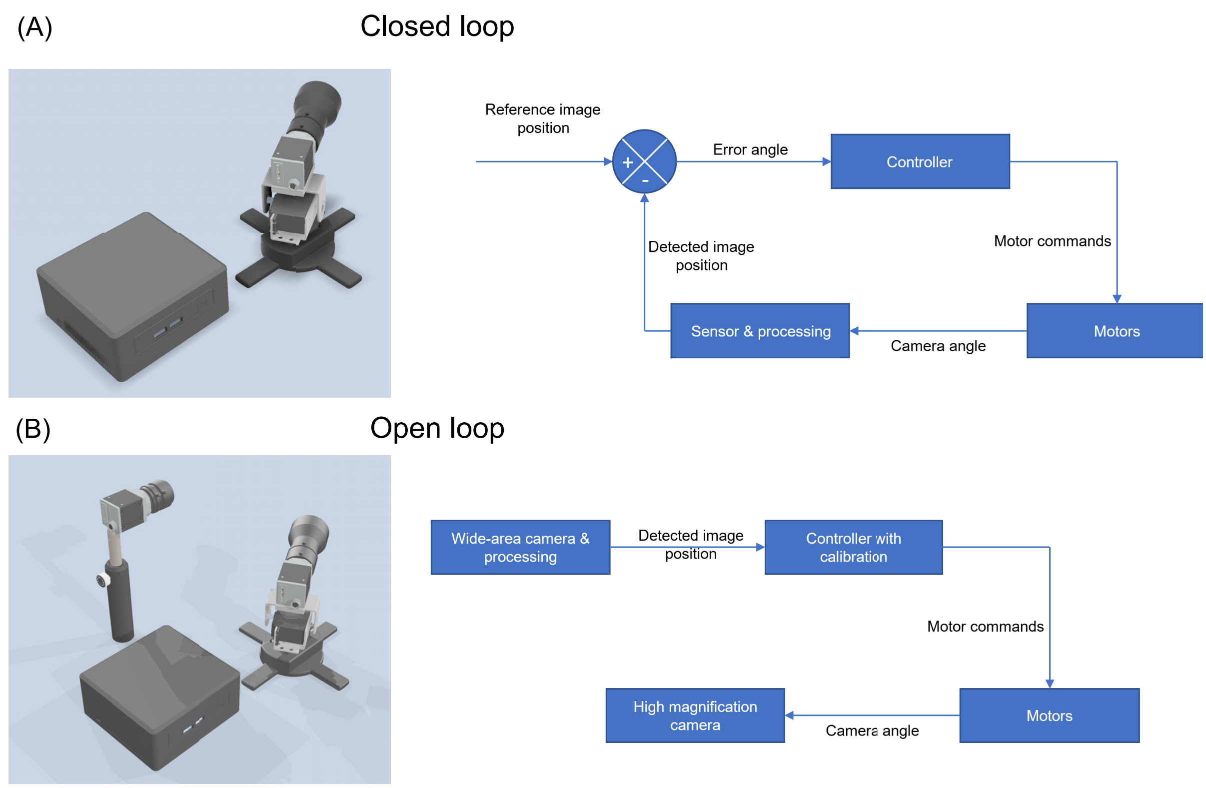


**Fig. 1** Closed-loop and open-loop configurations for tracking animals with a moving camera. (**A**) An example closed-loop configuration consists of a camera mounted on a pan-tilt motor system and a computer to perform low-latency image processing and steer the motors (left). A block diagram indicates a system in which the controller seeks to keep minimal error angle by constantly updating steering the camera via motors such that the detected object in each new image is near the reference position (right). If tracking is lost with a closed-loop system, it may be difficult to recover. (**B**) In an example open-loop configuration, and additional wide-area camera is added to the system (left). The algorithmic structure of the open-loop system steers the high-magnification camera via motors in a feed-forward way such that the image processing from the wide-area camera is not affected by moving the motors, which steer the high-magnification camera (right). This system is robust to loss of tracking but requires an additional camera and calibration of the system.

## Camera systems that move automatically to keep moving organisms in view

### Closed-loop systems based on digital image analysis

Our examples in this section track an animal with a moving camera whose output is used both for closed-loop motor control and saved for later analysis. The camera itself may move or may look through a moving optical system. Nowadays this approach is enabled by the comparatively low cost and simple infrastructure required to perform high-speed image processing using digital cameras as sensors and is referred to as *visual servoing* in the robotics community.

The first historical example of automatically moving the optical system based on live tracking from a moving organism is a *Paramecium* tracker that used an early high-speed digital camera developed in collaboration with the company Hamamatsu Photonics KK (Ogawa et al. 2005). In this work, the authors described a novel high-speed sensor readout mechanism used to analyze a high-magnification view of the ciliate through a microscope at a framerate of one kilohertz (1000 frames per second). Fast linear voice coil motors were used drove an XY microscope stage with micron precision and allowed the authors to achieve a closed-loop "lock on" of this microscopic organism. The authors further established closed-loop control of the micro-organism's behavior by using electrical stimulation to induce animal turning and could trap it in a 1 mm area. This system thus implemented a form of the virtual reality described in the introduction in which artificial sensory input is generated based on the fine details of the organism's behavior. A few years later, a similar system was used to track spermatozoa of ascidians (sea squirts) in a microscope (Oku et al. 2008).

The idea of using a high-speed camera to implement a closed-loop "lock on" tracking was used in the macroscopic world by applying similar image acquisition and processing techniques to a new optical system. Okumura and colleagues built a lens system that transferred the optical path through a pair of fast-moving mirrors that could thus rapidly shift the field of view and they named their system the "saccade mirror" in analogy to saccadic human eye movements (Okumura et al. 2011). By actuating only mirrors, the authors eliminated the mass of the optics and sensors that would have otherwise been moved and thus increased the motion speed possible. As motors, voice coil galvanometers were used to rotate the mirrors and shift the field of view by 30° in 3 ms. Custom optics shifted the pupil near the mirrors, enabling the system to have a wider field of view. This system was used to create stunning images of a rapidly moving ball during which the ball is fixed in the center of the frame while the background changes as a player hits it with a racquet (Okumura et al. 2011). Although not demonstrated, the authors proposed the interesting idea of acquiring images from multiple targets by rapidly shifting gaze between them. Shortly after the Okumura paper, a similar concept was employed by another group to make a high-resolution video of a flying *Muscina stabulans* fly, with a simpler optical system in which two cameras shared an optical path by combining views with a beam splitter and one camera was used for tracking while the other was used to record high-resolution videos (Sakakibara et al. 2012). Okumura and colleagues also later extended their "pupil shift" optical system to incorporate an additional camera sharing the optical path and this system was also capable of recording videos using alternative hardware than that used for tracking (Okumura et al. 2015).

In the middle-ground between the microscopic world of single celled organisms and macroscopic world of flying insects, the "Flyception" and "Flyception 2" systems have been developed to perform brain imaging in freely walking *Drosophila melanogaster* (Grover et al. 2016, 2020). Based on similar ideas as above, these systems use galvo mirrors and low-latency digital image processing to achieve a tracking lock of the head of a fly walking in a 4 cm arena. Using an additional camera and illumination system operating in a different part of the spectrum from the infrared wavelengths used for tracking, the authors were able to record epi-fluorescence images of genetically encoded calcium indicators and thus record brain activity during unrestrained behavior.

Beyond microscopes and galvo mirrors, these ideas have been explored using alternative motion platforms such as a pan-tilt motor gimbal system with a tracking algorithm that removed visual motion expected from background movement due to motor commands to detect and lock on to the subject, in this case an unmanned aircraft system (Doyle et al. 2014). An XY gantry driven with servo motors can be used to maintain a camera directly above a larval zebrafish swimming in a shallow tank (Johnson et al. 2020). Three-dimensional recording of freely flying insects was achieved through the use of an open frame onto which cameras and infrared illumination are mounted and then moved via motorized winches (Pannequin et al. 2020). Drones represent an attractive platform for automatic animal tracking using closed-loop image analysis, for example to combat poaching, but present a number of challenges (Olivares-Mendez et al. 2015).

In the last few years, the success of artificial neural networks trained via "deep learning" in computer vision has begun to displace the type of heuristic-based image analysis approaches used by the works cited above in this section. While many neural networks are too computationally heavy to be used for low-latency tracking, there nevertheless are several software systems optimized for reduced latency such as the "You Only Look Once" (YOLO) framework (Redmon et al. 2016; Redmon and Farhadi 2018). This YOLO framework is used in a recent microscope-mounted tracking system for *Caenorhabditis elegans* (Dong et al. 2020). In the past few years, drones sold for the consumer and professional videography market have introduced image-based tracking called a "follow me" mode. Although the principles of operation of these systems are proprietary, they command the drone's flight trajectory based on real-time image processing, likely via the use of neural networks, and thus also qualify here as closed-loop systems.

## Closed-loop systems based on photodiode and photomultiplier sensors

Due to their almost ubiquitous use for many purposes, we introduced digital cameras as the first sensor type used for automatic animal tracking with a moving camera system. However, they were not the first technology employed, nor are they the fastest. In 1971, Howard Berg published a system for tracking *Escherichia coli* bacteria in 3D that achieved tracking lock at the focus of a microscope (Berg 1971). This system was the basic technology used to collect the data for the important run and tumble model of bacterial chemotaxis (Berg and Brown 1972). In this case, the center of the microscope optical path was focused at the center of four fiber optic fibers arranged in a cross. One photomultiplier tube at the focus of each optic fiber transduced light to electrical signals with negligible delay and Berg's system detected deviation from the center of the microscope system as an imbalance of the luminance signal across the arms of the cross. A similar opto-electric system collected light from focal planes slightly above and below the primary focal plane of the microscope and was used to adjust focus. Based on these XYZ error signals computed using analog electronics from the raw photomultiplier outputs, the microscope stage was moved using linear voice coil motors.

A related system based on quadrant photodiodes—a prepackaged array of four photodiodes that can be directly placed at the focus of an optical system—was used to track *C. elegans* in a microscope. This system separated light into spectral regions and optical paths to enable simultaneous lock on tracking and calcium imaging based recording of neuronal activity in an unrestrained worm (Faumont et al. 2011) and is the historical precedent for brain imaging based on automatic tracking of a freely moving animal. A derivative of this system is now commercially available as the Phototrack device from Applied Scientific Instrumentation. Although not described in the publication, the commercial website describing the system describes an astigmatism-based autofocus system (http://asiimaging.com/docs/phototrack_tracking_theory). Using astigmatism to compute depth based on photodiode output is an optical technique used in CD players and microscopy (Pohlmann 1992; Li et al. 2012).

Photodiodes have also been used in the macroscopic domain. A system based on quadrant photodiodes as the sensor and galvo-actuated mirrors was used in conjunction with a spectrally separate optical system to make videos of freely flying insects onto which a tracking lock had been established (Vo-Doan and Straw 2020). Due to the use of fast photodiodes, the dominant source of latency in this system was the movement of the galvo mirrors, and the entire system was characterized to have total latency in the millisecond range.

### Closed-loop systems based on animal-mounted radio emitters

Commercially available camera-carrying drones can film while automatically pointing the camera at a "beacon" consisting of a GPS location receiver and a wireless network access point. Based on data from the beacon and an onboard GPS receiver, the drone-mounted camera is steered to record video of the beacon. This approach avoids the challenges of image-based processing while introducing requirements of good GPS coverage and animals large enough to carry such a transmitter. Smaller radio transmitters that transmit only a carrier signal can be tracked from a drone if the receiving system can provide bearing or range signals. This approach has recently been validated to track iguanas in field conditions (Hui et al. 2021), although so far no drone-mounted camera was used to image the animal carrying the transmitter.

### Open-loop tracking systems that direct moving cameras

In contrast to the closed-loop systems described above where motion commands come from sensors looking through the moving optical path, open-loop designs use an alternative sensor whose output drives the motor system with camera. Such an open-loop system thus requires a calibration between the sensor used for tracking and the motors used to localize the high-magnification camera. Iosifidis and colleagues built a human tracking system in which live image analysis from a wide-angle stationary camera was used to command a high-magnification PTZ system (Iosifidis et al. 2011). A similar idea was used in the "Fly Mind Alteration Device" in which a freely walking *Drosophila* fly was tracked via a stationary camera. The low-latency tracking information was used to command a galvo mirror system to align a laser and high-magnification system to activate neurons through thermo- and opto-genetics while recording high-resolution videos (Bath et al. 2014). An open-loop system based on large motorized mirrors with commands from a separate wide-area camera was used to obtain high-resolution videos of freely flying birds at an airport (Hatori et al. 2016). Here, too brain imaging has made an appearance with the system by Kim and colleagues aiming a microscope objective at the brain of a freely swimming zebrafish larvae through the use of a stationary infrared camera that drives movement of an XY stage with a nearly planar swimming arena (Kim et al. 2017). Moving beyond 2D, when the terrain of the animal and its movement is more complex, a multi-camera 3D tracking system can be used as the source of motor commands. This was the approach used by Nourizonoz and colleagues to track mice and mouse lemurs in their EthoLoop system that commands servo motors and an electronic tunable lens to control focus also based on low-latency 3D tracking data (Nourizonoz et al. 2020). Earlier work adjusted focus of a high-magnification, high frame rate camera to record a high-resolution video of a landing fly (van Breugel and Dickinson 2012) based on input commands from a low-latency 3D tracking system (Straw et al. 2011).

## Discussion

Above we provided a brief review of the technology of camera systems that automatically follow a moving organism with a moving optical system by use of automated tracking. Both closed-loop and open-loop designs have been used to automatically control the motion of an optical system based on low-latency sensor output. Such systems allow bypassing the geometric and physical tradeoffs associated with stationary cameras and thus enable the study of animal movement in ways not otherwise possible.

One limitation of the above automatic tracking systems is that most of these systems operate within a laboratory environment. While laboratories are convenient for methodological development, many important animal behaviors do not occur indoors. Hence, one clear direction for future research is to move these systems outdoors. We can anticipate this will happen soon, because computer vision techniques for object detection and tracking in complex environments have made extremely rapid progress in the last few years. Historically, the heuristic based image-processing algorithms used in most of the camera-based tracking systems described above have difficulty when confronted with dynamic outdoor scenes. Uncontrolled movement of leaves from wind, time-varying illumination conditions, and visually complex backgrounds all make these traditional algorithms challenging or impossible to use outdoors. Because the recently developed deep learning approaches are more robust in outdoor conditions, we might expect progress here if latency from such image processing can be short enough. As discussed above, one system is already able to close the feedback loop from a camera using an artificial neural network for processing images and ultimately to move a microscope stage to track a worm in the lab (Dong et al. 2020). It seems likely that similar systems will soon take advantage of the increased image processing robustness offered by such neural networks to track animals with moving cameras in an outdoor situation. Commercial drones already offer image-based tracking to record videos of humans and vehicles. Drones and other fully mobile platforms offer dramatically more flexibility for camera movement than the other systems discussed here, and developments in this space offer great promise.

For an ideal, outdoor-capable system, we could imagine capabilities that have not yet been demonstrated. To study animal communication and collective behavior, it would be useful to track more than one animal simultaneously. Okumura et al. (2011) proposed to do this by rapidly switching a single camera system between multiple targets sequentially in time. For a single closed-loop system, this would require substantial refinement of the real-time processing to maintain estimates of animal position while gaze was directed elsewhere. An alternative approach might be to use multiple individual systems operating in parallel but targeting different individuals.

Multiple systems operating in parallel could also be used on a single individual to provide additional perspectives, literally, on the animal's behavior. Such multiple views would facilitate reconstructing three-dimensional representations of the animal's position and pose. Fusing information from multiple perspectives is a subject that has received much attention in the field of photogrammetry (Hartley and Zisserman 2004). Most of this could be adapted to moving cameras by considering camera calibrations as dynamic functions of time. In addition to providing additional information regarding the 3D position of animals, multiple cameras allow tracking despite occlusions in single camera views and to increase the volume over which the combined system functions (Straw et al. 2011). With appropriate software, similar benefits could be expected with moving viewpoints, as well.

Regarding three-dimensional information, even a single camera view can provide quantitative three-dimensional position estimates when viewing objects of known dimension (Hartley and Zisserman 2004). Other optical techniques for obtaining depth information might also be employed. Astigmatism introduces different focal lengths along different image axes and is used for focus control in some applications as described above. Lidar (light detection and ranging) is a rapidly developing technology seeing use in mobile phones and automobiles that enables generating depth data using an active light source and light sensor and measuring the time of flight of light. Both astigmatism and lidar offer opportunities for future improvement from even a single viewpoint.

Beyond moving outdoors, other further innovations in this space appear possible in the next few years. While automated pose estimation from pre-recorded and real-time video is being facilitated by several recent software packages (Mathis et al. 2018; Graving et al. 2019; Pereira et al. 2019; Kane et al. 2020), the tracking data output from these packages have so far not been integrated with movement data from automatically moving cameras to compute world coordinates of animals. Such a step seems feasible, and with approaches that reconstruct animal trajectories in addition to environmental surfaces (Haalck et al. 2020; Brady 2021), this could prove a productive future direction that would allow high-resolution views of animals as they travel over large areas while also building precise 3D models of those large areas. As mentioned, automatically following animals with drones would also be useful, and while there are a number of nontrivial technical and nontechnical challenges with such approaches such as how robust tracking is implemented, whether the drone itself may disturb animal behavior, and regulatory and social issues (Olivares-Mendez et al. 2015), the benefits of overcoming these challenges will be substantial.

One challenge confronting anyone wishing to employ these technologies is the financial and technical investment that must be made to get a system operational. With a single known exception mentioned above, none of the systems in this review are available commercially. Consequently, despite their potential utility in addressing scientific questions, potential users must decide to incur substantial risk to build a system themselves. The current trend in favor of releasing source code for software and detailed hardware instructions is helpful here and should be encouraged. For their work to be widely utilized, developers of such systems should also seek to facilitate reproducibility. However, without appropriate support from funding agencies and incentives for career advancement, it can be difficult for the developers to justify working on technical documentation and user support.

To conclude, while automatically tracking the animal with a camera has already opened up some exciting experimental capabilities, it is a technique with substantial potential to improve. Notably among these, automatically making high-resolution videos of animals moving outdoors remains a tantalizing but challenging goal that will likely be greatly assisted by recent advances in using artificial neural networks for computer vision and other technologies. Achieving this goal is likely to be worthwhile because it will allow new insights into animal behavior at multiple scales.

## Acknowledgments

The author wishes to thank T. Thang Vo-Doan for insightful conversations on the subject.

## Funding

This work was supported by internal funding from the Albert-Ludwigs-Universität Freiburg.

## Data availability statement

No new data were generated or analysed in support of this research.

## References

Bath DE, Stowers JR, Hörmann D, Poehlmann A, Dickson BJ, Straw AD. 2014. FlyMAD: rapid thermogenetic control of neuronal activity in freely walking *Drosophila*. Nat Methods 11:756–62.

Berg HC. 1971. How to track bacteria. Rev Sci Instrum 42:868–71.

Berg HC, Brown DA. 1972. Chemotaxis in *Escherichia coli* analysed by three-dimensional tracking. Nature 239:500–4.

Brady (2021) Three-dimensional measurements of animal paths usinghandheld unconstrained GoPro cameras and VSLAM software Bioinspiration & Biomimetics, 10.1088/1748-3190/abe346,

Dong S, Liu X, Li P, Tang X, Liu D, Kojima M, Huang Q, Arai T. 2020. Automated tracking system with head and tail recognition for time-lapse observation of free-moving *C. elegans*. In 2020 IEEE international conference on robotics and automation. ICRA. p. 9257–62.

Doussot C, Bertrand OJN, Egelhaaf M. 2021. The critical role of head movements for spatial representation during bumblebees learning flight. Front Behav Neurosci 14:1–16.

Doyle D, Jennings A, Black J. 2014. Optical flow background estimation for real-time pan/tilt camera object tracking. Measurement 48:195–207.

Faumont S, Rondeau G, Thiele TR, Lawton KJ, McCormick KE, Sottile M, Griesbeck O, Heckscher ES, Roberts WM, Doe CQ et al.. 2011. An image-free opto-mechanical system for creating virtual environments and imaging neuronal activity in freely moving *Caenorhabditis elegans*. PLoS One 6:e24666.

Graving JM, Chae D, Naik H, Li L, Koger B, Costelloe BR, Couzin ID. 2019. DeepPoseKit, a software toolkit for fast and robust animal pose estimation using deep learning. ELife 8:e47994.

Grover D, Katsuki T, Greenspan RJ. 2016. Flyception: imaging brain activity in freely walking fruit flies. Nat Methods 13:569–72.

Grover D, Katsuki T, Li J, Dawkins TJ, Greenspan RJ. 2020. Imaging brain activity during complex social behaviors in *Drosophila* with Flyception2. Nat Commun 11:1–10.

Haalck L, Mangan M, Webb B, Risse B. 2020. Towards image-based animal tracking in natural environments using a freely moving camera. J Neurosci Methods 330:108455.

Hartley R, Zisserman A. 2004. Multiple view geometry in computer vision. Cambridge: Cambridge University Press.

Hatori T, Hirota J, Sakakibara J. 2016. Automatic optical tracking of a flying bird. T Visual Soc Jpn 36:1–7.

Hedrick TL, Pichot C, Margerie E. 2018. Gliding for a free lunch: biomechanics of foraging flight in common swifts (*Apus apus*). J Exp Biol 221:jeb186270.

Hobbs PCD. 2009. Building electro-optical systems: making it all work. Wiley, Hoboken NJ.

Hui NT, Lo EK, Moss JB, Gerber GP, Welch ME, Kastner R, Schurgers C. 2021. A more precise way to localize animals using drones. J. Field Robot. 38.

Iosifidis A, Mouroutsos S, Gasteratos A. 2011. A hybrid static/active video surveillance system. Int J Optomechatron 5:80–95.

Johnson RE, Linderman S, Panier T, Wee CL, Song E, Herrera KJ, Miller A, Engert F. 2020. Probabilistic models of larval zebrafish behavior reveal structure on many scales. Curr Biol 30:70–82.

Kane GA, Lopes G, Saunders JL, Mathis A, Mathis MW. 2020. Real-time, low-latency closed-loop feedback using markerless posture tracking. ELife 9:e61909.

Kim DH, Kim J, Marques JC, Grama A, Hildebrand DGC, Gu W, Li JM, Robson DN. 2017. Pan-neuronal calcium imaging with cellular resolution in freely swimming zebrafish. Nat Methods 14:1107–14.

Lambert TJ, Waters JC. 2014. Assessing camera performance for quantitative microscopy. In: Waters JC, Wittman T, editors. Methods in cell biology. New York: Academic Press. p. 35–53.

Li L, Kuang C, Luo D, Liu X. 2012. Axial nanodisplacement measurement based on astigmatism effect of crossed cylindrical lenses. Appl Opt 51:2379–87.

Margerie E, Simonneau M, Caudal J-P, Houdelier C, Lumineau S. 2015. 3D tracking of animals in the field using rotational stereo videography. J Exp Biol 218:2496–504.

Mathis A, Mamidanna P, Cury KM, Abe T, Murthy VN, Mathis MW, Bethge M. 2018. DeepLabCut: markerless pose estimation of user-defined body parts with deep learning. Nat Neurosci 21:1281.

Muybridge E. 1878. The horse in motion.

Nourizonoz A, Zimmermann R, Ho CLA, Pellat S, Ormen Y, Prévost-Solié C, Reymond G, Pifferi F, Aujard F, Herrel A et al. 2020. EthoLoop: automated closed-loop neuroethology in naturalistic environments. Nat Methods 17:1052–9.

Ogawa N, Oku H, Hashimoto K, Ishikawa M. 2005. Microrobotic visual control of motile cells using high-speed tracking system. IEEE Trans Rob 21:704–12.

Oku H, Ishikawa M, Ogawa N, Shiba K, Yoshida M. 2008. How to track spermatozoa using high-speed visual feedback. Annu Int Conf IEEE Eng Med Biol Soc 2008:125–8.

Okumura K, Oku H, Ishikawa M. 2011. High-speed gaze controller for millisecond-order pan/tilt camera. IEEE international conference on robotics and automation 2011:6186–91.

Okumura K, Yokoyama K, Oku H, Ishikawa M. 2015. 1 ms Auto Pan-Tilt—video shooting technology for objects in motion based on Saccade Mirror with background subtraction. Adv Robot 29:457–68.

Olivares-Mendez MA, Fu C, Ludivig P, Bissyandé TF, Kannan S, Zurad M, Annaiyan A, Voos H, Campoy P. 2015. Towards an autonomous vision-based unmanned aerial system against wildlife poachers. Sensors 15:31362–91.

Pannequin R, Jouaiti M, Boutayeb M, Lucas P, Martinez D. 2020. Automatic tracking of free-flying insects using a cable-driven robot. Sci Robot 5:eabb2890.

Pereira TD, Aldarondo DE, Willmore L, Kislin M, Wang SS-H, Murthy M, Shaevitz JW. 2019. Fast animal pose estimation using deep neural networks. Nat Methods 16:117.

Pohlmann KC. 1992. The compact disc handbook. A-R Editions, Inc, Madison WI.

Redmon J, Divvala S, Girshick R, Farhadi A. 2016. You only look once: unified, real-time object detection. ArXiv:1506.02640.

Redmon J, Farhadi A. 2018. YOLOv3: an incremental improvement. ArXiv:1804.02767.

Sakakibara J, Kita J, Osato N. 2012. Note: high-speed optical tracking of a flying insect. Rev Sci Instrum 83:036103.

Straw AD, Branson K, Neumann TR, Dickinson MH. 2011. Multi-camera realtime 3D tracking of multiple flying animals. J R Soc, Interface 8:395–409.

van Breugel F, Dickinson MH. 2012. The visual control of landing and obstacle avoidance in the fruit fly *Drosophila melanogaster*. J Exp Biol 215:1783–98.

Vo-Doan TT, Straw AD. 2020. Millisecond insect tracking system. ArXiv:2002.12100.

Wachowski L, Wachowski L. 1999. The matrix. Warner Bros., Los Angeles CA.