



HAL
open science

Automated image-based tracking and its application in ecology

Anthony I Dell, John A Bender, Kristin Branson, Iain D Couzin, Gonzalo G de Polavieja, Lucas P.J.J. Noldus, Alfonso Pérez-Escudero, Pietro Perona, Andrew D Straw, Martin Wikelski, et al.

► **To cite this version:**

Anthony I Dell, John A Bender, Kristin Branson, Iain D Couzin, Gonzalo G de Polavieja, et al.. Automated image-based tracking and its application in ecology. *Trends in Ecology & Evolution*, 2014, 29 (7), pp.417-428. 10.1016/j.tree.2014.05.004 . hal-04440000

HAL Id: hal-04440000

<https://hal.science/hal-04440000v1>

Submitted on 5 Feb 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Automated image-based tracking and its application in ecology

Anthony I. Dell^{1*}, John A. Bender², Kristin Branson³, Iain D. Couzin⁴, Gonzalo G. de Polavieja⁵, Lucas P.J.J. Noldus⁶, Alfonso Pérez-Escudero⁵, Pietro Perona⁷, Andrew D. Straw⁸, Martin Wikelski^{9,10}, and Ulrich Brose¹

¹ Systemic Conservation Biology, Department of Biology, Georg-August University Göttingen, Göttingen, Germany.

² HasOffers Inc., 2220 Western Ave., Seattle, WA, USA.

³ Howard Hughes Medical Institute, Janelia Farm Research Campus, Ashburn, VA, USA.

⁴ Department of Ecology and Evolutionary Biology, Princeton University, Princeton, NJ, USA.

⁵ Instituto Cajal, CSIC, Av. Doctor Arce, 37, Madrid, Spain.

⁶ Noldus Information Technology BV, Nieuwe Kanaal 5, 6709 PA Wageningen, The Netherlands.

⁷ Computation and Neural Systems Program, California Institute of Technology, Pasadena, CA, USA.

⁸ Research Institute of Molecular Pathology (IMP), Vienna, Austria.

⁹ Max Planck Institute for Ornithology, Radolfzell, Germany.

¹⁰ Biology Department, University of Konstanz, Konstanz, Germany.

* Corresponding author: adell@gwdg.de

Abstract

The behavior of individuals determines the strength and outcome of ecological interactions, which drive population, community, and ecosystem organization. Bio-logging, such as telemetry and animal-borne imaging, provides essential individual viewpoints, tracks, and life histories, but requires capture of individuals and is often impractical to scale. Recent developments in automated image-based tracking offers opportunities to remotely quantify and understand individual behavior at scales and resolutions not previously possible, providing an essential supplement to other tracking methodologies in ecology. Automated image-based tracking should continue to advance the field of ecology by enabling better understanding of the linkages between individual and higher-level ecological processes, via high-throughput quantitative analysis of complex ecological patterns and processes across scales, including analysis of environmental drivers.

Measuring behavior

Individual behavior underlies almost all aspects of ecology [1-5]. Accurate and highly resolved behavioral data are therefore critical for obtaining a mechanistic and predictive understanding of ecological systems [5]. Historically, direct observation by trained biologists was used to quantify behavior [6, 7]. However, the extent and resolution to which direct observations can be made is highly constrained [8] and the number of individuals that can be observed simultaneously is small. In addition, an exact record of events is not preserved, only the

43 biologist's subjective account of them.

44 Recent technological advances in tracking now make it possible to collect large amounts
45 of highly precise and accurate behavioral data. For many organisms equipment can be attached
46 that provide information about the individuals spatiotemporal position, orientation, and
47 physiology. This 'bio-logging' allows remote reconstruction of behavior over large
48 spatiotemporal extents, providing essential individual viewpoints, tracks and life histories, and
49 thus important ecological and evolutionary insights [9-11]. Image-based tracking, for example
50 with video, is another tracking method that shows great potential in ecology. Like bio-logging,
51 image-based tracking involves digital recording of data. This means an objective view of events
52 is maintained, increasing repeatability of studies, and allowing biologists to mine data for
53 quantities not originally considered. Image-based tracking can be used when individuals are too
54 small to attach bio-loggers, or if the equipment itself changes behavior, and it can track all
55 visible and sufficiently resolved individuals within the imaged area, not just those with loggers
56 attached. Also, image-based tracking generally allows for higher spatiotemporal resolution of
57 behavioral data than bio-logging, and many imaging methods allow extraction of quantitative
58 information about the environment, such as its temperature, topography or the presence of
59 physical structure. In reality, constraints on the acquisition, processing and storage of digital
60 information currently limits image-based tracking to smaller spatiotemporal extents, and
61 extracting the position and pose of every individual in each image is difficult in complex habitat
62 or at high densities. Nonetheless, these constraints are rapidly being overcome and image-
63 based tracking now provides a valuable tool for ecologists to undertake rigorous hypothesis-
64 driven research (Box 1).

65 Here we review the state-of-the-art of image-based tracking, its strengths and limitations
66 when applied to ecological research, and its application to solve relevant ecological questions.
67 Techniques to determine whether a particular organism is present in an image have been
68 thoroughly reviewed recently [12], so we focus on the problem of tracking individuals as they
69 move between images.

70

71 **Automated image-based tracking**

72 Initial applications of image-based tracking required manual analysis [13, 14], which is
73 extremely effort intensive, often leads to poor spatiotemporal resolution, and is open to observer
74 effects such as subjective decisions about which information to record. Recent advances in
75 automation are overcoming these issues [15-17], and there now exist several image-based
76 systems capable of extracting individual behavior with minimal or zero manual intervention
77 (Table S1). Tracking over ecologically relevant spatiotemporal scales is becoming easier, due to
78 advances in imaging and computing technologies, and by the development of software capable
79 of tracking in real-time [18-20] and of recognizing individuals across image-sequences [21, 22].
80 Biologists now employ a wide range of imaging methods (e.g., near infrared, thermal infrared,
81 sonar, 3D) that permit tracking in environments where optical video is unsuitable (Box 2). To
82 date, automated image-based tracking has primarily been undertaken in the laboratory, where
83 biologists have examined genetic and physio-chemical drivers of behavior in model species
84 (Table S1) (Box 1). However, the past decade has seen expansion of these methods into the

85 field, and automated image-based tracking has now been undertaken on a wide diversity of
86 species, including plants, worms, spiders, insects, fish, birds, mammals and more (Table S1).

87 Automated image-based tracking involves three main steps (Figure 1): i) acquisition of
88 image sequences (Box 2), ii) detection of individuals and their pose in each image and
89 appropriate 'linking' of detections in consecutive images to create trajectories through time (Box
90 3), and iii) analysis of behavioral data (Box 4). Real-time tracking is performed as images are
91 acquired, removing the need for storing large amounts of digital information [18-20] and allowing
92 researchers to influence the animal's environment in real-time through virtual reality, robotics or
93 other dynamical stimulus regimes [23-25]. Even under relatively controlled and simple
94 laboratory conditions with small numbers of individuals, automated image-based tracking is a
95 difficult computer vision problem. Biological organisms are highly deformable objects which
96 behave in unconstrained and variable ways [26] and the environmental landscapes within which
97 they exist are complex and dynamic.

98 Ultimately, in automated image-based tracking there is a trade-off between the difficulty
99 of the tracking problem (horizontal axis in Figure 2) and the quality of tracking output (vertical
100 axis in Figure 2).

101

102 **Difficulty of the tracking problem**

103 Tracking is easiest in laboratory-based systems with a simple environmental landscape
104 and low numbers of individuals (left side of Figure 2), and most difficult in the field, where many
105 individuals from many different species interact across a complex environmental landscape
106 (right side of Figure 2).

107

108 *From individuals to interactions*

109 Monitoring the behavior of individuals as they interact with each other is difficult for
110 several reasons. First, organisms often move rapidly when interacting (Movie S13), requiring
111 data with high spatiotemporal resolution. Second, because multiple individuals are involved,
112 interactions are prone to occlusions, made especially worse because interactions often involve
113 close physical contact. Occlusions cause identity errors, which are not local but propagate
114 throughout the remaining image sequence. Manual corrections of these errors are labor
115 intensive. Customized automated algorithms which predict identity based on the relative speed
116 and direction of movement of individuals can reduce mistakes, and thus dramatically reduce the
117 number of manual interventions needed [27, 28], but error propagation is still unavoidable
118 because of the stochastic behavior of organisms [16] (Box 3). 'Fingerprinting' somewhat
119 resolves this problem (see below), but maintaining identities always becomes more difficult as
120 the number of close individuals scales with increasing density. Tracking individuals during
121 occlusions is an additional problem, and can be partly overcome when prior knowledge about
122 the shape of the organisms is incorporated into the system [27-29]. Recent approaches utilizing
123 multiple 3D depth cameras are especially useful in this regard [30] (Movie S22), and could
124 eventually be integrated with fingerprinting to assist in resolving identities during occlusions.

125 Most current attempts to track multiple individuals involve organisms that are similar in
126 size and shape (Table S1). In nature, however, interactions between species often involve

127 individuals that differ greatly in size and shape [31] (Movie S13). While such differences can be
128 useful for distinguishing individuals [8, 21, 22], many tracking systems rely on knowledge about
129 the typical shape of individuals, to aid in the segmentation and analysis of images [28, 29, 32].
130 Even if shape issues are overcome, it remains a difficult task for computer vision algorithms to
131 separate small animals from the body and appendages of larger animals. Algorithm features
132 allowing tracking of differently sized and shaped organisms, such as more sophisticated contour
133 representations or fingerprinting, would greatly enhance the usefulness of image-based tracking
134 to ecologists.

135

136 *Tracking in three dimensions*

137 Automated image-based tracking in two-dimensional (2D) environments is substantially
138 more straightforward than in 3D (Figure 2). Therefore, many tracking systems are limited to
139 simple 2D arenas and either involve organisms that naturally move in 2D, or quasi-2D, or work
140 by constraining normally 3D individuals to only move in 2D. This latter method can be achieved
141 by modifying organisms directly, such as by wing-clipping [28], or by using physical boundaries
142 to constrain behavior to near 2D [1, 21, 22, 28, 33, 34] (Movie S1, Movie S4, Movie S5, Movie
143 S10). In nature, however, most organisms incorporate at least some degree of movement in 3D,
144 which influences ecological interactions [3]. Tracking systems designed for 2D can provide
145 some resolution for behavior in a third spatial dimension [35], but ultimately developers must
146 produce tracking systems that can successfully track large numbers of animals in 3D space
147 (Movie S8).

148 Tracking unconstrained flying or swimming animals can be achieved in several ways, but
149 most often multiple cameras are employed [19, 30, 36-46] (Movie S8, Movie S6, Movie S22).
150 Although only two calibrated cameras taking images of the same point in space are required for
151 triangulation, information from additional cameras can incrementally improve localization,
152 especially if some cameras are limited by occlusion or low contrast [19]. Synchronizing multiple
153 cameras requires additional hardware and more complicated software that relates equivalent
154 objects between image sequences; however, this complexity can be hidden from the user by
155 dedicated multi-camera systems [19]. Triangulation is optimized when cameras are positioned
156 with maximally divergent locations, which in the field can introduce problems because arranging
157 unobstructed cameras at multiple locations can be difficult, as can be obtaining multiple views of
158 every location of interest.

159 Some technologies allow 3D tracking from a single imaging device, which could solve
160 many of these issues. For example, 3D images can be reconstructed from a single image of
161 reflections or shadows on a 3D surface [47, 48], although this is computationally challenging
162 and certainly some time away from use in tracking multiple moving targets. Other more recent
163 and promising developments in hardware are single-point 3D imaging technologies. RGB-D
164 (red, green, blue, depth) cameras, such as the Microsoft Kinect ([www.microsoft.com/en-
165 us/kinectforwindows/](http://www.microsoft.com/en-us/kinectforwindows/)), achieve this by combining a color video camera either with an infrared
166 projector to create a split infrared laser light field from which depth can be obtained (first
167 generation) or by using time of arrival of the photons themselves (second generation). Light-field
168 video cameras provide another promising technology (Box 2), where composite optics are used

169 to simultaneously capture images focused at multiple distances from the lens, thus allowing for
170 post-hoc selection of focus and ultimately 3D construction of the scene. As in 2D, multiple 3D
171 imaging cameras can be employed simultaneously to provide additional resolution and to cope
172 with occlusion [30].

173

174 *Into the field...*

175 Ecological systems are naturally embedded within environmental landscapes that are
176 considerably more complex than laboratory arenas, such as within streams, coral reefs, or the
177 forest floor (Movie S13). Although salient questions can be addressed in the laboratory, it is
178 critical that tracking can be undertaken in the field because environmental drivers in their natural
179 context—such as light, temperature, physical habitat, and spatial dimensionality—have
180 profound influences on behavior and thus ecosystem organization [1, 3, 49, 50] (Box 1). Many
181 of the techniques that enable automated image-based tracking in the field are similar to those
182 that enable 3D tracking, such as multiple cameras and single-point 3D imaging devices (see
183 above).

184 One of the primary constraints in the field is the ability to distinguish individuals within
185 each image from the background (Box 2), which often varies unpredictably due to such factors
186 as wind, water, and sunlight. The simplest method to track in complex environmental
187 landscapes is to employ an imaging method that provides clear contrast between the organisms
188 and the background. The growing number of imaging technologies now available (Box 2) means
189 there is a corresponding growing range of environmental contexts within which individuals can
190 be tracked. For example, it is now straightforward to image independently of visible light (Box 2),
191 meaning that tracking is no longer limited to environments with sufficient and homogenous
192 visible light [1, 4, 51-53] (Box 2). Another alternative is to use computer vision technologies that
193 detect animals even when their color pattern is statistically indistinguishable from the
194 background, based for example on their shape or movement [12]. Finally, it is possible to mark
195 individuals [54] or integrate with other tracking methods such as bio-logging – combining the
196 robustness of bio-loggers for detecting individuals in complex habitat with the high
197 spatiotemporal resolution of imaging [55].

198 Physical structure, like plant cover or soil, is more difficult to track within because it
199 increases the number of occlusions. Fingerprinting allows the addition of habitat structure
200 without increasing assignment errors, as identities are recovered following occlusion [12, 21, 22]
201 (Movie S5). Again, use of multiple cameras (Movie S6), marking [54], or integration with bio-
202 logging [55] can enable tracking and identity maintenance during or after occlusions. Some
203 imaging methods can even pass through physical structure, revealing the position of organisms
204 either behind the structure or embedded within it, thus removing the problem of occlusions from
205 physical habitat altogether. The behavior of small invertebrates within soil, for example, has
206 been successfully quantified using high-resolution X-ray microtomography, which works
207 because biological tissue attenuates X-rays less than the surrounding soil matrix [56] (Box 2).
208 Acoustic imaging (sonar) also permits imaging through relatively complex habitat, and is
209 especially effective in aquatic environments. Modern high-resolution sonar has allowed
210 biologists to investigate predator-prey interactions in habitats that would be impossible with

211 other imaging methods [4] (Movie S9) (Box 2). Additional technologies for imaging through
212 complex physical habitat are on the horizon [57], although these are still likely years away from
213 being successfully integrated into automated tracking systems.

214 For many ecological questions it is necessary to obtain quantitative information about
215 the environmental landscape, which can be integrated with tracking data to understand how the
216 environment influences behavior [1, 5, 18, 58] (Box 1). Remote quantification of the environment
217 is a key advantage of imaging over bio-logging, which only provides environmental information
218 in the immediate vicinity of the individual to which the logger is attached. Remote quantification
219 of the environment can easily be accomplished by imaging in the appropriate sensory regime,
220 such as optical video cameras for quantifying light conditions and thermal cameras for
221 quantifying the thermal landscapes. Methods for quantifying the physical structure of 3D
222 landscapes are rapidly advancing [59-61], and can be used for rendering features of natural
223 habitats, such as trees or streams. When combined with behavioral data, this environmental
224 information should allow biologists to represent an animal's cognitive map of its environment,
225 and thus understand the relationship between behavior and fitness [62]. This should be
226 especially rewarding when combined with methods that reconstruct the sensory fields of
227 individuals, providing knowledge about the sensory information on which animals base
228 decisions [27].

229

230 **Quality of the tracking output**

231 Ideally, the final output of tracking is the trajectory of each individual, spanning the entire
232 image sequence and including detailed information about body posture and positioning of
233 appendages. Realistically, however, this is a difficult outcome to obtain. We recognize two
234 primary factors determining the quality of the tracking output: i) how well identities are
235 maintained throughout the image sequence, and ii) whether only the mid-point point or the
236 detailed body posture of each individual is tracked.

237

238 *Identity maintenance*

239 We recognize three broad categories for how well identity is maintained by automated
240 tracking systems (vertical axis in Figure 2). In the first category, identity is not maintained
241 following occlusion, and instead new trajectories are produced each time a new individual is
242 recognized (top row in Figure 2). In the second category, algorithms link identities across
243 occlusions, based, for example, on the predicted movement of individuals (middle row in Figure
244 2). When the number of individuals is not constant, as in the field, or if occlusions are too
245 complex to link trajectories, the output can become similar to the first category. To prevent
246 identity switch errors the researcher must manually review uncertain events in the tracking data
247 and make appropriate corrections (see above) [28, 29, 34, 63] (Movie S1, Movie S10, Movie
248 S11).

249 In the third scenario, each trajectory always belongs to a single individual (bottom row in
250 Figure 2), similar to bio-logging. Here, individual organisms are recognized in each image, so
251 that following occlusion correct identities are always maintained and identity errors never
252 propagate. This often involves application of artificial markings, however natural variation in the

253 morphology of individuals can also be used to maintain identities throughout image sequences,
254 even following occlusion (Table S1). The simplest method involves using very general traits to
255 identify individuals, such as body size or body shape. General traits can be sufficient for
256 maintaining identities at low densities or when individuals vary greatly in size or shape, but in
257 many other instances in ecology individuals are likely to be similarly sized or shaped.
258 'Fingerprinting' uses a more comprehensive set of traits to recognize individuals, even when
259 individuals are indistinguishable to the human eye [12, 21, 22] (Table S1, Movie S5).
260 Fingerprinting is currently limited to small numbers of individuals and controlled laboratory
261 conditions, and while these limitations will be reduced as image quality increases and
262 fingerprinting is combined with other segmentation and recognition methods [12, 64-69],
263 untimely fingerprinting will always be limited for very large groups due to inevitable 'overlaps' in
264 fingerprints.

265

266 *Spatiotemporal position or detailed pose?*

267 Knowledge about the spatiotemporal position of organisms relative to the environmental
268 landscape, or to each other, is sufficient for many questions in ecology [1, 3, 4, 34, 54] (Box 1).
269 For other questions, such as those about the mechanics of locomotion or ecological
270 interactions, it is necessary to know about the positioning of appendages or other specific points
271 along an individual's body [8, 18, 20, 29, 30, 70] (Movie S2, Movie S11, Movie S14, Movie S22)
272 (Box 1). Estimating the center of body mass (position) is much simpler than detecting the
273 detailed body posture and position of appendages (pose). Obtaining detailed pose information is
274 not only a technically difficult computer vision problem, but requires higher spatial resolutions
275 which brings with it associated costs, such as reduced spatial extent of imaging and increased
276 data management and processing requirements [18].

277 Many tracking systems automatically identify individuals in images by fitting contour
278 models to foreground pixels, once they have been isolated from the background (Box 3). These
279 contour models can be simple, providing position and sometimes orientation [28] (Movie S1,
280 Movie S3), or they can be more complex and thus provide detailed information about body
281 posture or the position of legs, wings, or tails [8, 20, 29, 30, 69, 71] (Movie S2, Movie S11,
282 Movie S14, Movie S18, Movie S22) (Box 3). Automated tracking systems that use approaches
283 other than contour fitting [18, 21, 22, 69] (Movie S5, Movie S17) are more generalizable to
284 different sized and shaped organisms, a key feature ecologists will ultimately require from
285 tracking systems.

286

287 **Automated behavioral analysis**

288 In addition to estimating the position and pose of individuals, automated tracking
289 systems can also analyze individual and between-individual behaviors (behavior between
290 individuals is analyzed in much the same way as for single individuals, except the context
291 becomes behavioral correlations between individuals) (Box 4). Automatically annotating
292 behavior produces large quantities of consistently defined and highly resolved behavioral data,
293 providing biologists with unprecedented power to quantitatively understand general mechanisms
294 and principles underlying behavior [69, 72] (Box 1). Automated behavioral analysis is possible

295 with trajectory information alone, such as differentiating between an individual being stationary,
296 walking, or running, or with more detailed pose information, such as head position, contour
297 shape, or appendage position [8, 63, 69, 73] (Movie S2, Movie S14, Movie S11).

298 Identification of behavior into categories is a long-standing tradition in ethology and can
299 be achieved by automated behavioral phenotyping, where the complex trajectory and pose data
300 output by automated tracking is categorized into simpler, interpretable categories that best
301 characterize biologically and ecologically relevant behavioral phenotypes (Box 4). Automated
302 behavioral phenotyping can be undertaken with either supervised or unsupervised machine
303 learning, with the core goal of both approaches being to condense the very rich and complex
304 trajectory and pose data into a simpler form that is biologically and ecologically relevant (Box 4).

305 In supervised learning, the human expert identifies and categorizes patterns in the data
306 by informing the software of categorical behavior annotations [8, 16, 72, 74] (Movie S11, Movie
307 S12, Movie S14). For example, 'wing grooming' could represent when a fly rubs one or both
308 metathoracic legs over the top or the underside of the wing(s) [74]. Categorical annotations like
309 these simultaneously take into account many different features of the trajectory and pose data,
310 and result in categories that are generally easily interpretable by biologists and often have a
311 clear physiological or ecological significance. Besides allowing higher throughput than manual
312 annotation, the subjective *a priori* definitions of behavior chosen by a researcher can be
313 expressed in precise mathematical terms and, once behaviors are defined, the analysis of any
314 dataset is fully repeatable.

315 Unsupervised learning methods, on the other hand, apply computational techniques to
316 the raw data themselves to reveal what degrees of freedom are relevant in the data and then
317 automatically detect any stereotyped patterns (Movie S12) [69]. These behavioral patterns
318 might, or might not, be evident to human observers. Unsupervised techniques offer the
319 advantage of decreased subjectivity, and increased throughput, repeatability, and the chance of
320 finding rare behaviors [69, 75, 76]. Unsupervised methods use statistical methods to reduce the
321 dimensionality of trajectory and pose data, and of course these statistical methods themselves
322 depend on mathematical, human-generated assumptions inherent in the algorithms. It is
323 therefore critical to compare the output of these unsupervised methods to what biologists
324 already know about the inherent structure of behavior. As unsupervised methods become
325 increasingly powerful and more objective [69] they will become an important development for
326 community ecologists, who often study many different species which otherwise would require
327 manual behavioral labeling across taxa, which is time consuming and prone to inaccuracies.

328

329 **Summary and future directions**

330 Automated image-based tracking provides detailed information about the behavior of
331 individuals at local scales, such as how they move, with which other individuals they interact,
332 the sensory information available to them, and the role of internal and environmental drivers in
333 shaping their behavior (Box 1). This information should prove integral in mapping linkages
334 between genes, brain function, behavior, and species interactions (thus linking molecular
335 biology, neurobiology and ecology). Although development of automated image-based tracking
336 has been primarily laboratory-based, focused on model organisms in controlled conditions,

337 studies using this method now exist from a diversity of taxa and habitats, including in the field
338 (Table S1, Figure 2) (Box 1). Automation of data collection permits high levels of replication,
339 substantially increasing the amount and quality of behavioral data available to biologists. For
340 example, automated tracking of behavior within functional response experiments which
341 ecologists use to quantify species interaction strength, would allow researchers to run large
342 numbers of replicate trials where the behavior of every predator and prey is known, providing
343 unparalleled power to uncover novel patterns in the functional response data [77].

344 Like all technologies, automated image-based tracking has its limitations. Only
345 individuals within the imaged area can be tracked – meaning larger animals cannot be tracked
346 over their entire home range – and tracking in complex physical habitat or at high densities is
347 difficult. In addition, the storage and management issues that arise from the huge amounts of
348 digital data that are easily produced by imaging must be addressed. Emerging methods, such
349 as fingerprinting or real-time tracking, alleviate this problem somewhat [21, 22], but more needs
350 to be done. As limitations are overcome, and imaging and computational technologies advance,
351 automated image-based tracking should become firmly established as a powerful tool for
352 quantitative research in ecology [9]. These methods are already providing conceptual advances
353 on diverse topics like predator-prey interactions, collective behavior of animal groups, social
354 hierarchy, and population dynamics, and their continued application will only broaden this list of
355 topics (Box 1, Table S1). Eventually, automated tracking should influence the field of ecology
356 similarly to how it is influencing the genetic and behavioral sciences [5], allowing ecologists to
357 uncover mechanisms and principles that shape ecological systems, leading to a more general
358 and predictive science of community ecology with significant basic and applied benefits.

359

360 **Acknowledgements**

361 We thank Paul Craze, Lena Rohde, Andrew Spink, and one anonymous reviewer for
362 their detailed comments that improved our review, and especially Alex Gomez-Marin whose
363 insightful and detailed comments significantly enhanced the clarity and quality of our review. We
364 thank Michael Dickinson (Box 2; Figure 1b), Nickolay Hristov, Louise Allen and Thomas Kunz
365 (Box 2; Figure 1c), Scott Johnson (Box 2; Figure 1d), Nils Handegard and Simon Leblanc (Box 2;
366 Figure 1e), Brian Hicks (Box 2; Figure 1f), and Steve Feller (Box 2; Figure 1g) for their images,
367 and those people who kindly contributed information and movies for our Supplementary
368 Information (names provided at the beginning of each movie). Figure 1d-e,i (Box 4) and Movie
369 S14 modified from [8] with permission from Nature Publishing Group. Figure 1g (Box 4) and raw
370 sonar clip in Movie S9 reprinted from [4] with permission from Elsevier. Figure 1 (middle panel)
371 and Figure 1m (Box 3) modified from [19]. Figure 1e-i, l, n (Box 3) and Figure 1d-f (Box 4)
372 modified from [28] with permission from Nature Publishing Group. Figure 1j (Box 3), Figure 1i
373 (Box 4), and Movie S11 reprinted from [29] with permission from Nature Publishing Group.
374 Movie S16 reprinted from [32] with permission from Elsevier. Movie S18 reprinted from [71].
375 Movie S22 reprinted from [30].

376

377 **Author contributions**

378 A.I.D conceived the project and wrote first draft of the paper; all other authors
379 contributed significantly. A.I.D made final editorial decisions, including the inclusion or exclusion
380 of any technique or vendor covered in this review.

381

382 **Conflict of interest**

383 L.P.J.J. Noldus is director of Noldus Information Technology BV, developer of one of the
384 commercial tracking systems described in our paper (Table S1).

385

386 **Supplementary Information**

387 Supplementary Information can be found online at [XXXX](#), consisting of Table S1, and
388 multimedia highlighting the imaging, tracking and analysis of a variety of species in a variety of
389 habitats.

390

391 **References**

- 392 1 Berdahl, A., *et al.* (2013) Emergent sensing of complex environments by mobile animal
393 groups. *Science* 339, 574-576
- 394 2 Simpson, S.J., *et al.* (2010) Modelling nutritional interactions: from individuals to
395 communities. *Trends Ecol Evol* 25, 53-60
- 396 3 Pawar, S., *et al.* (2012) Dimensionality of consumer search space drives trophic interaction
397 strengths. *Nature* 486, 485-489
- 398 4 Handegard, N.O., *et al.* (2012) The dynamics of coordinated group hunting and collective
399 information transfer among schooling prey. *Curr Biol* 22, 1213-1217
- 400 5 Nathan, R., *et al.* (2008) A movement ecology paradigm for unifying organismal movement
401 research. *Proc Natl Acad Sci U S A* 105, 19052-19059
- 402 6 Altmann, J. (1974) Observational study of behavior: sampling methods. *Behaviour* 49, 227-267
- 403 7 Tinbergen, N. (1963) On aims and methods of Ethology. *Zeitschrift für Tierpsychologie* 20,
404 410-433
- 405 8 Dankert, H., *et al.* (2009) Automated monitoring and analysis of social behavior in *Drosophila*.
406 *Nat Methods* 6, 297-303
- 407 9 Rutz, C. and Bluff, L.A. (2008) Animal-borne imaging takes wing, or the dawn of 'wildlife
408 video-tracking'. *Trends Ecol Evol* 23, 292-294; author reply 294
- 409 10 Cooke, S.J., *et al.* (2004) Biotelemetry: a mechanistic approach to ecology. *Trends Ecol Evol*
410 19, 334-343
- 411 11 Krause, J., *et al.* (2013) Reality mining of animal social systems. *Trends Ecol Evol* 28, 541-
412 551
- 413 12 Kuhl, H.S. and Burghardt, T. (2013) Animal biometrics: quantifying and detecting phenotypic
414 appearance. *Trends Ecol Evol* 28, 432-441
- 415 13 Elliott, J.P., *et al.* (1977) Prey capture by the African lion. *Canadian Journal of Zoology* 55,
416 1811-1828
- 417 14 Bayne, B.L. and Scullard, C. (1978) Rates of Feeding by Thais-(Nucella)-Lapillus (L).
418 *Journal of Experimental Marine Biology and Ecology* 32, 113-129
- 419 15 Anonymous (2007) Geneticist seeks engineer: must like flies and worms. *Nat Meth* 4, 463-
420 463

- 421 16 Balch, T., *et al.* (2006) How multirobot systems research will accelerate our understanding of
422 social animal behavior. *Proceedings of the IEEE* 94, 1445-1463
- 423 17 Noldus, L.P.J.J., *et al.* (2002) Computerised video tracking, movement analysis and behaviour
424 recognition in insects. *Comput Electron Agr* 35, 201-227
- 425 18 Gomez-Marin, A., *et al.* (2012) Automated tracking of animal posture and movement during
426 exploration and sensory orientation behaviors. *PLoS ONE* 7, e41642
- 427 19 Straw, A.D., *et al.* (2011) Multi-camera real-time three-dimensional tracking of multiple
428 flying animals. *J R Soc Interface* 8, 395-409
- 429 20 Swierczek, N.A., *et al.* (2011) High-throughput behavioral analysis in *C. elegans*. *Nat*
430 *Methods* 8, 592-598
- 431 21 Pérez-Escudero, A., *et al.* (2013) idTracker: Tracking individuals in a group by automatic
432 identification of unmarked animals. *In Review*
- 433 22 Pérez-Escudero, A. (2013) Optimization principles in neurobiolgy and collective animal
434 behaviour. PhD thesis (directed by G. G. de Polavieja). Universidad Autónoma de
435 Madrid, Spain
- 436 23 Krause, J., *et al.* (2011) Interactive robots in experimental biology. *Trends Ecol Evol* 26, 369-
437 375
- 438 24 Ofstad, T.A., *et al.* (2011) Visual place learning in *Drosophila melanogaster*. *Nature* 474,
439 204-207
- 440 25 Straw, A.D., *et al.* (2010) Visual control of altitude in flying *Drosophila*. *Curr Biol* 20, 1550-
441 1556
- 442 26 Branson, K. and Belongie, S. (2005) Tracking multiple mouse contours (without too many
443 samples). *IEEE Computer Vision and Pattern Recognition* 1, 1039-1046
- 444 27 Strandburg-Peshkin, A., *et al.* (2013) Visual sensory networks and effective information
445 transfer in animal groups. *Current Biology* In press
- 446 28 Branson, K., *et al.* (2009) High-throughput ethomics in large groups of *Drosophila*. *Nat*
447 *Methods* 6, 451-457
- 448 29 de Chaumont, F., *et al.* (2012) Computerized video analysis of social interactions in mice. *Nat*
449 *Methods* 9, 410-417
- 450 30 Matsumoto, J., *et al.* (2013) A 3D-Video-Based Computerized Analysis of Social and Sexual
451 Interactions in Rats. *PLoS ONE* 8, e78460
- 452 31 Brose, U., *et al.* (2006) Consumer-resource body-size relationships in natural food webs.
453 *Ecology* 87, 2411-2417
- 454 32 Ohayon, S., *et al.* (2013) Automated multi-day tracking of marked mice for the analysis of
455 social behaviour. *J Neurosci Methods* 219, 10-19
- 456 33 Simon, J.C. and Dickinson, M.H. (2010) A new chamber for studying the behavior of
457 *Drosophila*. *PLoS ONE* 5, e8793
- 458 34 Katz, Y., *et al.* (2011) Inferring the structure and dynamics of interactions in schooling fish.
459 *Proc Natl Acad Sci U S A* 108, 18720-18725
- 460 35 Robie, A.A., *et al.* (2010) Object preference by walking fruit flies, *Drosophila melanogaster*,
461 is mediated by vision and graviperception. *J Exp Biol* 213, 2494-2506
- 462 36 Lacey, E.S. and Carde, R.T. (2011) Activation, orientation and landing of female *Culex*
463 *quinquefasciatus* in response to carbon dioxide and odour from human feet: 3-D flight
464 analysis in a wind tunnel. *Med Vet Entomol* 25, 94-103
- 465 37 Spitzen, J., *et al.* (2013) A 3D analysis of flight behavior of *Anopheles gambiae sensu stricto*
466 malaria mosquitoes in response to human odor and heat. *PLoS ONE* 8, e62995

- 467 38 Cachat, J., *et al.* (2011) Three-dimensional neurophenotyping of adult zebrafish behavior.
468 *PLoS ONE* 6, e17597
- 469 39 Ballerini, M., *et al.* (2008) Interaction ruling animal collective behavior depends on
470 topological rather than metric distance: evidence from a field study. *Proc Natl Acad Sci*
471 *U S A* 105, 1232-1237
- 472 40 Attanasi, A., *et al.* (2013) Tracking in three dimensions via multi-path branching.
473 *arXiv:1305.1495*
- 474 41 Attanasi, A., *et al.* (2013) Superfluid transport of information in turning flocks of starlings.
475 *arXiv:1303.7097*
- 476 42 Butail, S., *et al.* (2012) Reconstructing the flight kinematics of swarming and mating in wild
477 mosquitoes. *J R Soc Interface* 9, 2624-2638
- 478 43 Fontaine, E.I., *et al.* (2009) Wing and body motion during flight initiation in *Drosophila*
479 revealed by automated visual tracking. *J Exp Biol* 212, 1307-1323
- 480 44 Viscido, S.V., *et al.* (2004) Individual behavior and emergent properties of fish schools: a
481 comparison of observation and theory. *Marine Ecology Progress Series* 273, 239-249
- 482 45 Veeraraghavan, A., *et al.* (2006) Motion based correspondence for 3D tracking of multiple
483 dim objects. *IEEE* 2
- 484 46 Attanasi, A., *et al.* (2013) Wild swarms of midges linger at the edge of an ordering phase
485 transition. *arXiv:1307.5631*
- 486 47 Kanbara, M., *et al.* (2006) 3D scene reconstruction from reflection images in a spherical
487 mirror. *18th International Conference on Pattern Recognition, Vol 4, Proceedings* 4,
488 874-879
- 489 48 Chen, Z.H., *et al.* (2011) Single-View Reconstruction from an Unknown Spherical Mirror.
490 *2011 18th Ieee International Conference on Image Processing (Icip)*, 2677-2680
- 491 49 Kalinkat, G., *et al.* (2013) Habitat structure alters top-down control in litter communities.
492 *Oecologia* 172, 877-887
- 493 50 Dell, A.I., *et al.* (2011) Systematic variation in the temperature dependence of physiological
494 and ecological traits. *Proc Natl Acad Sci U S A* 108, 10591-10596
- 495 51 Hristov, N.I., *et al.* (2008) Applications of thermal infrared imaging for research in
496 aeroecology. *Integr Comp Biol* 48, 50-59
- 497 52 Chiu, C., *et al.* (2010) Effects of competitive prey capture on flight behavior and sonar beam
498 pattern in paired big brown bats, *Eptesicus fuscus*. *J Exp Biol* 213, 3348-3356
- 499 53 Betke, M., *et al.* (2008) Thermal imaging reveals significantly smaller Brazilian free-tailed
500 bat colonies than previously estimated. *Journal of Mammalogy* 89, 18-24
- 501 54 Mersch, D.P., *et al.* (2013) Tracking individuals shows spatial fidelity is a key regulator of ant
502 social organization. *Science* 340, 1090-1093
- 503 55 Weissbrod, A., *et al.* (2013) Automated long-term tracking and social behavioural
504 phenotyping of animal colonies within a semi-natural environment. *Nat Commun* 4, 2018
- 505 56 Johnson, S.N., *et al.* (2007) Non-invasive techniques for investigating and modelling root-
506 feeding insects in managed and natural systems. *Agricultural and Forest Entomology* 9,
507 39-46
- 508 57 Bertolotti, J., *et al.* (2012) Non-invasive imaging through opaque scattering layers. *Nature*
509 491, 232-234
- 510 58 Gomez-Marin, A., *et al.* (2011) Active sampling and decision making in *Drosophila*
511 chemotaxis. *Nat Commun* 2, 441

- 512 59 Izadi, S., *et al.* (2011) KinectFusion: real-time 3D reconstruction and interaction using a
513 moving depth camera. *ACM*, 559-568
- 514 60 Newcombe, R.A., *et al.* (2011) KinectFusion: Real-Time Dense Surface Mapping and
515 Tracking. *10th IEEE International Symposium on Mixed and Augmented Reality*
- 516 61 Yang, X., *et al.* (2013) Three-dimensional forest reconstruction and structural parameter
517 retrievals using a terrestrial full-waveform lidar instrument (Echidna®). *Remote Sensing*
518 *of Environment* 135, 36-51
- 519 62 Cagnacci, F., *et al.* (2010) Animal ecology meets GPS-based radiotelemetry: a perfect storm
520 of opportunities and challenges. *Philos Trans R Soc Lond B Biol Sci* 365, 2157-2162
- 521 63 Buhl, J., *et al.* (2006) From disorder to order in marching locusts. *Science* 312, 1402-1406
- 522 64 Fergus, R., *et al.* (2003) Object class recognition by unsupervised scale-invariant learning.
523 *2003 Ieee Computer Society Conference on Computer Vision and Pattern Recognition,*
524 *Vol Ii, Proceedings 2*, 264-271
- 525 65 Dollár, P., *et al.* (2010) The fastest pedestrian detector in the west. *British Machine Vision*
526 *Conference*
- 527 66 Burghardt, T. and Calic, J. (2006) Real-time face detection and tracking of animals. In *Real-*
528 *time face detection and tracking of animals*, pp. 27-32, IEEE
- 529 67 Hannuna, S.L., *et al.* (2005) Identifying quadruped gait in wildlife video. In *Image*
530 *Processing, 2005. ICIIP 2005. IEEE International Conference on*, pp. I-713-716
- 531 68 Burgos-Artizzu, X.P., *et al.* (2012) Social behavior recognition in continuous video. *Proc*
532 *Cvpr Ieee*, 1322-1329
- 533 69 Berman, G.J., *et al.* (2013) Mapping the structure of drosophilid behavior. *arXiv*
- 534 70 Bender, J.A., *et al.* (2011) Kinematic and behavioral evidence for a distinction between
535 trotting and ambling gaits in the cockroach *Blaberus discoidalis*. *J Exp Biol* 214, 2057-
536 2064
- 537 71 Mirat, O., *et al.* (2013) ZebraZoom: an automated program for high-throughput behavioral
538 analysis and categorization. *Front Neural Circuits* 7, 107
- 539 72 van Dam, E.A., *et al.* (2013) An automated system for the recognition of various specific rat
540 behaviours. *J Neurosci Methods* 218, 214-224
- 541 73 Dollár, P., *et al.* (2010) Cascaded pose regression. *IEEE*, 1078-1085
- 542 74 Kabra, M., *et al.* (2013) JAABA: interactive machine learning for automatic annotation of
543 animal behavior. *Nat Methods* 10, 64-67
- 544 75 Brown, A.E., *et al.* (2013) A dictionary of behavioral motifs reveals clusters of genes
545 affecting *Caenorhabditis elegans* locomotion. *Proc Natl Acad Sci U S A* 110, 791-796
- 546 76 Braun, E., *et al.* (2010) Identifying prototypical components in behaviour using clustering
547 algorithms. *PLoS ONE* 5, e9361
- 548 77 Dell, A.I., *et al.* (In review) The role of individual behaviour in determining trophic
549 interaction strength.
- 550 78 Fontaine, E., *et al.* (2008) Automated visual tracking for studying the ontogeny of zebrafish
551 swimming. *J Exp Biol* 211, 1305-1316
- 552 79 Censi, A., *et al.* (2013) Discriminating external and internal causes for heading changes in
553 freely flying *Drosophila*. *PLoS Comput Biol* 9, e1002891
- 554 80 Maimon, G., *et al.* (2008) A simple vision-based algorithm for decision making in flying
555 *Drosophila*. *Curr Biol* 18, 464-470
- 556 81 Pizzo, A.B., *et al.* (2013) The membrane raft protein Flotillin-1 is essential in dopamine
557 neurons for amphetamine-induced behavior in *Drosophila*. *Mol Psychiatry* 18, 824-833

558 82 van Breugel, F. and Dickinson, M.H. (2012) The visual control of landing and obstacle
559 avoidance in the fruit fly *Drosophila melanogaster*. *J Exp Biol* 215, 1783-1798
560 83 Wu, Z., *et al.* (2011) A combinatorial semaphorin code instructs the initial steps of sensory
561 circuit assembly in the *Drosophila* CNS. *Neuron* 70, 281-298
562 84 Tunstrom, K., *et al.* (2013) Collective states, multistability and transitional behavior in
563 schooling fish. *PLoS Comput Biol* 9, e1002915
564 85 Herbert-Read, J.E., *et al.* (2011) Inferring the rules of interaction of shoaling fish. *Proc Natl*
565 *Acad Sci U S A* 108, 18726-18731
566 86 Fan, J., *et al.* (2010) Automatic video-based analysis of animal behaviors. *IEEE*, 1513-1516
567 87 Wang, L. and Anderson, D.J. (2010) Identification of an aggression-promoting pheromone
568 and its receptor neurons in *Drosophila*. *Nature* 463, 227-231
569 88 Spey, I.-K., *et al.* (In review) How behavior mediates the effect of temperature and body size
570 on use of a dispersal corridor. *PLoS ONE*
571 89 Brady, D.J., *et al.* (2012) Multiscale gigapixel photography. *Nature* 486, 386-389
572 90 Straw, A.D. and Dickinson, M.H. (2009) Motmot, an open-source toolkit for realtime video
573 acquisition and analysis. *Source Code Biol Med* 4, 5
574 91 Stuurman, N., *et al.* (2007) Micro-Manager: Open Source software for light microscope
575 imaging. *Microscopy Today* 15, 42-43
576 92 Noldus Information Technology (2013) Media Recorder. [http://www.noldus.com/media-](http://www.noldus.com/media-recorder)
577 [recorder](http://www.noldus.com/media-recorder)
578 93 Norpix (2013) StreamPix. <http://www.norpix.com/products/streampix/streampix.php>
579 94 Bar-Shalom, Y. and Fortmann, T.E. (1988) *Tracking and data association*. Academic Press
580 95 Delcourt, J., *et al.* (2013) Video multitracking of fish behaviour: a synthesis and future
581 perspectives. *Fish and Fisheries* 14, 186-204
582 96 Hartley, R. and Zisserman, A. (2004) *Multiple view geometry in computer vision*. Cambridge
583 University Press
584 97 Zabala, F., *et al.* (2012) A simple strategy for detecting moving objects during locomotion
585 revealed by animal-robot interactions. *Curr Biol* 22, 1344-1350
586 98 Butail, S. and Paley, D.A. (2012) Three-dimensional reconstruction of the fast-start
587 swimming kinematics of densely schooling fish. *J R Soc Interface* 9, 77-88
588 99 Hinz, R., *et al.* (In Preparation) idSocial: a toolbox for social interactions.
589 100 Hrolenok, B. and Balch, T. (2013) Learning Executable Models of Multiagent Behavior
590 from Live Animal Observation. *ICML 2013 Workshop on Machine Learning For System*
591 *Identification*
592
593

594 **Box 1. Ecological insights from automated image-based tracking**

595 We see three key areas where considerable intellectual progress has been made in
596 ecology using automated image-based tracking. First, the kinematics of animal behavior [18-20,
597 24, 25, 35, 43, 58, 67, 70, 71, 75, 78-83], including the role of the internal state of animals, such
598 as their physiology or genes, and the external environment. Recent breakthroughs in remote
599 quantification of physical landscapes [59-61] and 3D imaging [30] should be especially helpful
600 for these questions. Second, collective behavior in animal groups [1, 27, 34, 39, 41, 44, 46, 63,
601 84, 85], including understanding how information about the physical and biological environment
602 transfers between individuals. Generally, this research centers on intraspecific groups
603 comprised of larger numbers of similar sized individuals. Third, determinants of social behavior
604 [8, 28-30, 32, 54, 55, 68, 72, 74, 86]. Research in this last category usually focuses on a small
605 number of individuals, because identifying the detailed pose required for automated behavioral
606 analysis is difficult in larger groups. Tracking over short durations (minutes) has aided in our
607 understanding of the genetic basis of social behavior, such as aggression or courtship [8, 87],
608 where the high throughput that automation allows provides enhanced power for uncovering
609 patterns in behavioral data [28]. Research over longer times can uncover complex temporal
610 linkages between social behaviors [8, 29], and experiments over the order of weeks are
611 providing unique insight into the social and behavioral development of individuals in intraspecific
612 groups [32, 54, 55].

613 Enormous potential exists for automated image-based tracking to address other key
614 issues in ecology. One area we expect significant growth is in the study of interspecific
615 interactions, which are critical to ecological systems [1-5]. For example, biologists recently used
616 automated analysis of sonar images to reveal how coordinated hunting by predators leads to
617 increased fragmentation and irregularities in the spatial structure of prey groups, and thus
618 inhibition of information transfer among prey [4]. Laboratory research alone provides much
619 scope for experimentally testing basic ideas about ecology, such as the role of body size or
620 predator density in determining trophic interaction strength (Movie S3) [77]. Image-based
621 tracking can also address more applied questions, such as the role of fragmentation in
622 population dynamics [88] or determining the size of animal populations that are historically
623 difficult to measure [53]. Integrating automated tracking techniques into images already
624 collected by trigger-based cameras to assess species occurrence and population abundances
625 [12] would provide important information about the behavior of organisms in natural
626 ecosystems.

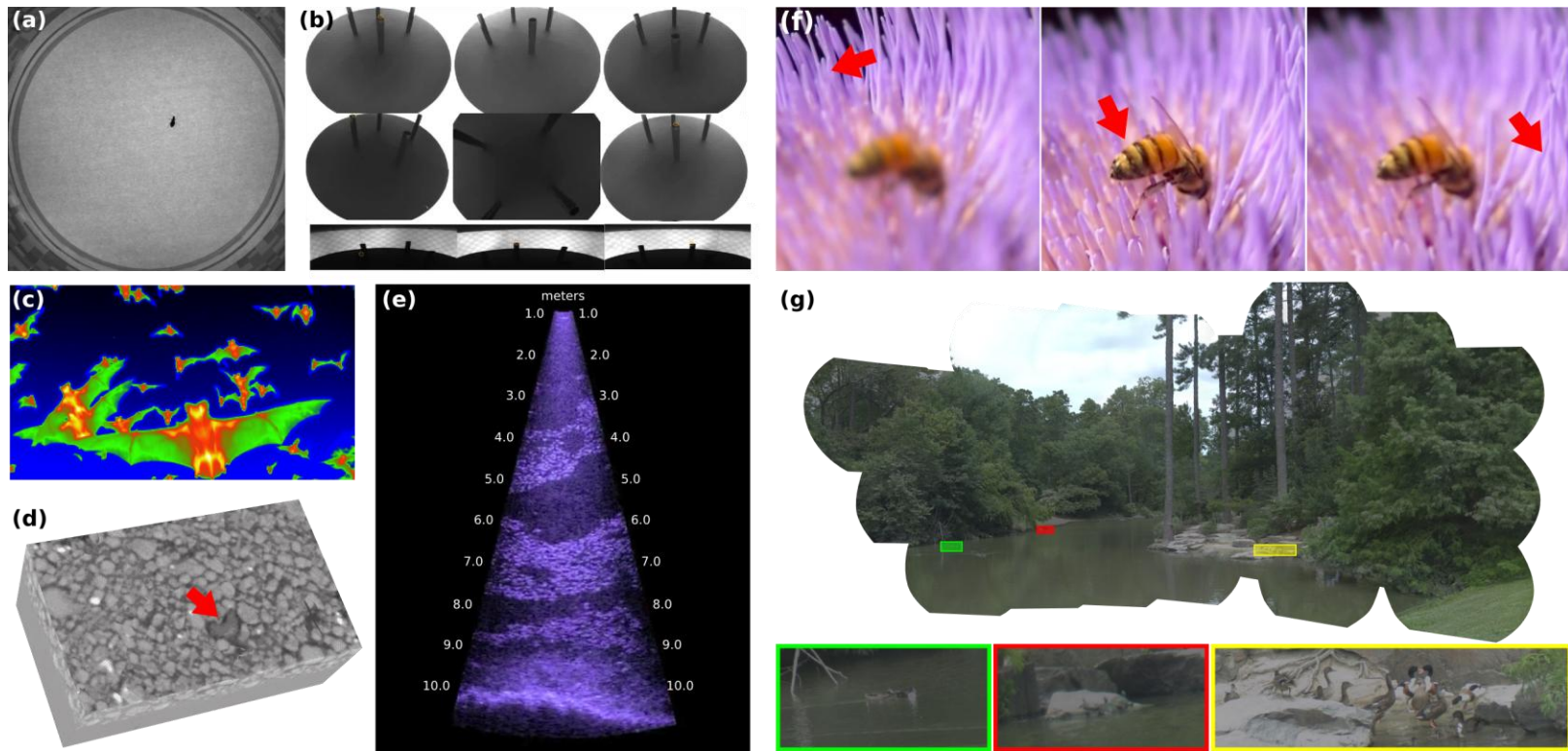
627

628 **Box 2. Obtaining an image**

629 The first step in automated image-based tracking involves obtaining a machine-readable
630 sequence of images that accurately represents the real world. This translation between the real
631 and digital world is a critical step, and time spent optimizing the image (such as ensuring
632 sufficient contrast between foreground and background) pays substantial dividends during
633 subsequent steps (Figure 1). Optical video is commonly used due to its accessibility and low
634 cost, but other imaging technologies considerably expand the range of environmental contexts
635 within which tracking can be undertaken (Figure I). These include infrared (Figure Ia-b), thermal
636 infrared [51] (Figure Ic; Movie S7), X-ray microtomography [56] (Figure Id), and sonar [4] (Figure
637 Ie; Movie S9). Light-field (Figure If) and multi-scale gigapixel [89] (Figure Ig) imaging should
638 permit tracking and scene reconstruction in 3D from a single image viewpoint. Although frame-
639 rates of giga-pixel cameras are increasing (S.D. Feller, unpublished), at three frames per minute
640 [89] they are currently too slow for most automated tracking applications. Light-field cameras
641 work at higher frame rates and there are several laboratories exploring if they can be
642 successfully incorporated into automated tracking systems (I.D. Couzin, G. G. de Polavieja,
643 unpublished). Ultimately, decisions about which imaging method to use should be determined
644 by the specific needs of the project.

645 Automated tracking generally requires a high-contrast image so that computer vision
646 algorithms can adequately discern organisms and their appendages from the surrounding
647 background (Box 3). A common and low-cost method of obtaining such images is to construct
648 an artificial arena for tracking experiments, which is often colored in contrast with the animals,
649 and brightly and uniformly lit with diffuse lighting (Figure Ia-b). Deciding on the spatial and
650 temporal resolution of images is also a key consideration. Higher resolutions generally result in
651 better tracking results and more precise quantification of behavior, but bottlenecks during the
652 transmission, storage, and processing of digital information can limit high temporal resolution to
653 low spatial resolution and/or short durations. Constraints on low spatial resolutions can be
654 overcome by integrating output from multiple cameras [19], and should become less important
655 as technology advances. Recording software is another important consideration, such as the
656 choice of codec for encoding and compressing digital data or ensuring that accurate time
657 stamps are obtained and that frames are not silently dropped, and robust open-source [90, 91]
658 and commercial [92, 93] options are available.

659
660



661
662
663
664
665
666
667
668
669
670
671
672
673
674
675

Figure I. A growing number of technologies allow capture of digital images for automated image-based tracking. **(a)** The most common is optical or near-IR video, most often used in simple 2D laboratory settings (left panel in Figure 1) (Movie S1, Movie S2, Movie S3, Movie S4, Movie S5, Movie S10, Movie S11, Movie S14, Movie S17). **(b)** Images from multiple cameras allow tracking in 3D, even with some degree of habitat complexity present (Movie S6, Movie S15). **(c)** Thermal imaging allows tracking in complete darkness, but requires that tracked animals have a surface temperature different from the surrounding landscape (Movie S7). **(d)** High-resolution X-ray microtomography permits imaging through complex habitat structure, such as soil (burrowing invertebrate highlighted by red arrow). **(e)** Acoustic imaging (sonar) can also image in habitats where optical video would be unusable, such as this image of predators foraging for schooling bait fish in a turbid estuary [4] (Movie S9). **(f)** Light-field cameras allow for post-hoc selection of focal points, thus potentially allowing tracking and construction of the scene in 3D from a single image point. The three panels in **(f)** were obtained from a single light-field image – each panel representing different focal points (highlighted by red arrow). **(g)** Newly developed gigapixel technologies also permit capture of images from a single image point with very high spatial resolutions and at multiscales, again allowing for 3D tracking from a single image point [89]. The three lower panels in **(g)** are enlarged sections of the main image. See Acknowledgments for credits and permissions.

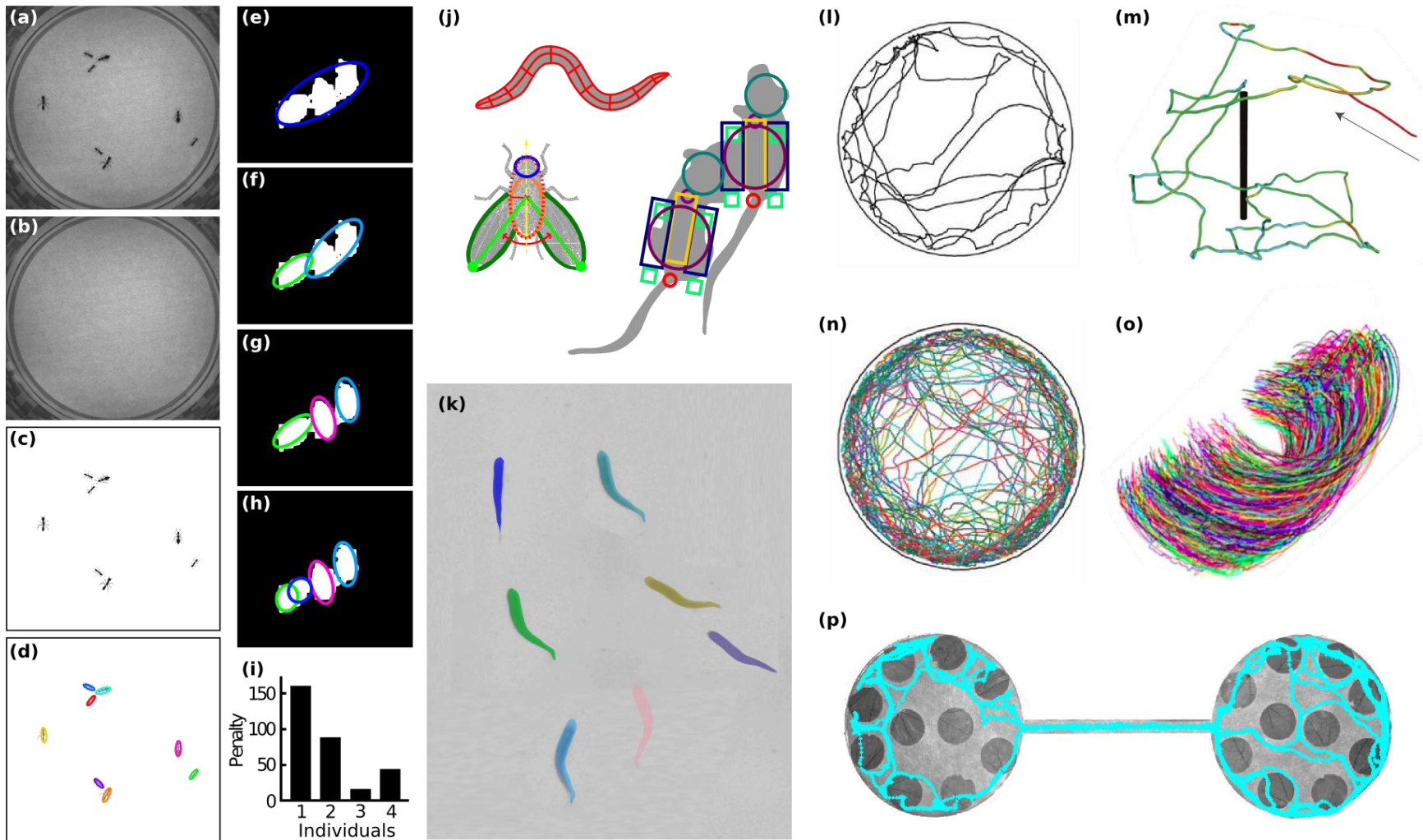
676 **Box 3. Identifying individuals and behaviors in images**

677 Once a set of suitable images has been obtained (Box 2), the position of individuals, and
678 often their pose, must be automatically computed to form trajectories through time. First, the
679 software must determine whether and where individuals are present in each image. How easily
680 this is done varies with the type and quality of images (Box 2), as well as how accurately each
681 individual's position can be predicted from its previous behavior (see below). Detection is
682 straightforward when the contrast between individuals and the background is substantial, and
683 when the background is known or does not change throughout the entire image sequence is
684 most easily performed by background subtraction (Figure 1a-c). The physical complexity of
685 natural systems will ultimately require more advanced techniques, such as those which
686 constantly update their background image [19], or through visual recognition methods [12, 64-
687 68], where the distinctive pattern associated with an individual's body and its motion can be
688 recognized against the clutter of the background.

689 The output of the detection stage is an estimate of the pixels associated with individuals
690 in each image. The position and pose of organisms with stiff and simple-shaped bodies can be
691 computed by fitting a shape contour to the image of the organism [8, 28] (Figure 1d), including
692 determining whether clumps of pixels should be separated into multiple individuals (Figure 1e-i).
693 The situation is more complex when the body is flexible and multiple degrees of freedom are of
694 interest, such as wing angles or head orientation (Figure 1j). Algorithms for learning and
695 computing an individual's pose is an active area of research, and involves either explicit
696 modeling of the body, or learning associations between image brightness patterns and pose
697 parameters [69, 73, 78].

698 Finally, the position of each individual must be linked over multiple frames to form
699 trajectories (Figure 1l-p). This is relatively simple for single individuals, although false and
700 missed detections become more likely when detection is problematic. Constructing trajectories
701 for multiple individuals often involves parameterization of a movement model which includes
702 information from previous frames, such as the acceleration of each individual or their preferred
703 direction of motion [94, 95]. Movement models also improve the detection phase of tracking, but
704 ultimately suffer from error propagation and thus can be labor intensive. Fingerprinting identifies
705 individuals from their image structure (see main text) and therefore recovers identities after
706 occlusion [21, 22] (Figure 1k, Movie S5).

707



708

709

710

711

712

713

714

Figure I. After imaging (Box 2) computer vision software must automatically detect the position, and sometimes pose, of individuals in the image to create trajectories. **(a-c)** A common approach for detecting individuals is background subtraction, where detection of individuals in raw images is achieved by removing an estimated background-only image, resulting in isolation of foreground pixels. **(d)** Contours, denoting individuals, can then be mapped on to clusters of these foreground pixels. How many individuals are within a pixel cluster can be determined in a number of ways. The cluster of pixels in **(e-h)** can be grouped as one, two, three, or four individuals, with **(i)** the optimal grouping being three individuals based on some quantifiable measure. When overlaps are large or

715 body shapes are non-rigid, other methods using past and future dynamics are more suitable (see main text). **(j)** More complex
716 contours can precisely map the pose of individuals, such as swimming in *C. elegans* [20] (Movie S2), wing positioning in *Drosophila*
717 [8] (Movie S14), or body posturing of mice during social interactions [29] (Movie S11). **(k)** Fingerprinting allows for maintenance of
718 identities through time by analyzing the complete image structure, often using differences between individuals that are undetectable
719 to the human eye, such as these zebra fish [21, 22] (Movie S5). Once individuals are detected and identified, their positions are
720 linked across frames to form trajectories. **(l)** This could be a single individual in a 2D landscape [28], **(m)** a single individual in a 3D
721 landscape (shown here with some habitat complexity) [19] (Movie S6), **(n)** multiple individuals in a simple 2D landscape [28] (Movie
722 S1), or **(o)** multiple individuals in a 3D landscape (Movie S8). **(p)** Trajectories through complex habitat structure can also be
723 produced, such as this woodlice navigating for 1 h between two habitat patches connected by a dispersal corridor [88]. See
724 Acknowledgments for credits and permissions.

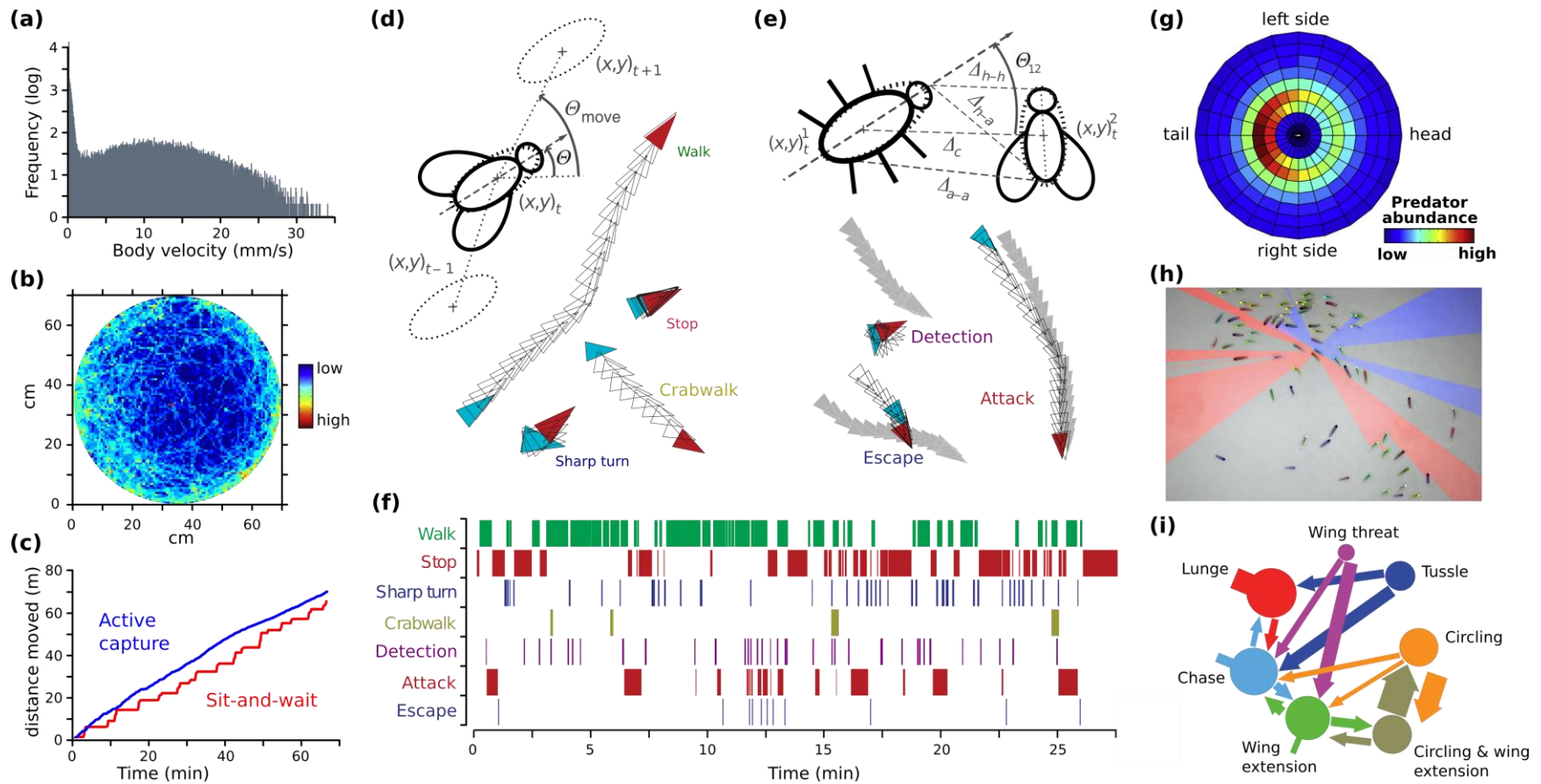
725 **Box 4. Analysis of tracking data**

726 Automation results in vast quantities of high-quality behavioral data, which not only
727 makes data management a key consideration, but presents major challenges in crystalizing this
728 information into tractable and meaningful statistics. This problem is not unique, and it is possible
729 to borrow data management and analysis techniques developed within other “big data” fields,
730 like molecular biology and bioinformatics.

731 The most basic output from automated tracking is the coordinates of the center of body
732 mass of one or more individuals through time (Box 3). Converting from pixel values into real
733 world coordinates is often not as simple as using a pixel-to-distance scale factor, because even
734 in situations with little depth variation the effects of perspective and foreshortening can be
735 important. These issues can be readily overcome with standard photogrammetric techniques
736 [96], or if the filming arena is not flat by integrating a 3D model of the surface into these
737 calculations [35]. Once coordinates (and pose estimates if available) are produced, then even
738 very simple analysis can address basic ecological questions such as where and how animals
739 behave and interact [4, 8] (Figure 1a-c).

740 Higher order patterns in position and pose data can identify individual or between-
741 individual behaviors [69] (Figure 1d-e). Investigation of the relative position and behavior
742 between individuals, such as conspecifics [21, 39, 63, 84] (Movie S1, Movie S2, Movie S5,
743 Movie S7, Movie S8, Movie S11) or interacting predators and prey [4] (Movie S3, Movie S9),
744 provide significant insights into mechanisms underlying the strength and outcome of ecological
745 interactions, and the role of the physical environment [1] (Movie S10). For instance, how much
746 time do animals spend grooming, courting, searching, or chasing, and where and when do they
747 perform these actions, and for interacting individuals, what are the relationships between their
748 subjective sizes, body angles, and relative directions of motion? [97]. Ultimately, automated
749 behavioral analysis is limited only by our ability to quantitatively define behavior, or in our ability
750 to develop machine learning algorithms that can do this for us [74]. One popular procedure is to
751 have a human identify behaviors in video, such as grooming or eating, without defining them.
752 Given several examples of each behavior, a computer algorithm can learn distinguishing data
753 features, creating a classifier or internal model of each behavior. This classifier can then be
754 applied to new datasets, including new individuals and potentially other species. User-friendly,
755 automated tools such [74] make these highly technical analyses accessible to non-experts.

756



757

758 **Figure I.** The final step in automated image-based tracking is analysis, where position and pose data are analyzed to understand
 759 relevant biological, and ecological, patterns and processes. Simple statistics of positional data for individuals include (a) frequency
 760 distributions of body velocity, (b) location intensity maps revealing where throughout the landscape individuals spend their time, and
 761 (c) distance-time plots, which can highlight foraging strategies employed by predators. (a-b) are data from an adult wolf spider
 762 (*Lycosa*) moving around a 65 cm diameter circular arena for one hour, and (c) is a wolf spider (*Lycosa*, red line) and a centipede
 763 (*Lithobius*, blue line) under the same conditions (A.I. Dell, unpublished). This behavioral data can be automatically condensed into
 764 simpler, interpretable categories that characterize real behavioral phenotypes, either for (d) single individuals or (e) between
 765 individuals. The top-left panel in (d, e) show a contour model of the individual/s, together with the quantities (or traits) that can be
 766 measured, such as (d) a fly or (e) an insect predator feeding on a fly [8, 28]. Symbol definitions are: x, y , spatial coordinates of the
 767 midpoint of the individual; t , time; Θ , orientation; Θ_{move} , moving direction; Δ , distance; a , abdomen; h , head; c , midpoint of the animal

768 [8]. Also shown in (d, e) are some example behaviors that can be derived from these basic quantities [28]. Blue and red triangles are
769 plotted at the start and end of each behavior example. For (e) behaviors between individuals the position of the non-target individual
770 is plotted in gray. **(f)** From these behavioral assignments, time budgets can reveal how individuals spend their time or the
771 components of ecological interactions (A.I. Dell, unpublished data). Analysis of between-individual position data can highlight
772 patterns in the spatial arrangement of individuals within groups, and aid in understanding collective behavior. **(g)** shows the angular
773 density of predators around a representative focal prey, with high predation density behind the prey revealing that piscivorous
774 predators tend to attack from behind their prey [4], and **(h)** shows a reconstruction of the visual field of an individual fish embedded in
775 the center of a large school, which provides novel insight into information flows within animal groups [27]. Behavioral information can
776 also reveal the temporal linkage between complex behaviors, such as **(i)** this transitional behavioral graph for social interactions in
777 *Drosophila* [8]. See Acknowledgments for credits and permissions.
778

779 **Glossary**

780 **Background subtraction.** A method used by software to compare the current video frame
781 with a stored picture of the background; any pixel of the current frame that is significantly
782 different from the corresponding pixel in the background is likely to be associated to the
783 body of an animal. Useful in situations where the background is unchanging, such as a
784 stationary camera, the surface of the background is rigid and lighting does not change.

785 **Behavior.** The actions of individuals, often in response to stimuli. Behavior can involve
786 movement of the individual's body through space, such as walking or chasing, or can
787 occur while the animal is stationary, such as grooming or eating.

788 **Bio-logging.** Attachment or implantation of equipment to organisms to provide information
789 about their identity, location, behavior, or physiology (e.g., global positioning systems,
790 accelerometers, video cameras, telemetry tags).

791 **Ecological interaction.** Any interaction between an organism and its environment, or
792 between two organisms (i.e., including interactions between conspecifics).

793 **Fingerprinting.** A method used to identify unmarked individuals using natural variation in their
794 physical and/or behavioral appearance. The method works by transforming the images
795 of each individual into a characteristic "fingerprint", which can then be used to distinguish
796 individual organisms both within and across videos.

797 **FPS.** The number of frames in an image sequence collected per second (frames-per-second).

798 **Image.** A digital representation of the spatiotemporal location, identity, or pose of an individual
799 or set of individuals (see Box 2).

800 **Machine learning.** A set of techniques that allow computer software to learn from empirical
801 data, user assumptions or manual annotation. These approaches are becoming
802 increasingly common in the analysis of behavior, where users can tag behavior in short
803 sequences of images and the software can predict occurrences of these behaviors
804 throughout the entire image sequence.

805 **Marking.** The attachment of artificial "marks" to organisms to maintain their identity, such as
806 paint or barcodes.

807 **Occlusion.** When the view of any individual in an image is disrupted either by another
808 individual or physical habitat (i.e., the occluding object lies in a straight line between the
809 focus individual and the camera).

810 **Pixel.** A physical point in a 2D digital image, and therefore the smallest controllable element of
811 a picture represented on the screen. The equivalent of a pixel in 3D space is a voxel.

812 **Pose.** Any additional geometrical quantity of interest other than the center of the main body of
813 the animal, such as orientation, wing positions, body curvature, etc.

814 **Position.** The center of body mass of an individual in time and space.

815 **Resolution.** The number of pixels/voxels in a digital image.

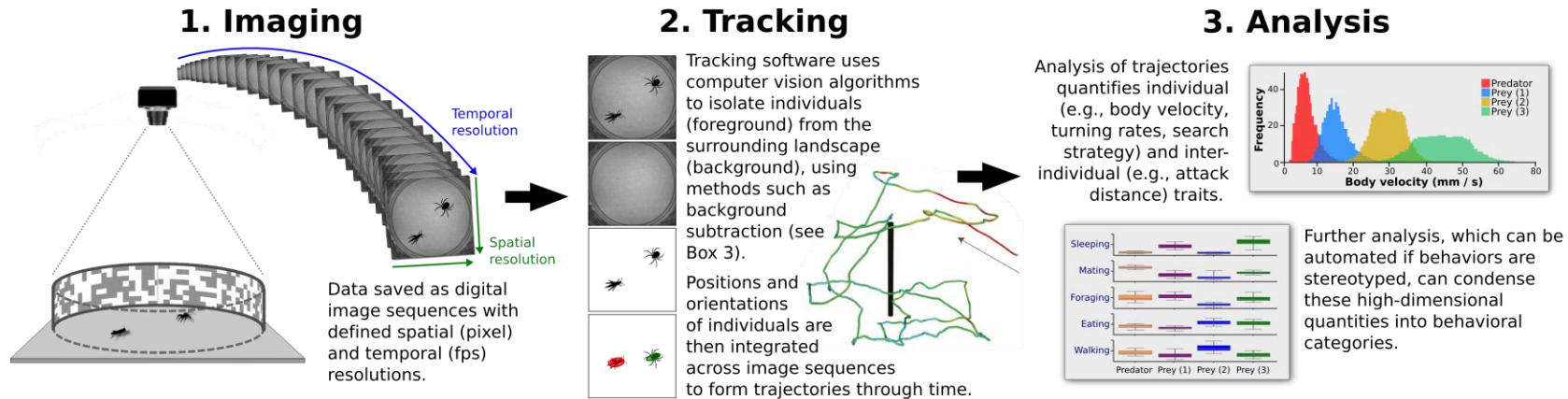
816

817 **Call to developers: the ideal automated image-based tracking system for ecologists**

818 Ecologists have a clear need for systems that can robustly and quickly gather large
819 amounts of precise data on the behavior of multiple individuals from a diversity of species.
820 Development of an automated image-based tracking system like this will require integrating
821 disparate pieces of technology and software into a coherent, and user-friendly, package. This
822 appears possible in the near future because of the continued increase in power, speed and
823 capabilities of hardware and software technologies. Any system that does this should have a
824 number of key features.

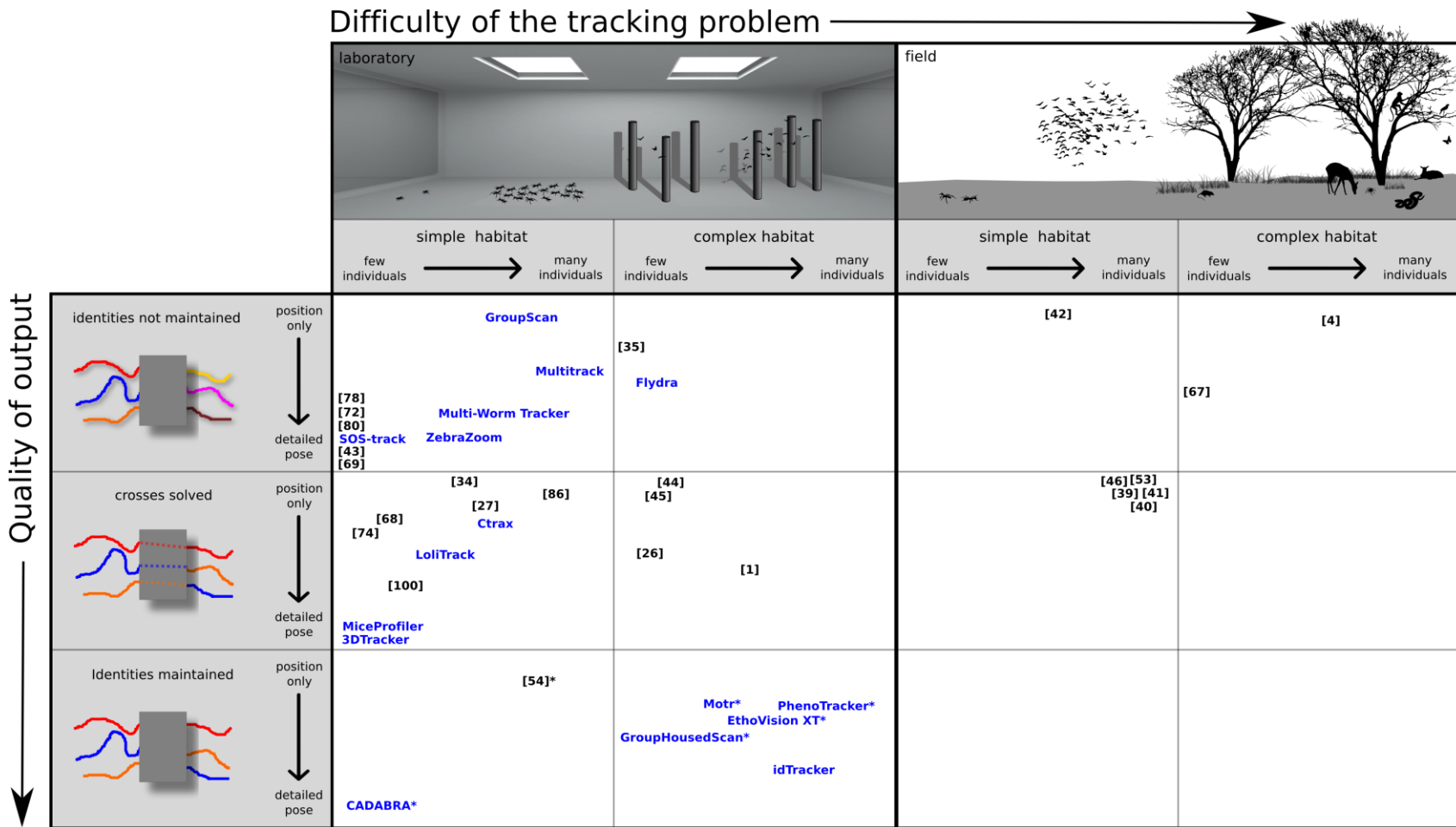
- 825 1) Simple to use. Automated image-based tracking systems involve a complex integration of
826 software and hardware (Figure 1), but many ecologists have insufficient funding to purchase
827 expensive imaging and computational equipment, nor the technical expertise required to use
828 them. While there is some responsibility for ecologists to acquire these skills, tracking
829 system development is ultimately the role of researchers and engineers with expertise in
830 computer vision and informatics.
- 831 2) Marking should be unnecessary, as this is time-consuming, requires capture of individuals,
832 and can alter their behavior and how other organisms interact with them.
- 833 3) Flexible enough that it can successfully track in a diverse range of experimental conditions,
834 and individuals that vary greatly in their size, shape, and patterns of behavior.
- 835 4) Tracking large numbers of individuals is necessary because ecological systems are often
836 characterized by high densities.
- 837 5) Ecological systems are naturally embedded within a diverse range of environmental
838 contexts, so it is essential that automated tracking systems function within diverse and
839 complex habitats. The development of tracking software able to isolate animals from
840 complex backgrounds, together with increasing automation of behavioral analysis, means
841 the capabilities of automated image-based tracking for field use will only increase in the
842 coming years.
- 843 6) The system must overcome the significant data storage and data management issues that
844 inevitably arise when tracking at larger spatial and temporal scales. This can be partly
845 overcome by real-time tracking, which reduces the need for processing and storing large
846 amounts of digital data.
- 847 7) A single image point (i.e., camera) would be preferred over requiring multiple cameras,
848 which can be difficult to integrate into a coherent system and can introduce disturbance
849 effects that alter natural behavior.
- 850 8) Be mostly automated so that tracking and analysis is quick and consistent, including
851 identifying individuals and their interactions, and quantifying their behavior in meaningful
852 ways. However, still provide flexibility in the ways in which users can extract data, including
853 at all levels of data acquisition, tracking and analysis.

854
855
856
857



858
859
860
861
862
863
864
865
866

Figure 1. The three general steps involved in automated image-based tracking of behavior are: imaging (Box 2), detection of individuals and their pose in the image and appropriate 'linking' of detections to create separate tracks through time for each individual (Box 3), and analysis of trajectory and behavioral data (Box 4). To date, imaging is often done in the laboratory (left panel), which can more easily provide a clean, crisp image that minimizes tracking errors. Each of these steps are strongly interlinked and time spent optimizing one step (e.g., imaging) can pay huge dividends in time and effort saved at later steps (e.g., reducing tracking errors).



867
868
869
870
871
872
873
874
875
876

Figure 2. A number of key automated image-based tracking systems and studies relevant to ecologists, ranging on the horizontal axis from low numbers of individuals interacting in a simple landscape in the laboratory (left of figure) to a natural system within a complex biological and physical landscape (right of figure). The vertical axis separates studies and tracking systems by the type of output provided, specifically whether and how identity is maintained and whether position or detailed pose are tracked. In some cases positions of labels have been slightly moved for visual clarity. Tracking systems can cover multiple categories, but are only shown once as tracking in simpler habitats with smaller numbers of individuals and less detailed pose will almost always be possible. See Table S1 for more details about each tracking system. Publications that use any of the tracking systems in Table S1 are not listed separately here. Numbers in square brackets are citations (see reference list). * denotes the use of bio-logging or marking.

Supplementary Information

Automated image-based tracking and its application in ecology

Anthony I. Dell*, John A. Bender, Kristin Branson, Iain D. Couzin, Gonzalo G. de Polavieja, Lucas P.J.J. Noldus, Alfonso Pérez-Escudero, Pietro Perona, Andrew D. Straw, Martin Wikelski, and Ulrich Brose

*Corresponding author: adell@gwdg.de

Table S1. Key automated image-based tracking systems available to ecologists (pg. 2)

Table S2. Criteria for describing the automated image-based tracking systems in Table S1 (pg. 6)

Movie S1. Multiple *Drosophila* moving in a simple 2D landscape.

Movie S2. Tracking of multiple *C. elegans* moving in a simple 2D landscape using Multi-Worm Tracker.

Movie S3. Wolf spider consuming a moth in a simple 2D landscape.

Movie S4. Tracking shoaling behavior of zebrafish using EthoVision XT.

Movie S5. Tracking of different animal groups maintaining identities using idTracker.

Movie S6. *Drosophila* moving throughout a 3D landscape with physical structure tracked using Flydra.

Movie S7. Flying bats imaged using multiple thermal cameras.

Movie S8. 3D tracking of large swarms of flying animals in the field.

Movie S9. Schooling fish and their predators in a turbid estuary imaged using high-resolution sonar.

Movie S10. Tracking of schooling fish in a complex light environment.

Movie S11. Tracking of social interactions in mice using MiceProfiler.

Movie S12. Automated behavioral annotations of social interactions in *Drosophila*.

Movie S13. Field video of a rattlesnake striking at prey.

Movie S14. Social interactions in *Drosophila* tracked with CADABRA.

Movie S15. Position and pose tracking systems from CleverSys.

Movie S16. Long-term tracking of social behavior in mice using Motr.

Movie S17. Tracking the posture of a single individual for a range of species using SOS-track.

Movie S18. Tracking of zebrafish larvae with ZebraZoom.

Movie S19. Position and pose tracking with Lolitrack.

Movie S20. Tracking a single bee and multiple ants with Multitrack.

Movie S21. Position and pose tracking with Phenotracker.

Movie S22. Tracking social behavior in rats using 3D video.

Table S1. Some key automated image-based tracking systems (in alphabetic order) that are readily available to research ecologists, with focus on those suited for tracking behavior during species interactions. See Table S2 for detailed descriptions of each category.

3DTracker [30] (Movie S22)

Availability and usability: open-source, simple to use, specific¹

Raw output: 3D trajectories, 3D pose²

Requirements: Win, Matlab, multiple 3D cameras³

Maximum individuals: 2

Shape requirements: rodent-like⁴

Identity method: crosses solved⁵

Analysis tools: basic kinematics, social behaviour⁶

Taxa studied: rats⁷

Required resolution per individual: 300 points⁸

Notes: ¹Executables and sources available online at <http://matsumotoi.github.io/> [30]. Requires uncluttered background. Requires multiple 3D cameras. ²Estimates 3D trajectories and detailed 3D pose of four body parts (head, neck, trunk, and hip). ³Currently uses 4 x Microsoft Kinect cameras. Matlab required for behavioral analysis. ⁴Developed for rats, but could be applied to mice with higher resolution cameras. ⁵3D video enables more stable tracking during close contact. On average, automation makes errors in 20% of 1 min videos. Includes tool to facilitate manual correction. ⁶Includes Matlab scripts extracting basic movement parameters (such as velocity or angle) and social behaviors (e.g., approaching, mounting, head-hip contact) based on 3D trajectories of body parts. ⁷Social interactions in rats [30]. ⁸3D points covering the surface of each individual.

CADABRA [8] (Movie S14)

Availability and usability: open-source, simple to use, flexible¹

Raw output: 2D trajectories, detailed pose²

Requirements: Win/Linux/Mac, Matlab³

Maximum individuals: 2

Shape requirements: fly-shaped⁴

Identity method: maintained (size/markings)⁵

Analysis tools: extensive⁶

Taxa studied: flies⁷

Required resolution per individual: 20 pixels

Notes: ¹Executables available online [8]. Source code available upon request. Flexible to different setups. Can use conventional cameras. Can track several arenas simultaneously. Requires constant and uncluttered background. ²Estimates 25 traits of the pose and position of each individual (orientation, velocity, size, wing pose). ³Matlab Compiler Runtime library (free). ⁴Developed for *Drosophila*. ⁵Uses size difference between pairs (e.g., male or female). Similar sized individuals must be marked (otherwise system makes highest probability matches). ⁶Uses estimates of 25 traits of both flies (position, orientation, velocity, size, wing pose) to automatically quantify 8 social behavior (e.g., lunging, wing extension, copulation). ⁷Genetic and environmental influences on social behavior in *Drosophila* [8, 87].

Ctrax [28] (Movie S1, Movie S3)

Availability and usability: open-source, simple to use, flexible¹

Raw output: 2D trajectories, orientations

Requirements: Win/Linux/Mac, Matlab²

Maximum individuals: 50

Shape requirements: elliptical³

Identity method: crosses solved⁴

Analysis tools: extensive, see JAABA⁵

Taxa studied: flies, cockroaches, fish, spiders⁶

Required resolution per individual: 15 pixels

Notes: ¹Ctrax online at <http://ctrax.sourceforge.net>. Flexible to different setups. Can use conventional cameras. Requires constant and uncluttered background. Ongoing maintenance, enhancements, and support provided. ²Matlab required for capabilities beyond tracking, which are important (i.e., error fixing, analysis). ³Developed for *Drosophila*, works well with any species of elliptical shape. Individuals must have similar size. ⁴For *Drosophila*, with full automation system makes an identity error on average once every 5 fly-hours with density of 10, once every 1.5 fly-hours with density of 20, and once every 40 fly-minutes with density of 50. Error rate is zero with minimal user supervision. Includes application to facilitate manual correction. ⁵JAABA [74] supersedes Ctrax's behavioral analysis package (free online at <http://jaaba.sourceforge.net>). JAABA is a machine-learning system to create automatic behavior classifiers, which allows exploration of differences in thousands of behavior statistics between large numbers of individuals. JAABA requires tracking data from other tracking systems (e.g., Ctrax, CADABRA, Multi-Worm Tracker, Motr). ⁶Genetic and environmental influences on *Drosophila* social behavior [24, 28], cockroach locomotion [70], collective behavior in fish [85], spider foraging behavior (Dell et al. unpublished).

EthoVision XT (Movie S4)

Availability and usability: commercial, simple to use, flexible¹

Raw output: 2D trajectories, orientations²

Requirements: Win

Maximum individuals: 16^(note 3)

Shape requirements: flexible⁴

Identity method: maintained (size/markings)⁵

Analysis tools: extensive⁶

Taxa studied: rodents, fish, insects, pigs, spiders, more⁷

Required resolution per individual: 15 pixels⁸

Notes: ¹Developed by Noldus Information Technology BV. Available online at <http://www.noldus.com/ethovision>. Free trial version available. Software only or with integrated hardware. Valid for conventional cameras and laboratory setups, including multiple arenas. A constant and uncluttered background not required. Shadows, bedding material and objects can be present in the arena. Dedicated systems available for fish (DanioVision) and rodents (PhenoTyper). ²Track3D (<http://www.noldus.com/innovationworks/products/track3d>) extension allows 3D tracking of single individuals using 2 cameras. Standard software tracks center of body mass, specialized modules recognize body shape and multiple body points (e.g. nose point, tail base) of rodents. Simple and easy-to-use interface. ³Can track 100 arenas (16 individuals in each arena) with a single camera. ⁴Can track many shapes, from small insects to large mammals. ⁵Similar-sized individuals require color marking (when maintenance of identity required). With small numbers of individuals the system can often maintain identities of unmarked animals. ⁶Includes visualization, data editing and animation options. Module for automated behavior recognition (rodents: rearing, grooming, sniffing), trial and hardware control, behavior-physiology integration. ⁷See all publications at <http://www.noldus.com/ethovision-xt/selected-publications>. ⁸Requires 3 pixels per individual for position only.

Flydra [19] (Movie S6)

Availability and usability: open-source, difficult to use, specific¹

Raw output: 3D trajectories (real-time)

Requirements: Linux, multiple cameras

Maximum individuals: 3^(note 2)

Shape requirements: flexible³

Identity method: not maintained

Analysis tools: none

Taxa studied: flies, birds⁴

Required resolution per individual: 1 pixel

Notes: ¹Core 3D algorithms available under open-source license at <https://github.com/strawlab/pymvg> (image acquisition and synchronization code, extended kalman filter for tracking, and analysis tools to verify tracking is working are not currently open-source). Requires synchronization of multiple computers and high-speed cameras. Requires constant and uncluttered background (some degree of habitat complexity possible). Advanced Linux skills needed. ²Limited by computer speed. ³Individuals should be roughly the same size, and have independent behavior (i.e. not valid for social species, aggression, courtship, predation, etc.). ⁴Flight control in *Drosophila* and hummingbirds [19, 25, 79, 80, 82].

GroupHousedScan (Movie S15)

Availability and usability: commercial, simple to use, flexible¹

Raw output: 3D trajectories, detailed pose (real-time)²

Requirements: Win, two cameras³

Maximum individuals: 2^(note 4)

Shape requirements: rodent-like⁵

Identity method: maintained (size/markings)

Analysis tools: extensive⁶

Taxa studied: rats, mice, primates⁷

Required resolution per individual: 200 pixels⁸

Notes: ¹Developed by CleverSys Inc. Available online at http://cleversysinc.com/?csi_products=grouphousedscan. Software only or with integrated hardware. Valid for conventional cameras and laboratory setups. Requires contrasting background. Ability to easily characterize spatial components of the physical environment for integration with tracking data. Bedding material, food and water containers, and other objects can be present in the arena. Useful for long-term tracking over longer durations, such as multiple days, as is capable of adjusting between day and night conditions automatically. Simple and easy-to-use interface. ²Dedicated 2D systems can monitor the movement and detailed pose of single animals either from the side (HomeCageScan - http://cleversysinc.com/?csi_products=homecagescan and PrimateScan - http://cleversysinc.com/?csi_products=primatescan) or top (TopScan - http://cleversysinc.com/?csi_products=topscan-suite). TopScan can be extended to 4 individuals with SocialScan add-on. Pose data include 8 different body points on a rodent: head or nose, ears, forelimbs, hindlimbs, upper back, lower back, abdomen, tail. HomeCageScan identifies same 8 points. PrimateScan identifies same 8 points on a primate. TopScan and SocialScan identify 4 points on a rodent: nose, forelimb, center of mass, tailbase. ³Single-camera options available (see note 2). ⁴see note 2. ⁵Developed for mice and rats, has dedicated system for a single primate (see note 2). ⁶Kinematic measurements such as speed, velocity, orientation, shape, etc. Automatically outputs individual (eat, drink, sleep, walk, jump, rear up, hang, groom, sniff, twitch, stretch, etc.) and social interaction (contact, sniff, follow, leave, approach, etc.) behaviors and events. ⁷Genetic and environmental disorders in rats and mice (unpublished). ⁸Can be set as a threshold.

GroupScan (Movie S15)

Availability and usability: commercial, simple to use, flexible¹

Raw output: 2D trajectories

Requirements: Win

Maximum individuals: 100^(note 2)

Shape requirements: flexible³

Identity method: not maintained

Analysis tools: basic kinematics⁴

Taxa studied: *Drosophila*, fish⁵

Required resolution per individual: 10 pixels⁶

Notes: ¹Developed by CleverSys Inc. Available online at http://cleversysinc.com/?csi_products=groupscan. Software only or with integrated hardware. Valid for conventional cameras and laboratory setups. Requires contrasting background. Ability to easily characterize spatial components of the physical environment for integration with tracking data. Simple and easy-to-use interface. ²Preset at 100, but can be varied. ³Developed for fruit flies and fish, but extendible to any species. ⁴Basic population-level statistics, including count, and average distance travelled, velocity, inter-frame body pixel change, etc. ⁵Genetic and environmental disorders in *Drosophila* and fish (unpublished). ⁶Can be set as a threshold.

idTracker [21, 22] (Movie S5)

Availability and usability: open-source, simple to use, flexible¹

Raw output: 2D trajectories, orientations

Requirements: Win/Linux/Mac, matlab²

Maximum individuals: 20^(note 3)

Shape requirements: flexible⁴

Identity method: maintained (fingerprinting)⁵

Analysis tools: extensive, see idSocial⁶

Taxa studied: mice, fish, flies, ants⁷

Required resolution per individual: 150 pixels

Notes: ¹Available at www.idtracker.es. Simple to use. Adaptable to different laboratory setups. Requires image with good contrast between animals and background. Valid for conventional cameras. Can track compressed videos. ²Development version requires Matlab (free compiled binaries do not). Fingerprinting computationally heavy, runs on conventional computer with at least 4GB RAM. ³Depends on species and conditions. ⁴No inherent contour. Can track wide range of size and shaped individuals. ⁵Maintains identities automatically in videos of any length, enabling tracking with complex crossings, occlusions, or perturbations. Ability to identify individuals across videos. ⁶idSocial [99] is an accompanying open source analysis package giving mean distances, distribution of distances, kinematic parameters of interactions, probabilities for relative distances, aggression leadership or followership hierarchies and collective behavior. ⁷Movement and collective behavior in fish, flies, mice, and ants [21, 22].

LoliTrack (Movie S19)

Availability and usability: commercial, simple to use, flexible¹
Raw output: 2D trajectories, simple pose²
Requirements: Win

Maximum individuals: 24
Shape requirements: flexible³
Identity method: crosses solved⁴

Analysis tools: basic kinematics⁵
Taxa studied: fish, rodents, birds, insects, crustaceans, more⁶
Required resolution per individual: 9 pixels

Notes: ¹Developed by Loligo Systems. Available online at http://www.loligosystems.com/?action=shop_show&varenr=AB10190. Free trial version available. Valid for conventional cameras and laboratory setups, including multiple arenas. Software only or with integrated hardware. Does not require constant and uncluttered background. Ability to easily characterize spatial components of the physical environment for integration with tracking data. Simple and easy-to-use interface. Shuttlesoft (http://www.loligosystems.com/?action=shop_show&varenr=AB10202) is a dedicated multi-chamber system for analysis of preference or avoidance to environmental drivers, such as temperature or dissolved compounds. ²Three points along body axis. ³No inherent contour. Can track wide range of size and shaped individuals. ⁴Error rate data not available. ⁵Basic measurements per individual, such as body velocity, distance moved, turning rate, time spent in zone. Dedicated solutions available upon request. ⁶See all publications at http://www.loligosystems.com/index.php?action=references_show_all&menu=14&keyword=VIDEO%20TRACKING.

MiceProfiler [29] (Movie S11)

Availability and usability: open-source, simple to use, flexible¹
Raw output: 2D trajectory, detailed pose
Requirements: Win/Linux/Mac

Maximum individuals: 2^(note 2)
Shape requirements: rodent-like³
Identity method: crosses solved⁴

Analysis tools: extensive⁵
Taxa studied: rodents⁶
Required resolution per individual: 350 pixels

Notes: ¹Available online at http://icy.bioimageanalysis.org/plugin/Mice_Profiler_Tracker. Easy to use and adaptable to different laboratory setups. Requires unchanging and uncluttered background. Tracker struggles when individuals are in very close contact. ²Behavioral analysis limited to two individuals. ³Developed for mice. ⁴Switches identities on average twice per minute. Includes tool to facilitate manual correction. ⁵System includes a behavioral chronogram generator (e.g., contact events, sniffing, chase, escape) and a temporal behavioral analysis module. ⁶Social interactions in mice [29, 55], rats, guinea pigs (unpublished).

Motr [32] (Movie S16)

Availability and usability: open-source, simple to use, flexible¹
Raw output: 2D trajectory, simple pose²
Requirements: Win/Linux/Mac, Matlab³

Maximum individuals: 6^(note 4)
Shape requirements: elliptical⁵
Identity method: maintained (marking)⁶

Analysis tools: extensive, see JAABA⁷
Taxa studied: mice⁸
Required resolution per individual: 5000 pixels⁹

Notes: ¹Available at <http://motr.janelia.org>. Can use conventional cameras. Requires unchanging and uncluttered background and uniform illumination. ²Pose represented as an ellipse, including x,y coordinates, major and minor axis of ellipse, and ellipse orientation. ³Matlab 2009b or later. ⁴System validated up to 6 individuals, possibly can manage more. ⁵Developed for rodents, but generalizable to other elliptical-shaped organisms. ⁶Requires that individuals are marked – might also work with natural differences in pattern, but this has not been tested. The system automatically learns external appearance of individuals from training videos and generalizes to identify them in a group. ⁷JAABA [74] was used to learn complex behaviors such as following and chasing (for details on JAABA [74] see Ctrax - note 6). ⁸Social development in groups of mice [32]. ⁹In their standard configuration (1024 x 768 pixel resolution), they fit a small rectangle (50 x 100 pixels) around each mouse to extract features needed for individual identification. With these specifications, each mouse body formed an ellipse with a long axis ~30-85 pixels, and the small axis ~10-45 pixels.

Multitrack (Movie S20)

Availability and usability: open-source, simple to use, flexible¹
Raw output: 2D trajectory, orientation
Requirements: Linux, additional libraries²

Maximum individuals: 1000^(note 3)
Shape requirements: flexible⁴
Identity method: not maintained⁵

Analysis tools: none⁶
Taxa studied: ants, bees⁷
Required resolution per individual: 3 pixels⁸

Notes: ¹Available online at www.bio-tracking.org/category/software. Flexible to different setups. Can use conventional cameras. Requires constant background. System packed with companion programs for creating shape models and backgrounds for different experimental setups. ²OpenCV and PCL (standard computer vision libraries), and Qt 4.0 (for creating graphical user interfaces). All three are cross platform and freely available. ³Maximum tested ~1000 individuals, but can track an arbitrarily large number. ⁴Individuals are assumed to be rigid (see note 1). Different types of body shapes can be tracked simultaneously by creating multiple models. ⁵A proportion of crossings can be solved with movement models. ⁶Only produces tracking data. ⁷Spatial dynamics of ant [100] and bee (unpublished) behavior. ⁸Can be set as a threshold.

Multi-Worm Tracker [20] (Movie S2)

Availability and usability: open-source, simple to use, flexible¹
Raw output: 2D trajectory, detailed pose (real-time)²
Requirements: Win, LabVIEW Vision (NI) run-time license³

Maximum individuals: 80^(note 4)
Shape requirements: worm-like⁵
Identity method: not maintained⁶

Analysis tools: basic kinematics, basic behaviour⁷
Taxa studied: nematodes, fly larvae⁸
Required resolution per individual: 160 pixels

Notes: ¹Available online at <http://sourceforge.net/projects/mwt/>. Requires unchanging and uncluttered background. Can use conventional cameras. ²Includes position of centroid, body size (number of pixels), vectors defining long (and orthogonal) axes of the shape, skeleton (11 point line along midline), and outline of body (compressed bitcode stored as a string). Summary file can store when a stimuli occurred, and 13 simple statistics (number of objects, mean speed, mean size, etc.). ³Choreography runs on Win/Linux/Mac. ⁴Up to ~500 individuals with lower frame rate (~10-15 fps). ⁵Optimized for *C. elegans*. ⁶Identities not maintained – trajectories lost when animals touch. ⁷Choreography is an offline basic behavioral analysis package (direction, velocity). Plugins can compute organism- or condition-specific behaviors. ⁸Determinants of behaviour in *C. elegans* [20], neural biology of *Drosophila* larvae [81, 83].

PhenoTracker (Movie S21)

Availability and usability: commercial, easy to use, flexible¹

Raw output: 2D trajectories, basic pose²

Requirements: Win³

Maximum individuals: 50^(note 4)

Shape requirements: flexible

Identity method: maintained (size/markings)

Analysis tools: basic kinematics, social behaviour⁵

Taxa studied: flies, ants, fish, rodents⁶

Required resolution per individual: 200 pixels⁷

Notes: ¹Developed by TSE Systems. Available online at <http://www.tse-systems.com/products/behavior/video-tracking-software/phenotracker/index.htm>. Free trial version available. Software only or with integrated hardware. A constant and uncluttered background not required. Shadows, bedding material and objects can be present in the arena. Valid for conventional cameras and laboratory setups, including multiple arenas. Online support available. ²Identity not maintained if individuals not marked. Basic pose includes head, center of body, and tail. ³Server version also available, requiring web browser. ⁴Maximum tested, but in principle could do more. ⁵Basic measurements per individual, such as body velocity, distance moved, turning rate, time spent in zone. Basic analysis for social interactions, including orientation towards or distance from other individuals. Can do statistical analyses between videos. ⁶See all publications at <http://www.tse-systems.com/cgi-bin/refsearchn.pl?type=pub>. ⁷Requires 4 pixels per individual for position only.

SOS-track [18] (Movie S17)

Availability and usability: open-source, simple to use, flexible¹

Raw output: 2D trajectory, detailed pose (real-time)²

Requirements: Win/Linux/Mac, Matlab

Maximum individuals: 1^(note 3)

Shape requirements: flexible⁴

Identity method: not maintained⁵

Analysis tools: basic kinematics, sensory biology⁶

Taxa studied: flatworm, larvae, fly, fish, rodent⁷

Required resolution per individual: variable⁸

Notes: ¹Available online at <http://sourceforge.net/projects/sos-track/>. Valid for conventional cameras and laboratory setups. Suited for different organisms and arenas. Can track several arenas simultaneously. Ability to add and track accurate information about the environment-organism interaction. Ability to easily correct errors (e.g., head-tail swaps). Requires constant uncluttered background. ²Includes position of centroid, head, tail, and midpoint, and skeleton, curvature, and total area. Real-time data does not include pose information. ³Main package can track a single individual. Extendable to multiple individuals (Movie S17), but loses identity when individuals are similar (code available from authors upon request). ⁴Wide range of size and shaped individuals can be tracked, including elliptical, worm-shaped, and legged organisms. ⁵see note 3. ⁶Low-level motorsensory measurements (body posture, kinematic variables – angles, velocities, distances) and sensory information relevant to the individual (i.e., system can map relevant points along the individuals body to the sensory landscape, such as local orientation relative to the sensory gradient, or stimulus intensity at the front of the animal). Automated basic behavior classification available for worm-shaped animals: turns and head casts. ⁷Environmental control of behavior in *Drosophila* larva [18, 58], other taxa (unpublished). ⁸Can be set as a threshold, although pose tracking impossible at low resolutions.

ZebraZoom [71] (Movie S18)

Availability and usability: open-source, simple to use, specific¹

Raw output: 2D trajectory, basic pose

Requirements: Linux, Matlab, OpenCV, C++²

Maximum individuals: 7^(note 3)

Shape requirements: zebrafish-like⁴

Identity method: crosses solved⁵

Analysis tools: basic kinematics, basic behaviour⁶

Taxa studied: zebrafish⁷

Required resolution per individual: 350 pixels

Notes: ¹Free online at <http://sourceforge.net/p/zebrazoom/wiki/Home/>. Limited flexibility to other setups. High-speed camera (~330 Hz). Probably requires tuning of several parameters to adapt it to a new set-up. Not ideal for interactions involving close encounters between individuals. ²Tested on Linux. ³Validation for up to 7 individuals, possibly system can deal with more. ⁴Optimized for zebrafish larvae, probably not generalizable to many other fish species. ⁵Tracking algorithm not robust to crossings - switching of identification between two larvae estimated to occur every ~109 s (density of 7 individuals). Solves crossings, making on average one mistake per animal every 110 seconds. ⁶For each behavior detected, extraction of output parameters per swim burst (i.e, number of oscillations, tail beat frequency, duration, orientation, distance travelled, mean speed) and parameters also calculated for population. Attached behavioral cluster package includes an automatic (trained) classifier of behaviors in three classes: slow forward swim, routine turn, and scape. ⁷Genetic and environmental drivers of behavior in zebrafish [71].

Table S2. Criteria for describing the automated image-based tracking systems in Table S1.

Category	Description
Analysis tools	Does the system automatically (or semi-automatically) output higher-dimensional indices of trajectory and pose data, such as individual kinematics (e.g., body velocity, distance to wall) or between individual behaviors (nearest neighbor, relative velocity). Is there behavioral phenotyping available? What packages or add-ons are available to assist with behavioral analysis?
Availability and usability	Where can the system be obtained? How easy is the system to adapt to different experimental setups and how easy is the system to use?
Shape requirements	What body shape is the tracking system optimized for?
Taxa studied	List of taxa that have already been tracked with the system, and corresponding publications.
Identity method	Is the identity of individuals maintained through occlusions, or are they solved using movement models (see Box 3)? If they are maintained, then is this through i) markers, ii) body size, or iii) other more specific morphological or behavioral traits that we group under the umbrella of fingerprinting. See Figure 2 (main text) for more information about identity maintenance.
Required resolution per individual	The approximate minimum size of individuals required for the system to perform in the way described in Table S1 (e.g. maximum number of individuals, ability to solve crossings, etc.). It is measured as the number of pixels that cover the area of an individual, averaged across a video in which the animals are as small (relative to arena size) as possible for the system.
Maximum individuals	Maximum number of individuals that can be tracked.
Raw output	Does the output of tracking include trajectories in 2D or 3D? Is any pose data collected, such as orientation or more detailed pose? Does the system run in real-time.
Requirements	What OS and any other software does the system require? Do you need any special imaging equipment, such as multiple cameras?