

## Wing and body motion during flight initiation in *Drosophila* revealed by automated visual tracking

Ebraheem I. Fontaine<sup>1,\*</sup>, Francisco Zabala<sup>2</sup>, Michael H. Dickinson<sup>2</sup> and Joel W. Burdick<sup>1</sup>

<sup>1</sup>Mechanical Engineering, California Institute of Technology, Pasadena, CA 91125, USA and <sup>2</sup>Bioengineering, California Institute of Technology, Pasadena, CA 91125, USA

\*Author for correspondence (e-mail: ebraheem@robotics.caltech.edu)

Accepted 17 January 2009

### SUMMARY

The fruit fly *Drosophila melanogaster* is a widely used model organism in studies of genetics, developmental biology and biomechanics. One limitation for exploiting *Drosophila* as a model system for behavioral neurobiology is that measuring body kinematics during behavior is labor intensive and subjective. In order to quantify flight kinematics during different types of maneuvers, we have developed a visual tracking system that estimates the posture of the fly from multiple calibrated cameras. An accurate geometric fly model is designed using unit quaternions to capture complex body and wing rotations, which are automatically fitted to the images in each time frame. Our approach works across a range of flight behaviors, while also being robust to common environmental clutter. The tracking system is used in this paper to compare wing and body motion during both voluntary and escape take-offs. Using our automated algorithms, we are able to measure stroke amplitude, geometric angle of attack and other parameters important to a mechanistic understanding of flapping flight. When compared with manual tracking methods, the algorithm estimates body position within  $4.4 \pm 1.3\%$  of the body length, while body orientation is measured within  $6.5 \pm 1.9$  deg. (roll),  $3.2 \pm 1.3$  deg. (pitch) and  $3.4 \pm 1.6$  deg. (yaw) on average across six videos. Similarly, stroke amplitude and deviation are estimated within 3.3 deg. and 2.1 deg., while angle of attack is typically measured within 8.8 deg. comparing against a human digitizer. Using our automated tracker, we analyzed a total of eight voluntary and two escape take-offs. These sequences show that *Drosophila melanogaster* do not utilize clap and fling during take-off and are able to modify their wing kinematics from one wingstroke to the next. Our approach should enable biomechanists and ethologists to process much larger datasets than possible at present and, therefore, accelerate insight into the mechanisms of free-flight maneuvers of flying insects.

Supplementary material available online at <http://jeb.biologists.org/cgi/content/full/212/9/1307/DC1>

Key words: *Drosophila*, escape response, estimation, flight, geometric modeling, tracking.

### INTRODUCTION

The fruit fly, *Drosophila melanogaster* has emerged as one of the most important model genetic organisms used in modern biology. Although their nervous system contains only 300,000 neurons, *Drosophila* exhibit an array of complex behaviors and offer an ever-increasing and widely accessible set of methods for altering genes and controlling their temporal and spatial expression. One limitation in the use of *Drosophila* as a model system linking genes to behavior is the difficulty in rapidly quantifying body kinematics. Flight behaviors, in particular, pose many problems for any attempts at automated digitization of movement and behavior. First, wing motion during flight is very rapid, which necessitates the use of high-speed imaging and thus large data streams. Second, flight is intrinsically three-dimensional, which necessitates the coordination of data from two or more cameras. Previous studies of *Drosophila* flight maneuvers have required laborious manual methods to capture the body and wing kinematics (Altshuler et al., 2005; Card and Dickinson, 2008; Fry et al., 2003; Fry et al., 2005). The time-consuming nature of this approach limits a researcher's ability to thoroughly characterize individual components of flight behavior such as take-off and landing. Thus, development of an automated tracking technique that estimates the complete wing and body posture during flight would greatly facilitate future research on flight control in flies and other organisms.

To address the concerns above, we developed an automated model-based tracking technique that can capture the 3D body and wing motion of *Drosophila* from a high-speed multi-camera video sequence. Previously, many studies in *Drosophila* flight control measured the relative wing motion during tethered flight by shining an infrared light upon the fly and measuring the resulting shadow with a photodiode receptor (Dickinson et al., 1993; Gotz, 1987). In that approach, 3D wing motion is reduced to a 1D photodiode voltage signal. Recently, Graetzel and colleagues developed a real-time computer vision system to measure the wing motion of a tethered fly (Graetzel et al., 2006). A single camera view was used to track the angular position of the wing's leading and trailing edge in the projected camera view. Zanker and colleagues measured the full 3D motion of flies during tethered flight using stroboscopic video and mirrors to capture multiple views (Zanker, 1990). However, the 3D reconstruction of the fly's geometry relied on manual digitization of six key points on the wing in each camera view. More recently, Fry developed customized software to simplify the manual fit of 3D wing models to free-flight *Drosophila* in multiple camera views (Fry et al., 2003). This technique was expanded to analyze hovering and take-off behaviors in fruit flies and honey bees (Altshuler et al., 2005; Card and Dickinson, 2008; Fry et al., 2005). The algorithm proposed in this paper extends the work of Fry and colleagues

(Fry et al., 2003) by developing visual tracking techniques to automatically fit a 3D fly model to images captured from multiple calibrated camera views.

Our approach is motivated by and builds upon computer vision techniques that estimate the 3D rigid motion of a human from multiple calibrated camera views (Moeslund et al., 2006). In model-based approaches, a 3D human model containing kinematic chains is given, and the goal is to estimate the body posture and joint angles using image measurements (e.g. silhouettes, appearance textures, optical flow). Another complementary approach involves the direct reconstruction of the model shape and motion without use of a prior model (Ristroph et al., 2009). The best choice of approach will be dictated by the governing experimental conditions and aims. Model-based approaches facilitate accurate tracking in videos containing occlusions and environmental clutter. Model-based techniques are also better suited to image sequences with poor contrast and low lighting. Conversely, reconstruction-based approaches may not accurately estimate an organism's shape and motion unless the voxels (i.e. 3D pixels) are labeled correctly, which can prove difficult when images have low contrast or contain occlusions and clutter. Model-based approaches also allow the experimentalist to include those degrees of freedom in the model that are most relevant to experimental goals so that the state estimation process performs inference on relevant physical quantities. However, model-based approaches require an extra step (either manual or automated) to initialize the model on the first frame of the image sequence.

While we adopt the model-based idea from the human motion-tracking literature, there are several challenges peculiar to the problem of *Drosophila* tracking that require special attention. The large body and wing rotations exhibited by *Drosophila* during take-off require the use of unit quaternions to continuously parameterize the rotations. We present the first velocity-invariant motion prediction model that uses a quaternion parameterization [extending the approach of Rosenhahn and colleagues (Rosenhahn et al., 2007a)]. The near-cylindrical shape of the *Drosophila* body makes it difficult to estimate the roll angle about the body's highly symmetric central axis. The human tracking literature has considered such unobservable states (e.g. due to depth ambiguities and rotations about axes of symmetry in limbs) primarily within the context of monocular video (Sminchisescu and Triggs, 2003a; Sminchisescu and Triggs, 2003b), and these techniques are not applicable to our multi-camera setup. Recently, Kyrki and Kragic demonstrated tracking of the rotations of spherical objects and solids of revolution by integrating texture features into their CAD model (Kyrki and Kragic, 2006). Unfortunately, due to the low luminance associated with the high frame rate ( $6000 \text{ frames s}^{-1}$ ) that is needed to capture *Drosophila* wing kinematics, our video is void of any robust surface texture features except the silhouette (e.g. see Fig. 1). Instead, we rely upon the gross symmetrical motion of the wing beats to provide cues to the location of the body's dorsal edge.

We apply our method to flight initiation, a behavior that offers particular challenges for an automated tracking algorithm. During take-off, wingbeat frequency is high (Lehmann and Dickinson, 1998), the body may undergo large rotational movements (Card and Dickinson, 2008), and images of the body may be occluded by the substratum from which the fly launches itself. Flies are known to initiate flight in at least two ways: voluntary take-offs, which are elicited by attractive odors or unknown internal triggers, and escape take-offs, which are elicited by looming visual stimuli (Trimarchi and Schneiderman, 1995; Hammond and O'Shea, 2007; Card and Dickinson, 2008). Analysis of body motion indicates that during

voluntary take-offs animals jump slowly, but stably, into the air. In contrast, escape take-offs are characterized by both high translational and high rotational velocity (Card and Dickinson, 2008). In particular, escaping flies often begin flight with an elevated rotational velocity around the roll axis from which they must quickly recover or else crash into the ground. Presumably, flies are able to recover through feedback mediated by their eyes, ocelli, as well as mechanoreceptors on their wings and halteres. Although prior studies show that flies are indeed able to rapidly recover from extreme perturbations at the onset of flight, the means by which they modify wing kinematics to reach a stable equilibrium are unknown. Prior analyses using human-based digitization methods have provided a picture of *Drosophila* wing and body motion during stable hovering (Fry et al., 2005). Using an automated model-based algorithm for tracking wing and body motion we are able to provide the first picture of how flies transition from the highly unstable conditions at flight initiation to a stable pattern of motion.

The methods section first summarizes the model-based visual tracking approach. Next, a detailed *Drosophila* model is developed, including a velocity-invariant motion prediction model. Thereafter, we describe the technique for fitting the geometric model to the images while incorporating biomechanical constraints. We also test the results of our method against manually tracked and simulated data. The final section presents results obtained by applying this method to voluntary and escape take-offs.

## MATERIALS AND METHODS

The video subjects consist of 3 day old female *Drosophila melanogaster* (Meigen) filmed at  $6000 \text{ frames s}^{-1}$  with a shutter speed of  $50 \mu\text{s}$ . The video sequence was filmed by Card and Dickinson (Card and Dickinson, 2008) in a previous study that analyzed flight initiation using high-speed cameras (Photron Ultima APX, San Diego, CA, USA) to capture orthogonal views at a resolution of  $512 \text{ pixels} \times 512 \text{ pixels}$ . Flies entered the recording volume by crawling up a glass pipette, and the cameras were focused on the pipette tip to maximize the resolution at take-off. After ascending a few body lengths, the airborne flies typically became out of focus in one or more camera views (Fig. 1).

### Foreground segmentation

Nearly all conceivable approaches to automated tracking will rely upon accurate foreground segmentation, the process whereby the image pixels belonging to the organism and those belonging to the background are differentiated. Typical laboratory environments provide nearly constant background illumination during flight maneuvers. Hence, background subtraction is used to segment the pixels belonging to the fly. In addition, the appearance of *Drosophila* is very consistent during the video sequences. Fig. 2D shows a histogram of fly pixel intensity values over 200 frames from three different camera views. The characteristic bimodal shape is due to the opaque nature of the body cuticle, which consistently appears darker than other body parts (i.e. wings and legs) when back-lit. We utilize this appearance consistency to further segment fly pixels into body and appendage groups. At each frame and for each camera, we fit a 1D Gaussian mixture model with two members to the segmented fly pixels using the expectation-maximization (EM) algorithm.

A threshold, whose value is chosen as the local minima between the modes of the Gaussian densities, is used to segment the body and appendages (see Fig. 2C). This secondary segmentation is used in the image registration step (see Fig. 2B and 'Model registration' below).

### Model-based image tracking and non-linear estimation

To quantify the body and wing kinematics of *Drosophila* from video sequences recorded in multiple synchronized cameras, we first build a geometric model of the fly that is defined by the vector  $p$  of the parameters that encode the model's position, orientation and internal shape. Tracking over a sequence of images indexed by the positive integer  $k$  is performed by recursively estimating the parameters  $p_k$  from image measurements  $z_k$  at time step  $k$ . This tracking approach is based on a discrete time dynamic state space model:

$$p_k = f(p_{k-1}, \xi_{k-1}), \quad (1)$$

$$z_k = h(p_k, v_k), \quad (2)$$

where  $f$  describes the motion model,  $h$  is the measurement model, and  $\xi_k$  and  $v_k$  are independent and identically distributed noise processes. From a Bayesian perspective, tracking is based on the use of Bayes' rule to estimate the posterior probability density function,  $P$ :

$$P(p_k | z_{1:k}) = \frac{P(z_k | p_k) P(p_k | z_{1:k-1})}{P(z_k | z_{1:k-1})}, \quad (3)$$

where  $z_{1:k-1} = \{z_1, z_2, \dots, z_{k-1}\}$ . The optimal estimate of the fly's pose is the conditional mean of  $P(\cdot)$ :  $p_k = \mathbf{E}[p_k | z_k] = \int p_k P(p_k | z_{1:k}) \mathbf{d}p_k$ . Computationally, the estimates are obtained from a repeating two step process: (1) given the organism's estimated pose from the last frame, use the dynamic model (Eqn 1) to predict the animal's pose in the current frame; (2) use the image measurements from the current frame to further refine the estimate. In general, this recursive tracking solution is intractable and approximate solutions must be

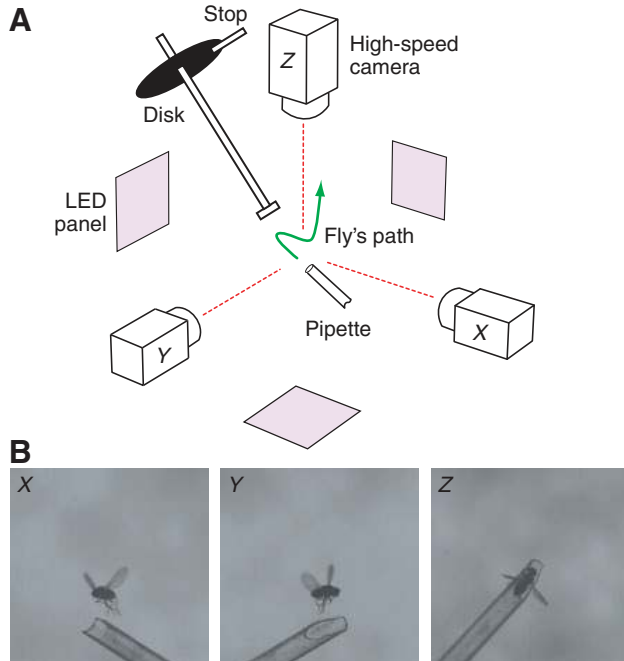


Fig. 1. Experimental set-up for capturing 3D high-speed sequences of take-off. (A) Arrangement of high-speed cameras and LED panels for back lighting. Individual flies emerged from inside a Pasteur pipette. To elicit escape responses, a stop was removed that released a black disk which fell toward the fly along a brass rod. (B) Images of *Drosophila* synchronously captured from three camera views. The high-speed video offers no strong visual features except the silhouette. Even with three camera views, the complex wing beat motion is difficult to capture due to low observability of the wings at certain postures and motion out of the camera's depth of field.

used instead. These approximation methods can be divided into two categories: those that assume the normally distributed density functions [e.g. Kalman filters (van der Merwe and Wan, 2003)] and those that allow arbitrary density functions [e.g. particle filters (Doucet et al., 2001)]. While particle filters can solve very general estimation problems, there are many visual tracking problems where the normal assumption holds, in which case Kalman filters provide accurate solutions and computational efficiency. The present study demonstrates that the normal assumption holds when tracking *Drosophila* via multiple cameras within a constrained laboratory that contains few environmental occlusions; hence we adopt the Kalman filtering approach. In particular, we employ the recently developed Sigma Point Kalman Filters (SPKFs), whose improved statistical linearizations have been shown to work well when applied to non-linear motion and measurement models (Ito and Xiong, 2000; van der Merwe and Wan, 2003; Nørgaard et al., 2000; Sibley et al., 2006). However, an appropriate motion model (Eqn 1) and measurement model (Eqn 2) are also needed in order to estimate the fly's motion parameters. First we construct a geometric model of *Drosophila*, whose outline forms the basis for the measurement equations. Second, we develop an empirical motion model. These models, when used in conjunction with the SPKF, allow automatic estimates of the fly parameters across an image sequence.

### Geometric model

The exoskeleton and wings of fruit flies exhibit various deformations during flight maneuvers. However, because the flies are filmed at a low magnification in order to capture the gross body and wing motion, we can reasonably assume that the fly's body and wing

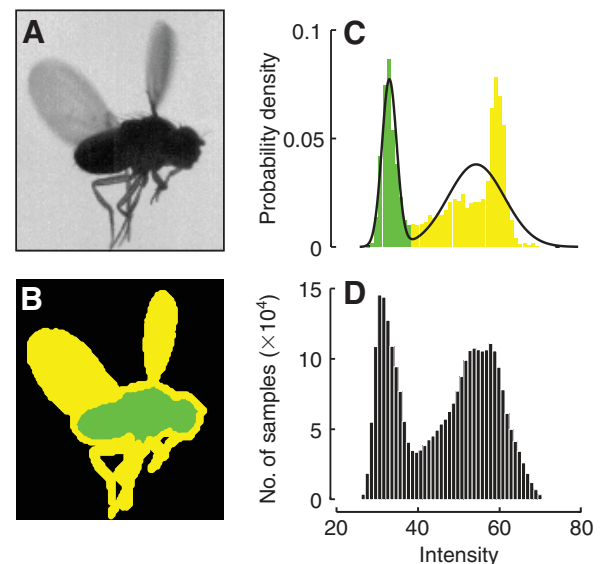


Fig. 2. Segmentation procedure for *Drosophila* images. (A) Typical image of *Drosophila* during flight initiation. (B) Image segmented into body (green) and appendage (yellow) pixels. (C) Histogram of pixel intensities (0–255) from A fitted with the sum of two Gaussians. The local minimum of the Gaussian sum is chosen as the threshold to classify body and appendage pixels. (D) Histogram of fly pixels calculated from background subtraction in over 200 frames across three different camera views. The characteristic bimodal shape is due to the opaque nature of the fly's body cuticle (lower intensity peak) versus the more translucent appendages (higher intensity peak).

parts undergo rigid body motion. Under this rigid body assumption, Dickson and colleagues constructed a polygonal model of the fruit fly from multiple calibrated images of the body and wing (Dickson et al., 2006). This triangular mesh model (Fig. 3A) is integrated into a high performance software library to simulate the Newtonian dynamics of flapping flight (see <http://www.ode.org>). We used this polygonal model to construct a parameterized generative model of the fly that contains three primitive shapes: the body, head and wings (Fig. 3B). The primitive shapes are assembled into an articulated model where each wing joint is modeled as a spherical joint (permitting arbitrary rotations about all three coordinate axes). In this paper, the head is rigidly attached to the thorax, though extra degrees freedom could easily be added. The shapes are constructed by applying continuous transformations to a B-spline curve. For instance, the fly's body (thorax and abdomen) and head are constructed by revolving a profile curve,  $R(u)$ , around a known centerline,  $C(u)$ , while the wings are constructed by transforming and scaling a closed curve (see Fig. 3A–C). This generative model offers a more compact representation of the fly's shape than the triangle mesh (i.e.  $\sim 80$  spline control points *versus*  $10^4$  mesh points).

One limitation of the current version of our automated tracker is the inability to detect and quantify deformations of the wing and body that violate the rigid body assumption. Wing and body deformations can be quite large in insects (Combes and Daniel, 2003a), and may play an important role in stability and maneuverability (Combes and Daniel, 2003b). There is nothing in our general methods that would preclude altering the geometric model of the insect to include deformation. However, such an effort would be worthwhile only if the deformation were large enough and the spatial resolution sufficient to capture them. Our imaging system was optimized to capture as large a portion of a fly's take-off behavior as possible, sacrificing spatial resolution for spatial extent. It should be possible for other researchers to modify our imaging arrangement and tracking algorithm in order to detect deformations of the body and wings, especially in larger insects in which such distortions are more pronounced.

### Coordinate transformations

To parameterize the rotations of the fly's body and wings relative to a fixed global frame, we utilize unit quaternions because their global representation does not suffer from the singularities inherent in Euler angle schemes. A unit quaternion, denoted  $Q=[q_1 q_2 q_3 q_0]^T$  can be equated to a  $3 \times 3$  rigid body rotation matrix by the relation:

$$R_Q = \begin{bmatrix} q_0^2 + q_1^2 - q_2^2 - q_3^2 & 2q_1q_2 - 2q_0q_3 & 2q_1q_3 + 2q_0q_2 \\ 2q_1q_2 + 2q_0q_3 & q_0^2 - q_1^2 + q_2^2 - q_3^2 & 2q_2q_3 - 2q_0q_1 \\ 2q_1q_3 - 2q_0q_2 & 2q_2q_3 + 2q_0q_1 & q_0^2 - q_1^2 - q_2^2 + q_3^2 \end{bmatrix}. \quad (4)$$

A rigid body coordinate transformation:

$$\mathbf{M} = \begin{bmatrix} R_Q & T \\ \mathbf{0}^T & 1 \end{bmatrix},$$

is a  $4 \times 4$  matrix that defines the position of a body-fixed reference frame relative to a world-fixed frame, with  $T$  being the  $3 \times 1$  vector that represents the translational distance between the reference frames. A kinematic chain of an articulated body is represented as the consecutive application of coordinate transforms. In this context, one coordinate transform defines the relative motion between the

fly body and the world reference frame, while two others define the relative motion between the wings and the body (see Fig. 4). The kinematic chain for one wing joint has the form:

$$\begin{bmatrix} X'_j \\ 1 \end{bmatrix} = \begin{bmatrix} R_{Q^b} & T \\ \mathbf{0}^T & 1 \end{bmatrix} \begin{bmatrix} R_{Q^w} & T_{bw} \\ \mathbf{0}^T & 1 \end{bmatrix} \begin{bmatrix} X_j \\ 1 \end{bmatrix}, \quad (5)$$

$$\begin{bmatrix} X'_j \\ 1 \end{bmatrix} = \mathbf{M}(p) \begin{bmatrix} X_j \\ 1 \end{bmatrix}, \quad (6)$$

where  $X_j$  represents the 3D coordinates of the  $j$ th model point of a wing in the local wing frame, and  $T_{bw}$  is the translation between the body-centered frame and the corresponding wing joint. The state of the fly model is  $p=[T Q^b Q^w]^T$ , where the superscripts refer to the body, left wing and right wing, respectively. These 15 parameters uniquely define the pose of the model fly at a particular instant of time.

An initial step is needed to initialize the fly model to the first frame of the video sequence, adjusting the model's dimensional parameters to the given organism, as well as approximating the animal's initial pose in the first video frame. This initialization process is enabled by a customized software package (Card and Dickinson, 2008). A user clicks on six locations (head, tail and joint/tip locations of both wings) on the fly's body in two out of three camera views to localize its 3D position. A wire frame model is next adjusted about its rotational axis until an approximate visual match is realized (see Fig. 5A). Thereafter, the body transformation found by the manual initialization is refined using the registration procedure described in 'Model registration' (below) applied to segmented images of the body shown in Fig. 2B. Finally, the body's shape profile is adjusted, while fixing the pose, to match the body-only segmented images. The entire width profile is modeled as the combination of two B-spline curves representing the head and the body. The B-spline control points are adjusted to best match the images, similar to a procedure recently published for zebrafish (Fontaine et al., 2008). Fig. 5B illustrates the results of such a shape refinement process for a particular *Drosophila*.

### Motion prediction via scaled motion dynamics

Practically speaking, the dynamic motion model (Eqn 1) predicts the fly's pose in the current frame being processed, based on the previously estimated poses. This prediction better initializes the fly parameter estimates before they are updated using the image measurement, restricting the search space for possible pose configurations. As a result, the tracking algorithm is able to cope with self-occlusion (e.g. the wing covers the body or *vice versa*) and other misleading visual clues (e.g. the fly partially covered by the pipette or other obstacles). While this approach assumes *a priori* knowledge of the animal's dynamics, a first principles construction of such a model is a daunting task, and the resulting model may be quite complicated. Techniques to craft empirical low-dimensional dynamical models have been developed within the human motion-tracking literature, where the motion model is calculated using principal component analysis (PCA) (Sidenbladh et al., 2002) and Gaussian processes (Urtasun et al., 2006; Sminchisescu and Jepson, 2004). However, these techniques are not invariant with respect to velocity (i.e. animals that have the same motion pattern but travel at different speeds, or animals traveling the same speed but filmed at different frame rates). Thus, low-dimensional models cannot provide accurate predictions when the training images (i.e. those sequences used to learn the motion pattern) are captured at different frame rates

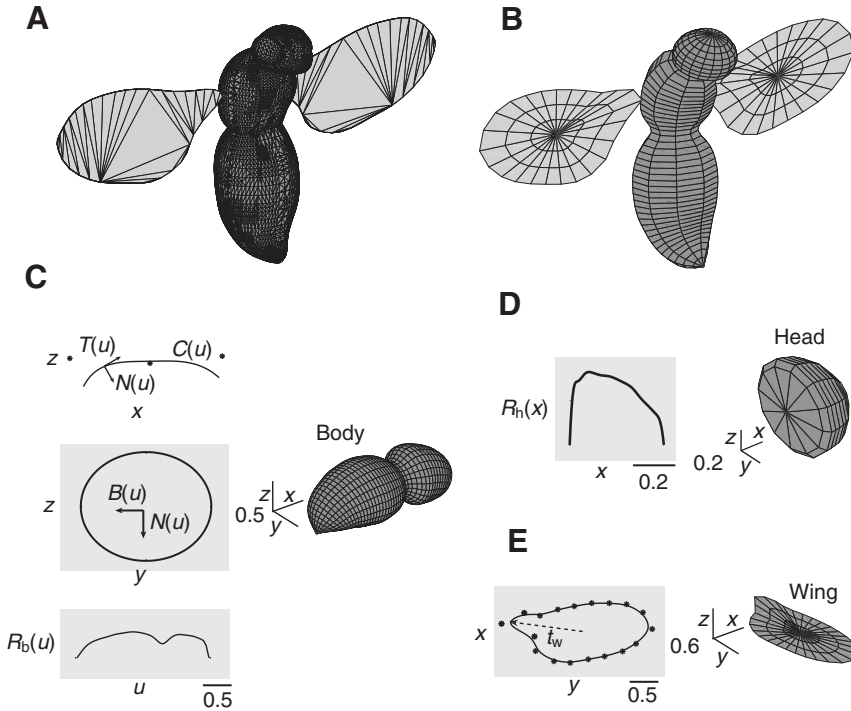


Fig. 3. Generative model of fly body. (A) Triangle mesh of *Drosophila* calculated from multiple calibrated images [courtesy W. Dickson (Dickson et al., 2006)]. (B) Complete generative model constructed from the data points shown in A. The model consists of three shape primitives: the body, head and wing. The generative modeling approach offers a more compact representation of the shape and motion of the fly than its triangle mesh counterpart. (C–E) Method for constructing components of the body shape primitive. (C) The centerline  $C(u)$  is a 3D B-spline curve with five control points (only three of them are visible in the axes). The curve of the centerline lies completely in the  $x$ – $z$  plane. The width profile,  $R_h(u)$ , is revolved around  $C(u)$  using an elliptical cross-section where the lateral direction is 20% wider than the dorsal–ventral direction. (D) Complete head model of the fly constructed identically to the method described for C using a different profile curve and the  $x$ -axis as the centerline. (E) Outline profile of the fly wing model constructed from a closed planar B-spline curve with 20 control points. For other definitions see ‘Table of abbreviations’ in text.

to the experimental images. In addition, dimensionality reduction works well for a single motion pattern, but a difficult-to-estimate mixture of regressors is needed in the case of multiple different motion patterns (Jacobs et al., 1991). We build upon the work of Rosenhahn and colleagues who model the motion patterns in the original high-dimensional space (Rosenhahn et al., 2007a). This approach can incorporate training data that are rescaled to different velocities, and that consist of completely different motion patterns.

We assume a set of temporally ordered training samples are available:

$$\{\tilde{p}_i := (\tilde{T}_i, \tilde{Q}_i^b, \tilde{Q}_i^{lw}, \tilde{Q}_i^{rw}) := (\tilde{T}_i, \tilde{Q}_i^b, \tilde{Q}_i^s) \mid i = 0 \dots N\}, \quad (7)$$

where  $\tilde{T}_i$ ,  $\tilde{Q}_i^b$  and  $\tilde{Q}_i^s$  respectively represent the translation, rotation (in quaternion form) and wing joint angle vectors (details on collecting the training samples are given in Results). This list of temporally ordered training samples is denoted  $\mathcal{P} = \langle \tilde{p}_0 \dots \tilde{p}_N \rangle$ , and the sublist in  $\mathcal{P}$  of length  $m$  ending at time  $i$  is denoted  $\langle \tilde{p}_{i-m+1} \dots \tilde{p}_i \rangle$ . To predict the state  $p_{k+1}$ , the training list is searched to find the location in the list that best matches the sublist of previous tracked states,  $\langle p_{k-m+1} \dots p_k \rangle$ . For the matching to be invariant with respect to the fly’s velocity, the matching is performed at different scalings  $s$  of  $\mathcal{P}$ . The different scalings of the training data, denoted  $\mathcal{P}^s$ , are calculated using two different techniques. Scaled body translations are obtained by linear interpolation and resampling. To produce valid rotations, spherical linear interpolation (Slerp) is employed (Shoemake, 1985) on the quaternion parameters. The resulting scaled lists are denoted  $\mathcal{P}^s = \{\tilde{p}_i^s := (\tilde{T}_i^s, \tilde{Q}_i^{b,s}, \tilde{Q}_i^s) \mid i = 0 \dots sN\}$ .

The best matching sublist of the training data is chosen as:

$$\operatorname{argmin}_{s,j} \sum_{v=0}^{m-1} \left( \left\| \mathbf{Q}_{k-v} - \tilde{\mathbf{Q}}_{j-v}^s \right\| \right). \quad (8)$$

Note that only the wing joint angles are considered in the initial matching process, as their motion will be invariant with respect to

the fly’s global orientation. Fig. 6 illustrates this technique. To calculate the wing displacement for the dynamic update step, Eqn 1, the predicted motion between frames is estimated from the training data set, and this relative motion is applied to the current state to predict the state variables in the next frame:

$$\partial \mathcal{Q}_{j+1}^{lw,s} = \tilde{\mathcal{Q}}_{j+1}^{lw,s} \cdot \left( \tilde{\mathcal{Q}}_j^{lw,s} \right)^{-1}, \quad (9)$$

$$\mathcal{Q}_{k+1}^{lw} = \partial \tilde{\mathcal{Q}}_{j+1}^{lw,s} \cdot \mathcal{Q}_k^{lw}, \quad (10)$$

where  $\cdot$  denotes quaternion multiplication. An identical calculation to Eqn 9 is also performed for the right wing and the body orientation. The predicted body translation is given by:

$$T_{k+1} = T_k + \left( \tilde{T}_{j+1}^s - \tilde{T}_j^s \right), \quad (11)$$

so the entire predicted state consists of  $p_{k+1} = [T_{k+1} \mathcal{Q}_{k+1}^b \mathcal{Q}_{k+1}^{lw} \mathcal{Q}_{k+1}^{rw}]$ .

### Model registration

The filter’s measurement update step, which is effectively an image registration procedure, is constructed as follows. To quantify the registration error between the 3D fly model and the image measurements, we minimize the distance between a set of 3D points on the model and a set of 3D lines that are constructed from the image. We assume a set of calibrated pin-hole cameras:

$$\lambda_j^i \begin{bmatrix} u_j^i \\ v_j^i \\ 1 \end{bmatrix} = K_i \begin{bmatrix} R_i & C_i \end{bmatrix} \begin{bmatrix} X_j^i \\ 1 \end{bmatrix} \quad i = 1 \dots M, \quad (12)$$

with known intrinsic parameters  $K_i$  and extrinsic parameters  $(R_i, C_i)$  (Svoboda et al., 2005).  $X_j^i$  denotes the  $j$ th 3D point in our fly model with respect to the fixed frame, and  $(u_j^i, v_j^i)$  are the pixel coordinates of this point in camera  $i$ . In order to create correspondences between the model and image silhouette features in a given camera view,

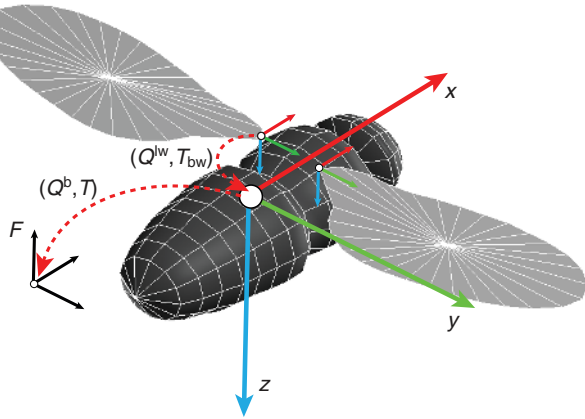


Fig. 4. Geometric generative model of *Drosophila*. Following aeronautical convention, the rotations about the  $x$ -,  $y$ - and  $z$ -axes are defined as roll, pitch and yaw, respectively. The downward pointing  $z$ -axis is chosen so that positive pitch angles correspond to pitching upwards. The model's kinematic chain includes a coordinate transformation from the left wing frame to the body frame [given by  $(Q^w, T_{bw})$ ] and a transformation from the body-fixed to the world-fixed frame  $F$ , denoted by  $(Q^p, T)$ . Analogous transformations exist for the right wing.

the model, whose pose is calculated in the dynamic prediction step, is first projected using Eqn 12 to produce the set of 2D image points corresponding to 3D points on the model surface (Fig. 7). Next, a closed B-spline curve is fitted to the 2D boundary points  $x_j^i$  to determine the local normal vector  $n_j^i$  at each boundary point. For each point  $x_j^i$ , a search in the data image (Fig. 2B) is performed along the normal  $n_j^i$  to locate edges. Because the 3D coordinates of the projected points  $x_j^i$  are known, one obtains a set of correspondences between edge locations  $e_j^i$  and 3D model locations  $X_j^i$  (Fig. 7A). These correspondences are recomputed at each iteration of the Kalman filter update, similar to the widely used iterated closest point (ICP) algorithms (Rusinkiewicz and Levoy, 2001).

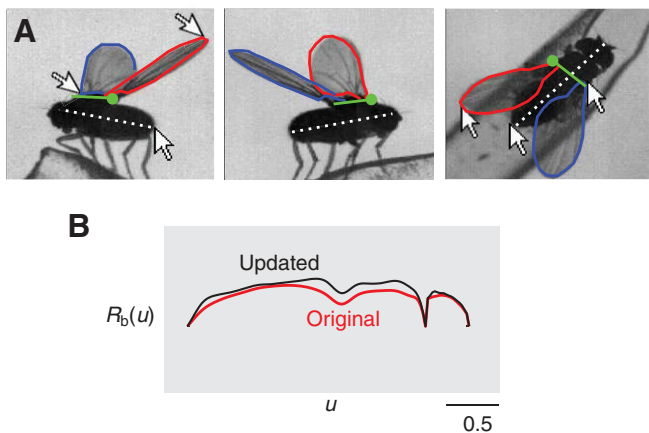


Fig. 5. Procedure for initializing automated tracker. (A) We used customized software for manual digitizations of *Drosophila* body kinematics from Card and Dickinson (Card and Dickinson, 2008). Points were clicked at the head, tail, wing joint and wing tip in multiple camera views to manually fit a geometric model to the images. The manually estimated body pose was then used as an initial guess for the automated algorithm. (B) At the initial frame, the profile of the body was refined, while holding the pose parameters fixed, to more closely match the actual shape of the fly by minimizing the error described in 'Model registration' (Materials and methods).

Next, a set of projection rays emanating from the 2D edge locations are reconstructed so that the Euclidean distance between the model points and corresponding rays can be minimized. The projection rays are represented in Plücker coordinates to permit an easy calculation of the distance between a point and a line. Let  $L_j^i=(n_j^i, m_j^i)$  denote the Plücker coordinates of the projection ray connecting edge point  $e_j^i$  with camera center  $C_i$ , where  $n_j^i$  is a unit vector colinear to the line and  $m_j^i=x \times n_j^i$  is for any point  $x$  in the line. Given a camera calibration, these coordinates are:

$$n_j^i = R_i^T K_i^{-1} \begin{bmatrix} e_j^i \\ 1 \end{bmatrix}, \quad m_j^i = C_i \times n_j^i. \quad (13)$$

A point  $x$  is incident with line  $L$  if  $x \times n - m = 0$ . The distance between a point  $x \notin L$  and line  $L=(n, m)$  is  $\|x \times n - m\|$ , and this quantity serves as a convenient error measure (Fig. 7B). Hence, the state is updated by collecting all of the correspondences across all camera views (Fig. 7C) and minimizing the error:

$$\min \sum_{i,j} \left\| \left( \mathbf{M}(p) X_j \right)_{3 \times 1} \times n_j^i - m_j^i \right\|^2. \quad (14)$$

This approach for 3D pose estimation is also utilized in Rosenhahn et al. (Rosenhahn et al., 2007b). By concatenating the error vectors, this error function can be put in the standard Kalman filter form  $\|z-h(p)\|^2$  where  $h(p)=[\mathbf{M}(p)X]_{3 \times 1} \times n$  and  $z=m$ . Minimizing the point-to-line distances offers considerable computational savings compared with an error function that minimizes point-to-point distance in the image plane. Such an error function involves the projection of the 3D model and the calculation of its occluding contour for each function evaluation.

#### *Drosophila* constraints

The *Drosophila* tracking algorithm must incorporate two constraints. The first simply insures that the quaternions maintain unit length, and the second addresses the practical unobservability of the body's roll angle due to the body's axial symmetry. The form of the latter constraint is suggested by anatomical principles. Because the wings are simultaneously actuated through oscillatory deformations of the exoskeleton, we assume that the body roll angle will remain (roughly) symmetrical between the two wing angles. This does not impose a condition that the wing motion is symmetrical. Instead, it repositions the body's roll angle, while allowing the wings to follow the image data. This technique is illustrated in Fig. 8, and the detailed calculations that quantify this constraint are presented in the Appendix. Both of these non-linear constraints can be expressed as a functional relation of the form  $c(p_k)=0$  that must be included within the estimation algorithm. Within the SPKF filter framework, the constraints are incorporated by using a Sigma Point transform applied to a projection operator which projects the state estimate onto the constraint surface (Julier and LaViola, 2007). This method presumes the existence of a projection function  $w(p_k)$  such that:

$$c(w(p_k)) = 0 \quad \forall p_k \in \mathbb{R}^n. \quad (15)$$

The quaternion constraint has the form  $w_1(p_k)=Q_k/\|Q_k\|=1$ , while the analogous result for the roll angle constraint is developed in the Appendix.

#### Quantifying tracker performance

To assess the accuracy of the proposed method, we compared body pose estimates in six video sequences with those reported in Card

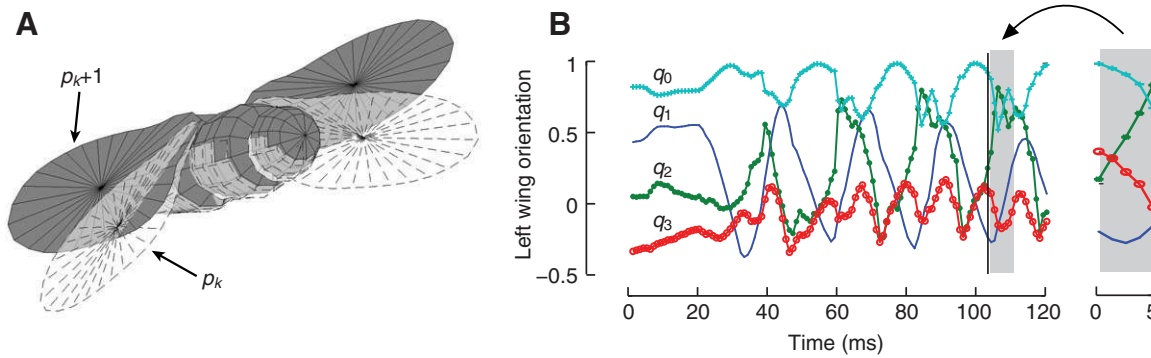


Fig. 6. Predictive component of tracker. (A) Motion model used to predict the location of the fly in the next frame,  $p_{k+1}$ , given estimates from the previous frame,  $p_k$ . Here, the displayed motion during the upstroke of the wingbeat is exaggerated to illustrate the concept. (B) Rotational motion of *Drosophila* left wing motion during take-off (120 out of 380 samples shown). Motion is parameterized by four quaternions which vary smoothly with time. The query of  $m=5$  previously calculated poses is matched with position 106 of the prior database. The relative motion to position 107 is used to calculate the prediction.

and Dickinson (Card and Dickinson, 2008), where the body pose was captured manually using the customized software described in ‘Coordinate transformations’ (above). For the manually tracked study, a reduced body model was fitted to the images typically every five frames and a spline was used to smoothly interpolate the location of the body model at intermediate frames (wing joint angles were not estimated in this study). This approach decreased the labor-intensive nature of manual digitization, and it removed some of the variance attributed to human fitting by providing temporal

smoothness. Fig. 9 demonstrates quantitatively that our automated method can realize estimates that are comparable with human visual inspection. In reality, there are no ‘ground truth’ estimates for these data sets. Whereas the performance of human motion-tracking systems is typically measured against a marker-based system, the tiny size of *Drosophila* only permits visual measurements subject to human interpretation. In this section all errors are reported as root mean square (r.m.s.) values. Wing errors represent an average r.m.s. value between the left and right.

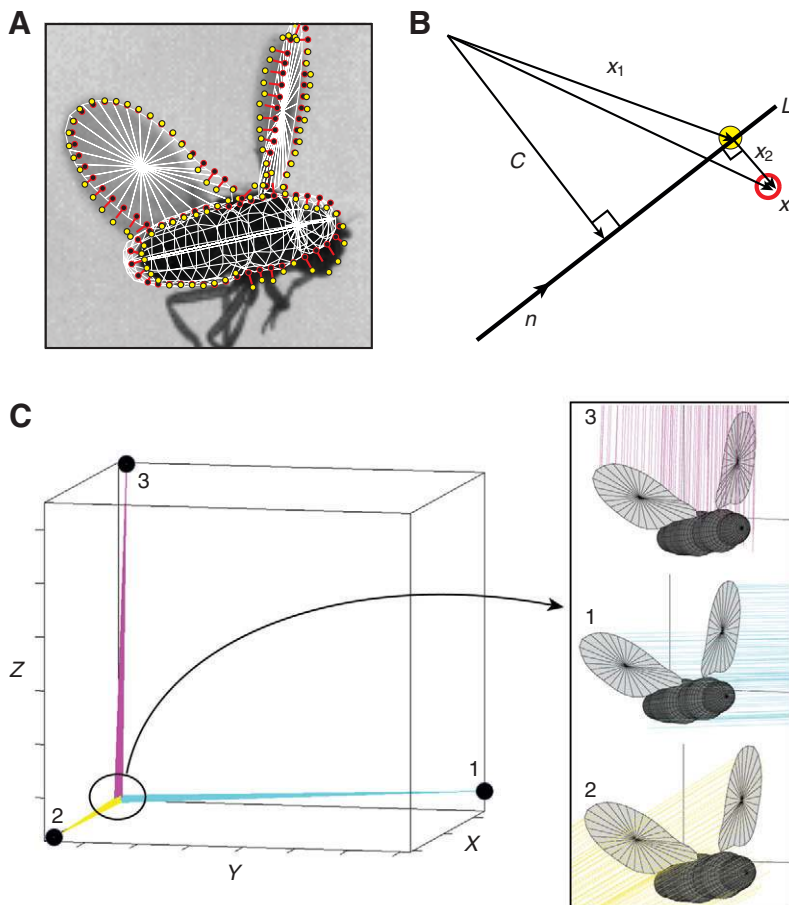


Fig. 7. (A) Correspondence between projected model points  $x_j^i$  and detected edge locations  $e_j^i$  shown in red and yellow, respectively. The edge locations are used to reconstruct the projection ray corresponding to that point on the image silhouette. (B) The distance,  $\|x_2\|$ , from a point  $x$  to a line  $L=(n, m)$  represented in Plücker coordinates is calculated using the relation  $\|x_2\|=\|x \times n - m\|$ . This distance provides the error vector,  $x_2$ , that is minimized to make the model points match the projection rays of the image silhouette. (C) In order to fit the geometric model to the images from multiple camera views, we reconstruct the projection rays from the image silhouette in each of the three camera views. The intersection of the projection rays from each camera and the model are displayed individually for illustration. We fit the model by minimizing the point to line distance across all projection rays in all cameras.

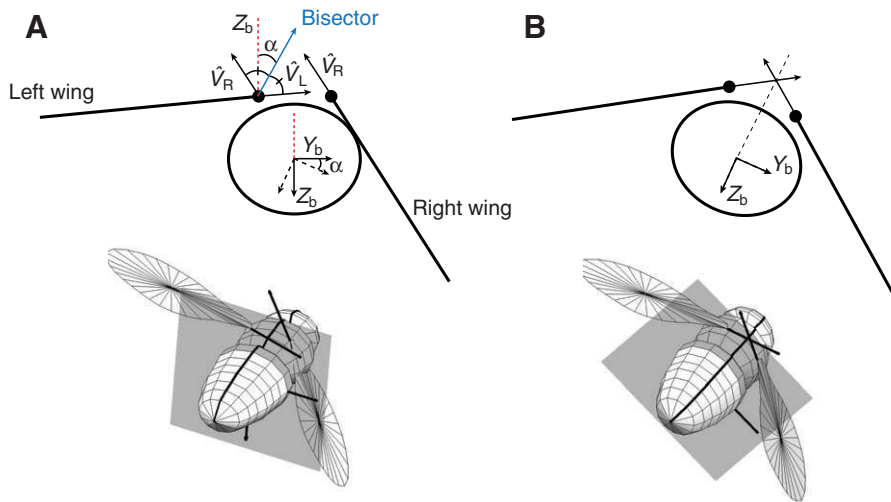


Fig. 8. Implementation of roll constraint. Because the roll angle of the body is unobservable from silhouette data in the images, a symmetry constraint within the transverse plane of the body must be incorporated. (A) Unconstrained estimate of the fly's pose; (top) projection of the model vectors into the transverse plane, (bottom) 3D pose with transverse plane illustrated in gray. This body configuration is highly unlikely given the biomechanics of *Drosophila*. (B) Constrained estimate after rotating the body by angle and updating joint angle vectors,  $Q$ .

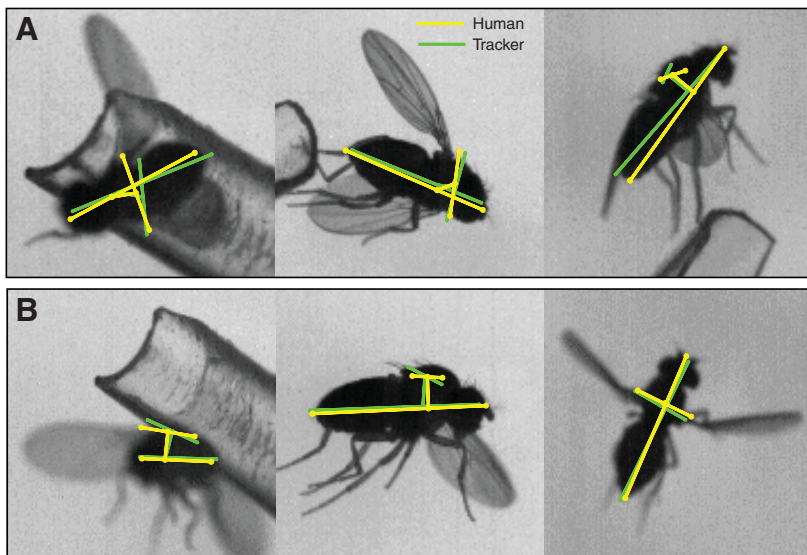


Fig. 9. Performance metrics of tracker compared to human digitizer. Only body kinematics are compared because human-tracked wing motion is unavailable. (A,B) Two frames where a large discrepancy in roll angle was observed between the human estimate and the algorithm. From visual inspection, the human estimate in A appears more accurate than the algorithm's estimate, while in B the algorithm appears to provide a better estimate and more accurate roll angle. (C) Time trace of entire video sequence with frames A and B indicated. Tracker values are solid lines, data from human are shown as open circles. (D) Root mean square (r.m.s.) deviations between the human estimates and our tracker for body orientation and translation. Each bar represents a separate video sequence. The roll angle shows the greatest deviation, as expected due to the symmetrical nature of the fly's body. Video sequence from C has the largest deviation amongst all videos.

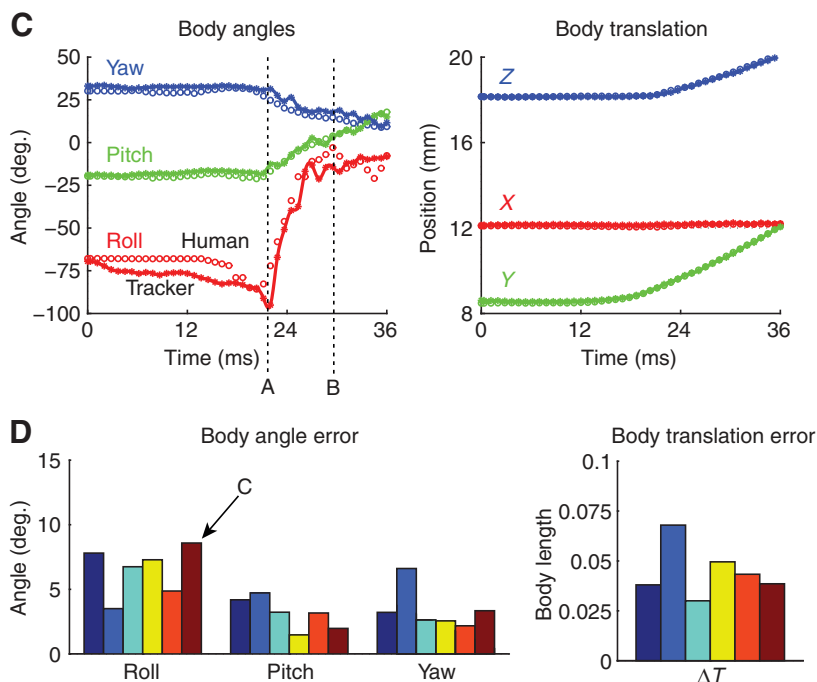




Fig. 9A,B illustrates two configurations where large differences between human and automated roll angle estimates were observed. Based on visual inspection, it appears that the human estimates were more accurate in Fig. 9A, while the automated estimates are slightly better in Fig. 9B. Both display the reduced body frame model used for manual fitting (the long axis indicates the head and tail locations, and the raised 'T' junction indicates the approximate wing joint locations which are the visual cues used to determine roll angle). Fig. 9C is the time trace of the body orientation and translation with frames A and B clearly marked. Fig. 9D summarizes the results for all six video sequences. The algorithm can typically estimate the body's center of mass location within 5% of the body length, an absolute distance that is of the order of 0.1 mm. Body orientation is also estimated well. As expected, the roll angle exhibits the largest deviations of  $6.5 \pm 1.9$  deg. on average due to the greater uncertainty associated with rotations about a highly symmetrical axis. The video sequence associated with Fig. 9A–C represents the sequence with the largest error (8.6 deg.) between the human and automated roll angle estimates.

Fig. 10 compares wing angle performance against a human digitizer for a representative voluntary take-off sequence. The results are nearly identical for stroke amplitude ( $\theta$ ; 3.3 deg. error) and stroke deviation ( $\phi$ ; 2.1 deg. error), although the two methods do differ for angle of attack ( $\alpha$ ; 8.8 deg. error), especially during stroke reversals. Such differences are expected, as the subjective choices that are required of a human digitizer are most difficult during stroke reversal when the wing is rapidly flipping and changing direction. This does not imply that a human operator is necessarily less precise, and because there is no absolute ground truth for a captured behavioral sequence it is impossible to determine which method yields more

accurate kinematic data. Automatic tracking is, however, more objective and repeatable. Thus, the algorithm will be useful in practical application because it achieves estimates comparable to human interpretation, while significantly decreasing the labor involved in measuring such data.

Another performance assessment was performed to compare the tracker estimates with an actual ground truth. We use the geometric model (Fig. 11A) and the known experimental camera calibration to construct a set of synthetic images (Fig. 11B) of the fly along a realistic trajectory involving a stable voluntary take-off. The algorithm is used to track these synthetic images, and the results are displayed in Fig. 11C. The difference between the estimate and the ground truth at each time step is displayed as a histogram of residuals. Body position and orientation accuracy are similar to those achieved when comparing with manual tracking (Fig. 9D). Estimates of the wing angles, however, exhibit a broader distribution of errors. Because our model is connected in a kinematic chain, errors in the wing angles are coupled to errors in the body position and orientation. Stroke amplitude ( $\theta$ ) and deviation ( $\phi$ ) display strong accuracy with errors of 3.3 deg. and 4.8 deg., respectively. Geometric angle of attack ( $\alpha$ ) is also estimated with errors of 17.2 deg. Higher errors in angle of attack are expected due to decreased camera resolution about this degree of freedom. In some instances, the synthetic image contained very few wing pixels due to our model's infinitesimal wing thickness. Also, when the wing speed is small the direction of motion is noisier, which causes the angle of attack measurements to be noisier. For this reason, we do not include the measurements before the initial downstroke in our error analysis.

The tracking algorithm's two failure modes are primarily seen during escape behaviors. These fast maneuvers can cause

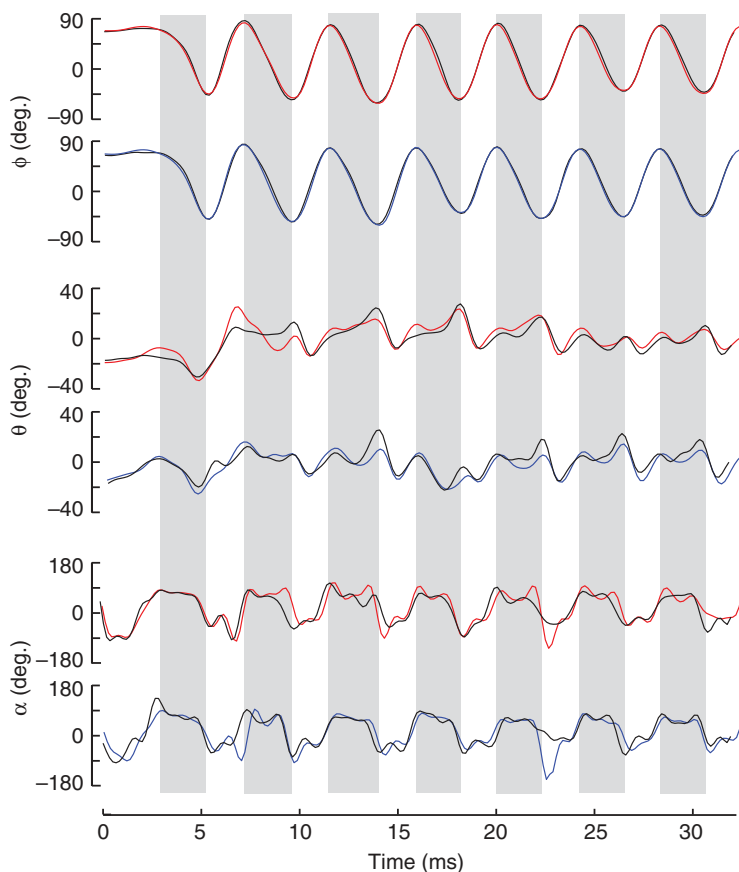


Fig. 10. Comparison of manual and automated tracking of wing kinematics. Each pair of traces (for  $\theta$ ,  $\phi$  and  $\alpha$ ) plots the kinematic angles for the right (red) and left (blue) wing. Automatically tracked data are shown in color; manual-digitized data are shown in black. The r.m.s. errors are 3.3 deg. ( $\theta$ ), 2.1 deg. ( $\phi$ ) and 8.8 deg. ( $\alpha$ ).

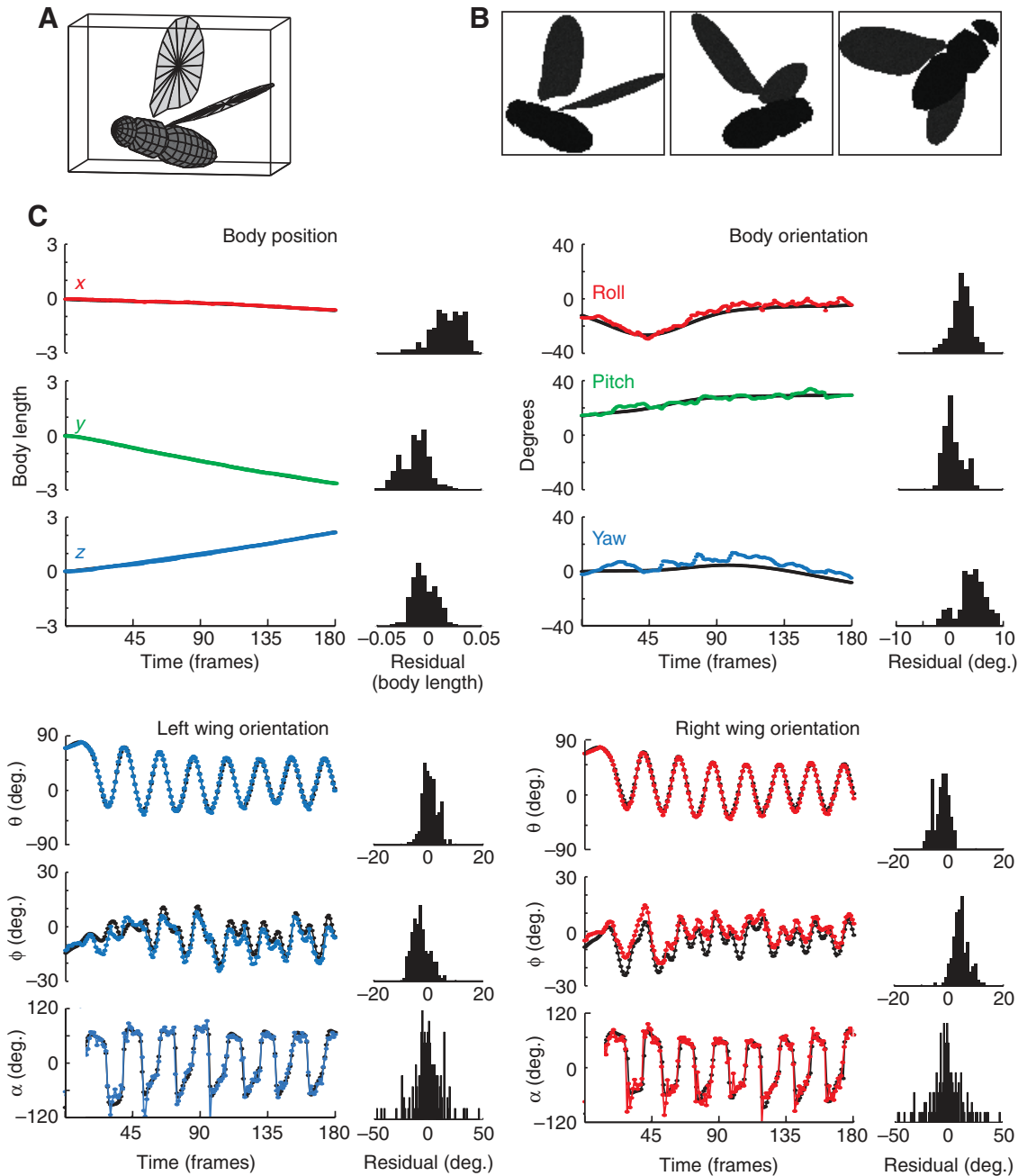


Fig. 11. Performance metrics of tracker compared with synthetic images. (A) The generative model of the fly and known camera calibration are used to construct (B) synthetic images of a realistic trajectory of stable voluntary take-off. (C) The difference between the estimate (colored line) and the ground truth (black line) at each time step is displayed as a histogram of residuals. Body position and orientation accuracy are similar to those achieved when comparing with manual tracking (Fig. 9D). Stroke amplitude ( $\theta$ ) and deviation ( $\phi$ ) have errors of 3.3deg. and 4.8deg., respectively. Angle of attack ( $\alpha$ ) error of 17.2deg. is due to higher errors when the wing speed is small because the wing velocity direction is a noisier signal. For this reason, we don't plot angle of attack before the initial downstroke.

significant wing deformations that are not captured by our current rigid model (see Fig. 12Aii). However, given the multiple camera views and scaled motion priors, the algorithm is able to continue tracking and provide good estimation once the wing assumes a less deformed configuration (Fig. 12B). Another failure mode of the current algorithm is shown in Fig. 12Ci, where the right wing of the fly is flipped. The frame is taken from an upstroke of the wing path, so the right wing should be undergoing supination like the left wing. Instead, the leading edge is facing downwards

as if during a downstroke. This failure primarily occurs during the escape maneuvers when complicated body motions self occlude one of the wings in one or more camera views. Despite the strong motion prior, this causes the state posterior distribution to violate the normal assumption. In addition, the incorrect orientation of the right wing could be caused by the inaccuracies in the body orientation estimate, which are primarily due to inaccuracies in the body shape (length, width, deforming centerline axis, etc.).

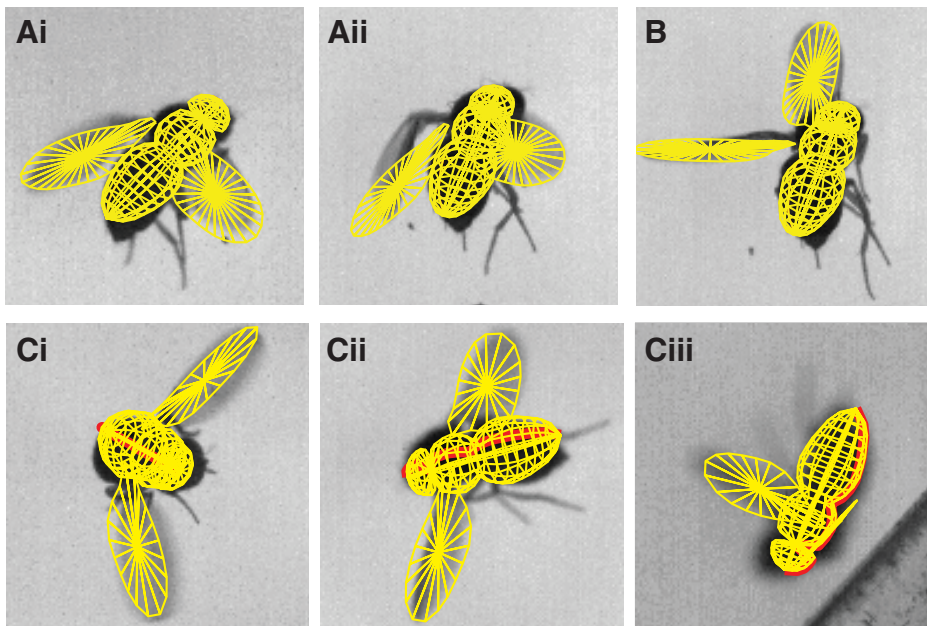


Fig. 12. Examples of gross errors in tracking algorithm. (A) During some escape maneuvers, the fly's wing can undergo large deformations (shown in Aii) that are not captured by our current rigid body model. In other camera views (Ai), this deformation is not apparent. (B) Despite this large error, the algorithm does not lose track and is able to continue successful estimation. (C) Another failure mode of the tracking algorithm. The fly as seen in Ci is facing towards the camera during an upstroke. The left wing (top) is in the proper configuration, but the right wing (bottom) is flipped in the wrong orientation (pronation instead of supination).

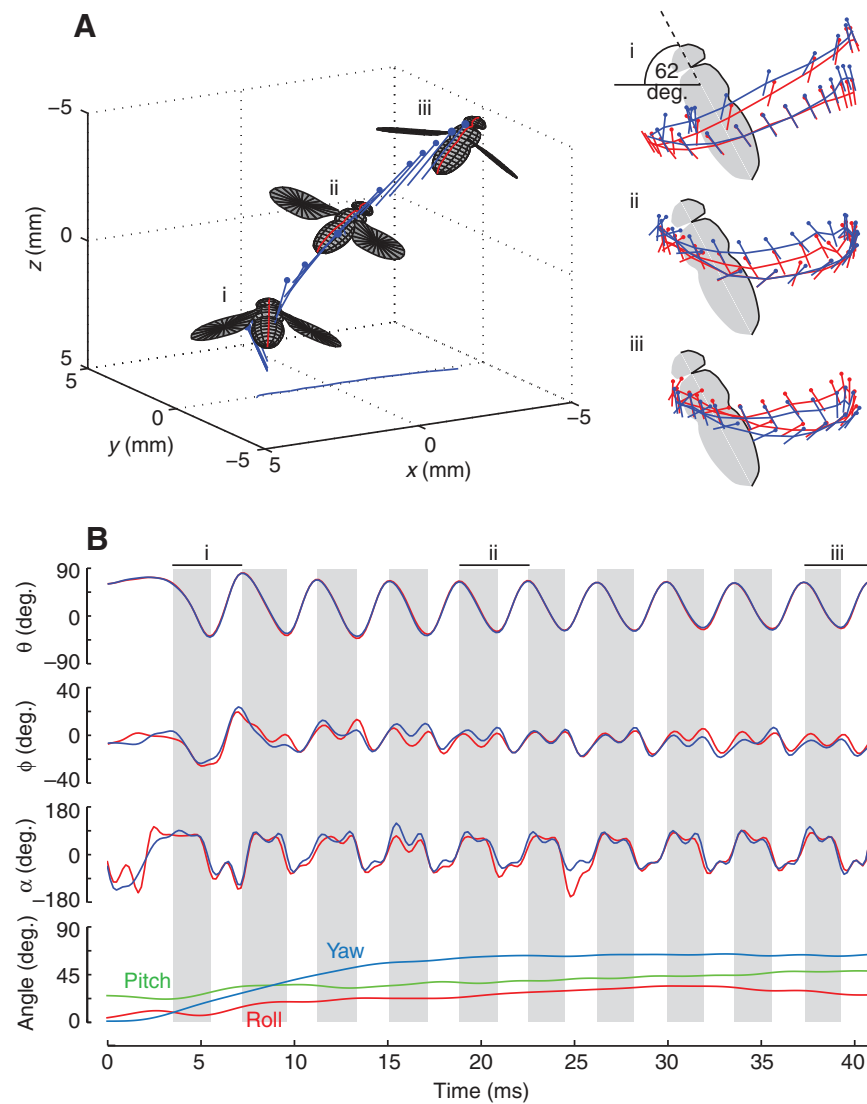


Fig. 13. Example of voluntary take-off. (A) 3D trajectory of fly during take-off sequence. Wing kinematics for stroke cycles at the beginning, middle and end of the sequence are shown to the right. The right wing is indicated in red, the left in blue. (B) Time history of angles describing wing and body kinematics throughout the take-off sequence. The wing angles were defined relative to a plane through the wing hinges that is inclined 62 deg. from the body axis (see Ai), which is the position of the mean stroke plane in hovering flies. Sequence is shown in supplementary material Movie 1. See text for details.

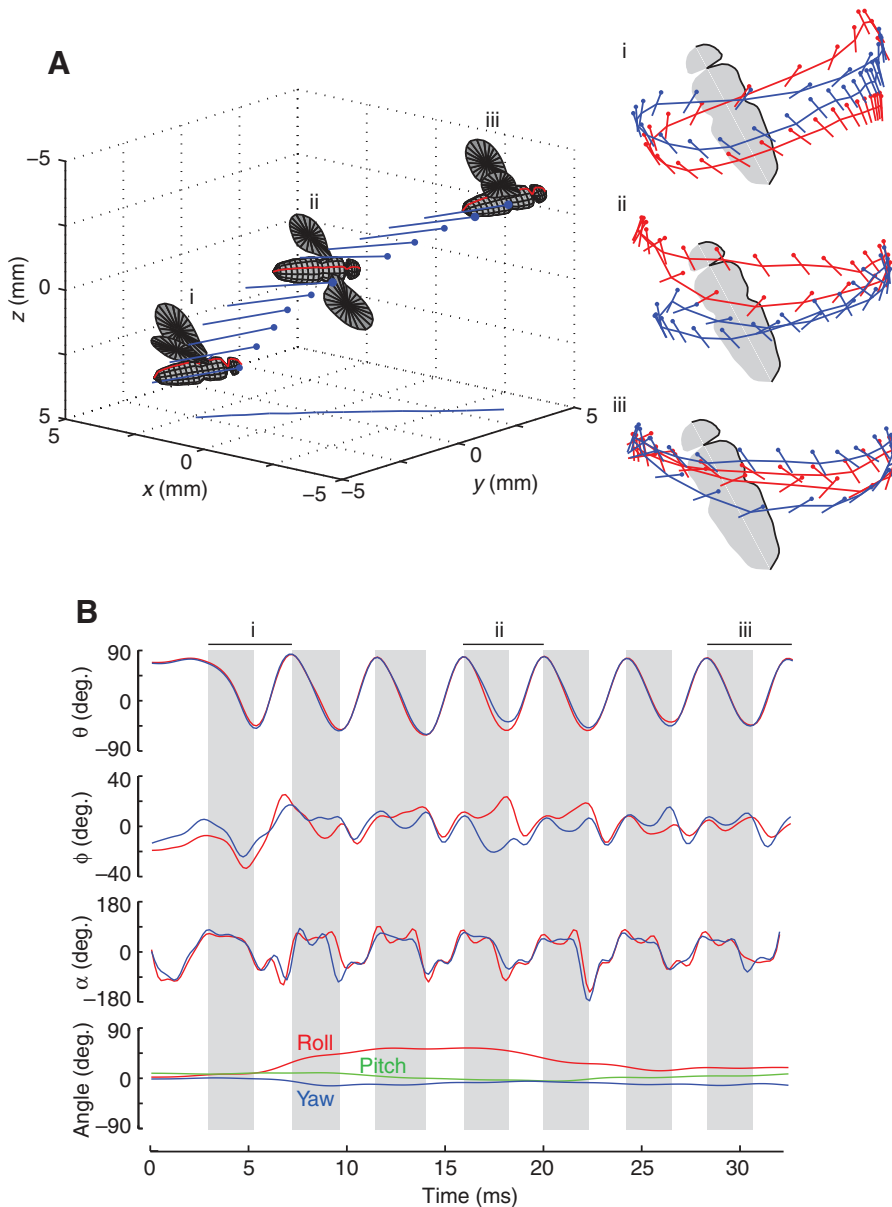


Fig. 14. Example of voluntary take-off. Plotting convention same as Fig. 13. Sequence is shown in supplementary material Movie 2. See text for details.

## RESULTS

The algorithm is written in MATLAB<sup>TM</sup> and has an average computation time of  $45 \pm 3$  s/frame<sup>-1</sup> on a 3.0 GHz Intel<sup>®</sup> Xeon processor. The source code will be made freely available upon request to the authors. To illustrate our algorithm's capabilities, we analyzed video of two different types of behavior: voluntary take-offs (flies remained on a pipette undisturbed until they flew away) and escape take-offs (flies were startled by the approach of a falling disk). For each video sequence, the geometric model was manually initialized to the first frame according to 'Coordinate transformations' (Materials and methods). The database of training samples ('Motion prediction', Materials and methods) consisted of 380 samples from a voluntary take-off and 111 samples from an escape take-off. Initially the training samples were captured manually, and then they were gradually replaced with the estimated values from the tracking algorithm. Wing angles are measured in a body-centered coordinate frame with the longitudinal axis pitched at 62 deg. with the horizontal (Fig. 13Ai), which is consistent with hovering flight (David, 1978). The orientation measurements are

smoothed with a zero phase lag fourth order Butterworth filter with a cut-off frequency of 1000 Hz and 250 Hz for the wings and body, respectively.

Using our tracking algorithms, we analyzed a total of nine take-off sequences. Of these, we describe in detail four sequences (two voluntary and two escape) that illustrate the range of changes in wing and body kinematics that occur at the onset of flight. Fig. 13 (and supplementary material Movie 1) shows a voluntary take-off in which the onset of flight was particularly smooth and stable. By the third downstroke, the fly reaches a consistent pattern of wing motion that is maintained with little change for the rest of the sequence. This basic pattern in which the wings follow gentle 'U-shaped' trajectories is quite similar to that previously described for stably hovering fruit flies using a manual method of digitization (Fry et al., 2005). The main change in stroke kinematics throughout the entire sequence is a slight gradual decrease in stroke amplitude ( $\theta$ ), which is accomplished primarily through a drop in the ventral extent of the wing stroke (compare kinematics at i, ii and iii). Once wing motion stabilizes, the fly maintains a constant body pitch of

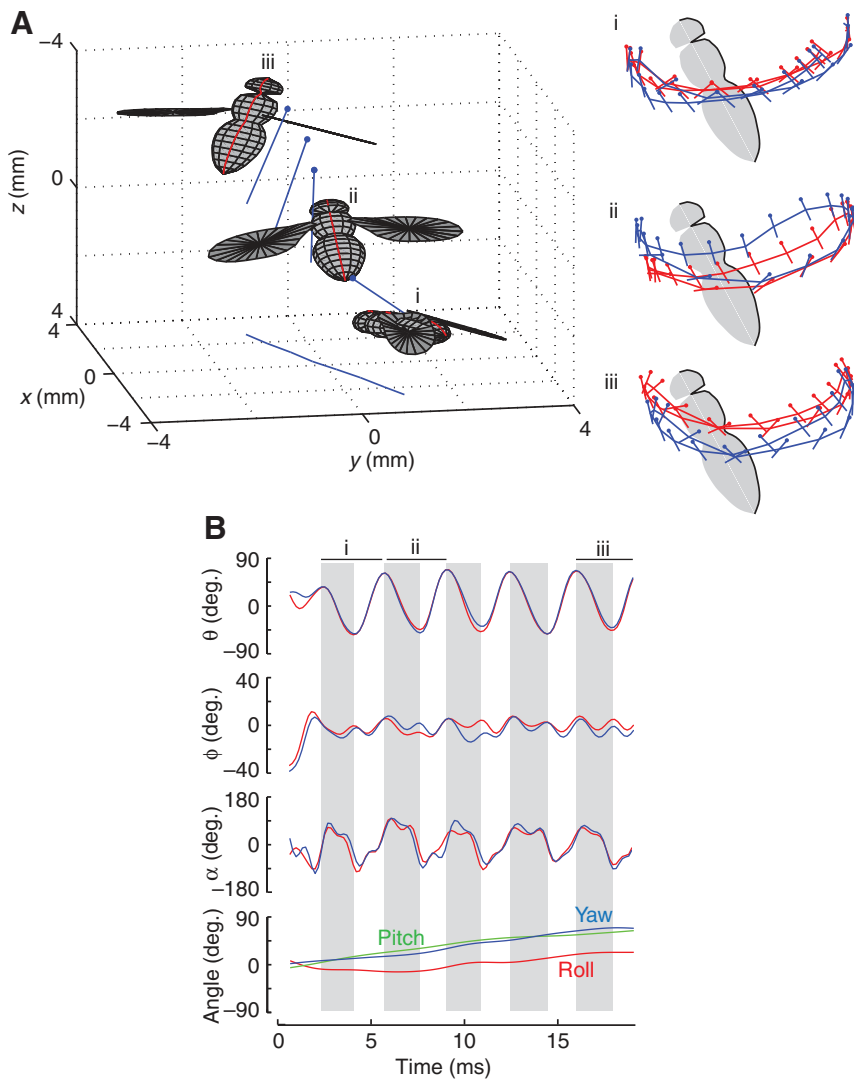


Fig. 15. Example of escape take-off. Plotting convention same as Fig. 13. Sequence is shown in supplementary material Movie 3. See text for details.

approximately 45 deg., a slight roll of about 20 deg., and a constant heading. The only major break in left–right symmetry is during the seventh upstroke in which the right wing shows a very high angle of attack. The fact that this change is maintained for just one wingstroke indicates that the fly has the ability to modulate wing kinematics on a stroke-by-stroke basis. There is, however, no obvious change in body orientation as a consequence of this one stroke. The stroke frequency averaged across the sequence is 268 Hz, which is somewhat higher than measured during stable hovering flight, but consistent with studies of flight initiation using tethered flies (Lehmann and Dickinson, 1998).

The kinematics over the first two strokes in Fig. 13 indicate how rapidly the pattern of wing motion can reach steady conditions following the jump that initiates flight. The flight sequence begins with the wings held constant in a dorsal location. The wing motion parallel to the stroke plane ( $\theta$ ) reaches the steady-state pattern almost instantly, as does the pattern for wing axial rotation ( $\alpha$ ). The main difference in wing motion during the first two strokes is that stroke deviation ( $\phi$ ) exhibits a sawtooth-like pattern of constant downward motion during the downstroke and constant upward motion during the upstroke. The result is that the wing follows a more ventral trajectory during the downstroke than during the upstroke (Fig. 13Ai), opposite to the pattern exhibited during steady flight (Fig. 13Aiii). Note that during the first downstroke the wing angle

of attack is nearly parallel to the body axis, which is itself roughly parallel to the horizontal plane. Such an arrangement would create very high vertical forces just as the animal takes off. As found previously for free flight (Fry et al., 2005), the fly in Fig. 13 did not exhibit clap and fling (Weis-Fogh, 1973), even during the initial strokes of take-off.

The sequence shown in Fig. 14 (and supplementary material Movie 2) shows another voluntary take-off. The first two strokes of the flight sequence are virtually identical to those shown in Fig. 12, suggesting that voluntary take-offs begin with a stereotyped pattern of wing motion. The final stroke of the sequence again resembles the ‘U-shaped’ pattern indicative of stable flight. In this sequence, however, the fly generates a brief, but extreme, maneuver starting with the fourth stroke (Fig. 14Aii). At this time, the left wing undergoes a shorter stroke ( $\theta$ ) and a large negative deviation ( $\phi$ ) while the right wing maintains the same stroke length but undergoes a positive deviation. The net result is a large left–right asymmetry in wing motion. This asymmetry continues, slightly attenuated, during the fifth stroke, but then reverses during the sixth such that the left wing undergoes a more positive deviation while the right wing undergoes a more negative deviation ( $\phi$  trace, Fig. 14B). Starting with the fourth stroke, the animal begins to roll at a rate of roughly 5500 deg. s<sup>-1</sup>. Rotation this fast is likely to activate the campaniform sensilla at the base of the halteres (Sherman and

Dickinson, 2003), which might be responsible for initiating the compensatory reaction observed during the sixth stroke, in which the pattern of wing motion exhibited in the fourth and fifth stroke reverses. Throughout this maneuver, the animal maintains a constant pitch and heading and a wingbeat frequency of 241 Hz. As with the other voluntary sequence, the fly does not exhibit clap and fling.

The sequence shown in Fig. 15 (and supplementary material Movie3) is an example of an escape response, elicited by a looming visual stimulus, that nevertheless resulted in a relatively stable take-off. By the fourth and fifth strokes the animal has achieved the 'U-shaped' pattern typical of stable hovering. During the second and third strokes of the sequence the fly exhibits a switch in the pattern of stroke amplitude and deviation ( $\phi$  trace, Fig. 15B) that is reminiscent of that seen in Fig. 14. The sequence differs from those shown in Figs 13 and 14 most notably at the start of wing motion. Inspection of the video sequence indicates that the thoracic flight motor begins oscillating before the animal has raised its wings, and as a consequence the wing stutters during the first stroke. The initial downstroke is not coordinated with a large negative deviation as it is in the voluntary take-off sequences. The initial wingbeat frequency is 300 Hz, substantially higher than that measured at the start of the

voluntary take-offs or in stable hovering flight (Fry et al., 2005). The animal starts the sequence with its body parallel to the ground, but pitches upward over the first five strokes to reach a posture typical of low stable flight. Again, the fly did not exhibit clap and fling.

The sequence shown in Fig. 16 (and supplementary material Movie4) is an escape take-off that was chosen because it represents an extreme case in wing and body kinematics. At the start of flight, the jump legs generate a large torque that pitches the animal backwards so that it is upside down for most of the sequence. As in the sequence in Fig. 15, the wings begin oscillating before they have been elevated to a proper start position and as a consequence they appear to 'unfurl' during the first stroke (Fig. 16Ai). By the end of the sequence, the animal has partially recovered, not by pitching down, but rather by rolling toward an upright position (see the rising roll trace at the bottom of Fig. 15). This strong rolling moment is correlated with an extreme asymmetry in wing motion in which the tip of the right wing follows a broad open loop (Fig. 16Aii). The angle of attack is very high during the upstroke, whereas it is much lower during the downstroke. It is presumably this large asymmetry during the upstrokes and downstrokes of the right wing that creates the moment which starts to roll the animal from an upside-down position.

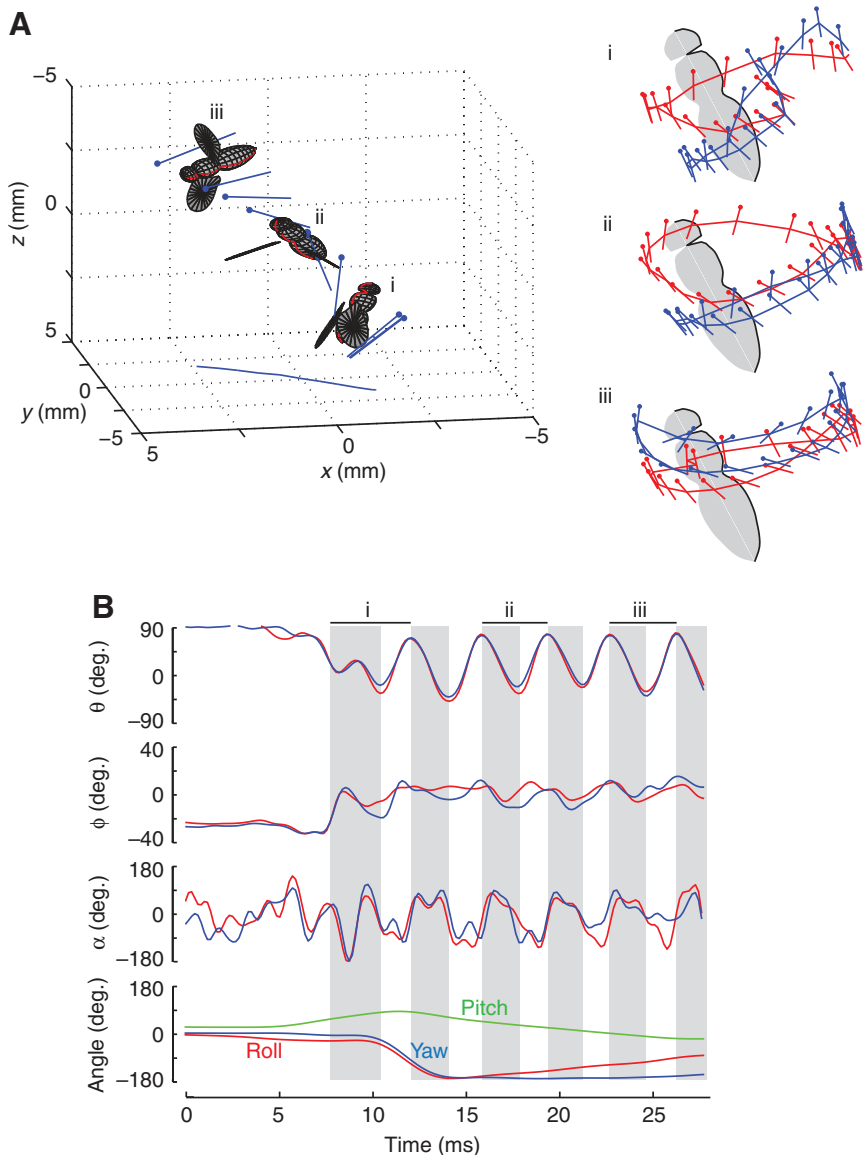


Fig. 16. Example of escape take-off. Plotting convention same as Fig. 13. Sequence is shown in supplementary material Movie 4. See text for details.

Although the fly is not fully recovered, the kinematics in the last stroke in the sequence begin to resemble the stable flight pattern seen in the other sequences. This sequence illustrates how quickly an animal can recover from an enormous perturbation at the onset of flight. As with the escape sequence shown in Fig. 15, the fly in Fig. 16 begins flight with a rather high wingbeat frequency of 285 Hz. Consistent with the other three sequences, the fly did not exhibit clap and fling during take-off.

A more extensive comparison of voluntary take-offs is shown in Fig. 17, which plots the wing kinematics from eight automatically tracked sequences. To align the sequences from eight different flies, the data were normalized by either stretching or contracting the time axis so that the first three stroke periods were equivalent. The sequences are remarkably similar indicating that voluntary take-offs are quite stereotyped, in contrast to escape responses. None of the flies exhibited clap and fling, and the take-off kinematics closely resemble those of hovering flies (Fry et al., 2005). A detailed analysis of escape take-offs, which are much more variable, will be the subject of a future study.

### DISCUSSION

We have presented a practical model-based visual tracking algorithm that estimates the 3D motion of free flying *Drosophila* from multiple camera views. The algorithm uses a dynamic state

estimation framework to provide robustness to self-occlusions and static background clutter such as that due to the glass pipette from which the animals launch. By simply matching motion patterns in a training set, the approach provided accurate prediction in video sequences containing multiple types of behaviors (voluntary *versus* escape take-off). Our examination of take-off behaviors showed that the algorithm can successfully track a fly through complex tumbling motions. Our method has comparable accuracy to manual-based human digitization, but offers the promise of allowing biologists to analyze much larger image-based databases of kinematics.

The first application of our method has already provided new insight into the complex dynamics of flight initiation. Our results suggest that voluntary take-offs begin with a stereotyped, feed-forward pattern of motion in which the wing creates large vertical forces during the first downstroke when the longitudinal body axis is parallel to the substratum (Fig. 13Ai; Fig. 14Ai). The pattern of wing motion then approaches that of stable flight within two or three strokes (Fig. 13Aii; Fig. 14Aii). In cases in which the fly must recover from instabilities introduced by the jump, the sequences reveal how quickly the sensorimotor system can respond to bring the animal towards a stable flight posture, even when the animal is initially flipped upside down (Fig. 16). The sequences also suggest that these animals do not rely on clap and fling kinematics to generate elevated lift, even at the onset of flight. This supports the notion that the clap and fling in *Drosophila*

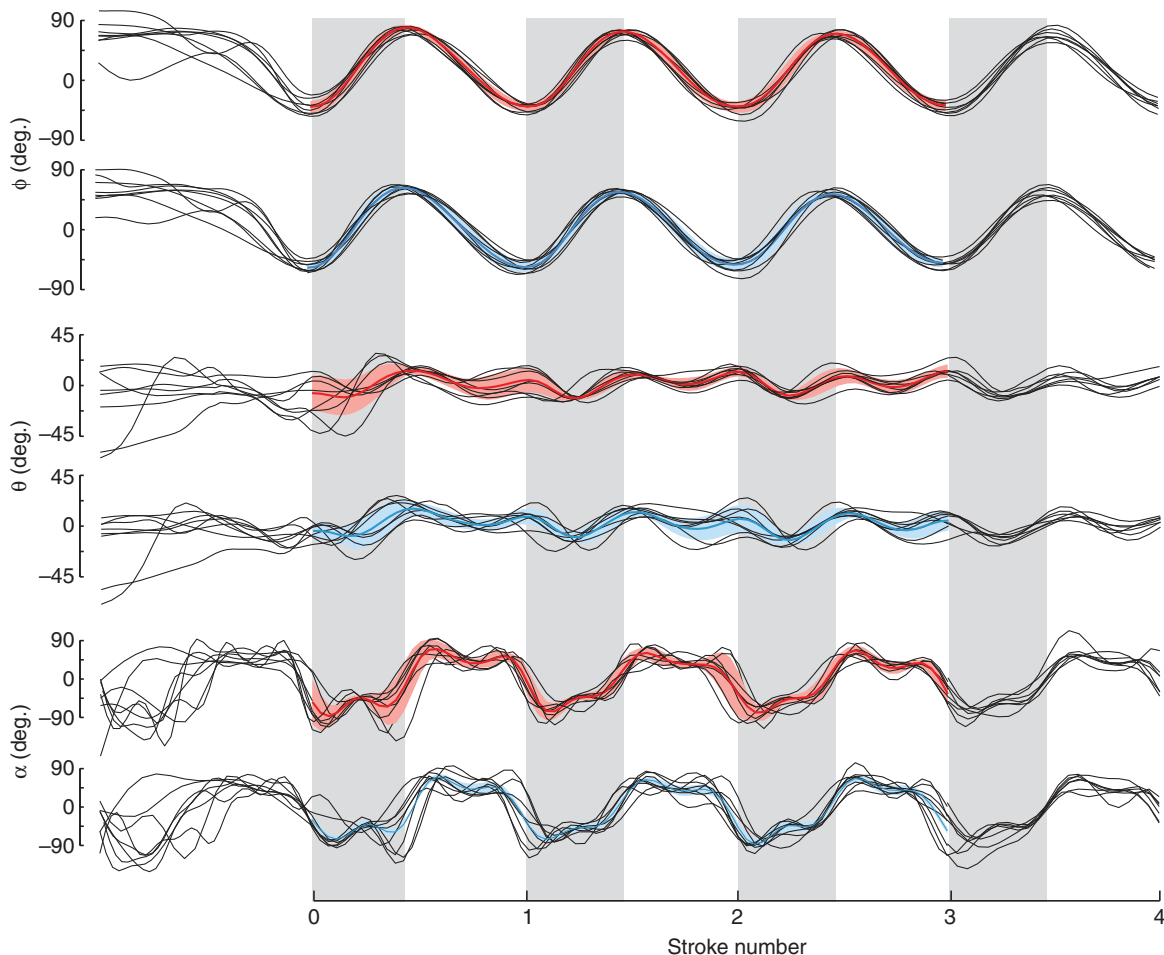


Fig. 17. Wing kinematics during voluntary take-offs. Each set of traces shows data from eight sequences. Raw data are shown in black; averages are shown in color for the right (red) and left (blue) wings. The standard deviation envelopes are plotted as light red and blue areas in each trace. To align the different flight sequences, the time axis of each trace was normalized so that the first three strokes took the same amount of time. Averages and standard deviation envelopes are shown only over this three-stroke interval.

may be in large part an artifact of tethering (Fry et al., 2005). The results confirm, however, that flies do rely on very high wingbeat frequencies at the onset of flight (Lehmann and Dickinson, 1998). All of the sequences show evidence that changes in wing kinematics may last for only one or a few wingbeat cycles, which suggests that the underlying neuromuscular circuits can operate on a stroke-by-stroke basis to alter aerodynamic forces and moments. Evidence for this rapidity is suggested by studies of tethered flight (Heide and Gotz, 1996; Balint and Dickinson, 2004), but is now supported by free-flight kinematics. In the future it will be possible to gain a richer insight into take-offs and other aspects of flight control through the application of model-based automated tracking.

### APPENDIX

A constraint on the body's roll angle, which is deployed after the image registration process, can be developed as follows. Conceptually, we assume that the dynamic prediction step produces a roughly correct candidate orientation of the model. The constraint rotates the body about its  $x$ -axis so that the  $z$ -axis bisects the angle formed by the wing vectors in the body's transverse plane. The equations are developed for the left wing only; an analogous calculation is carried out for the right wing. Let  $R_{Q_L}=[X^b Y^b Z^b]$  and  $R_{Q_L}R_{Q_w}=[X^{tw} Y^{tw} Z^{tw}]$  denote the coordinate axes of the body and left wing relative to the fixed frame at the current time step (the subscript  $k$  is omitted for brevity). Let  $V_L=Y^{tw}$  and  $V_R=-Y^{tw}$  be known as the wing vectors that point from the wing tip towards the wing joint.

The vectors  $\mathbf{Y}^b$  and  $\mathbf{Z}^b$  define an orthonormal basis in the planar subspace transverse to the fly's body. The symmetry constraint is imposed in this subspace. Let:

$$\hat{V}_L = \langle V_L, \mathbf{Z}^b \rangle \mathbf{Z}^b + \langle V_L, \mathbf{Y}^b \rangle \mathbf{Y}^b \quad (\text{A1})$$

$$= V_L^x \mathbf{i} + V_L^y \mathbf{j} \quad (\text{A2})$$

denote the projection of the left wing vector into the transverse plane. Next,  $\hat{V}_R$  is mirrored about the body's  $z$ -axis and the angle between them is calculated as:

$$2\alpha = \cos^{-1} \left( \left\langle \left( \begin{array}{cc} 1 & 0 \\ 0 & -1 \end{array} \right) \hat{V}_R, \hat{V}_L \right\rangle \right). \quad (\text{A3})$$

As  $\alpha$  is always positive, we change signs if  $|V_L^x| > |V_R^x|$ , which denotes a counter-clockwise rotation (Fig. 8 is a clockwise rotation,  $\alpha > 0$ ). The constrained body transformation is calculated by applying the coordinate transformation that encodes the roll update to the unconstrained transformation:

$$\begin{bmatrix} R_{Q^{b\ddagger}} & T^\ddagger \\ \mathbf{0}^T & 1 \end{bmatrix} = e^{\xi\alpha} \begin{bmatrix} R_{Q^{b*}} & T^* \\ \mathbf{0}^T & 1 \end{bmatrix}, \quad (\text{A4})$$

$$\text{where } \xi = \begin{pmatrix} -\mathbf{X}^b \times T \\ \mathbf{X}^b \end{pmatrix}, \quad (\text{A5})$$

is the geometric twist of the body's roll axis, and  $\ddagger$  and  $*$  denote the constrained and unconstrained estimates, respectively. As our model is a kinematic chain, this roll transformation also rotated the wings to an incorrect position. Let  $X_i^* \in \mathbb{R}^3$  denote the  $i$ th wing point in our model at the unconstrained estimate. The constrained value of the wing joint angles is calculated as:

$$\mathbf{Q}^\ddagger = \underset{\mathbf{Q}}{\operatorname{argmin}} \sum_i \left( X_i^* - \mathbf{M}(\mathbf{Q})X_i \right)^2, \quad (\text{A6})$$

where  $\mathbf{M}$  is defined in 'Coordinate transformation' in Materials and methods (i.e. the distance between the wings points at the unconstrained state and the constrained state is minimized, holding the body transformation fixed and modifying the wing joint angles). This calculation which imposes the roll angle symmetry constraint, denoted  $\mathbf{w}_2(p_k)$ , is applied after the quaternion projection such that the complete projection function is given by  $p_k^\ddagger = \mathbf{w}(p_k) = \mathbf{w}_2[\mathbf{w}_1(p_k)]$ .

### LIST OF ABBREVIATIONS

$B(u)$	bi-normal vector to geometric model's centerline curve
$C_i$	3D coordinates of $i$ th camera center
$C(u)$	centerline of generative fly body model
$E[\cdot]$	expectation
$e_j^i$	measured edge location corresponding with $x_j^i$
$\mathbf{F}$	world-fixed frame
$f(\cdot, \cdot)$	motion model
$h(\cdot, \cdot)$	measurement model
$K_i$	intrinsic parameters of $i$ th camera center
$L_j^i$	Plücker coordinates of projection ray connecting $e_j^i$ and $C_i$
$\mathbf{M}$	rigid body coordinate transformation
$\mathbf{n}_j^i$	local normal vector corresponding to $x_j^i$
$N(u)$	normal vector to geometric model's centerline curve
$\bar{p}$	fly parameters from training database
$p_k$	fly parameters at $k$ th time step
$q$	unit quaternion parameterizing rotation in 3D
$Q$	unit quaternion
$Q^b$	quaternion rotation from body to fixed frame
$Q^{Lw}$	quaternion rotation from left wing joint frame to body frame
$Q^{Rw}$	quaternion rotation from right wing joint frame to body frame
$R_h$	profile curve of head model
$R_i$	rotation matrix from fixed frame to camera-centered reference frame for $i$ th camera center
$R(u)$	profile curve of fly body model
$T$	translation from body-fixed to world-fixed frame
$T(u)$	tangent vector to geometric model's centerline curve
$T_{bw}$	translation from body-centered coordinate frame to wing joint frame
$t_w$	distance from wing center to joint attachment location on geometric model
$u$	centerline parameterization
$V_L$	left wing vector
$V_R$	right wing vector
$\mathbf{w}(\cdot)$	projection function to incorporate non-linear constraints
$X_j^i$	3D coordinates of $j$ th model point
$x_j^i$	$j$ th boundary point of generative model projected into $i$ th camera image plane
$z_k$	image measurement at $k$ th time step
$\alpha$	geometric angle of attack
$\theta$	stroke amplitude
$\phi$	stroke deviation

The authors would like to thank Gwyneth Card, who recorded all video illustrated in Results and made her customized manual tracking software and body kinematic data (Card and Dickinson, 2008) readily available. Will Dickson provided the digitized surface points of a *Drosophila* body and wings that were used to construct the generative model. We also thank the Beckman Institute of Caltech for providing financial support for this project. This work was also supported by an NSF FIBR grant (0623527) to M.H.D.

### REFERENCES

- Altshuler, D. L., Dickinson, W. B., Vance, J. T., Roberts, S. P. and Dickinson, M. H. (2005). Short-amplitude high-frequency wing strokes determine the aerodynamics of honeybee flight. *Proc. Natl. Acad. Sci. USA* **102**, 18213-18218.
- Balint, C. N. and Dickinson, M. H. (2004). Neuromuscular control of aerodynamic forces and moments in the blowfly, *Calliphora vicina*. *J. Exp. Biol.* **207**, 3813-3838.
- Card, G. and Dickinson, M. H. (2008). Performance trade-offs in the flight initiation of *Drosophila*. *J. Exp. Biol.* **211**, 341-353.
- Combes, S. A. and Daniel, T. L. (2003a). Flexural stiffness in insect wings. I. Scaling and the influence of wing venation. *J. Exp. Biol.* **206**, 2979-2987.
- Combes, S. A. and Daniel, T. L. (2003b). Flexural stiffness in insect wings. II. Spatial distribution and dynamic wing bending. *J. Exp. Biol.* **206**, 2989-2997.
- David, C. (1978). The relationship between body angle and flight speed in free-flying *Drosophila*. *Physiol. Entomol.* **3**, 191-195.



- Dickinson, M. H., Lehmann, F. and Gotz, K. (1993). The active control of wing rotation by *Drosophila*. *J. Exp. Biol.* **182**, 173-189.
- Dickson, W., Straw, A., Poelma, C. and Dickinson, M. (2006). An integrative model of insect flight control. In *Proceedings of the 44th AIAA Aerospace Sciences Meeting and Exhibit*. Reston, VA: American institute of Aeronautics and Astronautics.
- Doucet, A., de Freitas, N. and Gordon, N. (2001). *Sequential Monte-Carlo Methods in Practice*. New York: Springer-Verlag.
- Fontaine, E., Lentink, D., Kranenbarg, S., Muller, U. K., van Leeuwen, J. L., Barr, A. H. and Burdick, J. W. (2008). Automated visual tracking for studying the ontogeny of zebrafish swimming. *J. Exp. Biol.* **211**, 1305-1316.
- Fry, S., Sayaman, R. and Dickinson, M. (2003). The aerodynamics of free-flight maneuvers in *Drosophila*. *Science* **300**, 495-498.
- Fry, S. N., Sayaman, R. and Dickinson, M. H. (2005). The aerodynamics of hovering flight in *Drosophila*. *J. Exp. Biol.* **208**, 2303-2318.
- Gotz, K. G. (1987). Course-control, metabolism and wing interference during ultralong tethered flight in *Drosophila melanogaster*. *J. Exp. Biol.* **128**, 35-46.
- Graetzel, C., Fry, S. N. and Nelson, B. J. (2006). A 6000 Hz computer vision system for real-time wing beat analysis of *Drosophila*. In *Proceedings of the IEEE/RAS-EMBS International Conference on Biomedical Robotics and Biomechanics*, pp. 278-283. Piscataway, NJ: IEEE Computer Society Press.
- Hammond, S. and O'Shea, M. (2007). Escape flight initiation in the fly. *J. Comp. Physiol. A* **193**, 471-476.
- Heide, G. and Gotz, K. (1996). Optomotor control of course and altitude in *Drosophila melanogaster* is correlated with distinct activities of at least three pairs of flight steering muscles. *J. Exp. Biol.* **199**, 1711-1726.
- Ito, K. and Xiong, K. (2000). Gaussian filters for nonlinear filtering problems. *IEEE Trans. Automat. Contr.* **45**, 910-927.
- Jacobs, R. A., Jordan, M. I., Nowlan, S. J. and Hinton, G. E. (1991). Adaptive mixtures of local experts. *Neural Comput.* **3**, 79-87.
- Julier, S. and LaViola, J. (2007). On kalman filtering with nonlinear equality constraints. *IEEE Trans. Signal Process.* **55**, 2774-2784.
- Kyrki, V. and Kragic, D. (2006). Tracking unobservable rotations by cue integration. In *IEEE International Conference on Robotics and Automation*, pp. 2744-2750. Orlando, FL: ICRA.
- Lehmann, F. O. and Dickinson, M. H. (1998). The control of wing kinematics and flight forces in fruit flies (*Drosophila* spp.). *J. Exp. Biol.* **201**, 385-401.
- Moeslund, T. B., Hilton, A. and Kruger, V. (2006). A survey of advances in vision-based human motion capture and analysis. *Comput. Vis. Image Underst.* **104**, 90-126.
- Nørgaard, M., Poulsen, N. and Ravn, O. (2000). New developments in state estimation for nonlinear systems. *Automatica* **36**, 1627-1638.
- Ristroph, L., Berman, G. J., Bergou, A. J., Wang, Z. J. and Cohen, I. (2009). Automated hull reconstruction motion tracking (HRMT) applied to sideways maneuvers of free-flying insects. *J. Exp. Biol.* **212**, 1324-1335.
- Rosenhahn, B., Brox, T. and Seidel, H. P. (2007a). Scaled motion dynamics for markerless motion capture. In *IEEE Conference on Computer Vision and Pattern Recognition*. Piscataway, NJ: IEEE Computer Society.
- Rosenhahn, B., Brox, T. and Weickert, J. (2007b). Three-dimensional shape knowledge for joint image segmentation and pose tracking. *Int. J. Comput. Vis.* **73**, 243-262.
- Rusinkiewicz, S. and Levoy, M. (2001). Efficient variants of the ICP algorithm. In *International Conference on 3D digital Imaging and Modeling (3DIM)*. Piscataway, NJ: IEEE Computer Society.
- Sherman, A. and Dickinson, M. H. (2003). A comparison of visual and haltere mediated equilibrium reflexes in the fruit fly, *Drosophila melanogaster*. *J. Exp. Biol.* **206**, 295-302.
- Shoemaker, K. (1985). Animating rotation with quaternion curves. *SIGGRAPH Comput. Graph* **19**, 245-254.
- Sibley, G., Sukhatme, G. and Matthies, L. (2006). The iterated sigma point kalman filter with applications to long range stereo. In *Robotics Science and Systems*. Philadelphia, PA: MIT Press.
- Sidenbladh, H., Black, M. J. and Sigal, L. (2002). Implicit probabilistic models of human motion for synthesis and tracking. In *Proceedings of the 7th European Conference on Computer Vision, Part 1*. London: Springer-Verlag.
- Sminchisescu, C. and Triggs, B. (2003a). Estimating articulated human motion with covariance scaled sampling. *Int. J. Robot. Res.* **22**, 371-391.
- Sminchisescu, C. and Triggs, B. (2003b). Kinematic jump processes for monocular 3D human tracking. In *IEEE Conference on Computer Vision and Pattern Recognition*, vol. 1, pp. 69-76. Piscataway, NJ: IEEE.
- Sminchisescu, C. and Jepson, A. (2004). Generative modeling for continuous non-linearly embedded visual inference. In *Proceedings of the International Conference on Machine Learning*. Piscataway, NJ: IEEE.
- Svoboda, T., Martinec, D. and Pajdla, T. (2005). A convenient multi-camera self-calibration for virtual environments. *Presence* **14**, 407-422.
- Trimarchi, J. and Schneiderman, A. (1995). Different neural pathways coordinate *Drosophila* flight initiations evoked by visual and olfactory stimuli. *J. Exp. Biol.* **198**, 1099-1104.
- Urtasun, R., Fleet, D. and Fua, P. (2006). 3D people tracking with gaussian process dynamical models. In *IEEE Conference on Computer Vision and Pattern Recognition*, vol. 1, pp. 238-245. Piscataway, NJ: IEEE.
- van der Merwe, R. and Wan, E. (2003). Sigma-point kalman filters for probabilistic inference in dynamic state-space models. In *Proceedings of the Workshop on Advances in Machine Learning*. Piscataway, NJ: IEEE.
- Weis-Fogh, T. (1973). Quick estimates of flight fitness in hovering animals, including novel mechanisms for lift production. *J. Exp. Biol.* **59**, 169-230.
- Zanker, J. M. (1990). The Wing Beat of *Drosophila melanogaster*. I. Kinematics. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **327**, 1-18.