3D tracking of animals in the field, using rotational stereo videography

Emmanuel de Margerie, Manon Simonneau, Jean-Pierre Caudal, Cécilia
Houdelier, Sophie Lumineau

METHODS & TECHNIQUES

3D tracking of animals in the field using rotational stereo videography

Emmanuel de Margerie*, Manon Simonneau, Jean-Pierre Caudal, Cecilia Houdelier and Sophie Lumineau

ABSTRACT

We describe a method for tracking the path of animals in the field, based on stereo videography and aiming-angle measurements, combined in a single, rotational device. In open environments, this technique has the potential to extract multiple 3D positions per second, with a spatial uncertainty of <1 m (rms) within 300 m of the observer, and <0.1 m (rms) within 100 m of the observer, in all directions. The tracking device is transportable and operated by a single observer, and does not involve any animal tagging. As a video of the moving animal is recorded, track data can easily be completed with behavioural data. We present a prototype device based on accessible components that achieves about 70% of the theoretical maximal range. We show examples of bird ground and flight tracks, and discuss the strengths and limits of the method, compared with existing fine-scale (e.g. fixed-camera stereo videography) and large-scale tracking methods (e.g. GPS tracking).

KEY WORDS: Animal tracking, Stereo video, Spatial behaviour, Movement ecology, Flight kinematics

INTRODUCTION

Tracking the path of wild animals in the field yields information about multiple aspects of a species' biology. Long-term tracks, over days or more, inform ecologists on large-scale space use (e.g. home range, migration, dispersal). Locally, short-term tracks with higher sampling frequency and finer spatial resolution, allow biologists to observe the animal's path during a given activity phase (e.g. foraging), addressing questions about the animal's exploratory strategy, orientation skills or even biomechanical interaction with its physical milieu. Nathan (2008) synthesized the existing approaches to the study of organismal movement, and proposed an integrated 'movement ecology' framework.

Here, we present a method for local tracking of animal movement in 3D, based on stereo videography and aiming-angle measurements, from a single observation point. We aimed for: (i) tracking free-moving animals in the field; (ii) no animal tagging; (iii) a spatial uncertainty finer than GPS; (iv) omnidirectional tracking around the observer; (v) video recording the animal's behaviour; and (vi) a transportable, single-operator, affordable device.

The general principle of our method is to measure the position of an animal through its spherical coordinates, relative to the stationary observer (Fig. 1A). An angle measuring base (AMB), similar to a theodolite, records azimuth (a) and inclination (i) angles while the observer frames the moving animal in a viewfinder. Supported by

Laboratory of Animal and Human Ethology, UMR CNRS 6552, University of Rennes 1, Rennes 35042, France.

*Author for correspondence (emmanuel.demargerie@univ-rennes1.fr)

the AMB, a stereo-videography device (SVD) records stereo images of the animal, from which the distance (d) from the observer is calculated. Altogether, the device is similar to a surveying tacheometer (or total station), but works at a higher sampling frequency (up to the video frame rate). Moreover, the embedded video record of the animal is used to extract additional behavioural data that can be combined with the tracking data.

There are two expected limits to this tracking method. First, the animal must remain visible during its movement; hence, the method only applies to terrestrial and aerial paths in open environments. The second limit results from the stereo-image-based distance evaluation: as uncertainty in terms of the distance measure increases quadratically with distance from the observer (Cavagna et al., 2008), the range of the tracking device will be finite, restricting precise tracking to a given radius around the observer.

In order to assess the usefulness and limits of this method, we investigated its theoretical aspects, constructed a prototype device and tracked various bird species during their locomotor activities.

RESULTS AND DISCUSSION
Quantization resolution and position uncertainty

The main theoretical results, which are essential to understanding the field results, are reported here (see Appendix for details).

On a standard dual-camera SVD (Fig. 1B), the distance measurement resolution (Δd, i.e. the smallest measurable distance variation, in m) is proportional to the square of the distance to the animal (d, in m), divided by the base length separating the two image sensors (BL, in m), the image width (IW, in pixels) and the focal length (eqFL, 35 mm-equivalent focal length, in m).

$$\Delta d = 0.036 \, \frac{d^2}{\text{BL IW eqFL}}. \tag{1}$$

The AMB angular resolution for azimuth (Δa, rad) and inclination (Δi, rad), measured with N bits digital encoders is:

$$\Delta a = \Delta i = \frac{2\pi}{2^N}. \tag{2}$$

Perpendicular to the observer–animal radial direction, Δi translates into a linear 'meridian' resolution Δm (in m) that is well approximated by:

$$\Delta m \approx d \, \tan(\Delta i). \tag{3}$$

Similarly, Δa translates into a 'parallel' resolution Δp (m):

$$\Delta p \approx d \, \tan(\Delta a) \cos(i). \tag{4}$$

Unlike Δm, Δp depends on i. It is maximal and equals Δm when $i=0$ (assumed below for simplicity).

Fig. 1C shows how resolutions Δd, Δm and Δp change with increasing distance. As Δd grows quadratically, it becomes vastly superior to Δm and Δp at larger distances. As Δi and Δa are very

List of symbols and abbreviations

a	azimuth angle (rad)
AMB	angle measuring base
BL	base length between cameras (m)
d	observer–animal distance (m)
d_{max}	maximal range (m)
D_{max}	maximum distance at which animals move (m)
D_{min}	minimum distance at which animals move (m)
D_{typ}	typical distance at which animals move (m)
DOF	depth of field (m)
eqFL	35 mm-equivalent focal length (m)
f	quadratic polynomial model
FL	focal length (m)
FOV	camera field of view (rad)
h	inverse-curve model
i	inclination angle (rad)
IW	digital image width (pixels)
k	error multiplying factor
lhFOV	camera linear horizontal field of view (m)
NI	noise index
POI	point of interest
QPU	quantization position uncertainty (m)
rms	root-mean-square
RSV	rotational stereo videography
s	lateral shift between stereo images (m)
s_c	s at the centre of the image (m)
SF	sampling frequency (Hz)
SVD	stereo-videography device
SW	sensor physical width (m)
TSL	track step length (m)
V	animal speed (m s^{-1})
VOI	volume of interest (m^3)
Δa	azimuth angle resolution (rad)
Δd	distance resolution (m)
Δi	inclination angle resolution (rad)
Δm	meridian resolution as per Δi (m)
Δp	parallel resolution as per Δa (m)
Δs	lateral shift resolution (m)
ε	residual difference between s and s_c (m)

Fig. 1. Geometric principles of the tracking method. (A) The position of the animal (An) relative to the observer (O) is measured using spherical coordinates, i.e. azimuth angle (a), inclination angle (i) and distance (d). (B) The basic principle of distance (d) measurement from stereo images. BL, base length separating dual cameras; FL, focal length; x_l, x_r, position of the animal image on the left and right images, respectively. The lateral shift (s) between stereo images is an inverse function of distance. (C) Plots of the distance resolution (Δd), as a function of distance (d), for BL=1 m, IW=1920 pixels and various focal lengths (eqFL). Δd grows quadratically with increasing distance, limiting the range of precise tracking. The linear resolutions due to the angular digital measurement (Δm and Δp, for 13 bit digital rotary encoders) are also plotted for comparison.

small (8×10^{-4} rad for a 13-bit encoder), we may assume that an orthogonal quantization of space is performed around the animal position. We are interested in the positional uncertainty (i.e. the expected distance between the measured point and the true point) associated with this space quantization. For each quantized dimension, the mean square uncertainty is 1/12 the square of resolution (Bennett, 1948). Moreover, uncertainties along orthogonal dimensions sum quadratically (Seeber, 2003). Hence, in 3D space, the quantization positional uncertainty (root-mean-square, QPU$_{rms}$, in m) is given by:

$$\text{QPU}_{rms} = \frac{1}{\sqrt{12}}\sqrt{(\Delta d)^2 + (\Delta m)^2 + (\Delta p)^2}. \quad (5)$$

Table 1 gives the distance at which a given value of QPU$_{rms}$ is attained, i.e. the maximal range of the technique (d_{max}) for an acceptable quantization positional uncertainty. When $\Delta d \gg \Delta m$, Eqns 1 and 5 yield a simplified formula for estimating d_{max}:

$$d_{max} \approx 10\sqrt{\text{QPU}_{rms}\,\text{BL}\,\text{IW}\,\text{eqFL}}. \quad (6)$$

Device implementation

Aiming at a large range without compromising transportability, we set BL to 1 m (Fig. 2). Videos are recorded in the high definition available on most current retail digital video cameras
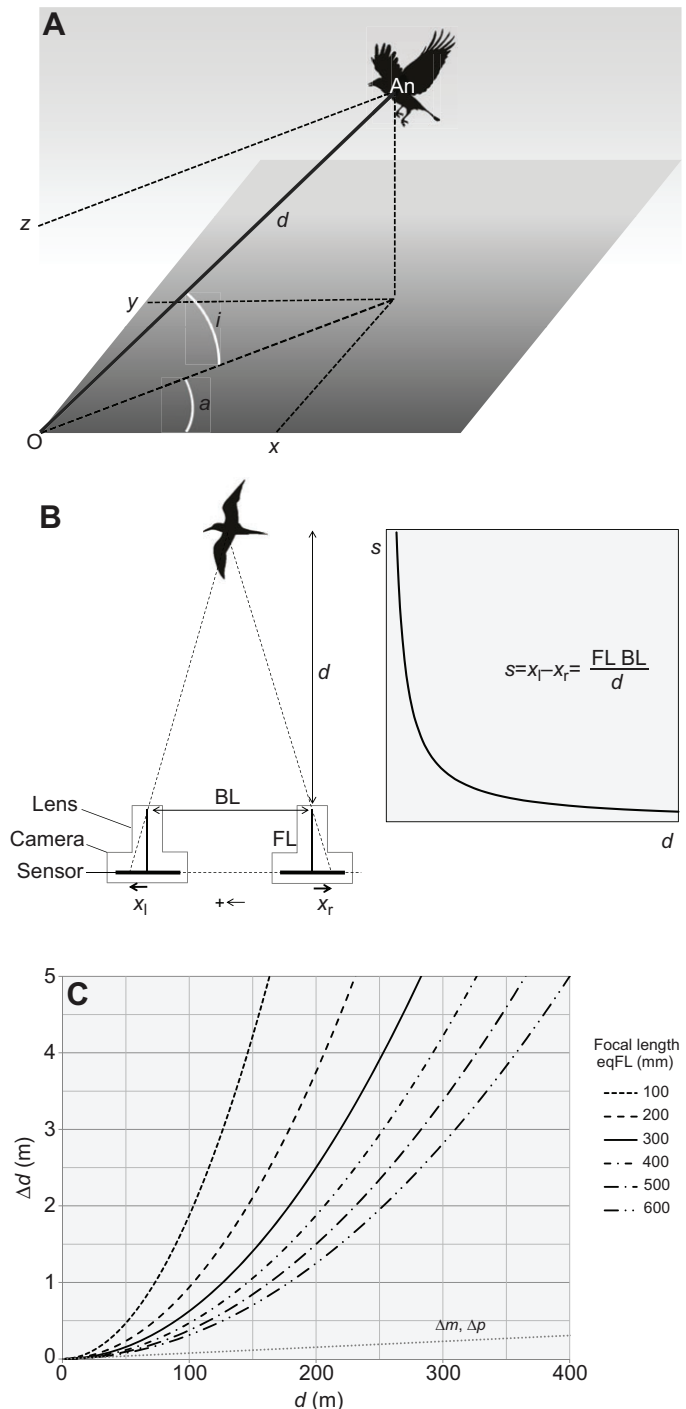
(IW=1920 pixels), at 25 Hz. To avoid synchronization issues between dual cameras, we rely on a single camera and a set of mirrors, projecting stereo images side by side on the sensor (Inaba

Table 1. Maximal range (d_{max}) of tracking

QPU$_{rms}$ (m)	Range d_{max} (m)					
	eqFL=100 mm	200 mm	300 mm	400 mm	500 mm	600 mm
0.01	12	17	20	23	24	25
0.05	30	42	51	58	64	70
0.1	42	60	73	84	93	102
0.5	96	135	166	191	213	234
1	135	192	235	271	303	332
5	303	429	526	607	679	744
10	429	607	744	859	960	1052

d_{max} is shown as a function of the maximal acceptable quantization position uncertainty (QPU$_{rms}$) and the focal length (eqFL).
Values hold for a tracking device with base length between cameras (BL)=1 m and digital image width (IW)=1920 pixels, equipped with 13-bit digital encoders.

et al., 1993). Telephoto lenses of eqFL=323 or 646 mm are used, depending on the animal proximity. Azimuth and inclination angles are measured continuously by a pair of 13-bit digital rotary encoders and recorded on a data logger.

Our prototype device weighs ~20 kg, and when folded can be transported by a single operator on a hand trolley. The cost of the device components amounts to approximately €5000 (including camera: €1500, lenses: €1200, tripod and head: €1000, rotary encoders: €500; laser rangefinder for calibration: €300; AMB and SVD materials and components: €500).

True error of the device

While a perfect device would measure positions with an error equal to QPU$_{rms}$, a real system (physical device+video analysis) will inevitably make larger errors. Two types of error can occur: (1) systematic error, which shifts successive positions along the track by a similar vector – this type of error is of importance to users aiming at positioning the track in its absolute environment (e.g. landscape map); (2) random error, which scatters successive positions in unpredictable directions – this is of particular importance to users interested in relative measurements between positions (e.g. distance, speed, angle). We focus on this scenario below.

At the same time, there are several possible sources of error: (i) space quantization; (ii) point of interest (POI) placement error in stereo images; (iii) calibration error, in particular static device optical or structural distortion, not fully corrected by the calibration procedure (see Appendix); (iv) in-motion device structural distortion, caused by mechanical load during active tracking; and

(v) time-stamping errors, causing diachrony between a, i and d measurements.

A series of error tests should be performed with any new device in order to assess its real error characteristics. For our prototype device, we performed static and dynamic tests (see supplementary material Figs S1, S2). The results show that, overall, random error is about twice the error expected from space quantization alone. We call this random error multiplying factor k. For our current prototype device, $k \approx 2$.

Predicting the tracking range for a given species and locomotor activity

If the acceptable positional error is clearly known (e.g. indexed on animal size; Theriault et al., 2014), Table 1 directly gives the theoretical maximal range of the tracking method. The k error factor of the device should be accounted for, either by dividing the acceptable error by k before entering the data in the table, or by multiplying the output d_{max} range value by $1/\sqrt{k}$ (i.e. 0.7 for k=2, see Eqn 6).

We also propose a 'noise-to-signal' approach based on the distance between two track points (track step length, TSL, in m), which is equal to the animal speed (V, in m s^{-1}) divided by the sampling frequency (SF, in Hz).

$$\text{TSL} = \frac{V}{\text{SF}}. \tag{7}$$

For a given animal speed, the SF will determine the TSL. The TSL value will in turn determine the smallest path pattern that will
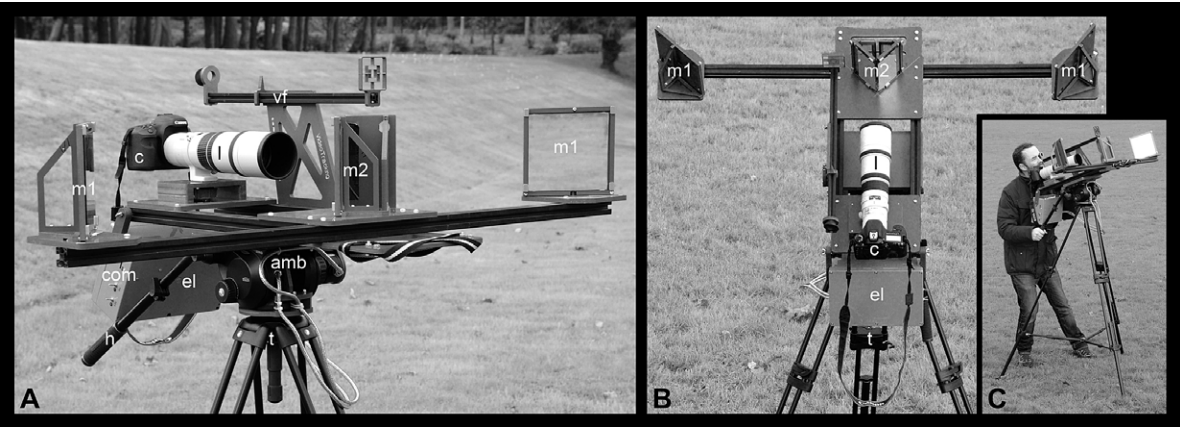

Fig. 2. Prototype device. (A) Device components. m1, primary mirrors; m2, secondary mirrors; l, 646 mm eqFL lens; c, camera; vf, viewfinder; amb, angle measuring base with two digital rotary encoders; h, handle; el, electronics case with amb commands (com); t, tripod. (B) Top view showing the geometry of the stereo videography device. (C) Device in operation during flight tracking.

be resolved in the track (e.g. a circular loop of radius TSL will have a length of 2π TSL, and contain only seven positions). Hence, the user should choose as high a SF as possible to minimize TSL and resolve fine track patterns (see e.g. Rowcliffe et al., 2012). In contrast, if the TSL is too small, random positional error will deteriorate the track (distance between points, and angles between track segments). The amount of noise (noise index, NI) can be quantified by the ratio of the random error to the TSL:

$$NI = \frac{k \, QPU_{rms}}{TSL}. \qquad (8)$$

For NI=1, the rms random error (i.e. the standard deviation of position) is equal to the interval between two track points, resulting in a very noisy track. The acceptable NI will depend on the aim of the study (path pattern description versus biomechanics), on the scale of relevant path patterns relative to TSL, and on the intent of data smoothing. If the user awaits a raw, unsmoothed track containing readable spatial patterns, we suggest keeping NI<0.5, and monitoring the real NI value along the measured tracks. Setting upper bounds on TSL and NI allows calculation of an acceptable QPU_{rms} (Eqn 8), and in turn a maximal range d_{max} for the device (Table 1, Eqn 6). In the end, d_{max} is the radius of a spherical volume of interest (VOI) within which the animal should be reliably trackable.

Magpie walk track

We tracked a common magpie (*Pica pica*) walking and feeding on a flat grass lawn, using the 323 mm eqFL lens. The 345 s track was sampled at 1 Hz (i.e. once every 25 video frames). The bird moved at a mean speed of 0.25 m s^{-1}, covering a distance of about 90 m. The mean TSL was 0.25 m. The random position error (2 QPU_{rms}) was 0.04 m at 30 m from the device, 0.06 m at 40 m and 0.09 m at 50 m. Based on the mean TSL, NI was 0.16, 0.24 and 0.36, respectively. Fig. 3A shows a trajectory that is indeed smoother at shorter distances, and noisier beyond 50 m. NI could potentially be lowered by using a lower SF/larger TSL (i.e. downsampling). The video record allowed identification of moments when the bird pecked in the grass (most of which were immediately followed by a trophic interaction with another, younger magpie rejoining the focal individual). With these behavioural data combined with the track (and many replications), it would be possible to study the spatial strategy underlying the foraging activity.

Swift flight track

We recorded the flight of a common swift (*Apus apus*), for 45 s, and sampled its 3D track at 6.25 Hz (i.e. once every 4 video frames). We used our longer lens for this track (eqFL=646 mm), and it was sometimes difficult to keep the bird within the frame, resulting in some missing data along the track. The mean speed from raw positions was 10.76 m s^{-1}, for a travelled distance of about 470 m. The mean TSL was 1.72 m. Random error (2 QPU_{rms}) was 0.18 m at 100 m, 0.38 m at 150 m and 0.68 m at 200 m. NI was 0.10, 0.22 and 0.40, respectively, and again the track appears less smooth at greater distances (Fig. 3B). Speed data obtained from raw position subtraction contain important noise. As an alternative to downsampling, we performed spline smoothing (Garcia, 2010). The smoothed path (Fig. 3C,D) shows speed data that could potentially be used for a kinematic analysis. The speed from smoothed data ranged from 4.77 to 14.31 m s^{-1} (mean 10.22 m s^{-1}). We detected a probable prey capture at the upper-right of the track (lowest speed, protracted head). Hence, both flight (flapping/gliding) and aerial feeding behaviour data can be

combined with the positional and speed data. Note that the wind speed would have to be subtracted from the ground speed to yield the air speed of the animal, as needed in a biomechanical perspective. Depending on the bird distance and height, wind measurements from stationary anemometers or balloon launch tracking should be integrated with the tracking data (see Henningsson et al., 2009; Pennycuick et al., 2013). With these complementary data, and many replications, one could provide reliable foraging speeds of a swift, to be compared with migration, roosting and display flight speeds (Henningsson et al., 2009, 2010).

Woodpecker flight track

We recorded a brief (5 s) flight bout of a European green woodpecker (*Picus viridis*) at close range, with the same device configuration and error as for the magpie walk. Because the woodpecker was moving much faster (mean speed from raw data, 9.63 m s^{-1}), we could sample its track at 25 Hz (i.e. on every video frame), with a TSL large enough (mean 0.38 m) to maintain acceptable NI values (0.11, 0.16 and 0.24 at 30, 40 and 50 m, respectively). A side view of the track (Fig. 3E) shows a typical undulating pattern, with alternating flapping and bounding (fully retracted wings) phases. The ground speed during these phases can be estimated after spline smoothing.

Comparison with existing tracking methods

A first comparable tracking method is the 'Ornithodolite' of Pennycuick (1982), and subsequent implementations (Tucker, 1995; Hedenström et al., 1999). Those systems measure the same variables (*a*, *i*, *d* from a single point). However the distance measure is not based on recorded stereo images, but rather on the manual actuation of an optical rangefinder by the operator. A downside is that the tracking accuracy depends on the operator skill in aiming exactly at the moving bird, while simultaneously adjusting the rangefinder knob. Our method corrects aiming errors as long as the animal remains within the recorded images, and postpones distance measurement to later image analysis. Although this is time consuming, it enables the possibility of extracting accurate, corrected positions at high frequencies, with less user-skill dependency. Another downside of the Ornithodolite is the lack of an embedded record of the animal behaviour, unless the system is augmented with secondary behavioural data acquisition (e.g. video).

Recently, Pennycuick et al. (2013) used a pair of military binoculars equipped with a laser rangefinder, a magnetic compass and an inclinometer. Although this system is very portable, it has limited SF (<0.5 Hz) and is much less affordable than our system.

Delinger and Willis (1988) proposed a device measuring only the aiming angles (*a*, *i*) of a video camera. Two distant systems and a triangulation method are used to measure position. The requirement for dual operators is a downside and implies synchronization issues, but this system potentially offers low uncertainty at long distances, and behavioural records. Tucker and Schmidt-Koenig (1971) had used a similar dual-theodolite system, without the video record.

Image-based tracking using fixed cameras is another, more widespread method. A single fixed camera can record 2D movements, in the laboratory (e.g. Aureli et al., 2012) or even outdoors (Pillot et al., 2010; Collett et al., 2013). 3D tracks in the field have been measured using multiple fixed cameras (Major and Dill, 1978; Pomeroy and Heppner, 1992; Ikawa et al., 1994; Budgey, 1998; Ballerini et al., 2008; Corcoran and Conner, 2012; Shelton et al., 2014). The VOI is defined by the fixed intersection of the cameras' field of view (FOV). To cover a large VOI, and track
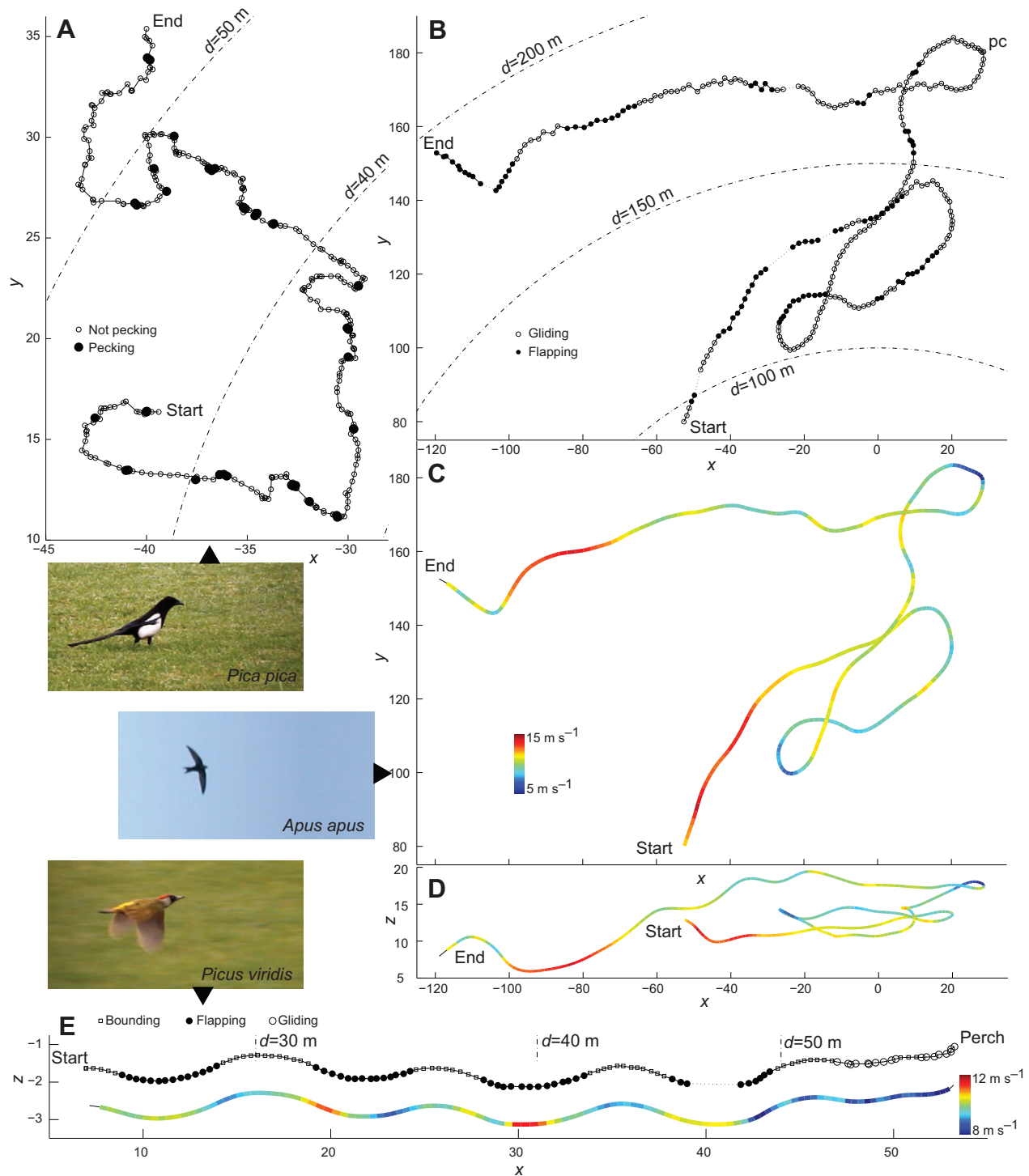
Fig. 3. Example tracks. (A) Magpie walk track sampled at 1 Hz, top view. (B–D) Swift flight track at 6.25 Hz: (B) unsmoothed track, top view; pc, prey capture; (C) spline-smoothed track with speed data, top view; (D) spline-smoothed track with speed data, side view. (E) Woodpecker flap-bounding flight track at 25 Hz, side view. Both the unfiltered and spline-smoothed track with speed data (Z-axis offset: −1 m) are shown. Scales as per axes values (m). Note that the vertical scale is twice the horizontal scale in side views (D,E). Dashed arcs indicate distance from the observer and device.

animals for a significant duration, this technique usually requires wider angle lenses (eqFL≈50 mm), which has a few drawbacks. First, a larger between-cameras distance is required to maintain a low position uncertainty (Eqns 1,5), which can limit the system's portability (Cavagna et al., 2008; see Theriault et al., 2014, for recent progress). Moreover, the animal projects a small image on the camera sensor, which can limit positional and behavioural analysis

(Theriault et al., 2014). However, as a benefit, fixed cameras capture the entire VOI continuously, hence multiple animals present in the VOI can be tracked simultaneously. The size of the VOI depends on the desired spatial uncertainty: recent studies have monitored VOIs from 10^2 m³ (Corcoran and Conner, 2012) up to 10^4 m³ (Theriault et al., 2014; Shelton et al., 2014) or even 10^6 m³ (Ballerini et al., 2008; Cavagna et al., 2008). Vertebrate flight bouts of a few seconds

can usually be recorded. In comparison with fixed-camera stereo videography, our method is based on a short BL/long eqFL, rotational configuration. The short BL allows for a single, easily transportable device. The long eqFL allows a greater magnification of the animal image, but can limit the possibility of tracking multiple animals. The rotational, omnidirectional device yields a virtually spherical VOI, which in some conditions allows for longer tracking bouts (e.g. 45 s in Fig. 3B, in a field VOI$\approx 10^7$ m^3). However, as with other stereo-videography techniques, the size of the VOI remains strongly dependent on the tolerated spatial uncertainty.

Aside from optical systems, GPS tracking (Cagnacci et al., 2010) has as a main benefit its unlimited, global range. The position uncertainty of GPS is about 6.5 m in 2D (distance rms, drms; Seeber, 2003) and more than 10 m in 3D (mean radial spherical error, MRSE). It can be increased by various environmental factors, and field errors of 30 m are often assumed (Frair et al., 2010). Although GPS tags can sample positions at up to 1 Hz (e.g. Dell'Ariccia et al., 2008; Vyssotski et al., 2009), they are often used at much lower SF, to preserve the tag's battery life (e.g. Debeffe et al., 2013). These specifications make GPS tracking well adapted to large-scale/long-term tracking, but less so to fine-scale local path investigations (Frair et al., 2010; Rowcliffe et al., 2012). Other radiowave-based tracking methods, such as VHF tracking (smaller tags than GPS; Daniel Kissling et al., 2014), scanning harmonic radar (even smaller passive tags; Ovaskainen et al., 2008; Lihoreau et al., 2012) and surveillance or tracking radars (no tag; Gauthreaux and Belser, 2003; Henningsson et al., 2009) each have specific advantages over GPS (especially for tracking small species), but lack the global range, and usually do not provide lower spatial uncertainty than GPS tracking, nor a SF above 1 Hz.

The present tracking method attains GPS-like uncertainty (QPUrms$\approx$10 m) around 500–1000 m from the device (Table 1). This finite range suggests that the present method should not be considered as an alternative to GPS for long-term tracking (an animal flying forward at 10 m s^{-1} crosses such a VOI within a few minutes), but rather as a valuable complementary technique at the local scale. Within its range, it is capable of much finer – metres to centimetres – uncertainty, combined with higher SF. Animal follow-up is based on continuous visibility rather than tagging, which has both downsides (limited to open environments, pseudoreplication) and benefits (no animal capture, sample size not limited by the cost of the tags). Lastly, the embedded record of animal behaviour provides supplementary data that help with understanding the mechanisms at play along the animal path, and reveal both movement patterns and processes (Nathan, 2008).

In conclusion, by allowing animal image magnification and omnidirectional tracking, the present method expands the range of operation – and the potential track duration – of field stereo videography, with minimal field deployment difficulties. It cannot match the range of a GPS tracking system, but within its operational range provides richer information (fine-scale spatio-temporal and behavioural data), non-invasively. We hope that this comparatively accessible tracking method (we propose the acronym RSV for rotational stereo videography) will allow biologists to develop new spatial behaviour and movement ecology studies, at intermediate spatial scales.

MATERIALS AND METHODS
Prototype components
The AMB is composed of a Manfrotto™ (Cassola, Italy) 545B tripod (25 kg payload) and 509HD head, coupled with two AKIndustrie™ (Thal-Marmoutier, France) CHO5 13 bit encoders. The 26 parallel encoder outputs are wired to an Arduino MEGA microcontroller board (www.arduino.cc), through a latch interface based on four SN74LS374N octal flip-flops (Texas-Instruments™, Dallas, TX, USA). The angular SF is 50 Hz. Angle values are converted from Gray code to steps (0–8191), time-stamped to the closest millisecond, and recorded on a SD memory card in a Data Logging Shield (Adafruit™, New York, NY, USA). The AMB is powered by a 7.2 V, 2700 mAh battery.

The SVD has a BL of 1 m. We use 6 mm-thick first surface mirrors (FSM, Toledo, OH, USA) of dimensions 150×150 mm (outer, primary mirrors) and 70×150 mm (W×H; inner, secondary mirrors). Mirrors and camera are supported by 30×30 mm aluminium beams, assembled with 9 mm-thick PVC machined plates. The angular position of outer mirrors can be adjusted, allowing for FOV convergence adjustment. We use a Canon™ (Tokyo, Japan) EOS 7D camera, recording full HD (1920×1088 pixels, W×H) frames at 25 Hz. The lens is either a Nikon™ (Tokyo, Japan) 200 mm f/4 Ai, or a Canon™ EF 400 mm f/5.6 L. As the camera has a 22.3 mm-wide sensor, the eqFL is 323 mm and 646 mm, respectively (see Appendix).

Image analysis
We use Matlab™ (MathWorks™, Natick, MA, USA) to analyze individual video frames. The lateral distance between the left and right images of the animal is measured and converted to distance using a reference curve inferred from a calibration video. The horizontal and vertical position of the animal in the frame is used to correct recorded angles for aiming errors. See Appendix and supplementary material Fig. S5 for details.

APPENDIX
Distance resolution
Let two cameras with identical focal length FL (m), separated by base length BL (m), simultaneously capture the image of an animal at a distance d (Fig. 1B). The image of the animal on the right camera's sensor plane is shifted laterally by an amount s (m) compared with the left image (Cavagna et al., 2008). The relationship between these variables is:

$$s = x_l - x_r = \frac{\text{FL BL}}{d}. \qquad (A1)$$

There is an inverse relationship between d and s, which allows us to calculate the distance to the animal, based on the measurement of the lateral shift between the pair of images:

$$d = \frac{\text{FL BL}}{s}. \qquad (A2)$$

As shown in Fig. 1B, at shorter distances the image shift is large and varies steeply with a change in distance, whereas at long distances, the shift is small and remains much more stable. By deriving Eqn A2, we have:

$$\frac{\delta s}{\delta d} = -\frac{\text{FL BL}}{d^2}. \qquad (A3)$$

Or, in absolute values:

$$\Delta d = \frac{d^2}{\text{FL BL}} \Delta s. \qquad (A4)$$

The unit shift variation Δs is the physical width of a pixel, which is determined by the sensor physical width (SW, in m) and the image width (IW, in pixels).

$$\Delta s = \frac{\text{SW}}{\text{IW}}. \qquad (A5)$$

From Eqns A4 and A5, we obtain:

$$\Delta d = \frac{d^2 \, \text{SW}}{\text{FL BL IW}}.$$ (A6)

To normalize the results across various camera sensor sizes, we replace FL with the 35-mm-equivalent focal length (eqFL in m), with reference to the photography standard (SW=0.036 m).

$$\frac{\text{FL}}{\text{SW}} = \frac{\text{eqFL}}{0.036}.$$ (A7)

Eqns A6 and A7 lead to the final distance resolution formula (see Eqn 1 in Results and discussion).

Device design issues
Focal length choice
The main benefit of using longer lenses is the reduction in Δd (Fig. 1C) and increase in the device range (Table 1, Eqn 6). Also, the animal image on the video record is magnified, which can ease the behavioural analysis. However, there are downsides. First, a higher FL implies a narrower camera FOV. The linear horizontal field of view (lhFOV, in m) at a given distance is:

$$\text{lhFOV} = \frac{d \, \text{SW}}{\text{FL}} = \frac{0.036 d}{\text{eqFL}}.$$ (A8)

If the FOV is very narrow, continuously framing an erratically moving animal is not easy, which gives rise to missing data. Moreover, for a given sensor size (SW), longer lenses provide less depth of field (DOF), suggesting the lens should be used at a smaller aperture to obtain a sharp image throughout the range of the device. Lastly, very long lenses are heavy; hence, a stiffer and vibration-dampened support is needed. Note that a solution for obtaining a higher eqFL without the DOF and weight downsides is to use a camera with a smaller sensor (Eqn A7). See supplementary material Figs S3 and S4 for FOV and DOF plots that can help identify the appropriate FL value.

Convergence
A small angle of convergence (typically <1 deg), tilting the optical axes of each camera inwards, is needed to maximize superposition of the two FOVs, especially at shorter distances. Although the formal relationship between d and s (Eqn A2) gets more complex (see Woods et al., 1993), the small angle implied does not significantly affect the subsequent calculation results.

Video versus photo
Current retail cameras are both photo and video capable, contrary to what was previously available (Cavagna et al., 2008). Most cameras provide a video mode, recording 1920×1080 pixel frames at 30 or even 60 Hz. In photo mode, higher definition images can be recorded, at a lower frame rate (e.g. 5184×3456 pixels at 8 Hz for our camera). Hence, when the tracking does not need a very high SF, using the camera in photo mode instead of video mode can increase IW and hence the range of the device (Eqn 6). However, photo frame rate is usually less stable than video frame rate, and a series of photo files contains less behavioural information than a video file.

Mirrors versus dual cameras
With the mirrors/single camera configuration, left and right images are each projected on one half of the same sensor, which solves synchronization issues. The range of the apparatus remains unchanged (as SW/IW in Eqn A6), but the captured FOV is

halved compared with a dual camera system (multiply the results of Eqn A8 by 0.5).

Rolling shutter effect
On the widely available CMOS sensor cameras, each video frame is captured progressively from top to bottom, usually within 1/100 to 1/30 s ('rolling' electronic shutter), such that different parts of a single frame are actually not recorded perfectly simultaneously. This can contribute to time-stamping errors (see Results and discussion, 'True error of the device', error source v) when rotating the SVD very quickly (fast, close movements), but could be corrected in a refined analysis method. Note that the rolling shutter effect is much less pronounced but still exists in photo mode (about 1/250 s for a mechanical shutter). CCD sensors are free from this effect ('global' shutter).

Ways to increase range
According to the noise-to-signal approach of maximal range (Eqn 8), the first way to increase the range, as already discussed, is to choose a larger TSL (lower SF, i.e. downsampling). If this is not possible without losing relevant path information, Eqn 6 states that doubling BL or IW (e.g. '4K' video standard) or eqFL of the device will multiply the range by a factor of $\sqrt{2}$. These effects are multiplying, hence the ranges given in Table 1 could be increased about 3-fold by doubling all three parameters (but with cost, portability and data storage consequences).

Operation in the field
To set the device in the field: (i) choose an unobstructed point of view and evaluate a typical distance (D_{typ}) and distance range (D_{min} to D_{max}) at which animals move; (ii) select a focal length that will provide enough distance resolution ($d_{\text{max}} \geq D_{\text{max}}$), and check that the FOV is not too narrow to reliably frame the moving animal, even at D_{min}; (iii) install the tripod and AMB using a spirit level, then place the SVD on top; set the tripod head friction and counterbalance so that the SVD can move smoothly; (iv) set the camera to full manual video mode; (v) focus the lens to D_{typ}, and close the lens aperture until the captured image is sharp from D_{min} to D_{max}; an aperture as small as f/16 or smaller might be needed; leave the focus ring untouched afterwards; (vi) set the mirrors' convergence so that a point at D_{typ} is projected on the centre of each stereo image; then check that FOV superposition is effective from D_{min} up to D_{max}; (vii) set the camera exposure: set a shutter speed that will stop animal motion on each video frame (1/200 s or faster) and then adjust camera sensitivity (ISO) to get a properly exposed image.

The procedure for tracking an animal is as follows: (i) start the video record; (ii) start the angle record; (iii) perform a brief angular oscillation with the SVD, for angular/video synchronization purposes; (iv) track the animal(s) by keeping it in both right and left images; (v) at the end of tracking, perform a second quick angular movement; (vi) stop the video and angle records.

Image analysis
The main steps of the analysis are: (i) extract still frames from the video file, at the desired SF; (ii) for each frame, measure the lateral shift (s) and the position of the animal in the image (x_{m}, y_{m}) (see supplementary material Fig. S5); (iii) convert s to distance (d), using the reference curve from the calibration video (see below); (iv) synchronize distance and angle data, and for each d value obtain an associated azimuth (a) and inclination (i) value; (v) correct a and i for aiming errors using the position of the animal in the image

(x_m, y_m); (vi) convert spherical coordinates (a, i, d) to Cartesian coordinates (x, y, z); (vii) plot the track.

Calibration

A calibration video is performed for each field session, recording images of static points situated at known distances (we use a Nikon™ Forestry Pro hand laser rangefinder to independently measure the distance to ~6 reference points between D_{min} and D_{max}). The goal is to build a reference curve for the relationship between d and s. We cannot simply use Eqn A2 because (i) the returned d value holds only for a point exactly in front of the apparatus, (ii) it does not account for convergence and (iii) other factors such as lens optical distortion and device structural distortion can interfere. In reality, s mainly depends on d, but also slightly on the (x_m, y_m) position of the point in the image. Hence, each reference point needs to be filmed at various (x_m, y_m) positions in the image, by 'scanning' with the apparatus. The analysis of the calibration video frames yields a large set of (d, s, x_m, y_m) values that are used to build a reference model. First, we note ε the residual difference between s at any position in the image and s_c at the centre of the image $(x_m=0, y_m=0)$, and fit a quadratic polynomial model f to the observed variation of ε with x_m, y_m and s:

$$s - s_c = \varepsilon = f(x_m, y_m, s). \tag{A9}$$

Then, we fit a 3-coefficient (C_{1-3}) inverse-curve model h to the variation of d with s_c:

$$d = h(s_c) = \frac{C_1}{(s_c - C_2)} + C_3. \tag{A10}$$

Using this calibration reference model to compute the distance to a tracked animal is a three-step process: (i) extract (s, x_m, y_m) from the video frame; (ii) compute ε using the f model, and subtract ε from s to obtain s_c (i.e. the lateral shift if the animal was perfectly centred); (iii) compute d using the h model.

References
Aureli, M., Fiorilli, F. and Porfiri, M. (2012). Portraits of self-organization in fish schools interacting with robots. *Physica D* **241**, 908-920.

Ballerini, M., Cabibbo, N., Candelier, R., Cavagna, A., Cisbani, E., Giardina, I., Orlandi, A., Parisi, G., Procaccini, A., Viale, M. et al. (2008). Empirical investigation of starling flocks: a benchmark study in collective animal behaviour. *Anim. Behav.* **76**, 201-215.

Bennett, W. R. (1948). Spectra of quantized signals. *Bell Syst. Tech. J.* **27**, 446-472.

Budgey, R. (1998). *Three Dimensional Bird Flock Structure and its Implications for Birdstrike Tolerence in Aircraft*. Stara Lesna, Slovakia: International Bird Strike Commitee.

Cagnacci, F., Boitani, L., Powell, R. A. and Boyce, M. S. (2010). Animal ecology meets GPS-based radiotelemetry: a perfect storm of opportunities and challenges. *Philos. Trans. R. Soc. B Biol. Sci.* **365**, 2157-2162.

Cavagna, A., Giardina, I., Orlandi, A., Parisi, G., Procaccini, A., Viale, M. and Zdravkovic, V. (2008). The STARFLAG handbook on collective animal behaviour: 1. Empirical methods. *Anim. Behav.* **76**, 217-236.

Collett, T. S., de Ibarra, N. H., Riabinina, O. and Philippides, A. (2013). Coordinating compass-based and nest-based flight directions during bumblebee learning and return flights. *J. Exp. Biol.* **216**, 1105-1113.

Corcoran, A. J. and Conner, W. E. (2012). Sonar jamming in the field: effectiveness and behavior of a unique prey defense. *J. Exp. Biol.* **215**, 4278-4287.

Daniel Kissling, W., Pattemore, D. E. and Hagen, M. (2014). Challenges and prospects in the telemetry of insects. *Biol. Rev.* **89**, 511-530.

Debeffe, L., Morellet, N., Cargnelutti, B., Lourtet, B., Coulon, A., Gaillard, J.-M., Bon, R. and Hewison, A. J. M. (2013). Exploration as a key component of natal dispersal: dispersers explore more than philopatric individuals in roe deer. *Anim. Behav.* **86**, 143-151.

Delinger, W. G. and Willis, W. R. (1988). High-precision portable instrument to measure position angles of a video camera for bird flight research. *Rev. Sci. Instrum.* **59**, 797-801.

Dell'Ariccia, G., Dell'Omo, G., Wolfer, D. P. and Lipp, H.-P. (2008). Flock flying improves pigeons' homing: GPS track analysis of individual flyers versus small groups. *Anim. Behav.* **76**, 1165-1172.

Frair, J. L., Fieberg, J., Hebblewhite, M., Cagnacci, F., DeCesare, N. J. and Pedrotti, L. (2010). Resolving issues of imprecise and habitat-biased locations in ecological analyses using GPS telemetry data. *Philos. Trans. R. Soc. B Biol. Sci.* **365**, 2187-2200.

Garcia, D. (2010). Robust smoothing of gridded data in one and higher dimensions with missing values. *Comput. Stat. Data Anal.* **54**, 1167-1178.

Gauthreaux, S. A., Jr and Belser, C. G. (2003). Radar ornithology and biological conservation. *Auk* **120**, 266-277.

Hedenström, A., Rosén, M., Akesson, S. and Spina, F. (1999). Flight performance during hunting excursions in Eleonora's falcon Falco eleonorae. *J. Exp. Biol.* **202**, 2029-2039.

Henningsson, P., Karlsson, H., Bäckman, J., Alerstam, T. and Hedenström, A. (2009). Flight speeds of swifts (*Apus apus*): seasonal differences smaller than expected. *Proc. R. Soc. B Biol. Sci.* **276**, 2395-2401.

Henningsson, P., Johansson, L. C. and Hedenström, A. (2010). How swift are swifts *Apus apus*? *J. Avian Biol.* **41**, 94-98.

Ikawa, T., Okabe, H., Mori, T., Urabe, K.-i. and Ikeshoji, T. (1994). A method for reconstructing three-dimensional positions of swarming mosquitoes. *J. Insect Behav.* **7**, 237-248.

Inaba, M., Hara, T. and Inoue, H. (1993). A stereo viewer based on a single camera with view-control mechanisms. In *Proceedings of the 1993 IEEE/RSJ International Conference on Intelligent Robots and Systems '93* **3**, 1857-1865.

Lihoreau, M., Raine, N. E., Reynolds, A. M., Stelzer, R. J., Lim, K. S., Smith, A. D., Osborne, J. L. and Chittka, L. (2012). Radar tracking and motion-sensitive cameras on flowers reveal the development of pollinator multi-destination routes over large spatial scales. *PLoS Biol.* **10**, e1001392.

Major, P. F. and Dill, L. M. (1978). The three-dimensional structure of airborne bird flocks. *Behav. Ecol. Sociobiol.* **4**, 111-122.

Nathan, R. (2008). An emerging movement ecology paradigm. *Proc. Natl. Acad. Sci. USA* **105**, 19050-19051.

Ovaskainen, O., Smith, A. D., Osborne, J. L., Reynolds, D. R., Carreck, N. L., Martin, A. P., Niitepold, K. and Hanski, I. (2008). Tracking butterfly movements with harmonic radar reveals an effect of population age on movement distance. *Proc. Natl. Acad. Sci. USA* **105**, 19090-19095.

Pennycuick, C. J. (1982). The ornithodolite: an instrument for collecting large samples of bird speed measurements. *Philos. Trans. R. Soc. B Biol. Sci.* **300**, 61-73.

Pennycuick, C. J., Åkesson, S. and Hedenström, A. (2013). Air speeds of migrating birds observed by ornithodolite and compared with predictions from flight theory. *J. R. Soc. Interface* **10**, 20130419.

Pillot, M. H., Gautrais, J., Gouello, J., Michelena, P., Sibbald, A. and Bon, R. (2010). Moving together: incidental leaders and naïve followers. *Behav. Process.* **83**, 235-241.

Pomeroy, H. and Heppner, F. (1992). Structure of turning in airborne rock dove (*Columba Livia*) flocks. *Auk* **109**, 256-267.

Rowcliffe, J. M., Carbone, C., Kays, R., Kranstauber, B. and Jansen, P. A. (2012). Bias in estimating animal travel distance: the effect of sampling frequency. *Methods Ecol. Evol.* **3**, 653-662.

Seeber, G. (2003). *Satellite Geodesy*. Berlin: Walter de Gruyter.

Shelton, R. M., Jackson, B. E. and Hedrick, T. L. (2014). The mechanics and behavior of cliff swallows during tandem flights. *J. Exp. Biol.* **217**, 2717-2725.

Theriault, D. H., Fuller, N. W., Jackson, B. E., Bluhm, E., Evangelista, D., Wu, Z., Betke, M. and Hedrick, T. L. (2014). A protocol and calibration method for accurate multi-camera field videography. *J. Exp. Biol.* **217**, 1843-1848.

Tucker, V. A. (1995). An optical tracking device for recording the threedimensional paths of flying birds. *Rev. Sci. Instrum.* **66**, 3042-3047.

Tucker, V. A. and Schmidt-Koenig, K. (1971). Flight speeds of birds in relation to energetics and wind directions. *Auk* **88**, 97-107.

Vyssotski, A. L., Dell'Omo, G., Dell'Ariccia, G., Abramchuk, A. N., Serkov, A. N., Latanov, A. V., Loizzo, A., Wolfer, D. P. and Lipp, H.-P. (2009). EEG responses to visual landmarks in flying pigeons. *Curr. Biol.* **19**, 1159-1166.

Woods, A., Docherty, T. and Koch, W. (1993). Image distrortions in stereoscopic video systems. In *Proceedings of the SPIE: Stereoscopic Displays and Applications IV*, 1915.