JOURNAL OF THE ROYAL SOCIETY Interface

CrossMark click for updates

J. R. Soc. Interface (2012) **9**, 2624–2638
doi:10.1098/rsif.2012.0150
Published online 23 May 2012

Reconstructing the flight kinematics of swarming and mating in wild mosquitoes

Sachit Butail[1], Nicholas Manoukis[2], Moussa Diallo[3], José M. Ribeiro[4], Tovi Lehmann[4] and Derek A. Paley[1,*]

[1]*Department of Aerospace Engineering, University of Maryland, College Park, MD 20742, USA*
[2]*US Pacific Basin Agricultural Research Center, Agricultural Research Service, US Department of Agriculture, Hilo, HI 96720, USA*
[3]*Malaria Research and Training Center, Faculté de Médecine, de Pharmacie et d'Odontostomatologie, Université de Bamako, Bamako, Mali*
[4]*Laboratory of Malaria and Vector Research, National Institute of Allergy and Infectious Diseases, Bethesda, MD 20892, USA*

We describe a novel tracking system for reconstructing three-dimensional tracks of individual mosquitoes in wild swarms and present the results of validating the system by filming swarms and mating events of the malaria mosquito *Anopheles gambiae* in Mali. The tracking system is designed to address noisy, low frame-rate (25 frames per second) video streams from a stereo camera system. Because flying *A. gambiae* move at $1–4\,\mathrm{m\,s^{-1}}$, they appear as faded streaks in the images or sometimes do not appear at all. We provide an adaptive algorithm to search for missing streaks and a likelihood function that uses streak endpoints to extract velocity information. A modified multi-hypothesis tracker probabilistically addresses occlusions and a particle filter estimates the trajectories. The output of the tracking algorithm is a set of track segments with an average length of 0.6–1 s. The segments are verified and combined under human supervision to create individual tracks up to the duration of the video (90 s). We evaluate tracking performance using an established metric for multi-target tracking and validate the accuracy using independent stereo measurements of a single swarm. Three-dimensional reconstructions of *A. gambiae* swarming and mating events are presented.

Keywords: target tracking; *Anopheles gambiae*; mosquito swarms; mating mosquitoes

1. INTRODUCTION

Quantitative observations of the flight patterns of wild mosquitoes are critical to expanding our understanding of swarming and mating behaviour [1–6]. Female *Anopheles gambiae* find male swarms in order to mate [5,7]. A single mating event results in all of the fertilized eggs that a female mosquito lays in her lifetime [8,9]. Although the basis of mate selection has generated much interest [6–8,10,11], generation of three-dimensional trajectory data of mosquitoes in wild swarms has not been previously accomplished. As in earlier work on midges [12], such trajectory data can provide valuable insight into the dynamical aspects of collective behaviour [13,14]. Past studies on swarming insects [2,3,5,15] focused on two-dimensional trajectories or three-dimensional positions. Recent advancements in high-resolution filming, computer vision and estimation techniques have increased the degree of automation in data collection, and have made available large datasets for subsequent analysis, such as those developed for starlings [16] and fruitflies [17]. Similar analyses of malarial mosquitoes may inform the first steps towards strategies of vector control [7,8].

Multi-target tracking systems have been developed for other animals. In two dimensions, ants have been tracked using a joint-state particle filter and interaction models [18]. In three dimensions, up to a hundred bats have been tracked with three cameras using a Kalman filter in conjunction with a multi-dimensional assignment strategy [19,20]. Fruitflies have been tracked in an acrylic box by setting up the problem of data association across views and time in the form of a global optimization problem that is solved at every step [21]. Real-time tracking systems for flies were developed using an extended Kalman filter in Straw *et al.* [22] and Grover *et al.* [23]. Each of these tracking systems

*Author for correspondence (dpaley@umd.edu).

Electronic supplementary material is available at http://dx.doi.org/10.1098/rsif.2012.0150 or via http://rsif.royalsocietypublishing.org.

Received 28 February 2012
Accepted 26 April 2012

implements a nonlinear filtering or optimization method with specialized likelihood functions, data association strategies and/or experimental design. However, the targets are large with a dark centre or appear in an arena constructed to minimize noise, unlike wild mosquitoes.

Filming wild mosquitoes poses special challenges such as low natural lighting and a cluttered dynamic background. At least two cameras are needed to reconstruct the three-dimensional position of individual mosquitoes. A multi-camera set-up with a large baseline reconstructs positions accurately, but may be difficult to implement in the field. In a multi-target scenario, one must also address the data-association problem, which entails assigning image blobs to targets across multiple views and time steps. With the typical number of mosquitoes in the swarms we studied,[1] the data-association problem is non-trivial. The challenge lies in tracking small, highly manoeuvrable targets that appear as dots or faded streaks in noisy images with frequent occlusions.

This paper describes an automated multi-target tracking system that reconstructs the three-dimensional flight kinematics of individual mosquitoes in wild swarms. We collect data using two cameras operating synchronously at 25 frames per second. (The frame rate is limited by the ambient light.) The cameras and a laptop are powered by an uninterrupted power supply (UPS) for up to 30 min. The mosquitoes appear as dark streaks or dots on a light background. At high speeds, the mosquito streaks fade, making them hard to detect and even harder to track. Because the swarms are dense, occlusions are frequent and often appear in both camera frames. We tested the system by filming swarms and mating events of *A. gambiae* in a rural village in Mali in August 2010. Figure 1 shows a pair of magnified and enhanced sample frames from this field experiment.

In order to track each insect in a wild swarm, we implemented a probabilistic multi-target tracking framework. Specifically, the contributions of this paper are as follows: (i) we provide a measurement likelihood function that uses the properties of image streaks such as midpoint and endpoint locations to extract insect position and velocity; (ii) we provide methods to improve data association in noisy images by adaptively seeking missing measurements and splitting occluded blobs into individual measurements; and (iii) we present validated tracking results in the form of three-dimensional trajectories of wild mosquito swarming and mating events filmed in Mali in August 2010. Although we describe the experimental method and tracking algorithm for mosquito swarms, the techniques presented in this paper may be beneficial for generating trajectory data for other insect swarms in the field or laboratory.

The tracking system is implemented in MATLAB and consists of two parts: an automated component that outputs track segments called tracklets and a human-supervised component that is used to verify and combine the tracklets into full-length tracks. Tracklets

[1]Typical swarms in the field site where we filmed ranged between 30 and 100 mosquitoes, however other sites are known to have swarms of up to 1000 mosquitoes.

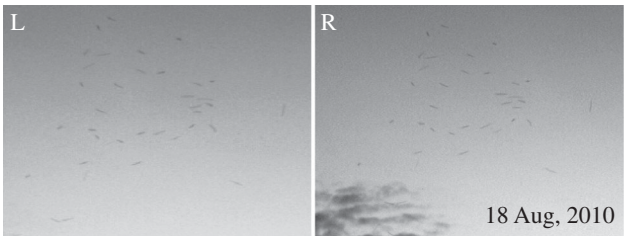

Figure 1. The pair of images above are magnified and enhanced versions of raw footage obtained from the authors' fieldwork in Mali.

produced by the automated component typically range between 15 and 25 frames (0.6–1 s) long and can be used to extract position and velocity data for 80 per cent of the swarm. The human-supervised component uses a particle filter to combine tracklets into individual mosquito tracks. It takes up to 20 min to generate a 10 s track (250 frames). When validated using data filmed in Mali in August 2010, the tracking system produced 30–40 s trajectories of individual mosquitoes in swarms of 6–25 mosquitoes. We have reconstructed six swarms and six mating events from these data. We evaluate the performance of the automated component of the tracking system using an established metric based on position error and the number of targets tracked (cardinality); tracking accuracy was also evaluated using two independent rigs to simultaneously track the same swarm.

The paper is organized as follows: §2 provides a background on multi-target tracking, data association and tracking performance evaluation. Section 3 presents the novel components of the tracking system, including the likelihood function and occlusion reasoning. Section 4 describes the data collection, validation methods and performance evaluation; it also includes representative kinematic data for wild mosquito swarming and mating events. Section 5 summarizes the paper and our ongoing analyses of the kinematic data.

2. PROBABILISTIC TRACKING AND DATA ASSOCIATION

Our aim in designing the mosquito tracking system was to combine nearly indistinguishable measurements available from stereo images recorded at discrete times into trajectories that represent real mosquitoes (targets). We represent target i at time step k by the state vector $\boldsymbol{X}_i[k] \in \mathbb{R}^6$, which contains the target's instantaneous three-dimensional position and velocity. In a Bayesian framework, a tracking algorithm recursively iterates through two steps: the update step and the predict step. The update step uses a measurement model to revise the estimate based on new observations. The predict step integrates a motion model to obtain the target state at the time of the next measurement. The measurement $\boldsymbol{Z}_i[k] \in \mathbb{R}^6$ in our case consists of the two-dimensional positions of the midpoint and two endpoints of an elongated blob in an image that corresponds to the motion-blurred silhouette of a flying mosquito in each of the two images. Assuming

motion model $\boldsymbol{F}$ and measurement model $\boldsymbol{H}$, the state of target i satisfies

$$\left.\begin{aligned} &\boldsymbol{X}_i[k] = \boldsymbol{F}(\boldsymbol{X}_i[k-1], \boldsymbol{w}) \\ \text{and} \quad &\boldsymbol{Z}_i[k] = \boldsymbol{H}(\boldsymbol{X}_i[k], \boldsymbol{n}), \end{aligned}\right\} \tag{2.1}$$

where $\boldsymbol{w}$ and $\boldsymbol{n}$ denote disturbance and noise values, respectively.

Because of noise, disturbances and approximations in $\boldsymbol{F}$ and $\boldsymbol{H}$, the state estimate is a random quantity represented in the form of a probability density function (pdf). We recursively construct the filtering pdf $p(\boldsymbol{X}[k]|\boldsymbol{Z}^k)$ of the joint state $\boldsymbol{X}[k]$ at time step k given the set $\boldsymbol{Z}^k$ of all measurements up to k using the conditional probabilities $p(\boldsymbol{Z}_i[k]|\boldsymbol{X}_i[k])$ and $p(\boldsymbol{X}_i[k]|\boldsymbol{Z}^{k-1})$. The probability $p(\boldsymbol{Z}_i[k]|\boldsymbol{X}_i[k])$, known as a likelihood function, is the conditional probability of measurement $\boldsymbol{Z}_i[k]$ given the state $\boldsymbol{X}_i[k]$. The filtering pdf can be obtained with minimum mean-square error provided the models $\boldsymbol{F}$ and $\boldsymbol{H}$ are linear and the noise values $\boldsymbol{w}$ and $\boldsymbol{n}$ are Gaussian. Otherwise, it is possible to use suboptimal methods such as an extended Kalman filter [24], which predicts and updates the estimate using a first-order linearization of (2.1), or a particle filter [25,26], which represents the target state as a point-mass distribution. A particle filter is attractive in that it relaxes most restrictions on the target and measurement models and the disturbance and measurement noise, but a particle filter is computationally burdensome for a large, joint, state space. We address the computation-size problem by making the assumption that the targets do not interact at short timescales (less than 40 ms), which allows us to use a separate filter for each target. Particle-filtering methods also allow us to encode extra information such as the velocity of the mosquito using the streak endpoints. (See the electronic supplementary material for a description of particle filtering.)

A multi-target tracking system must associate measurements and targets. A target-based method associates each target to a measurement [24], whereas a measurement-based method associates each measurement to a target [27]. A measurement-based method can inherently handle a variable number of targets, which may appear and disappear from the field of view. The reliability of the association depends on the proximity of the actual measurement to the predicted measurement, which is produced from the target estimate using the motion model and the measurement model; measurement proximity is determined using the position likelihood function.

Our choice of a data-association strategy is based on speed, variability and density of targets in the image. A nearest-neighbour association is target-based and associates the predicted measurement to the nearest measurement. It works well in low-target densities with high frame rates [22], but results in duplicate tracks and incorrect associations at high target densities. A global nearest-neighbour (GNN) association avoids duplicate assignments by minimizing a global assignment [28]. GNN has been successful in tracking dense aggregations [17,29] in which the number of targets are fixed and move in two dimensions (so that target overlap is rare), however, the possibility of a variable number of targets and frequent occlusions make it difficult to use GNN without additional heuristics. Short-duration occlusions can be addressed using motion coherence [19,21,30], whereas long-duration occlusions can be overcome by methods that minimize a global cost function over all measurements in a sliding window [20]. However, it is not clear how an offline global optimization method might address a low probability of detection, which is common in our datasets. Instead, we selected a measurement-based method called the multiple-hypothesis tracker (MHT), which looks into future assignment probabilities before making a decision on the current assignment [27]. Within the MHT, we use a motion model at each step to search for missing measurements.

A hypothesis in MHT is a combination of measurement-target assignments that satisfy the following two rules [27]: (i) a target is not associated to more than one measurement and (ii) a target is only associated to a measurement that lies within its gating volume. The gating volume or validation region is generated from the difference $\boldsymbol{\nu} = \boldsymbol{u} - f(\boldsymbol{r})$ between the position measurement $f(\boldsymbol{r}) \in \mathbb{R}^2$ and the predicted position measurement $f(\boldsymbol{r}) \in \mathbb{R}^2$. Let S be the covariance of $\boldsymbol{\nu}$, which is also called the innovation. If a measurement that is normally distributed about the true value lies within the validation region, the weighted norm $||\boldsymbol{\nu}||^2 = \boldsymbol{\nu}^T S^{-1} \boldsymbol{\nu}$ satisfies $||\boldsymbol{\nu}|| < t_{\text{gate}}$. (The quantity $||\boldsymbol{\nu}||$ is also called the Mahalanobis distance [31].) For example, a threshold value $t_{\text{gate}} = 16$ defines a region around a predicted measurement with 99.97 per cent probability of containing the actual measurement [24]. A measurement may be assigned to an existing target, a new target or a false alarm. As time progresses, each hypothesis gives rise to successive hypotheses resulting in an exponential growth in time. Hypothesis reduction strategies include applying a threshold on track probability, choosing a few best hypothesis [31], and clustering the targets [27]. (See the electronic supplementary material for a description of MHT.)

Measures of effectiveness to evaluate a multi-target tracking algorithm include the following [32]: track initiation delay (timeliness), position and velocity errors (accuracy), fragmentation and identity swaps (continuity) and number of targets tracked (cardinality). Performance evaluation also includes visual verification, running the algorithm on simulated data [19], comparison with manually generated ground truth and reconstructing the same event from independent camera systems [30]. We evaluate the performance of our tracking algorithm using manually generated ground-truth as well as filming using two independent camera systems. For comparing tracking results with ground truth, we use the optimal subpattern assignment (OSPA) metric [33], which is a well-established metric for evaluating multi-target tracking algorithms [32] that allows comparison of sets with differing cardinality. (See the electronic supplementary material for details on computing OSPA.)

3. THE MOSQUITO TRACKING SYSTEM

The mosquito tracking system takes a sequence of stereo image pairs as input and produces three-dimensional tracks as output. Figure 2 depicts a block diagram of

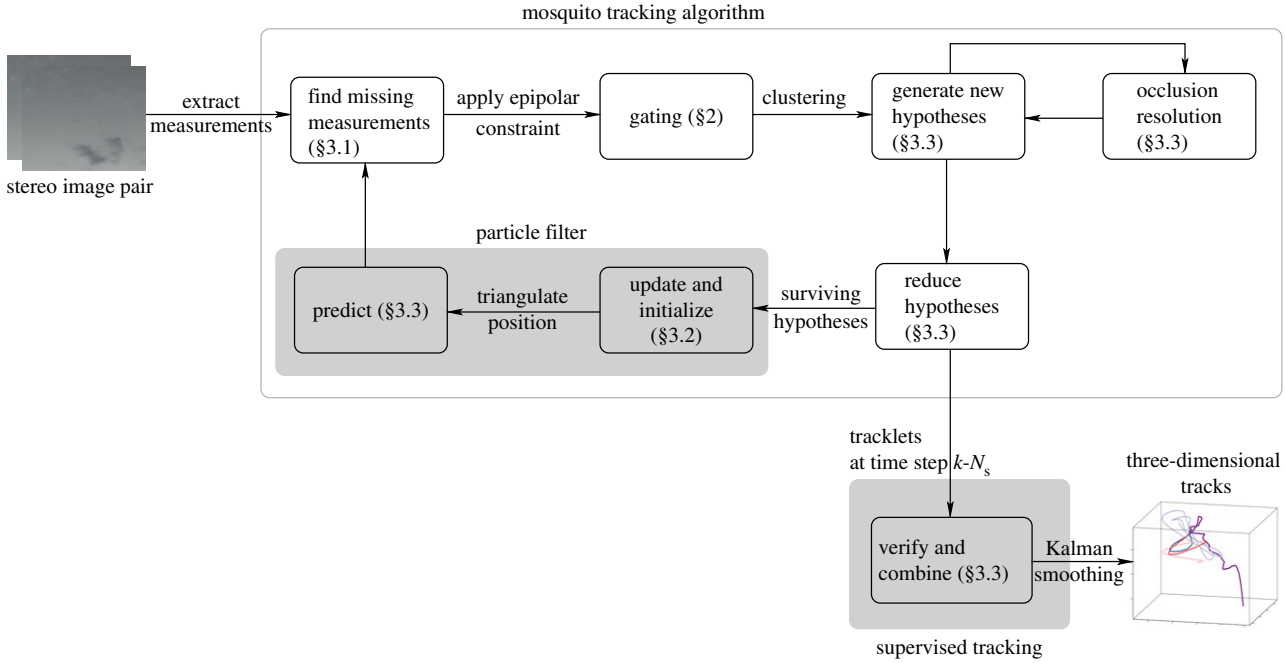

Figure 2. Block-diagram of the mosquito tracking system. (Online version in colour.)

the tracking system, which was created in MATLAB and includes an automated tracking algorithm and a human-supervised component. Image streaks are modelled as straight lines; we extract the midpoint and endpoints as measurements. We find missing measurements using a gating volume generated around predicted measurements. Measurement pairs, i.e. one from each camera, that satisfy the epipolar constraint [34] are selected for data association. We again use a gating volume to assign measurements and targets to independent sets called clusters. Instead of generating definite tracks, hypotheses connecting measurements to targets are propagated to the next step using a particle filter. Based on the probability of each hypothesis at the current time step, the number of hypotheses at a previous time step are reduced to a single assignment. A particle filter verifies and combines tracklets under human supervision and the combined tracks are passed through a Kalman smoother. The tracking algorithm is summarized in table 1.

The remainder of this section describes the novel aspects of the tracking system that we designed to improve its accuracy and level of automation. First, we describe an image-processing technique to find missing measurements during image segmentation. We then describe the measurement model that is used to extract velocity information from streaks. Finally, we present the data-association method, including the strategy to detect and address mosquito occlusions.

3.1. Extracting measurements

During observation of mosquito swarms, which typically appear silhouetted in front of swaying trees under a cloudy sky, it may not be possible to use a static (mosquito-free) background to segment the mosquitoes out of the image stream. Instead, we create a dynamic background by choosing the highest intensity point within a sliding window [35]. Let $B_{u,v}$ be the background image value at the pixel position (u,v) and $t_{\mathrm{win}} = 2d + 1$ be the width of the sliding window centred at time step k. The background value at time k is

$$B_{u,v}[k] = \max_{i \in [k-d, k+d]} B_{u,v}[i]. \tag{3.1}$$

The foreground F is obtained by subtracting the background B from the current image I and applying an intensity threshold t_{int}, i.e. $F_{u,v}[k] = \max(I_{u,v}[k] - B_{u,v}[k], t_{\mathrm{int}}) - t_{int}$. We automatically select the value of t_{int} by running the background subtraction algorithm recursively on different segments of the image sequence until the number of blobs detected are within an acceptable range of the expected number of mosquitoes. We extract blobs using the *regionprops* routine in MATLAB, which performs connected-component labelling to extract features such as centroid, area and bounding ellipse. We remove large insects and birds from the foreground by applying a threshold on the blob area. (See table 2 for the values of the threshold parameters.)

Owing to the duration of the camera exposure ($\delta t_{\mathrm{e}} = 25$ ms),[2] fast mosquitoes (1–4 m s^{-1}) appear as elongated image blobs or streaks. Depending on the mosquito speed, the streaks may fail to appear in the foreground for a given value of t_{int}. Existing strategies for low signal-to-noise environments include the track-before-detect approach [36], which permits raw sensor data as the input. The success for track-before-detect relies on the low target density and relatively straight movement of targets in the measurement space [37]. However, using raw sensor data is not a viable option for mosquito tracking, because it generates more false targets than observed in a single noisy image. Instead,

[2]The duration of exposure (25 ms) is less than the time between frames (40 ms). The remaining time (15 ms) is for image processing.

Table 1. Mosquito tracking algorithm.

Input: Sequence of synced images from a stereo-camera set-up, camera calibration matrices, parameters in table 2
Output: Estimated three-dimensional mosquito trajectories
For each time step k:

(1) *Extract measurements:* model each blob as a straight line and find the midpoint and endpoints.
(2) *Find missing measurements, if any:* ensure that each hypothesized target has at least one measurement within the gating volume; if not, lower the intensity threshold. If a measurement is found append it to the existing set of measurements.
(3) *Validate:* use the epipolar constraint (3.2) to generate valid measurement pairs, one from each camera view.
(4) *Cluster:* use gating volume of each target within a cluster to add measurements to that cluster. A cluster is the smallest set of measurements and targets that exist independently; combine/divide existing clusters as needed.
(5) *Compute hypotheses:* generate hypotheses for each cluster, and compute probabilities.
↪*Resolve occlusions:* if an occlusion is detected split the image blob into individual streaks as described in §3.3 and recompute the hypotheses.
(6) *Hypothesis reduction:* based on the most probable hypothesis at k and scanback range N_s, reduce the number of hypotheses at $k - N_s$ to a single assignment.
(7) *Initialize and update:* initialize tentative targets from unassociated measurement pairs; resample target states based on hypotheses using the three-dimensional estimate and velocity likelihood function (3.6). Each new target forms a new cluster.
(8) *Predict:* use the constant velocity motion model with random (Gaussian) disturbance to propagate hypotheses to time step $k + 1$.

Table 2. Parameter values used for data collection and tracking.

parameter	value	description
b	20 cm	stereo camera configuration baseline
t_{win}	7 frames	sliding window for segmentation
Σ_{ep}	diag{4,4} pixels^2	covariance of endpoint error
σ_w	100 m^2 s^{-4}	covariance of disturbance
δt_e	25 ms	duration of camera exposure
t_{gate}	16	threshold for gating volume
t_e	0.5	threshold on epipolar constraint
t_{area}	(20, 150)	minimum and maximum blob areas
N_s	1 frames	scanback for MHT
N_p	200	number of samples in particle filter

we search for the missing streak in a new foreground generated using a threshold equal to $t_{\text{int}} = 0.75 t_{\text{int}}$. The search is performed within the gating volume of the predicted measurement. If a missing measurement is found, it is added to the list of existing measurements.

A measurement $\boldsymbol{Z}^c = [\boldsymbol{e}^c_-, \boldsymbol{u}^c, \boldsymbol{e}^c_+]^{\mathrm{T}}$ from camera c contains the image locations of a streak's start $\boldsymbol{e}^c_-$, midpoint $\boldsymbol{u}^c$ and end $\boldsymbol{e}^c_+$. These values are extracted from a blob by modelling it as a straight line along the major axis of the bounding ellipse. The streak, therefore, represents a perspective projection of the mosquito trajectory for the duration of exposure δt_e. Let $\tilde{\boldsymbol{u}}^c = [(\boldsymbol{u}^c)^{\mathrm{T}}, 1]^{\mathrm{T}}$ be the homogeneous representation of $\boldsymbol{u}^c$. Assuming without loss of generality that camera 1 is the origin, a pair of measurements with midpoints $\boldsymbol{u}^1$ and $\boldsymbol{u}^2$, one from each camera, must satisfy the epipolar constraint [34]:

$$|(\tilde{\boldsymbol{u}}^2)^{\mathrm{T}} F \tilde{\boldsymbol{u}}^1| < t_e, \tag{3.2}$$

where $F \in \mathbb{R}^{3\times3}$ is the fundamental matrix for the stereo camera calibration and $t_e \ll 1$ is a value that depends on calibration accuracy. Measurement pairs from a true target satisfy the above constraint; clutter or mismatched measurement pairs should not. We use the midpoint and endpoint locations to define likelihood functions for position and velocity.

3.2. Position and velocity likelihood functions

A constant-velocity model suffices to describe the mosquito motion during the exposure, $\delta t_e = 25$ ms. (The streaks are well approximated by straight lines on the image plane.) Let $\boldsymbol{r} \in \mathbb{R}^3$ be the three-dimensional location of the midpoint of a streak. The start and end of the streak are located at $\boldsymbol{r}_- = \boldsymbol{r} - \dot{\boldsymbol{r}}\frac{\delta t_e}{2}$ and $\boldsymbol{r}_+ = \boldsymbol{r} + \dot{\boldsymbol{r}}\frac{\delta t_e}{2}$, respectively. The corresponding point on the image plane is given by the perspective projection model [34],

$$f^c(\boldsymbol{r}) = \left(\frac{w_1}{w_3}, \frac{w_2}{w_3}\right), \tag{3.3}$$

where $\boldsymbol{w} = P\tilde{\boldsymbol{r}} \in \mathbb{R}^3$, and P is the camera projection matrix. Let $\mathbb{N}(\boldsymbol{u}; f(\boldsymbol{r}), \Sigma)$ denote a normal density function evaluated at $\boldsymbol{u}$ with mean $f(\boldsymbol{r})$ and covariance matrix $\Sigma \in \mathbb{R}^{2\times2}$. Assuming that the measurement is normally distributed about the true value, the likelihood of midpoint $\boldsymbol{u}^c$ given $\boldsymbol{r}$ is

$$P^c_{\text{mp}}(\boldsymbol{u}^c|\boldsymbol{r}) = \mathbb{N}(\boldsymbol{u}^c; \mathrm{f}^c(\boldsymbol{r}), \Sigma_{\text{mp}}), \tag{3.4}$$

We set the diagonal entries of Σ_{mp} equal to the length of the major and minor axes of the streak's bounding ellipse in the streak frame; the off-diagonal entries are zero.

As with the midpoint likelihood function, we assume the endpoint likelihood function is based on a normal density function. However, owing to uncertainty in the labelling of the start and end of the streak, the endpoint likelihood function is bimodal. The directional ambiguity is described by a sum of conditional

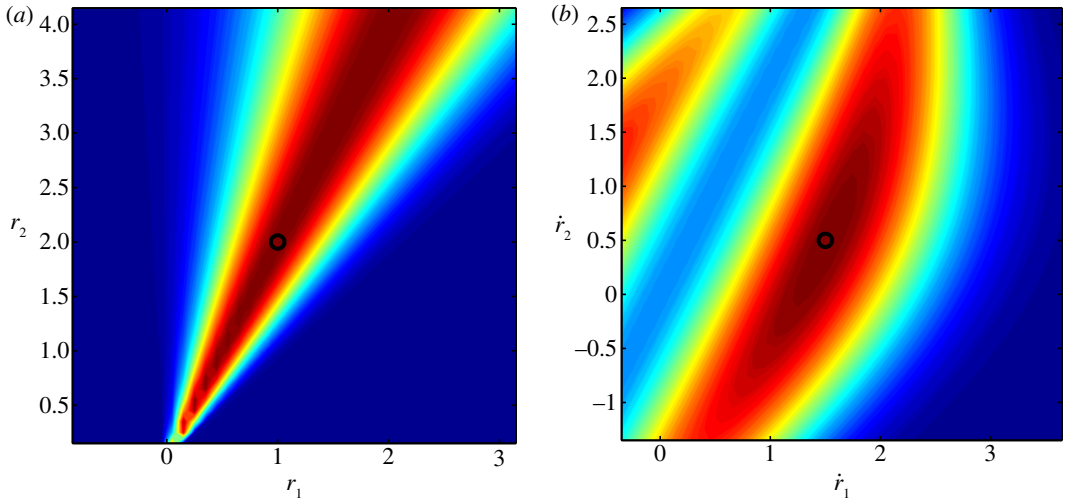

Figure 3. Top-down view of the (*a*) position and (*b*) velocity likelihood functions on a plane orthogonal to the image and parallel to the camera axis. The camera is located at $(r_1, r_2) = (0, 0)$; the black circle is the true value. (Online version in colour.)

probabilities on the order of endpoints. Let Σ_{ep} be the covariance of the endpoint position in pixels (computed empirically). The endpoint likelihood function is

$$P^c_{\mathrm{ep}}(\boldsymbol{e}^c_-, \boldsymbol{e}^c_+ | \boldsymbol{r}, \dot{\boldsymbol{r}}) = 1 - (1 - \mathbb{N}(\boldsymbol{e}^c_-; f^c(\boldsymbol{r}_-), \Sigma_{\mathrm{ep}})\mathbb{N}(\boldsymbol{e}^c_+; f^c(\boldsymbol{r}_+), \Sigma_{\mathrm{ep}})) (1 - \mathbb{N}(\boldsymbol{e}^c_-; f^c(\boldsymbol{r}_+), \Sigma_{\mathrm{ep}})\mathbb{N}(\boldsymbol{e}^c_+; f^c(\boldsymbol{r}_-), \Sigma_{\mathrm{ep}})), \tag{3.5}$$

where $\boldsymbol{r}_\pm = \boldsymbol{r} \pm \dot{\boldsymbol{r}}(\delta t_{\mathrm{e}}/2)$. The combined effect of using a pair of points in the endpoint likelihood function (3.5) is to reduce the set of velocity values along the camera axis, which is otherwise unobservable.

The combined position and velocity likelihood function is

$$P(\boldsymbol{Z}|\boldsymbol{X}) = \prod_{c=1,2} P(\boldsymbol{e}^c_-, \boldsymbol{u}^c, \boldsymbol{e}^c_+ | \boldsymbol{r}, \dot{\boldsymbol{r}}) = \prod_{c=1,2} P^c_{\mathrm{mp}}(\boldsymbol{u}^c|\boldsymbol{r}) P^c_{\mathrm{ep}}(\boldsymbol{e}^c_-, \boldsymbol{e}^c_+ | \boldsymbol{r}, \dot{\boldsymbol{r}}). \tag{3.6}$$

Figure 3 shows the combined position and velocity likelihood function. The likelihood function (3.6) is used to weight the particles in the resample step of the particle filter. We update the position estimates using triangulation [38], thereby effectively marginalizing out the position from the combined position and velocity filtering pdf.

A velocity likelihood function improves the reliability of data association by placing predicted measurements closer to the actual measurements. We compared the absolute velocity estimation error between a stand-alone position likelihood function and the combined position and velocity likelihood function (3.6). To create ground-truth data, we isolated a single mosquito track in both camera frame for 8 s. We then interpolate the position values to every 1/800th of a second. These values were then used to create an artificial mosquito streak during the time of exposure δt_{e} from a 1 cm sphere. We then project the streak on a pair of white synthetic left and right camera images with resolution 1392×1024 pixels. To achieve a faded-streak effect, we reduce the intensity value of a pixel by 30 every time it is visited on the screen during the time of exposure. Mosquito motion normal to the image plane results in darker, shorter streaks, where as motion parallel to the image plane results in lighter, longer streaks (figure 4*a*). We tracked this dataset using multiple Monte-Carlo runs of a particle filter. The combined position and velocity likelihood function performed better than the stand-alone position likelihood function, with an average improvement in mean absolute velocity error of 27 per cent (figure 4*b*).

3.3. Data association and occlusion resolution

Prior to weighting a target distribution with a likelihood function, we must first address the data-association problem. The mosquito data-association problem is challenging owing to the variable number of targets. To mitigate the uncertainty in association (for example, did the paths of two mosquitoes cross each other, or was it a close encounter?), we use a deferred-logic method called the MHT [27]. Each assignment of measurements to targets is set aside as a hypothesis and acted upon in a future time step when we are more certain. The certainty is computed using the probability of a hypothesis that depends on the innovation $\boldsymbol{\nu}^c = \boldsymbol{u}^c - f^c(\boldsymbol{r})$ of each measurement-target assignment in the hypothesis, the probability of detection of actual targets, and the covariance of the predicted measurement S.

We reduce the number of hypotheses by clustering and prune them by selecting a few best hypotheses based on their probability at each step. Clustering is performed by dividing the measurement and hypothesized targets at each step into independent sets. At each time step, measurements are associated with each cluster based on the combined gating volume of all targets within the cluster. Measurements that do not belong to any cluster form their own clusters. Two clusters that consist of the same measurement are combined to form a single cluster. Similarly, we split clusters that consist of targets only assigned to a single measurement. Hypotheses are computed for each cluster independently. Hypotheses within a large cluster (more than 10 measurements) are limited to a

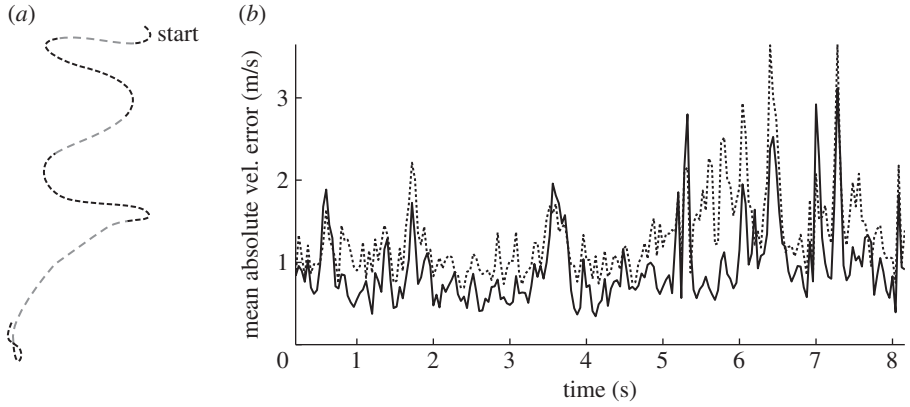

Figure 4. (*a*) Two-dimensional projection of a 1 cm sphere tracing the interpolated path of a mosquito. (*b*) Mean of absolute velocity estimate error over multiple Monte-Carlo runs using a stand-alone position likelihood function (dashed-dotted) compared with a combined position and velocity likelihood function (solid).

single localized GNN assignment [28]. Using a single scanback [27] at each step, we choose the hypothesis with the highest probability to reduce to one the number of hypotheses at the previous step. Child hypotheses resulting from a pruned parent hypotheses are also removed.

New targets are automatically initialized from unassigned measurements and confirmed if they are tracked for more than three frames. New target distributions are sampled from a normal distribution with a low standard deviation in position (5 mm) about the triangulated point, and a large standard deviation in velocity ($500\ \mathrm{mm\,s^{-1}}$) about zero. The combined likelihood function resamples the distribution to equally favour particles getting projected on either side of the streak in the next time step.

Occlusions are not directly addressed as part of any data-association strategy, because existing strategies assume that motion coherence will automatically associate the right tracks in a future time step. In our case, occlusions undermine the velocity estimate, making future associations less reliable. An occlusion is detected if (i) two measurement pairs within a hypothesis consist of the same measurement from a single camera, or (ii) multiple hypotheses assign the same measurement to two or more targets. We interpret an occlusion as a combination of individual streaks, which are then used to extract velocity information as described in §3.1.

In order to cluster the pixels in an occlusion blob, we use the information about the number of mosquitoes hypothesized in the occlusion as well as their position and velocity estimates to model the blob as a mixture of Gaussians. An expectation-maximization algorithm [39] uses position estimates for initial means and velocity estimates for covariance in each dimension to hard-cluster the pixels into individual streaks. This set of individual streaks is used as an initial guess to soft-cluster[3] the pixels into more accurate overlapping streaks. Using the shortest distance of a pixel from the line that passes through the split streak, we allow multiple assignments of each pixel to individual streaks. Figure 5 shows four instances of splitting an occluded blob into two individual mosquito streaks.

[3]Soft clustering allows a single pixel to be assigned to more than one cluster, whereas hard-clustering assigns each pixel to exactly one cluster.

4. DATA COLLECTION AND TRACKING RESULTS

To validate the mosquito-tracking system and film mosquito swarms in the field, we used a pair of phase-locked Hitachi KP-F120CL cameras in a stereo configuration. Each camera captured 10-bit images at 25 frames per second and 1392×1040 pixel resolution. Figure 6 shows a schematic of the data collection system. The video streams were recorded using a 2.8 GHz quad core laptop, an Imperx FrameLink Express frame grabber (Imperx Inc., Boca Raton, FL USA), and Streampix v. 5 software (Norpix Inc, Quebec, Canada). Each camera was calibrated onsite using a checkerboard and the Matlab Calibration Toolbox [40]. Reprojection error, which is a measure of calibration accuracy, was in sub-pixels for each camera. Relative camera orientation and position were determined by extrinsic calibration by taking multiple pictures of a stationary checkerboard with both cameras. During filming, the camera height, azimuth and elevation were recorded to create a ground-fixed reference frame. We used a Kestrel 4500 portable weather station (Nielsen-Kellerman, Boothwyn, PA, USA) to sample other environmental factors such as wind velocity and humidity at 0.1 Hz.

Filming was done in the village of Donéguébogou, Mali in Western Africa. Donéguébogou is 29 km north of Bamako and has been the site of previous research on *A. gambiae* mosquitoes [5,10]. Swarms formed approximately 20 min after sunset, initially with only one or two males then increasing in numbers, and lasted for 20 min. Most couples were seen 5–10 min after the swarm was first observed. Couples formed only for a few minutes during this period, then were no longer observed, though the males continued to swarm for many minutes after the last couple had formed. We filmed swarms of *A. gambiae* that formed over bare ground or markers.

Female mosquitoes are difficult to detect and track because they fly faster than the average male (see §4.3) and appear as a faint streak much of the time. However,

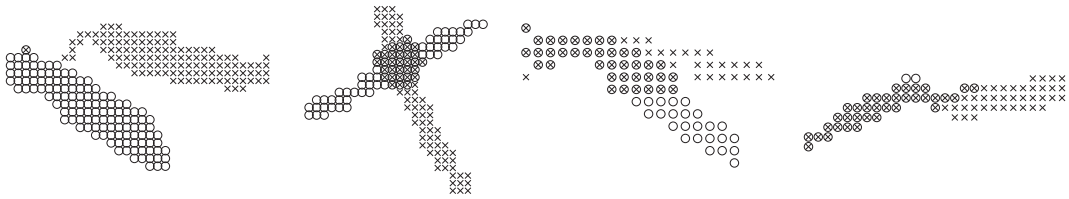

Figure 5. Four examples of occlusion resolution by soft-clustering the occluded blobs into two individual mosquito streaks. Each streak is denoted by a marker type.

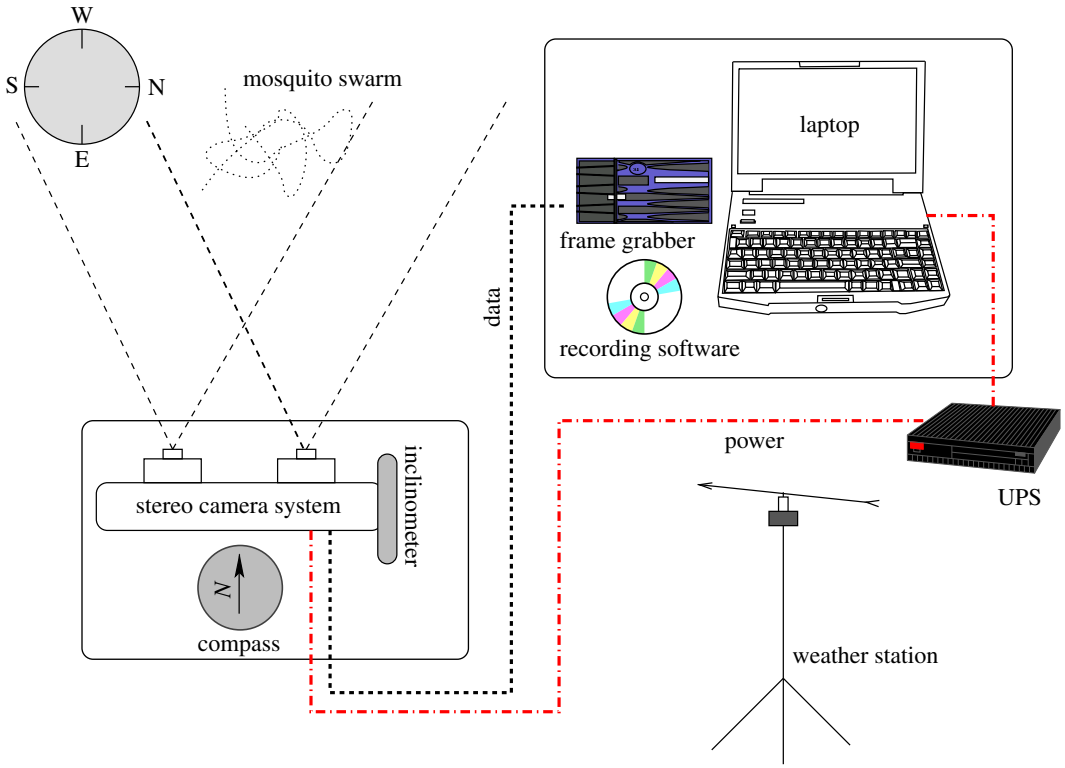

Figure 6. The data collection system consists of the stereo camera rig and a laptop powered by an uninterrupted power supply (UPS). A capture signal is sent from the frame grabber to record frames in sync. A Kestrel weather station records environmental factors such as temperature and wind velocity at 0.1 Hz. (Online version in colour.)

a mosquito couple is distinguishable to the human eye owing to its distinct flying pattern and darker appearance against the sky. Upon spotting a couple, we noted the frame number displayed on the laptop screen. The couples were located after filming by manually reviewing the video footage at the designated frame. Out of the two mating mosquitoes, the female mosquito was identified as the mosquito that entered the swarm last. We tracked the pair, first as a couple and then individually, by playing the sequence backwards. Parameter values used for data collection and tracking are described below and in table 2. The validation and evaluation of tracking performance follow.

4.1. Parameters used for data collection and tracking

The camera baseline b, i.e. the distance between cameras, affects the disparity Δu in pixel positions of an object in a stereo camera set-up [34]. A large disparity reduces uncertainty along the camera axis, which in turn improves accuracy as well as the ability to resolve occlusions. For a stereo-camera configuration with focal length f and no vertical offset between centres, the baseline and disparity are related according to $\Delta u = (bf)/z$ [34], where z is the distance along the camera axis of the target from the stereo set-up. The overlap between camera views is $(I_w - \Delta u)/I_w$, where I_w is the image width resolution in pixels. A large overlap is desirable for maximum coverage. Since the majority of swarms were filmed with $1.5 \leq z \leq 2.5$ m, we selected a baseline of 20 cm to achieve 80–90% overlap and three to five pixel separation between two mosquitoes that are 3 cm apart (approx. two body lengths) along the camera axis.

In addition to the intensity threshold t_{int} described in §3.1, foreground segmentation requires setting the sliding window t_{win} and a threshold on area of the blobs t_{area}. We selected $t_{win} = 7$ frames centred on the current frame, although swarms filmed at short ranges required a sliding window in the range of 3–5 frames. We computed the area-threshold limits $20 \leq t_{area} \leq 150$ from several different swarms to achieve the best rejection of noise as well as large insects.

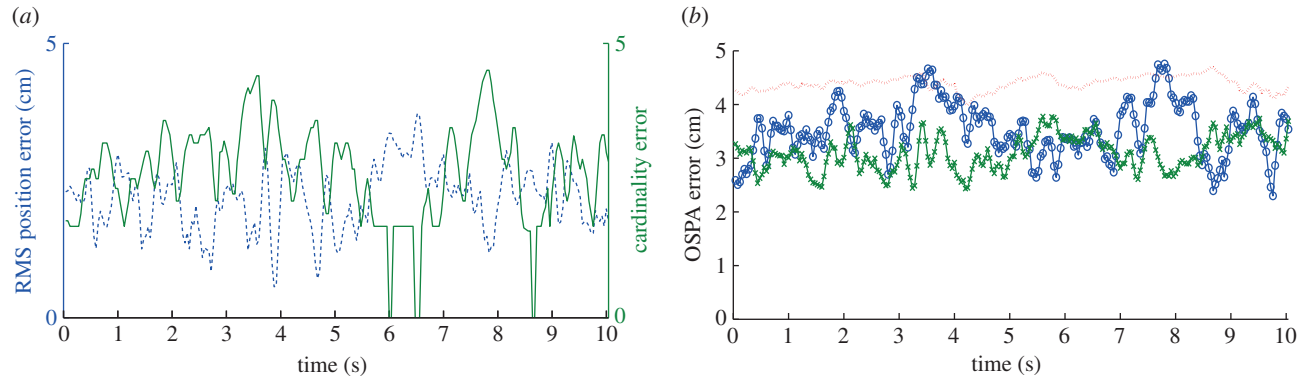

Figure 7. (*a*) Position (dashed blue) and cardinality error (solid green) for a swarm of 10 mosquitoes. (*b*) OSPA error for different methods and swarm sizes: nearest neighbour [41] (dotted) for a swarm of 20 mosquitoes and single-scan MHT for two swarms of sizes 10 (crosses) and 20 (circles), respectively. (Online version in colour.)

The covariance $\Sigma_{\text{ep}} = \text{diag}\{4, 4\}$ pixels^2 for endpoints was computed by manually selecting the endpoints of streaks in a random sampling of frames and comparing with the calculated value. The disturbance $\boldsymbol{w}$ for the constant-velocity model was sampled from $\mathbb{N}(0, 100\,\text{m}^2\,\text{s}^{-4})$, whose covariance was found by fitting a normal distribution to the acceleration values of manually generated tracks.

4.2. Validation and performance of tracking system

We tested the position accuracy of our tracking system using a calibration checkerboard with squares of known dimensions by manually clicking pairs of points whose separation distance was in the range 3–40 cm. This method yielded an error of 5 $\pm$ 5 mm for 50 pairs. Average position error by tracking an artificial mosquito projected on the stereo images as in §3.2 over multiple runs was 5 $\pm$ 4 mm. We also reconstructed tracks from a single swarming event on Aug. 29 using two independent stereo camera rigs. We created a common reference frame by measuring the height, azimuth and elevation of the cameras. The videos were time-synced using a laser pointer flashed at the end of the sequence. The mean distance between independent tracks of the same mosquito (200 data points) was 4.4 $\pm$ 1.3 cm, although up to 3 cm error can be attributed to the inter-frame time difference between the camera systems (caused by a possible mismatch in the pair of frames that contained the laser flash). A mosquito flying at an average speed of 1.5 m s^{-1} will cover 3 cm in 1/50th of a second.

Figure 7 shows the results of using the OSPA metric (see the electronic supplementary material, equation (3.3)) to compare tracks from the multi-target tracking system to the manually generated ground-truth. We tested two swarms with 10 and 20 mosquitoes, respectively. The order parameter and the cut-off parameter for computing OSPA values were set 2 and 50 mm, respectively. Decomposing OSPA into position and cardinality errors shows that the average root mean square (RMS) position errors in the 10- and 20-mosquito swarms were 2.17 $\pm$ 0.58 and 2.3 $\pm$ 0.46 cm, respectively. Correspondingly, average absolute position errors for the 10- and 20-mosquito swarms were

Figure 8. GPS measurements of filming locations in Mali, Africa ©2012 Google, ©2012 DigitalGlobe. (Online version in colour.)

1.74 $\pm$ 0.56 and 2.03 $\pm$ 0.47 cm, respectively. A low cardinality error was often accompanied by relatively high position error during periods when the swarm was dense, because of occlusions and false tracks. As would be expected for a stereo set-up, position error was highest (44%) in the range measurement (along the camera optical axis) when compared with either of the other two dimensions. OSPA was larger for the 20-mosquito swarm, mainly owing to cardinality errors. The position error is likely a consequence of image noise, which resulted in partially segmented streaks. (We mitigate this problem by filtering trajectory data using a Kalman smoother.) Average reprojection error on the images was less than two pixels.

The labelling error, which captures track continuity and identity swaps, was computed separately. An identity swap results in a labelling error of 2 before or after the swap in the sequence. Track fragmentation results in a labelling error of 1 after the disconnect occurs. We randomly selected 100 instances of 25 continuous frames in a swarm of 10 mosquitoes. The average labelling error (most of which was due to track fragmentation) was 2.1 $\pm$ 1.4 tracks. A simple average of track lengths across six swarms ranged between 15 and 25 frames corresponding to 0.6–1 s. Track fragmentation occurs owing to early terminations, which can be caused by the following:

— partially segmented streaks owing to noise, cloudy background and clutter. Partially segmented streaks

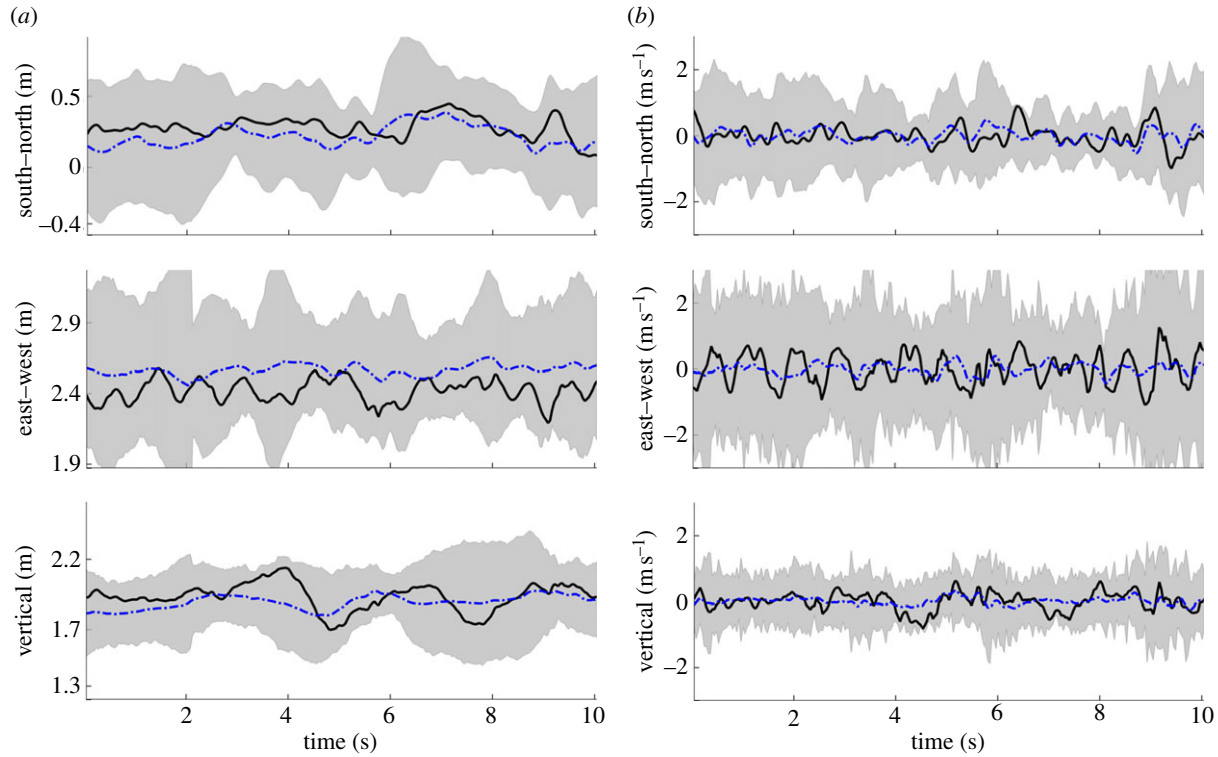

Figure 9. Instantaneous three-dimensional (*a*) position and (*b*) velocity tracks of a male mosquito (black solid) in the Aug. 29 sequence (S molecular form). Mean (dotted blue) and 3σ bounds (grey) for all mosquitoes also shown. Also see the electronic supplementary material, video S1. (Online version in colour.)

in one frame often violate the epipolar constraint. Decreasing the intensity threshold to get full streaks adds noise to the measurements. (A possible solution that we are exploring in ongoing work is to reconstruct the streak using velocity estimates.); and

— occlusion between a tracked and untracked target. Occlusions between a tracked (known) target and an as yet uninitialized target are not detected. The success rate of surviving such an occlusion depends on the motion of the tracked target after the occlusion. A manoeuvre or successive occlusion may terminate the track.

4.3. Tracking results

This section presents a subset of the three-dimensional trajectory data generated using the mosquito tracking system. We filmed 21 swarms and 13 mating events between 17 August 2010 and 3 September 2010. Out of the 21 swarms, 18 formed over bare ground and three formed over natural markers. (A natural marker is an area of high contrast with the rest of the ground such as a patch of grass.) *Anopheles gambiae* can be divided into two incipient species namely the M and S molecular forms [42]. In Diabaté *et al.* [10], a strong association between the swarming marker type and molecular form has been found. The M form was found to swarm over natural markers, whereas the S form swarms over bare ground [10]. We collected a few mosquitoes from each swarm and performed a polymerase chain reaction (PCR) test to determine the molecular form. All sequences presented in this paper were of type S. Each day two teams of three to five people with identical camera rigs selected separate swarming sites for filming. The swarming sites were usually within a few hundred metres of each other. Swarming sites were surveyed the day before to record average swarm size and location. Filming locations spread throughout the village (figure 8) were chosen based on swarm size (less than about 100 mosquitoes for tractability in tracking) and the presence of few trees or houses in the background (i.e. in the direction of the setting sun). Once filming began, 60–90 s stereo video sequences were recorded as 10-bit synchronized tiff images on separate solid-state drives. (The drives were backed up daily on to two separate discs.) A filming session typically produced five to eight video sequences before it became too dark to film.

To create representative trajectory dataset, we selected six video sequences that contain a mating event. We call these the mating sequences. We refer to the mating mosquitoes as the female and the focal male. We also selected six other video sequences with no female present, called the male-only sequences, to produce full-length trajectories of swarming behaviour. Trajectory data presented here are from swarms filmed on 20, 21, 25, 26, 28 and 29 August and 1 September. Male-only sequences last between 20 and 35 s, whereas mating sequences start a few seconds prior to the detection of female within the field of view and end when the couple flies out of field of view (0–5 s).

Figure 9 shows the position and velocity of a randomly selected male *A. gambiae* in the Aug. 29 male-only sequence, which was a swarm that formed over bare ground (S molecular form). The swarm consisted of 20 mosquitoes at the beginning of the

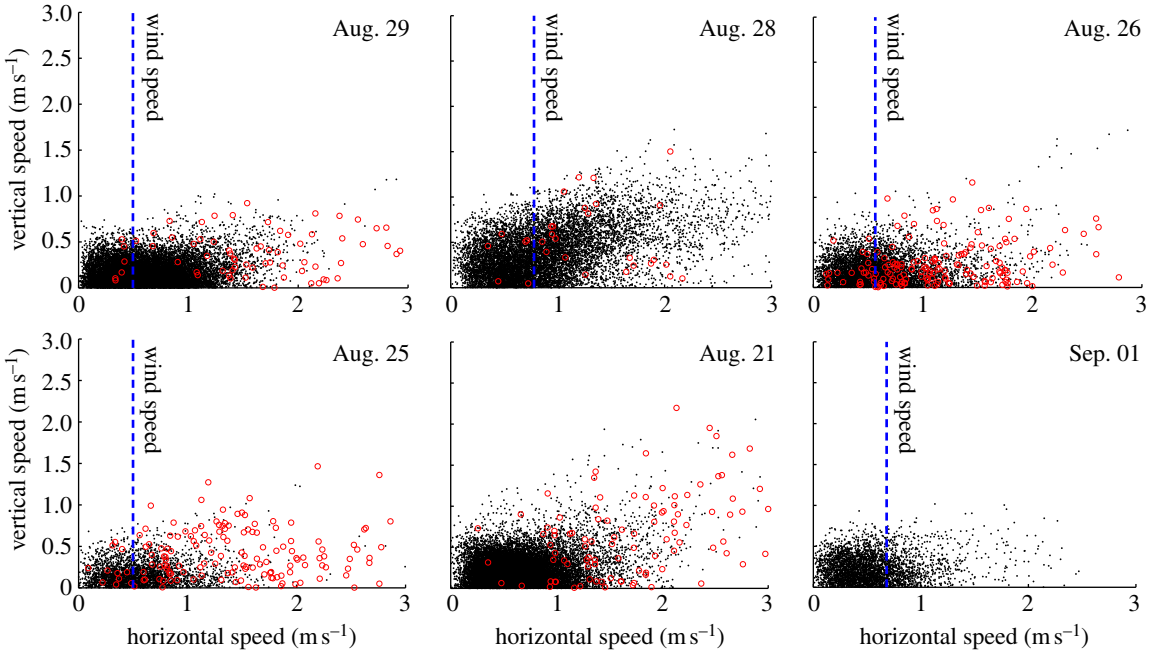

Figure 10. Instantaneous horizontal versus vertical speed of male mosquitoes in six sequences (black dots). Female speeds (red circles) and average wind speed (blue dashed lines) are shown, if available.

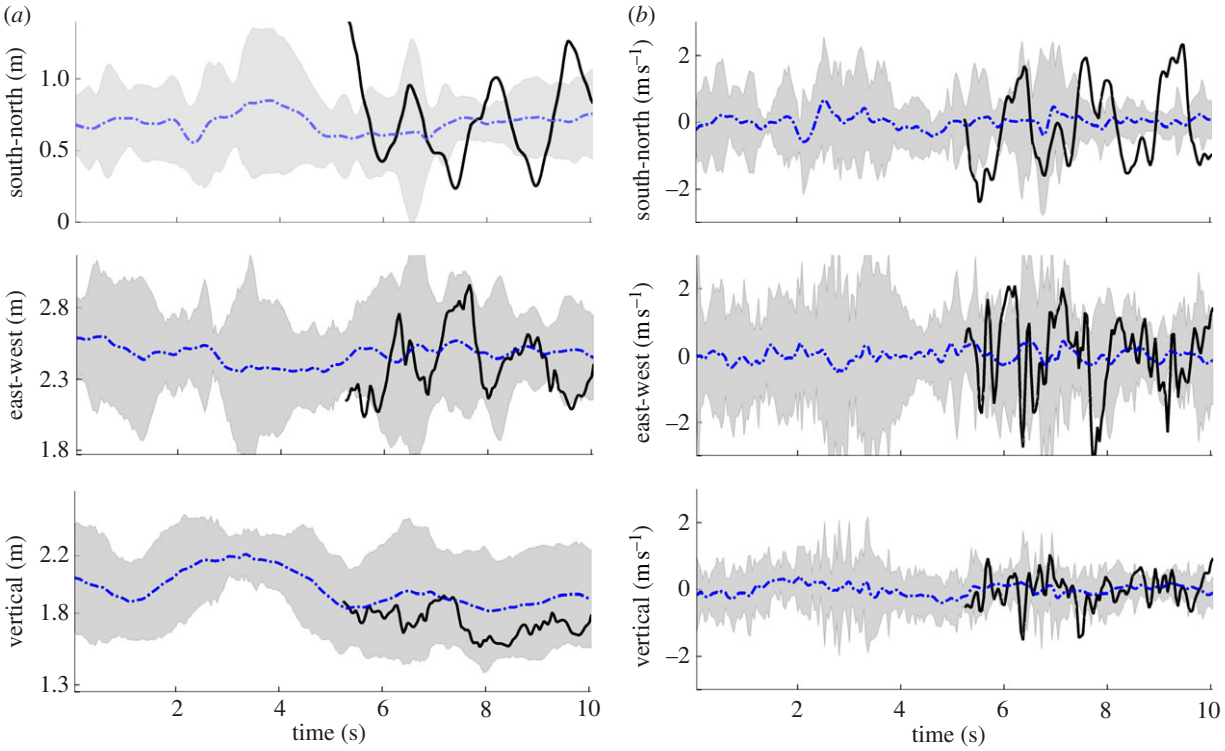

Figure 11. Instantaneous three-dimensional (*a*) position and (*b*) velocity of a female mosquito that coupled in the Aug. 29 sequence, filmed 70 s after the male position and velocity shown in figure 9. Mean (dotted blue) and 3σ bounds (grey) of all males are also shown. (Online version in colour.)

sequence and dropped to 19 after 10 s. The mosquito movement is characterized by quasi-periodic motion in each of the three spatial dimensions. The instantaneous mean position of the mosquitoes in the swarm, i.e. the swarm centroid, is also shown. The origin of the inertial frame is located at ground level under the camera rig; the inertial frame is oriented along east–west, north–south and vertical directions. The 3σ bounds for position and velocity of all of mosquitoes in this swarm are shown in grey. Swarm size (twice the 3σ bounds) averaged 1.17 m in the horizontal plane and 0.56 m in the vertical. The average swarm size across all planes ranged between 0.52 and 1.86 m. The average height of the swarm was 1.89 m. The average velocity along each dimension is close to zero with highest standard deviation in the east–west direction (0.514

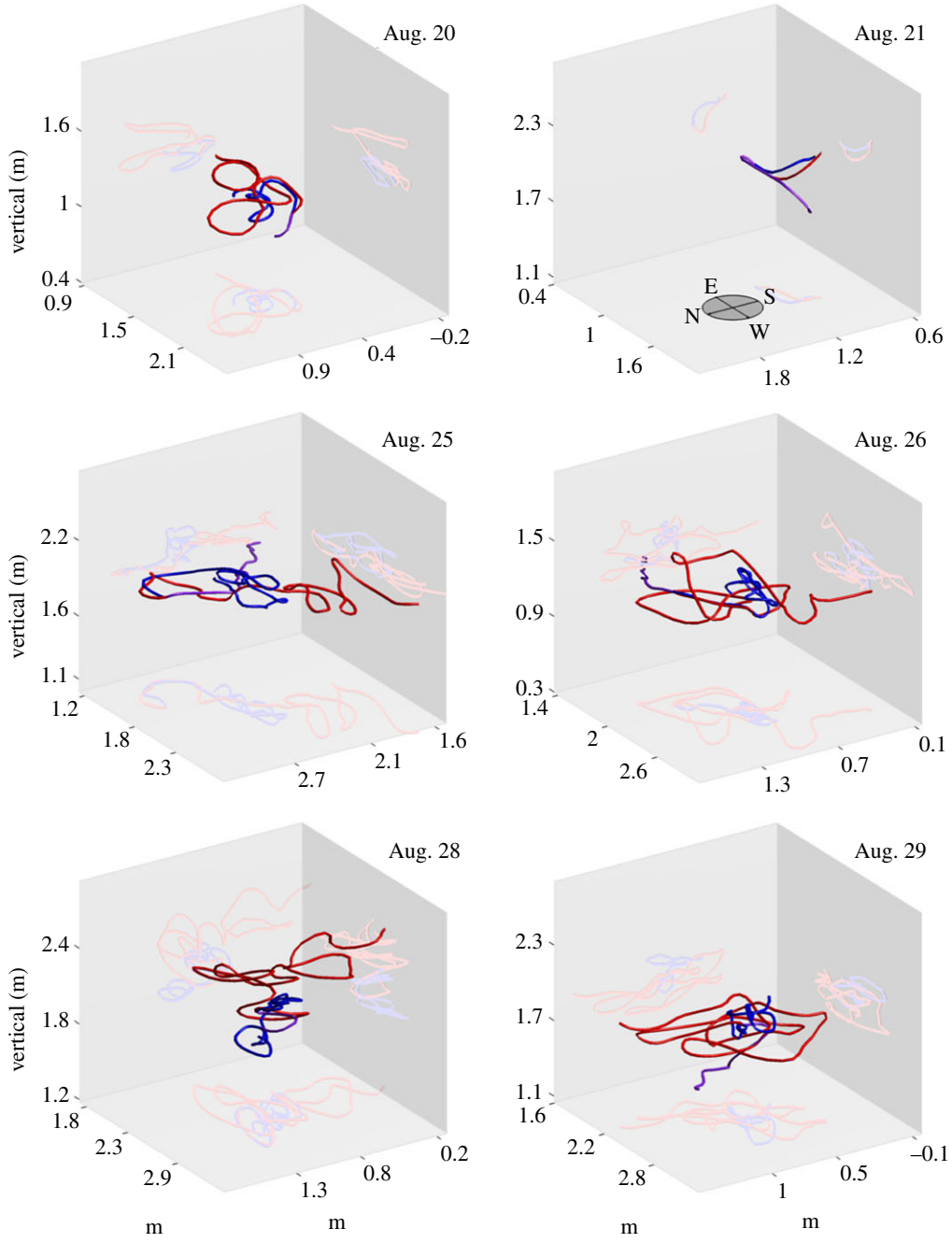

Figure 12. Three-dimensional reconstruction of *A. gambiae* mating events in the wild. The female mosquito track (red) and male mosquito track (blue) are shown. The couple is shown in purple. Pre-coupling tracks are projected onto two-dimensional planes on each side. (The Aug. 29 mating event is shown magnified in the electronic supplementary material. Also see electronic supplementary material, video S2.)

$m\,s^{-1}$) followed by the north–south ($0.332\,m\,s^{-1}$) and the vertical ($0.281\,m\,s^{-1}$).

Figure 10 shows the ratio between horizontal and vertical speed for each swarm. The Aug. 28 sequence was filmed on a day with relatively high wind (approx. $0.8\,m\,s^{-1}$) when compared with other sequences. The mosquito movement for that swarm was characterized by a rolling motion in the direction of the wind and relatively higher vertical velocities. In five out of the six swarms that we used to generate male-only sequences, we witnessed mating events at a later time. The horizontal and vertical speeds of female mosquito that formed couples are also plotted in figure 10. Non-parameteric Kruskal–Wallis tests on each dataset show that the average male and female speeds in the same sequence are significantly different for each sequence. The maximum p-value among all mating sequences was 0.0003. (In contrast, the maximum p-value for male speeds during the same mating sequence was 0.051.)

Figure 11 shows the position and velocity of a female mosquito that formed a successful couple in the Aug. 29 sequence. The mating sequence was filmed about a minute after the male-only sequence on the same date. The female appeared in the field of view 5 s prior to coupling. The movement of the female crosses the 3σ boundaries of the swarm in the north–south direction.

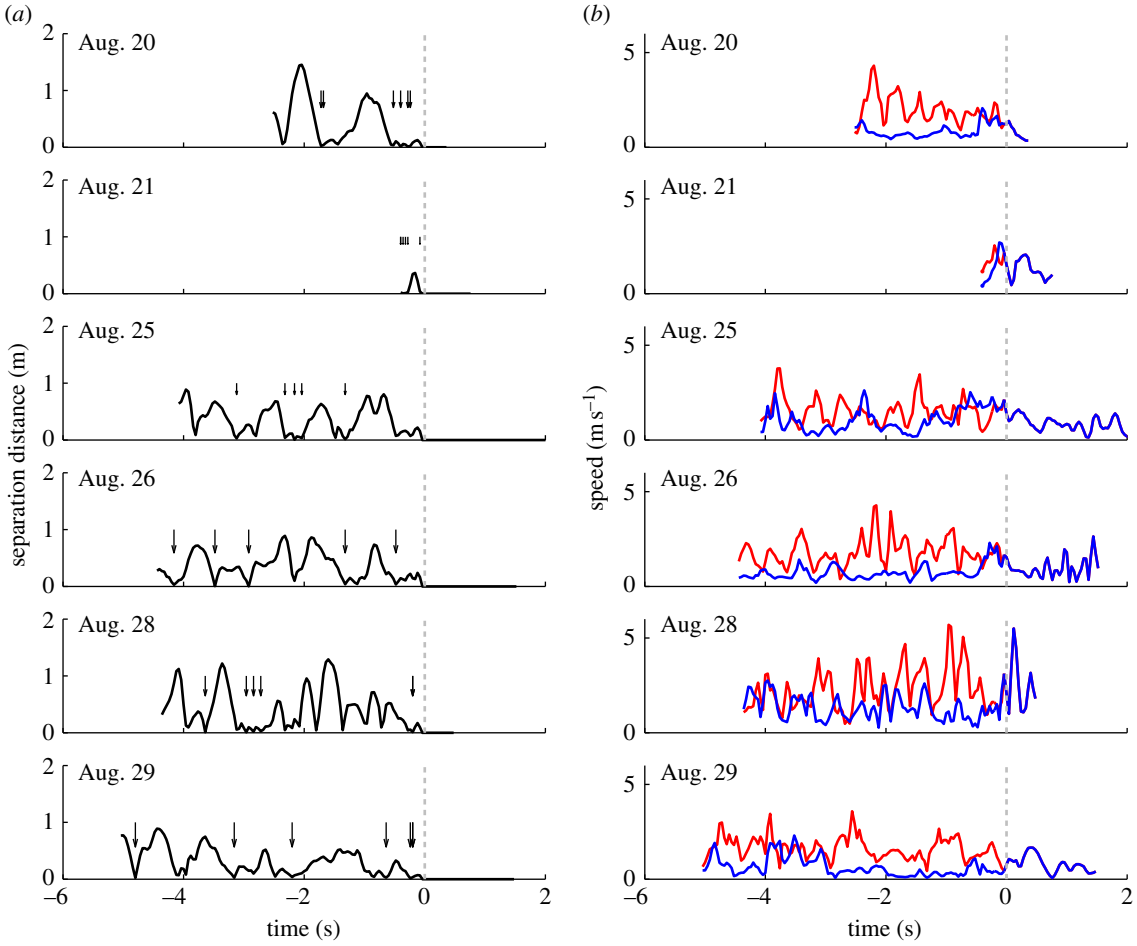

Figure 13. Relative distance (*a*) and speeds (*b*) of mating male and female *A. gambiae* mosquitoes in six mating sequences. Time 0 s occurs when the separation distance first drops below 2 cm. Arrows depict close encounters (separation distance less than 4 cm).

The average speed of the female was higher than the male mosquito until just before the couple forms, when the focal male speeds up. The vertical movement shows that the female stayed predominantly at heights corresponding to the lower half of the swarm. A three-dimensional reconstruction of the mating mosquitoes in six mating sequences is shown in figure 12. Across all mating sequences, the female mosquito covered an average 59 per cent more distance than the focal male during the same time interval.

Figure 13 shows the separation distance and speeds from six mating sequences. The amount of time we observed the females in the swarm before forming a couple was up to 5 s. In each mating sequence that lasted longer than 0.5 s, the number of close encounters (moments when the separation distance between the mating mosquitoes dropped below three body lengths, or 4 cm) with the successful male mosquito was in the range 3–6.

5. CONCLUSIONS AND ONGOING WORK

We describe a tracking system to reconstruct the three-dimensional trajectories of wild mosquito swarms. We address noisy images by adaptively seeking missing measurements and exploit streak orientation and length to extract velocity information. A probabilistic data association method that uses multiple hypotheses is modified to address occlusions. We evaluate the system using an established multi-target tracking metric and validate using independent measurements of the same swarm. Tracking results are presented in the form of three-dimensional trajectories of swarming and mating mosquitoes. To date, the data produced from the tracking system described in this paper are an order of magnitude larger (97 trajectories and 55 000 position points) than the last published result [5] on reconstruction of wild mosquito swarms, and the first to contain three-dimensional trajectories rather than three-dimensional positions. In ongoing work, we are investigating these trajectories to characterize swarming and mating behaviour.

As part of ongoing work on the tracking system, we are working to include the streak intensity in the image as part of the likelihood function. This will help predict the appearance of a mosquito on the image plane as a function of its velocity, thereby allowing the possibility of streak retrieval. Such an approach would, for example, reduce track terminations and create longer tracklets. Another aspect of the tracking system that we are investigating is the automatic detection of mating events. In order to avoid sifting through video streams to locate mating events, the distinct flying pattern and appearance of the mating couple

can be used for automatic detection and backwards tracking of the female. With enough mating events, a higher order motion model (that depends on more than one previous time step) will automatically predict and detect mating events.

This work is built on the prior experience of filming mosquito swarms at the University of Bamako in Mali. Specifically, we thank Alpha Yaro, Adama Dao and Sekou Traoré for their support. We gratefully acknowledge the travel support from Robert Gwadz of the Laboratory of Malaria and Vector Research at the National Institute of Allergy and Infectious Diseases. Thanks to Diana Huestis for performing the PCR tests. Finally, we would like to thank the residents of Donéguébogou for allowing us to film.

REFERENCES

1 Charlwood, J. & Jones, M. 1980 Mating in the mosquito, *Anopheles gambiae* sl. *Physiol. Entomol.* **5**, 315–320. (doi:10.1111/j.1365-3032.1980.tb00241.x)
2 Gibson, G. 1985 Swarming behaviour of the mosquito *Culex pipiens quinquefasciatus*: a quantitative analysis. *Physiol. Entomol.* **10**, 283–296. (doi:10.1111/j.1365-3032.1985.tb00049.x)
3 Ikawa, T., Okabe, H., Mori, T., Urabe, K. & Ikeshoji, T. 1994 A method for reconstructing three-dimensional positions of swarming mosquitoes. *J. Insect Behav.* **7**, 237–248. (doi:10.1007/BF01990084)
4 Charlwood, J., Pinto, J., Sousa, C., Madsen, H., Ferreira, C. & do Rosario, V. 2002 The swarming and mating behaviour of *Anopheles gambiae* ss (Diptera: Culicidae) from Sao Tome Island. *J. Vector Ecol.* **27**, 178–183.
5 Manoukis, N. C., Diabate, A., Abdoulaye, A., Diallo, M., Dao, A., Yaro, A. S., Ribeiro, J. M. C. & Lehmann, T. 2009 Structure and dynamics of male swarms of *Anopheles gambiae*. *J. Med. Entomol.* **46**, 227–235. (doi:10.1603/033.046.0207)
6 Diabate, A., Yaro, A. S., Dao, A., Diallo, M., Huestis, D. L. & Lehmann, T. 2011 Spatial distribution and male mating success of *Anopheles gambiae* swarms. *BMC Evol. Biol.* **11**, 184. (doi:10.1186/1471-2148-11-184)
7 Charlwood, J. & Pinto, J. 2002 Male size does not affect mating success (of *Anopheles gambiae* in Sao Tome). *Med. Vet. Entomol.* **16**, 109–111. (doi:10.1046/j.0269-283x.2002.00342.x)
8 Thailayil, J., Magnusson, K., Godfray, H., Crisanti, A. & Catteruccia, F. 2011 Spermless males elicit large-scale female responses to mating in the malaria mosquito *Anopheles gambiae*. *Proc. Natl Acad. Sci. USA* **108**, 13 677–13 681.
9 Bryan, J. 1968 Results of consecutive matings of female *Anopheles gambiae* species B with fertile and sterile males. *Nature* **218**, 489. (doi:10.1038/218489a0)
10 Diabaté, A., Dao, A., Yaro, A. S., Adamou, A., Gonzalez, R., Manoukis, N. C., Traore, S. F., Gwadz, R. W. & Lehmann, T. 2009 Spatial swarm segregation and reproductive isolation between the molecular forms of *Anopheles gambiae*. *Proc. R. Soc. B* **276**, 4215–4222. (doi:10.1098/rspb.2009.1167)
11 Pennetier, C., Warren, B., Dabiré, K. R., Russell, I. J. & Gibson, G. 2010 'Singing on the wing' as a mechanism for species recognition in the malarial mosquito *Anopheles gambiae*. *Curr. Biol.* **20**, 131–136. (doi:10.1016/j.cub.2009.11.040)
12 Okubo, A. 1986 Dynamical aspects of animal grouping: swarms, schools, flocks, and herds. *Adv. Biophys.* **22**, 1–94. (doi:10.1016/0065-227X(86)90003-1)
13 Shimoyama, N., Sugawara, K., Mizuguchi, T., Hayakawa, Y. & Sano, M. 1996 Collective motion in a system of motile elements. *Phys. Rev. Lett.* **76**, 3870–3873. (doi:10.1103/PhysRevLett.76.3870)
14 Parrish, J. K. & Hammer, W. M. 1997 *Animal groups in three dimensions*. Cambridge, UK: Cambridge University Press.
15 Okubo, A. & Chiang, H. C. 1974 An analysis of the kinematics of swarming of *Anarete pritchardi* Kim (Diptera: Cecidomyiidae). *Res. Popul. Ecol.* **16**, 1–42. (doi:10.1007/BF02514077)
16 Ballerini, M. *et al.* 2008 Interaction ruling animal collective behavior depends on topological rather than metric distance: evidence from a field study. *Proc. Natl Acad. Sci. USA* **105**, 1232–1237. (doi:10.1073/pnas.0711437105)
17 Branson, K., Robie, A. A., Bender, J., Perona, P. & Dickinson, M. H. 2009 High-throughput ethomics in large groups of *Drosophila*. *Nat. Methods* **6**, 451–457. (doi:10.1038/nmeth.1328)
18 Khan, Z., Balch, T. & Dellaert, F. 2005 MCMC-based particle filtering for tracking a variable number of interacting targets. *IEEE Trans. Pattern Anal. Mach. Intell.* **27**, 1805–1918. (doi:10.1109/TPAMI.2005.223)
19 Wu, Z., Hristov, N. I., Hedrick, T. L., Kunz, T. H. & Betke, M. 2009 Tracking a large number of objects from multiple views. In *Int. Conf. on Computer Vision, September 2009, Kyoto, Japan, IEEE*. pp. 1546–1553. Citeseer.
20 Wu, Z., Hristov, N., Kunz, T. & Betke, M. 2010 Tracking-reconstruction or reconstruction-tracking? Comparison of two multiple hypothesis tracking approaches to interpret 3D object motion from several camera views. In *Workshop on Motion and Video Computing*, pp. 1–8. IEEE.
21 Wu, H. S., Zhao, Q., Zou, D. & Chen, Y. Q. 2011 Automated 3D trajectory measuring of large numbers of moving particles. *Opt. Exp.* **19**, 7646–7663. (doi:10.1364/OE.19.007646)
22 Straw, A. D., Branson, K., Neumann, T. R. & Dickinson, M. H. 2010 Multi-camera real-time three-dimensional tracking of multiple flying animals. *J. R. Soc. Interface* **8**, 395–409. (doi:10.1098/rsif.2010.0230)
23 Grover, D., Tower, J. & Tavaré, S. 2008 O fly, where art thou? *J. R. Soc. Interface* **5**, 1181.
24 Bar-Shalom, Y. 1987 *Tracking and data association*. San Diego, CA: Academic Press Professional, Inc.
25 Gordon, N. J., Salmond, D. J. & Smith, A. F. M. 1993 Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEEE Proc. Radar Signal Process.* **140**, 107–113. (doi:10.1049/ip-f-2.1993.0015)
26 Freitas, N. D. & Gordon, N. 2001 *Sequential monte carlo methods in practice*. Basel, Switzerland: Birkhäuser.
27 Reid, D. 1979 An algorithm for tracking multiple targets. *IEEE Trans. Autom. Control* **24**, 843–854. (doi:10.1109/TAC.1979.1102177)
28 Kuhn, H. W. 1955 The Hungarian method for the assignment problem. *Naval Res. Logist. Q.* **2**, 83–97. (doi:10.1002/nav.3800020109)
29 Swain, D., Couzin, I. & Leonard, N. E. 2011 Real-time feedback-controlled robotic rish for behavioral experiments with fish schools. *Proc. IEEE* **100**, 150–163. (doi:10.1109/JPROC.2011.2165449)
30 Betke, M., Hirsh, D. E., Bagchi, A., Hristov, N. I., Makris, N. C. & Kunz, T. H. 2007 Tracking large variable numbers of objects in clutter. In *IEEE Conf. on Computer Vision and Pattern Recognition, June 2007, Minneapolis, MN, USA*, pp. 1–8. IEEE.
31 Cox, I. J. & Hingorani, S. L. 1996 An efficient implementation of Reid's multiple hypothesis tracking algorithm

and its evaluation for the purpose of visual tracking. *IEEE Trans. Pattern Anal. Mach. Intell.* **18**, 138–150. (doi:10.1109/34.481539)

32 Ristic, B., Vo, B., Clark, D. & Vo, B. 2011 A metric for performance evaluation of multi-target tracking algorithms. *IEEE Trans. Signal Process.* **59**, 3452–3457. (doi:10.1109/TSP.2011.2140111)

33 Schuhmacher, D., Vo, B. & Vo, B. 2008 A consistent metric for performance evaluation of multi-object filters. *Signal Process. IEEE Trans.* **56**, 3447–3457. (doi:10.1109/TSP.2008.920469)

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

35 Cavagna, A., Giardina, I., Orlandi, A., Parisi, G., Procaccini, A., Viale, M. & Zdravkovic, V. 2008 The STARFLAG handbook on collective animal behaviour: 1. Empirical methods. *Anim. Behav.* **76**, 217–236. (doi:10.1016/j.anbehav.2008.02.002)

36 Salmond, D. J. & Birch, H. 2001 A particle filter for track-before-detect. In *Proc. of American Control Conf.* vol. 5, *June 2001*, pp. 3755–3760. Arlington, VA, USA, IEEE.

37 Blackman, S. 2004 Multiple hypothesis tracking for multiple target tracking. *IEEE Aerospace Electron. Syst. Mag.* **19**, 5–18. (doi:10.1109/MAES.2004.1263228)

38 Hartley, R. & Sturm, P. 1997 Triangulation. *Comput. Vis. Image Underst.* **68**, 146–157. (doi:10.1006/cviu.1997.0547)

39 Dempster, A., Laird, N. & Rubin, D. 1977 Maximum likelihood from incomplete data via the EM algorithm. *J. R. Stat. Soc. Ser. B* (*Methodol.*) **39**, 1–38.

40 Bouguet, J.-Y. Camera calibration toolbox for Matlab. [Online]. See http://www.vision.caltech.edu/bouguetj/calib_doc/index.html

41 Butail, S., Manoukis, N., Diallo, M., Yaro, A. S., Dao, A., Traore, S. F., Ribeiro, J. M. C., Lehmann, T. & Paley, D. A. 2011 3D tracking of mating events in wild swarms of the malaria mosquito *Anopheles gambiae*. In *IEEE Conf. Engineering in Medicine and Biology Society*, pp. 720–723. Boston, MA, USA, IEEE.

42 Favia, G., della Torre, A., Bagayoko, M., Lanfrancotti, A., Sagnon, N., Touré, Y. T. & Coluzzi, M. 1997 Molecular identification of sympatric chromosomal forms of *Anopheles gambiae* and further evidence of their reproductive isolation. *Insect Mol. Biol.* **6**, 377–383. (doi:10.1046/j.1365-2583.1997.00189.x)