© 2016. Published by The Company of Biologists Ltd | Journal of Experimental Biology (2016) 219, 1608-1617 doi:10.1242/jeb.133751

METHODS & TECHNIQUES

Automated detection of feeding strikes by larval fish using continuous high-speed digital video: a novel method to extract quantitative data from fast, sparse kinematic events

Eyal Shamur[1,‡], Miri Zilka[2,*,‡], Tal Hassner[1], Victor China[3,4], Alex Liberzon[5] and Roi Holzman[3,4,§]

ABSTRACT

Using videography to extract quantitative data on animal movement and kinematics constitutes a major tool in biomechanics and behavioral ecology. Advanced recording technologies now enable acquisition of long video sequences encompassing sparse and unpredictable events. Although such events may be ecologically important, analysis of sparse data can be extremely time-consuming and potentially biased; data quality is often strongly dependent on the training level of the observer and subject to contamination by observer-dependent biases. These constraints often limit our ability to study animal performance and fitness. Using long videos of foraging fish larvae, we provide a framework for the automated detection of prey acquisition strikes, a behavior that is infrequent yet critical for larval survival. We compared the performance of four video descriptors and their combinations against manually identified feeding events. For our data, the best single descriptor provided a classification accuracy of 77–95% and detection accuracy of 88–98%, depending on fish species and size. Using a combination of descriptors improved the accuracy of classification by ~2%, but did not improve detection accuracy. Our results indicate that the effort required by an expert to manually label videos can be greatly reduced to examining only the potential feeding detections in order to filter false detections. Thus, using automated descriptors reduces the amount of manual work needed to identify events of interest from weeks to hours, enabling the assembly of an unbiased large dataset of ecologically relevant behaviors.

KEY WORDS: Automated classification, Feeding kinematics, High-speed video, Machine learning

INTRODUCTION

Quantitative analysis of animal movements constitutes a major tool in understanding the relationship between animal form and function, and how animals perform tasks that affect their chances of survival (Alexander, 1992; Dickinson et al., 2000; Marey, 1874). This discipline benefited greatly when filming technology enabled the freezing of fast movements and determination of the sequence of events that occur when animals move. Stroboscopic filming and multiple cameras, first used in the early 1900s, have evolved to designated 16 mm movie cameras capable of filming at hundreds of frames per second. In the last decades, digital high-speed videography has enabled the collection of detailed kinematics of animal motion. Because of technological and practical limitations such as camera memory and data analysis constraints, analysis is often focused on short video clips, usually <1 s. Commonly, events of interest, such as the movement of animals while jumping, landing or striking prey, are captured on video by manually triggering the camera at the right time, and saving the relevant range within each video sequence. The data are then digitized and analyzed to resolve temporal patterns in the sequence of events, variables such as speed and acceleration, and other quantitative kinematic data. This framework has enabled researchers to understand the mechanistic and behavioral aspects of diverse behaviors such as jumping, flying, running, gliding, feeding and drinking in many animal species (e.g. Altshuler et al., 2004; Holzman et al., 2007; James et al., 2007; Reis et al., 2010; Ribak and Swallow, 2007; Toro et al., 2004 among many others).

Manually triggering the camera to save short sequences is only suitable for events that can be easily identified in real time, are easy to induce, or are repetitive and frequent. For events that do not adhere to these criteria or that are unpredictable in space and time, manual triggering and saving short clips limits the possible scope of research. One example of the latter constraint is suction feeding by larval fish. Newly hatched fish subsist on a limited supply of yolk and thus must encounter and successfully capture food before their energy resources become depleted (Fyhn, 1989; Hunter, 1981). To capture their prey, larval fish swim towards it and then open their mouth while expanding the oral cavity. The expansion of the larvae's mouth generates a strong inward flow of water, and this flow is key to successful suction feeding, drawing the prey into the predator's mouth (Day et al., 2015; Lauder, 1980, 1985; Westneat, 2006). However, the body of a hatchling larva is a few millimeters long, and its mouth diameter is as small as 100 µm. The high magnification optics required to film these minute larvae leads to a small depth-of-field and limited visualized area. Actively swimming larvae remain in the visualized area for only a few seconds. A low feeding rate (especially in the first days post-hatching) results in a scarcity of feeding attempts in the visualized area (Holzman et al., 2015). Similar to adults, prey capture in larvae takes a few tenths of a millisecond (China and Holzman, 2014; Hernandez, 2000; Holzman et al., 2015), easily missed by the naked eye or conventional video.

Recently, continuous high-speed photography of long sequences (~100,000 frames) has shown that the prey capture success rates of early-stage larvae are substantially lower than those of their older counterparts (China and Holzman, 2014; Holzman et al., 2015).

[1]Department of Mathematics and Computer Science, The Open University of Israel, 1 University Road, P.O.B. 808, Raanana 43107, Israel. [2]Department of Physics, Faculty of Exact Sciences, Tel Aviv University, Tel Aviv 69978, Israel. [3]Department of Zoology, Faculty of Life Sciences, Tel Aviv University, Tel Aviv 69978, Israel. [4]The Inter-University Institute for Marine Sciences, POB 469, Eilat 88103, Israel. [5]School of Mechanical Engineering, Faculty of Engineering, Tel Aviv University, Tel Aviv 69978, Israel.
*Present address: Department of Physics, University of Warwick, Coventry CV4 7AL, UK.
‡These authors contributed equally to this work

§Author for correspondence (holzman@post.tau.ac.il)

Received 28 October 2015; Accepted 6 March 2016

List of symbols and abbreviations

ACC	accuracy
AUC	area under the curve
B_k	blob (segment) number k
$f_{desc}(V_k)$	feature descriptor of V_k
$I(k,i)_x$	horizontal gradient of pixel i at blob k
$I(k,i)_y$	vertical gradient of pixel i at blob k
MBH	dense trajectories and motion boundary histogram descriptor
MIP	motion interchange patterns descriptor
RBF	radial basis function
ROC	receiver operator curve
STIP	space–time interest points descriptor
SVM	support vector machine
T_{circ}	minimum allowed eigenvalue ratio that keeps out shapes that are too circular
T_{elong}	maximum allowed eigenvalue ratio that keeps out shapes that are too elongated
T_{max}	maximum number of pixels for a fish blob
T_{min}	minimum number of pixels for a fish blob
TN	true negative
TP	true positive
T_{Sobel}	gradient magnitude threshold
T_{txt}	texture threshold
VIF	violent flows descriptor
V_k	space–time volume number k (short video clip of fish head)
λ_{max}	maximum eigenvalue of a blob
λ_{min}	minimum eigenvalue of a blob
$\nabla B_{k,i}$	gradient of pixel i at blob k
$\|\|\nabla B_{k,i}\|\|$	gradient magnitude

This method was instrumental in testing the hypothesis that the hydrodynamic regime of low Reynolds numbers experienced by small larvae directly impedes the suction feeding mechanism, possibly leading to larval starvation and mortality (China and Holzman, 2014). Although these systematic observations of larval feeding attempts have proven crucial for understanding the feeding process, they were extremely labor intensive, limiting the widespread application of this method in larval fish research. For example, we estimate the data acquisition rate as 0.8–3 strikes h^{-1} (depending on larval age) when using traditional, burst-type high-speed cameras. Using continuous high-speed filming can mitigate some of these shortcomings by providing good spatio-temporal resolution by integrating over several minutes of feeding and thereby increase the probability of observing a prey-capturing strike. Still, the strikes then have to be identified by observing the videos at speeds 30 to 100-fold slower than the recorded speed, a time-consuming task. Our goal was therefore to develop a visualization method by which to computationally characterize rapid, sparse events in a non-intrusive, quantitative and objective way. Specifically, we set out to detect and classify prey-capture strikes from continuous high-speed movies of larval fishes. This procedure provides an unbiased, high-throughput method to measure feeding rates, feeding success, prey selectivity and handling time, as well as swimming speed and strike kinematics.

MATERIALS AND METHODS

Model organisms

We focused on three fish species: *Sparus aurata* Linnaeus 1758 [two age groups: 13 and 23 days post-hatching (dph); Sparidae, Perciformes, Actinopterygii], *Amatitlania nigrofasciata* (Günther 1867) (14–16 dph; Cichlidae, Perciformes, Actinopterygii) and *Hemichromis bimaculatus* Gill 1862 (8–15 dph; Cichlidae, Perciformes, Actinopterygii). *Sparus aurata* is a marine fish of high commercial importance, commonly grown in fisheries, whereas the two cichlid species are freshwater fish that are grown for the pet trade. *Sparus aurata* has a life history that is characteristic of pelagic and coastal fishes, whereas the cichlids provide parental care to their offspring. Thus, the cichlid larvae hatch at a much larger size and are more developed (Table 1).

The experiments described below complied with IACUC approved guidelines for the use and care of animals in research at Tel Aviv University, Israel.

Experimental set-up

During experiments, the larvae were placed in a small rectangular experimental chamber (26×76×5 mm). Depending on fish age and size, five to 20 larvae were placed in the chamber and were allowed several minutes to acclimate before video-recording began. Larval density was adjusted so that at least one larva would be present in the field of view throughout most of the imaging period. Typical feeding sessions lasted 5–10 min. Rotifers (*Brachionus rotundiformis*; ∼160 µm in length) were used as prey for all fish species as they are widely used as the standard first-feeding food in the mariculture industry.

Visualization of feeding larvae was done using a continuous high-speed digital video system (Vieworks VC-4MC-M/C180), operating at 240 frames s^{-1} with a resolution of 2048×1024 pixels (Holzman et al., 2015). The camera was connected to a PC, and controlled by Streampix 5 video acquisition software (Norpix, Montréal, Canada). A 25 mm *f*/1.4 C-mount lens (Avenir CCTV lens, Seiko Optical, Mong Kok, Hong Kong) was mounted on an 8 mm extension tube, providing a field of view of 15×28×3 mm (height×width×depth) at *f*=5.6. We used backlit illumination, using an array of 16 white LEDs (∼280 lumens) with a white plastic diffuser. The original videos were used for our core algorithm to capture every image detail; however, for the pre-processing stage the original videos were rescaled to 1024×512 pixels per frame to increase computation efficiency. This size was empirically determined to accelerate pre-processing computations while having a minimal impact on the final accuracy.

Manual identification of feeding strikes

Following recording, videos were played back at reduced speed (10 frames s^{-1}) to manually identify feeding attempts (Fig. 1). We defined feeding attempts as instances in which the mouth was opened at a time when a prey item was present at a distance of less than one-fifth of a body length in front of the larvae, while it was swimming towards the prey. Feeding attempts can be visually distinguished from breathing based on the size of the mouth opening

Table 1. Life-history traits for the species used in the present study

	Sparus aurata	*Amatitlania nigrofasciata*	*Hemichromis bimaculatus*
Egg diameter at hatching (mm)	∼1	∼1.3	∼1.3
Length of hatched larvae (mm)	3.5	5.0	4.9
Age at filming (days post-hatch)	13, 23	8, 11, 15	8, 14, 16
Length at filming (mm)	4.5, 6.5	5.6–6.1	5.5–5.9
Number of events used for classification	300	∼150	∼150

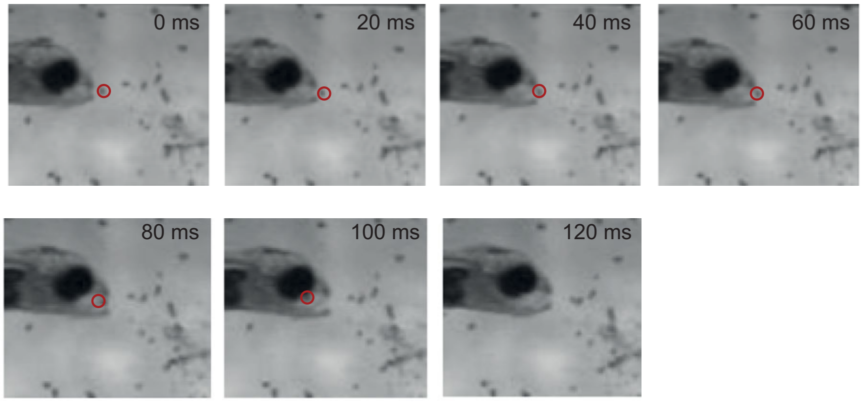

Fig. 1. Extracted spatio-temporal volume in canonical views (horizontal, right-facing views) of a feeding fish. The prey is marked by a red circle, and enters the mouth at 60 ms. The mouth closes at 120 ms.

and the opening speed. During a feeding attempt, the mouth opens fast and wide, typically >70% of its maximal diameter, whereas breathing is characterized by a slower and smaller mouth opening (<30%) (Brainerd and Ferry-Graham, 2006; Westneat, 2006). Overall, we obtained ∼75 feeding events for each of the four groups used in this study (Table 1; two *S. aurata* age groups, *A. nigrofasciata* and *H. bimaculatus*).

Classification of feeding strikes

In addition to the 75 feeding events identified for each group, short clips sampled at random space–time points were used to generate 75 non-feeding events. Each of these non-feeding events was viewed to verify the lack of feeding activity. These 600 clips were used as the underlying database for the machine learning classification algorithms (Table 1). The database was divided into database A, which comprised *A. nigrofasciata* and *H. bimaculatus*, and database B, which comprised the two age groups of *S. aurata*. Each database was analyzed separately.

A diagram describing the detection process of feeding events is provided in Fig. 2. Key to the process was the separation into two stages of the classification process. First, fish detection and pose normalization, i.e. adjusting the frame of view so that the larva

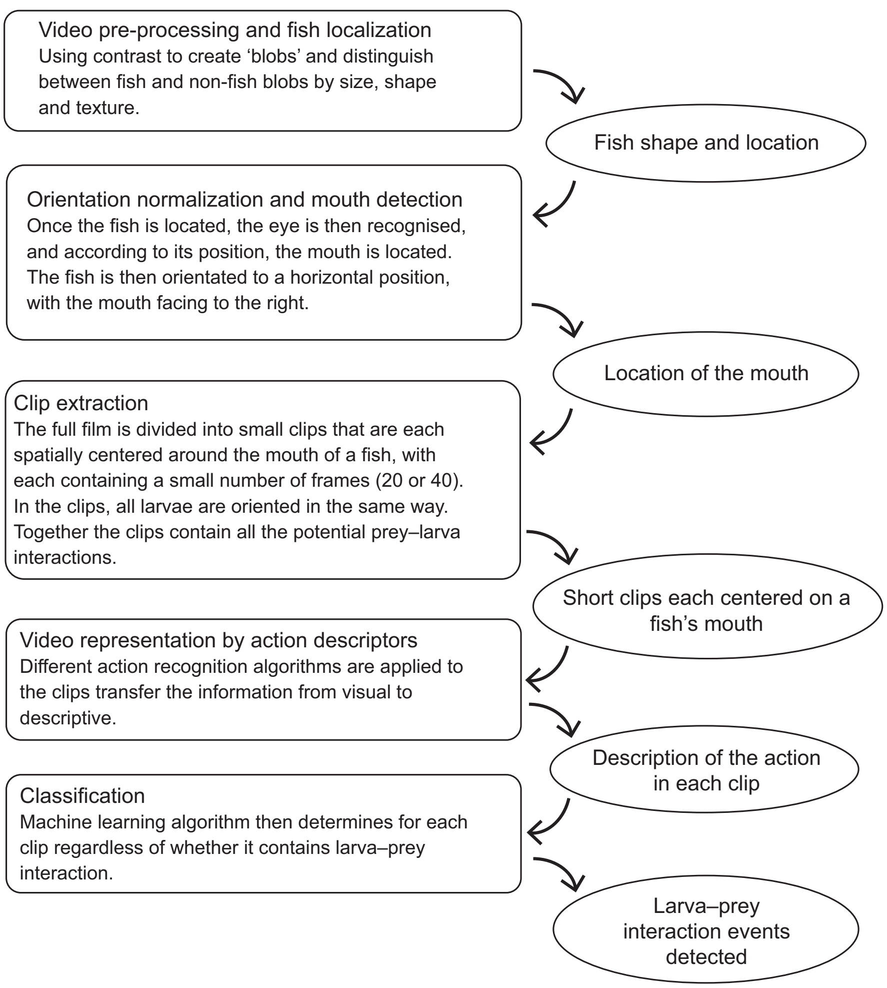

Fig. 2. Five main blocks of the classification algorithm (left) and their outputs (right).

Journal of Experimental Biology (2016) 219, 1608-1617 doi:10.1242/jeb.133751

would always be oriented in the same way. Second, classification of the local spatio-temporal regions, and the determination of either feeding or non-feeding events. We began by pre-processing the entire video to detect individual fish, discriminating between them and their background and other noise and artifacts in the video (Fig. 2; Stage A, below). Following this step, the shape of the detected fish was analyzed to determine the location of its mouth and to rotate it to a roughly horizontal position to provide orientation invariance (Stage B). These steps (detection and mouth localization) used the compressed 1024×512 pixel videos to locate spatio-temporal volumes ('clips') around each mouth. Clips were 21 frames, 121×121 pixels for database A, and 41 frames, 241×241 pixels for database B (~1 body length in both cases). Clips were extracted and represented using robust video descriptors (Stage C), using the original high-resolution 2048×1024 pixel videos. Finally, classification into feeding/non-feeding was performed using a radial basis function (RBF) support vector machine (SVM) classifier (Stage D).

Because of the high ratio between frame rate (240 frames s^{-1}) and the duration of feeding attempts (usually <60 ms), the classification processing did not need to be applied for every frame to reliably identify feeding attempts. We therefore empirically set the system to process 21 frame volumes only every 10th frame for *A. nigrofasciata* and *H. bimaculatus* or 41 frame volumes only every 20th frame for the slower feeding *S. aurata*. Extracted volumes overlapped by 11 and 21 for databases A and B, respectively. Because the duration of our clips was twice as long as the gap between the center frames, no frame was left unprocessed. Each larva was monitored for the entire duration in the field of view with every potential feeding event captured by at least two clips, as the extracted volumes overlapped. In the following sections we describe each of these steps in detail.

Stage A: video pre-processing and fish localization

In our data, typical video frames contained measurement noise, resulting from floating food particles, light/shadow speckles and dirt on the bottom of the chamber (Fig. 3A, Movie 1). Our processing thus began by attempting to remove much of this clutter. We first applied a standard image segmentation technique (Otsu, 1975), which provides a binary separation of the video to foreground/background pixels, used to separate the background from noise and fish blobs (Fig. 3A,B, Fig. S1).

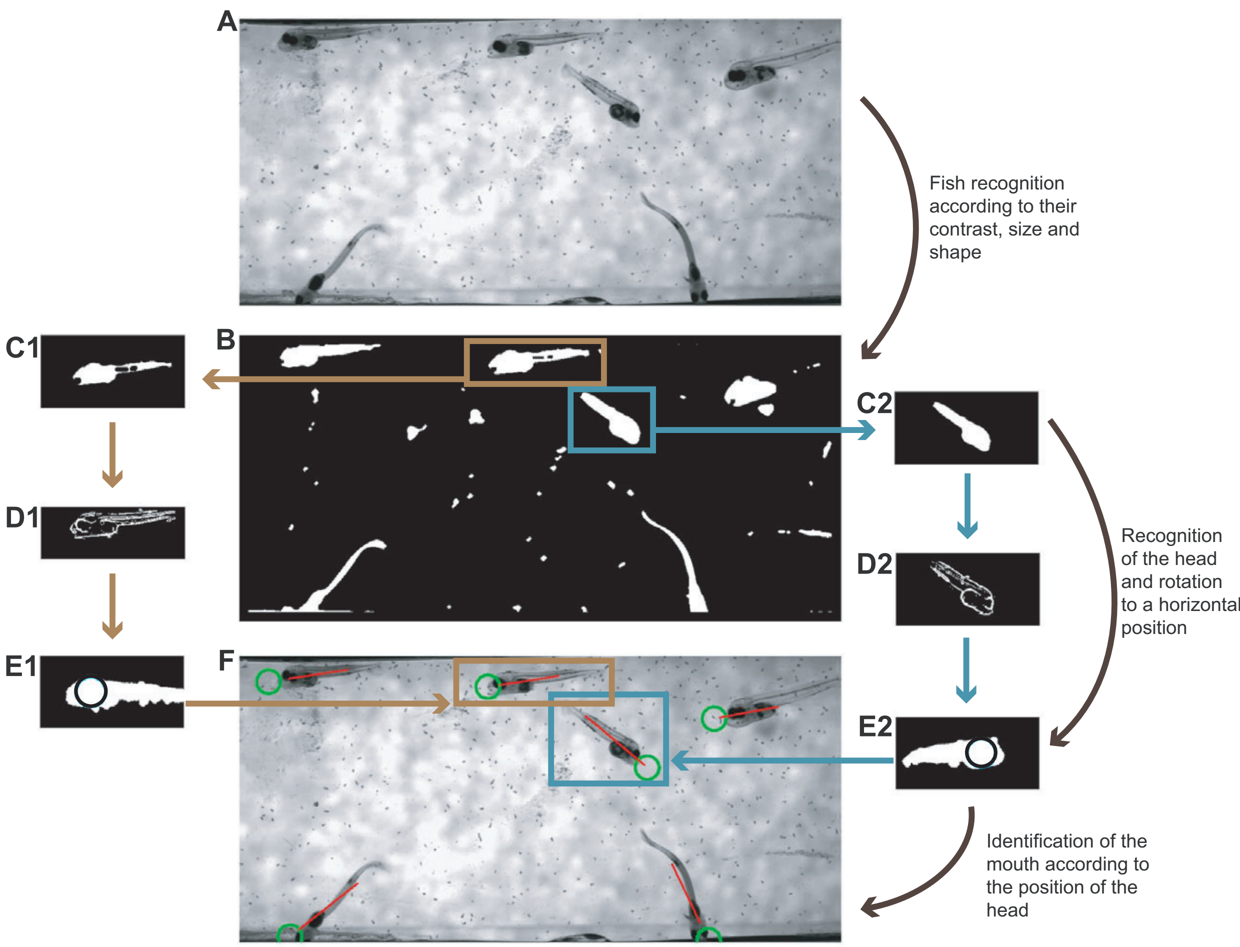

Fig. 3. Video processing to identify fish and determine mouth location (first two stages in Fig. 2, Movie 1). (A) An image is selected from the video (here, 23 days post-hatch *Sparus aurata*). (B) Binary separation of the foreground and background is followed by blob extraction (blue and brown insets in B). (C) Blobs qualified by an eigenvalue ratio test (having appropriate length/width ratios) are maintained, while small blobs are removed. (D) Gradient analysis is used to identify textured elements (fish) from non-textured ones (noise). (E) Pose normalization is applied to the blobs. The fish head is located by examining the radius of the maximum bounded circle. (F) The main axis of the fish body and the head are visualized, and projected onto the original image: green circles point to fish mouths, and red lines represent fish bodies' main (long) axis.

Journal of Experimental Biology (2016) 219, 1608-1617 doi:10.1242/jeb.133751

The fish species in our videos were of similar size and length-to-height (maximal dorso-lateral distance) ratio, making them geometrically different than most of the other shapes in the video. We therefore removed foreground blobs having less than a set minimum threshold number of pixels $T_{\min}$ or having more than a set maximum threshold number of pixels $T_{\max}$. Non fish-shaped blobs were then removed by considering the ratio between the two eigenvalues $\lambda_{\min}$ and $\lambda_{\max}$ of each foreground segment (i.e. the length along the longest and shortest axis of an equivalent ellipsoid). A blob B_k was removed if the following condition did not hold:

$$T_{\mathrm{circ}} < \frac{\lambda_{\min}}{\lambda_{\max}} < T_{\mathrm{elong}}. \quad (1)$$

The value for $T_{\min}$ was set to 350 pixels for 13 dph *S. aurata* and 800 pixels for the other groups. The value for $T_{\max}$ was set to 10,000 pixels. T_{elong} and T_{circ} were set to 100 and 1, respectively, where T_{circ} reflects the minimum allowed eigenvalue ratio that keeps out shapes that are too circular, whereas T_{elong} reflects the maximum allowed eigenvalue ratio that keeps out shapes that are too elongated. These values were determined by experimenting with several arbitrarily selected images, and remained unchanged throughout our experiments.

The above process eliminated most of the non-fish foreground blobs (Fig. 3C), but some blobs may still share the same size or shape of these fish. These blobs were identified by considering the texture within each blob (Fig. 3D, Fig. S1); blobs produced by noise typically present flat appearances compared with the textured fish bodies. Specifically, we evaluated the following expression for each foreground blob:

$$\sum_{i \in B_k} f\left(\| \nabla B_{k,i} \|\right) < T_{\mathrm{txt},} \quad (2)$$

where:

$$f(x) = \begin{cases} x > T_{\mathrm{Sobel}} & 1 \\ \text{else} & 0 \end{cases}. \quad (3)$$

Here, $\| \nabla B_{k,i} \| = \sqrt{I(k,i)_x^2 + I(k,i)_y^2}$, where I_x is the horizontal image gradient and I_y the vertical gradient, both at the *i*th pixel of the *k*th blob and both approximated using standard 3×3 Sobel filters. The values for T_{Sobel} and T_{txt} (the gradient magnitude and texture thresholds, respectively) were set to 120 and 140, respectively, and used throughout our experiments. These steps are visualized in Fig. 3.

Stage B: orientation normalization

As the fish swim freely in their tank, their heads may be oriented in any direction. This is quite different from standard action recognition applications, in which actions are typically performed oriented in the same manner: a video of a human actor walking would typically have the motion of the legs appearing at the bottom of the frame, below the rest of the body. Representations used to capture and discriminate between human actions are therefore not designed to be invariant to the rotational differences exhibited by our fish. Here, this invariance is introduced prior to feature extraction by rotating all fish-head spatio-temporal blobs to a canonical position, in a manner similar to that employed by low-level descriptors such as Scale Invariant Feature Transform (Lowe, 2004). Specifically, at the particular larval developmental stage considered here, the head is substantially bigger than any other part of the fish's anatomy. The head can therefore be detected simply by locating the max-bounded circle of the fish segment. The spatio-temporal volume around each head region is then rotated to align the *x*-axis of the entire fish blob with the frame's horizontal axis (Fig. 3) using standard principal components analysis. Additional invariance to reflection is then introduced by reflecting all spatio-temporal volumes to produce horizontally aligned, right-facing fish.

The two steps of detecting fish mouths and rotating the segments are visualized in Fig. 3 (see also Fig. S1). The result of this stage, mouth detection, is a defined area around each detected mouth (Movie 1). For dataset A (*A. nigrofasciata* and *H. bimaculatus*), we extracted 121×121 pixels centered on the mouth's central pixel for 21 frames, extracted from the compressed 1024×512 pixel video (10 frames before and after the central frame). For dataset B (*S. aurata*), we extracted 241×241 pixels centered on the mouth's central pixel for 41 frames (20 frames before and after the central frame), extracted from the original high-resolution 2048×1024 pixel video (Fig. 1). Extracted clips overlapped by 50% of their length. This choice of spatial dimensions allowed coverage of the entire head along with sufficient margins for possible food floating around the fish. The temporal dimension was empirically determined to be long enough to span feeding. Note that fish could appear in the frame for longer time frames. In such cases, several 41-frame-long clips would be generated and analyzed for each fish (i.e. long sequences were divided with overlapping between divisions, not trimmed).

Stage C: video representation

The pose-normalized video clips produced in the previous step are next converted to robust representations (descriptors), whose function is to represent actions appearing in videos as a set of floating point numbers (in our case 96–512 numbers). Each descriptor is produced by an algorithm that represents (describes) a video clip based on features of the image sequence (e.g. spatial or temporal derivatives or integral across the image sequence). Effectively, going from video to feature descriptor representations (i.e. a set of floating point numbers) allowed us to reduce the dimensionality of the analysis problem at hand. It further allowed us to represent videos in a manner that is invariant to confounding appearance variations (e.g. changes in illumination, imaging noise, etc.) yet varies with relevant appearance variations (e.g. is assigned with different values for feeding versus non-feeding events). In general, three low-level representation schemes have been central in action recognition systems. These are the local descriptors, optical flow and dynamic-texture based representation schemes. Local descriptors locate 'interesting points' in space–time and extract representations only for these points and their immediate surroundings. An entire video is represented by pooling these points in various manners (e.g. by counting how many times different representations appear in a video). Optical flow methods compute the per-pixel flow (the motion at that pixel, from one frame to the next) and represent a video by analyzing this motion. Finally, dynamic-texture-based methods apply low-level, space–time filters to the entire video (to all pixel locations in all frames) and represent videos by statistics of these filter responses. Because detection of larval feeding strikes is an unexplored computer vision problem, we felt it necessary to evaluate all three of these representation schemes (see below).

We used representations that are known and tested algorithms that have been designed to capture and recognize different actions by extracting discriminative information unique to each action, but remain robust to small differences in how each action is performed, the actor performing it, the viewing conditions, and more. We experimented with a number of recent video representations, previously shown to provide excellent action recognition performance, and chose three descriptors – the best performed

Journal of Experimental Biology (2016) 219, 1608-1617 doi:10.1242/jeb.133751

descriptors from each scheme. Thus, each pose-normalized clip of a larva's mouth was encoded using the following action descriptors: (1) space–time interest points (STIP), a local descriptor (Laptev, 2005); (2) motion interchange patterns (MIP), a dynamic-texture based descriptor (Kliper-Gross et al., 2012); (3) dense trajectories and motion boundary histogram (MBH), an optical-flow-based descriptor presented in Wang et al. (2011); and (4) violent flows (VIF), developed to particularly identify violent action (Hassner et al., 2012). The first three have been shown to provide excellent action classification performance on videos of humans performing a wide range of actions. Because feeding strikes could easily be categorized as violent action, it is but natural to check this VIF descriptor as well. All four have been shown in the past to be complementary of each other (e.g. Kliper-Gross et al., 2012). As we later show, combining these representations indeed substantially elevated detection accuracy. We note that there are other, more elaborate methods of comparing video representations (e.g. Kliper-Gross et al., 2011); however, we found their substantial computational overhead to be unnecessary for our purposes.

Stage D: classification

Binary classification of each clip, V_k, as representing either interaction or non-interaction with prey, was performed by first extracting feature descriptors $f_{\mathrm{desc}}(V_k)$, where the subscript 'desc' represents STIP, MIP, MBH or VIF, and then classifying these feature vectors using standard SVMs with RBF kernels (Cortes and Vapnik, 1995). SVM was directly applied to discriminate between descriptors $f_{\mathrm{desc}}(V_k)$ extracted from each clip. In addition, we performed tests with stacking SVM classifiers – of these descriptors – a machine learning paradigm in which multiple learners (of the four descriptors mentioned in our case: MIP, STIP, VIF and MBH) are combined to solve the same problem (classification as feeding or non-feeding). Multiple descriptors were evaluated by stacking SVM classifiers (Wolpert, 1992) as stacking SVM has been proven to outperform the single SVM. Specifically, decision values of SVM classifiers applied separately to each representation were collected in a single vector. These vectors of decision values were then classified using an additional linear SVM.

The final output of our analysis is a list of frame numbers and in-frame *x*–*y* locations, where larva–prey interaction occurs.

Evaluation

We conducted a two-step evaluation of our method. In the first step, we tested the classification scheme, which is the core of our identification method. In the second step, we tested our overall identification method. Classification tests of the first step were conducted to learn and evaluate the classification models while seeking to classify clips as feeding or as non-feeding events. The best models were kept and later used as the classification's core algorithm. Classification tests assess the probability of the classifier to make a correct classification of a clip; the accuracy it reports should be compared with a random guess of whether the clip shows a feeding or non-feeding event, which provides a baseline accuracy of 50% in our benchmark. Classification tests are the standard way to evaluate the performance of a classifier in the computer science machine learning literature (Hassner, 2013; Hassner et al., 2012; Kliper-Gross et al., 2012, 2011). Detection tests were performed in the second step to evaluate the entire pipeline, by testing the detection correctness of feeding/non-feeding events in the original videos. These tests demonstrate how the entire framework performs on a typical use case, where unseen new videos are provided for analysis. It is a different metric, which reflects the ability of the framework to detect relevant instances of an event in a movie. This is the practical implementation of the whole system, because it is related to the quality of results that the end user (who is interested in the organism) would want to evaluate. The detection tests used the models learned previously during the classification tests. Note that these models need to be learned only once, whereas they can be used multiple times. In both classification and detection, our tests were applied separately to the faster feeding fish, *A. nigrofasciata* and *H. bimaculatus*, and to the slower *S. aurata*.

Classification tests

Our classification benchmarks include clips that were extracted using the process described in Fig. 2. We measured binary classification rates for larva–prey interaction versus larva–prey non-interaction events. We then compared our system's performance versus manually labeled ground-truthing. We note that testing the classification in this manner is standard practice in evaluating action recognition systems (Hassner, 2013), particularly when positive events are very rare, as they are here.

This benchmark contains two databases. Database A contained 150 videos of feeding events and 150 videos of non-feeding events of *A. nigrofasciata* and *H. bimaculatus.* Both species have similar morphology and strike kinematics, and consequently were treated collectively in the same database. Database B contained 150 videos of feeding events and 150 videos of non-feeding events of *S. aurata.* We used a leave-one-out, six-fold, cross-validation test protocol. For each set of six clips, we took five as the training set used by the algorithm and employed the sixth event to test the algorithm. Each fold contained 50 video-exclusive clips; that is, a clip only belongs to one fold, thereby preventing biases from crossing over from training to testing. In total, for each of the six tests, the database is divided into two: one division contains 250 volumes and is used to train the SVM classifiers, and the second division contains 50 volumes and is used for testing. In each test division, half of the volumes portray feeding events and half portray non-feeding events.

We report the mean±s.e.m. accuracy (ACC) computed over all six divisions. Here, mean accuracy is the average number of times our system predicted a feeding versus non-feeding event on our sets of volumes. Standard error was measured across the six test divisions. We also provide the overall area under the curve (AUC) – the area under the receiver operator curve (ROC) – as well as the sensitivity (true positive/positives) and specificity (true negative/negative). ROC is a graphical plot that illustrates the performance of a binary classifier system, and the AUC is generally used as a statistic for model comparison (Metz, 1978).

Detection tests

We next measured the rate at which our workflow correctly detected feeding events in videos. Our tests were performed on a video with 6000 frames of *H. bimaculatus*, which included 14 manually labeled feeding events. Our pipeline decomposed this video into a total of 535 potential clips. Separate tests were performed on a video of 4200 frames depicting *S. aurata* larvae. Here, only five feeding events were manually labeled, compared with a total of 451 potential clips automatically extracted by our system.

In our detection tests, reported in the results, we provide the following performance measures for each video: true positive (TP) and true negative (TN), which are the number of times a larva–prey interaction and a larva–prey non-interaction were detected as such, respectively. Accuracy was defined as the percentage of clips correctly detected as either positive or negative. We also provide the confusion matrices for each test, showing the detection rates (in

Journal of Experimental Biology (2016) 219, 1608-1617 doi:10.1242/jeb.133751

Table 2. Breakdown of the time required for each of the components of our system

Step	Time (s)
Per frame	
Compression	0.042
Fish head detection	1.07
Per volume	
Pose normalization (rotation and mirroring)	0.21
STIP encoding	7.35
MIP encoding	7.01
MBH encoding	1.02
VIF encoding	4.01
SVM classification	0.01

All steps of our method were implemented in MATLAB except STIP and MBH encodings and the SVM classification, which were available as (much faster) pre-compiled code. The only element that performs differently in the learning (0.01 s) versus execution (<0.001 s) is the SVM classifier. Manual detection of feeding events took ∼20 min per 10,000 frames for a well-trained individual. MBH, dense trajectories and motion boundary histogram descriptor; MIP, motion interchange patterns descriptor; STIP, space–time interest points descriptor; SVM, support vector machine.

percentages) of predicted positive and negative events versus actual labels for each event. Here too, as with our classification tests, we report performance for all descriptors and their combinations.

Our tests were conducted on a standard Win7, 1 core Intel i7 4770 CPU 64 bit 3.4 GHz processor, 16 GB RAM machine. Table 2 provides a breakdown of the times required for each of the steps in our workflow.

RESULTS

In general, our classification and detection tests demonstrated our ability to automatically classify time–space visual information with fuzzy definitions (Tables 3–6). In terms of efficiency, out of all the action description algorithms incorporated, the major bottleneck is the MIP representation. This is because only a non-optimized MATLAB code exists for this descriptor. As we later show, the accuracy of the two fastest descriptors, MBH and VIF, is nearly as high as the accuracy obtained by combining all descriptors. These two descriptors may therefore be used on their own whenever computational costs must be considered.

Classification benchmark

Our benchmark results for *A. nigrofasciata* and *H. bimaculatus* are presented in Table 3. The highest performance was obtained by combining all the representations, with high accuracy of 92.7±1.4%, high AUC values (0.97; see also Fig. S2), high sensitivity (96.0) and high specificity (89.3). The fastest descriptor, MBH, performed almost as well on its own (ΔACC=1.7; ΔAUC=0.01; Δsensitivity=1.3; Δspecificity=2), making it an attractive option whenever computational resources are limited (Table 3, Fig. S2).

Table 3. Classification benchmark results for *Amatitlania nigrofasciata* and *Hemichromis bimaculatus*

Descriptor type	ACC (%)	AUC	Sensitivity	Specificity
STIP	69.7±3.9	0.81	69.3	70.0
MIP	86.0±2.1	0.93	75.3	66.7
MBH	91.0±1.1	0.98	94.7	87.3
VIF	74.7±2.3	0.78	71.3	78.0
MBH+VIF	91.0±1.2	0.96	94.0	88.0
STIP+MIP+MBH	90.0±2.0	0.97	94.0	86.0
MIP+MBH+VIF	92.0±1.0	0.97	96.0	88.0
STIP+MIP+MBH+VIF	92.7±1.4	0.97	96.0	89.3

Classification tests were conducted to evaluate the classification models while seeking to classify clips as feeding or as non-feeding events. The best models were kept and later used as the classification's core algorithm. Data shown are classification accuracy (ACC; mean±s.e.m.), area under the receiver operating characteristic curve (AUC), and sensitivity and specificity of each of the tested methods. Shaded row indicates the best result. MBH, dense trajectories and motion boundary histogram descriptor; MIP, motion interchange patterns descriptor; STIP, space–time interest points descriptor.

Table 4. Classification benchmark results for *Sparus aurata*

Descriptor type	ACC (%)	AUC	Sensitivity	Specificity
STIP	68.3±2.3	0.75	63.3	74.0
MIP	66.3±1.9	0.77	66.7	66.0
MBH	71.0±2.6	0.77	72.0	70.0
VIF	62.0±1.1	0.66	64.0	60.0
MBH+VIF	70.0±1.1	0.77	70.0	70.0
STIP+MIP+MBH	70.7±2.1	0.81	74.0	67.3
MIP+MBH+VIF	70.7±2.3	0.80	72.0	69.3
STIP+MIP+MBH+VIF	72.7±2.1	0.82	75.3	70.0

Classification tests were conducted to evaluate the classification models while seeking to classify clips as feeding or as non-feeding events. The best models were kept and later used as the classification's core algorithm. Data shown are classification accuracy (ACC; mean±s.e.m.), area under the receiver operating characteristic curve (AUC), and sensitivity and specificity of each of the tested methods. Shaded row indicates the best result. MBH, dense trajectories and motion boundary histogram descriptor; MIP, motion interchange patterns descriptor; STIP, space–time interest points descriptor.

Our benchmark results for *Sparus aurata* are reported in Table 4. These slower-feeding fish were harder to classify, as the differences in the descriptor encodings are more subtle. This was most evident in the VIF descriptor, originally designed to capture fast, violent actions, and which performed much better on the other sets (Table 4). The best performance was obtained by a combination of descriptors with an accuracy of 72.7±2.1%, AUC of 0.81 (see also Fig. S2), sensitivity of 75.3 and specificity of 70.0. Again, the fastest descriptor, MBH, had only marginally inferior performance (ΔACC=1.7; ΔAUC=0.05; Δsensitivity=3.3; Δspecificity=0; Table 3, Fig. S2).

Detection results

Detection results are provided in Table 5 for *H. bimaculatus* and in Table 6 for *S. aurata*. In both cases, MBH was the best representation compared with other representations and even representation combinations. For both cases, our system gave no false positives (upper right cells of confusion matrix) and very low rates of false negatives (lower left cells) of 5% and 25% for *H. bimaculatus* and *S. aurata*, respectively.

Our results indicate that no true larva–prey interaction events were missed, and only a negligible number of false detections (false negatives) are left over to examine and manually filter. The effort required by an expert to manually label videos is estimated at ∼20 min per 10,000 frames (40 s of raw video) for a well-trained individual, depending on the number of larvae in the frame and the number of feeding events. Thus, that effort can be reduced to examining only a few potential feeding detections, a process taking less than 1 min per feeding event.

DISCUSSION

Visualization of larval feeding is challenging because of size, time scale and rarity of feeding events at the early larval stages. However, visualization is essential for measuring the rate of feeding attempts and failed attempts. Here, we present a novel method that can be used to automatically identify and classify prey acquisition strikes in larval fishes, facilitating the acquisition of large datasets from swift,

Journal of Experimental Biology (2016) 219, 1608-1617 doi:10.1242/jeb.133751

Table 5. Detection results for a video of *Hemichromis bimaculatus*

Descriptor		Confusion matrix (%) Predicted feeding	Predicted non-feeding	TP (%)	TN (%)	ACC (%)
STIP	Feeding	100.0	0.0	100	66	83
	Non-feeding	34.2	65.8			
MIP	Feeding	92.8	7.14	93	83	88
	Non-feeding	17.2	82.77			
MBH	Feeding	100.0	0.00	100	95	98
	Non-feeding	5.5	95.0			
VIF	Feeding	92.9	7.1	93	70	81
	Non-feeding	30.3	69.7			
MBH+VIF	Feeding	100.0	0.00	100	91	95
	Non-feeding	9.0	91.0			
STIP+MIP+MBH	Feeding	100.0	0.0	100	86	93
	Non-feeding	13.5	86.5			
MIP+MBH+VIF	Feeding	100.0	0.0	100	89	94
	Non-feeding	11.4	88.6			
STIP+MIP+MBH+VIF	Feeding	100.0	0.0	100	83	92
	Non-feeding	16.6	83.4			

Detection tests evaluate the entire pipeline by evaluating how it performs on unseen new videos, reflecting the ability of the framework to detect a relevant event from a movie. Each row provides detection performance using a different video representation. Results include the confusion matrix for true versus predicted feeding and non-feeding events (shaded cells), true positive rate (TP), true negative rate (TN) and accuracy (ACC). MBH, dense trajectories and motion boundary histogram descriptor; MIP, motion interchange patterns descriptor; STIP, space–time interest points descriptor.

sparse events. This method can be used to facilitate the assessment of feeding rates and success, and to determine the fate of food particles during the feeding cycle. Following automatic identification, detailed kinematic analysis of prey acquisition strikes can be carried out. For example, the spatial resolution and frame rate reported here enable (manual) frame-by-frame digitization of landmarks on the fish's body to extract larval swimming speed during foraging and during prey acquisition strikes, determination of mouth size during prey acquisition strikes, and the distance between prey and predator during the strike (Holzman et al., 2015). Clearly, the frame rate we used (250 frames s^{-1}) may limit the resolution and accuracy of these measurements; however, better (already commercially available) hardware should now allow filming at 500–1000 frames s^{-1} at megapixel resolution for extended time periods and will improve the accuracy of such measurements.

The method we developed combines complex algorithms to classify time–space visual information with fuzzy definitions of the event for the post-manual review by human observers. This approach is therefore not limited to fish, and can be applied to any model system where specific tasks cannot be easily actuated. This could be especially important in studies of natural behaviors in field conditions, or when considering infrequent events. In bats, for example, the movement of the ears is fast and unpredictable, and is of special importance because of the bats' superior localization ability. Researchers have previously used high-speed video to capture this movement (Gao et al., 2011), but have not benefited from automated detection of events. Similarly, the method can be used to analyze interactions between cleaner fish and their clients (Bshary and Grutter, 2002; Bshary and Würth, 2001), which hitherto required laborious processing of videos and may be strongly biased by the subjectivity of the observer. In that system, important

Table 6. Detection results for a video of *Sparus aurata*

Descriptor		Confusion matrix (%) Predicted feeding	Predicted non-feeding	TP (%)	TN (%)	ACC (%)
STIP	Feeding	100.0	0.0	100	63	82
	Non-feeding	37.0	63.0			
MIP	Feeding	100.0	0.0	100	70	85
	Non-feeding	30.0	70.0			
MBH	Feeding	100.0	0.0	100	75	88
	Non-feeding	24.6	75.3			
VIF	Feeding	100.0	0.0	100	60	80
	Non-feeding	39.7	60.3			
MBH+VIF	Feeding	60.0	40.0	60	75	68
	Non-feeding	24.9	75.1			
STIP+MIP+MBH	Feeding	100.0	0.0	100	74	87
	Non-feeding	25.6	74.4			
MIP+MBH+VIF	Feeding	100.0	0.0	100	75	88
	Non-feeding	24.8	75.2			
STIP+MIP+MBH+VIF	Feeding	100.0	0.0	100	74	87
	Non-feeding	25.6	74.4			

Detection tests evaluate the entire pipeline by evaluating how it performs on unseen new videos, reflecting the ability of the framework to detect a relevant event from a movie. Each row provides detection performance using a different video representation. Results include the confusion matrix for true versus predicted feeding and non-feeding events (shaded cells), true positive rate (TP), true negative rate (TN) and accuracy (ACC). MBH, dense trajectories and motion boundary histogram descriptor; MIP, motion interchange patterns descriptor; STIP, space–time interest points descriptor.

parameters such as interaction time, frequency of interactions and identity of the initiator and terminator can be automated and save many human working hours. Our method can also be used for purely physical processes. For example, the resuspension of particles from the bottom by turbulent flows is a strongly stochastic process (Shnapp and Liberzon, 2015; Traugott et al., 2011), and therefore it is impossible to predict where and when particles dislodge from the surface. Yet, an understanding of the physical mechanism that leads to the event of dislodgement requires high spatial and temporal resolution to quantify the fluid field near the particle and solve the component forces that are exerted on it. Thus, it is necessary to visualize the close proximity of the particle and its own motion at high spatial (mm) and temporal (ms) resolution. Traditionally, enormous manual labor is needed to select all the relevant events from the videos that document them (Shnapp and Liberzon, 2015; Traugott et al., 2011). Automatic image processing methods, such as those presented here, can be designed to identify the first moment of particle movement, and mark the event for later processing. A very similar case is the development of a crack in solid surfaces in response to stress (Matsuyama et al., 2010), which is a highly non-linear and unpredictable physical process that should benefit from an automatic marking of the events for the consequent analysis of, for instance, initial crack size, its location and its speed of propagation.

High-speed cameras are a common tool in the study of feeding kinematics (Ferry-Graham et al., 2002; Oufiero et al., 2012; Wainwright and Bellwood, 2002; Wainwright et al., 2007, 2001; Westphal and O'Malley, 2013); they are often used to record short videos (lasting a few seconds) and the analysis is usually focused on feeding kinematics and prey response. Here, we use a digital video-recording system that is geared to collect continuous high-speed videos and facilitate the unbiased identification and isolation of behavioral events in the field of view. Combined with further analysis of strike kinematics performed on the isolated clips, our method will help provide a better understanding of how kinematics affects the larval feeding performance (a possible proxy of fitness). We believe that our approach can advance computational work for the modeling of larval feeding, leading to a better understanding of the specific larval failure mechanisms in the feeding process. Our method can be employed in a wide range of studies on larval feeding: the effect of inter- and intra-specific competition, food preferences and feeding selectivity, prey escape response and predator–prey co-evolution. All of these represent some of the enormous potential our approach can offer. Automatic software identification of feeding attempts will eliminate the current bottleneck when acquiring data. Identifying feeding attempts by means of the human eye is a time-consuming process; by automating this process, we will not only ensure objectivity but also enable data acquisition on a larger scale than obtained to date in the field of larval feeding.

Acknowledgements
We are indebted to T. Elmalich, N. Raab and M. Levi for their help with filming and with analysis of videos, Ardag Ltd for providing *Sparus* larvae for experiments, and N. Paz for editorial assistance.

Competing interests
The authors declare no competing or financial interests.

Author contributions
T.H. and R.H. designed and conceived the research. E.S. and T.H. developed and adapted the machine learning algorithms and computational pipeline. M.Z., V.C. and A.L. developed the high-speed continuous recording, and performed the filming. M.Z., V.C. and R.H. analysed and classified videos. All authors participated in writing the manuscript.

Funding
This study was supported by Israel Science Foundation grants 158/11 and 695/15 to R.H., and by Seventh Framework Programme (FP7) IRC award SFHaBiLF to R.H.

Data availability
The full MATLAB code for the framework is available at https://github.com/EyalShamur/Identification-of-Larval-feeding-strikes, including a short description and guide for this repository, and brief introduction of the code structure and use.

Supplementary information
Supplementary information available online at
http://jeb.biologists.org/lookup/suppl/doi:10.1242/jeb.133751/-/DC1

References

Alexander, R. M. (1992). *Exploring Biomechanics: Animals in Motion*. New York: Scientific American Library.
Altshuler, D. L., Dudley, R. and McGuire, J. A. (2004). Resolution of a paradox: Hummingbird flight at high elevation does not come without a cost. *Proc. Natl. Acad. Sci. USA* **101**, 17731-17736.
Brainerd, E. L. and Ferry-Graham, L. A. (2006). Mechanics of respiratory pumps. *Fish Physiol.* **23**, 1-28.
Bshary, R. and Grutter, A. S. (2002). Asymmetric cheating opportunities and partner control in a cleaner fish mutualism. *Anim. Behav.* **63**, 547-555.
Bshary, R. and Würth, M. (2001). Cleaner fish *Labroides dimidiatus* manipulate client reef fish by providing tactile stimulation. *Proc. R. Soc. B Biol. Sci.* **268**, 1495-1501.
China, V. and Holzman, R. (2014). Hydrodynamic starvation in first-feeding larval fishes. *Proc. Natl. Acad. Sci. USA* **111**, 8083-8088.
Cortes, C. and Vapnik, V. (1995). Support-vector networks. *Mach. Learn.* **20**, 273-297.
Day, S. W., Higham, T. E., Holzman, R. and Van Wassenbergh, S. (2015). Morphology, kinematics, and dynamics: the mechanics of suction feeding in fishes. *Integr. Comp. Biol.* **55**, 21-35.
Dickinson, M. H., Farley, C. T., Full, R. J., Koehl, M. A. R., Kram, R. and Lehman, S. (2000). How animals move: an integrative view. *Science* **288**, 100-106.
Ferry-Graham, L., Wainwright, P., Westneat, M. and Bellwood, D. (2002). Mechanisms of benthic prey capture in wrasses (Labridae). *Mar. Biol.* **141**, 819-830.
Fyhn, H. J. (1989). First feeding of marine fish larvae: are free amino acids the source of energy? *Aquaculture* **80**, 111-120.
Gao, L., Balakrishnan, S., He, W., Yan, Z. and Müller, R. (2011). Ear deformations give bats a physical mechanism for fast adaptation of ultrasonic beam patterns. *Phys. Rev. Lett.* **107**, 214301.
Hassner, T. (2013). A critical review of action recognition benchmarks. In *2013 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 245-250. IEEE.
Hassner, T., Itcher, Y. Kliper-Gross, O. (2012). Violent flows: real-time detection of violent crowd behavior. In *2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 1-6. IEEE.
Hernandez, L. P. (2000). Intraspecific scaling of feeding mechanics in an ontogenetic series of zebrafish, Danio rerio. *J. Exp. Biol.* **203**, 3033-3043.
Holzman, R., Day, S. W. and Wainwright, P. C. (2007). Timing is everything: coordination of strike kinematics affects the force exerted by suction feeding fish on attached prey. *J. Exp. Biol.* **210**, 3328-3336.
Holzman, R., China, V., Yaniv, S. and Zilka, M. (2015). Hydrodynamic constraints of suction feeding in low Reynolds numbers, and the critical period of larval fishes. *Integr. Comp. Biol.* **55**, 48-61.
Hunter, J. P. (1981). Feeding ecology and predation of marine fish larvae. In *Marine Fish Larvae: Morphology, Ecology, and Relation to Fisheries* (ed. R. Lasker), pp. 33-77. Seattle, WA: University of Washington Press.
James, R. S., Navas, C. A. and Herrel, A. (2007). How important are skeletal muscle mechanics in setting limits on jumping performance? *J. Exp. Biol.* **210**, 923-933.
Kliper-Gross, O., Hassner, T. and Wolf, L. (2011). One shot similarity metric learning for action recognition. In *Similarity-Based Pattern Recognition* (ed. M. Pelillo and E. R. Hancock), pp. 31-45. Springer.
Kliper-Gross, O., Gurovich, Y., Hassner, T. and Wolf, L. (2012). Motion interchange patterns for action recognition in unconstrained videos. In *Computer Vision–ECCV 2012* (ed. A. Fitzgibbon, S. Lazebnik, P. Perona, Y. Sato and C. Schmid), pp. 256-269. Springer.
Laptev, I. (2005). On space-time interest points. *Int. J. Comput. Vis.* **64**, 107-123.
Lauder, G. V. (1980). Hydrodynamics of prey capture in teleost fishes. In *Biofluid Mechanics*, Vol. II (ed. D. Schenck), pp. 161-181. New York: Plenum Press.
Lauder, G. V. (1985). Aquatic feeding in lower vertebrates. In *Functional Vertebrate Morphology* (ed. M. Hildebrand, D. M. Bramble, K. F. Liem and D. B. Wake), pp. 210-229. Cambridge, MA: Harvard University Press.
Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vis.* **60**, 91-110.

Journal of Experimental Biology (2016) 219, 1608-1617 doi:10.1242/jeb.133751

Marey, E. J. (1874). *Animal Mechanics: a Treatise on Terrestrial and Aerial Locomotion*. Henry S. King & Company.

Matsuyama, K., Yamada, M. and Ohtsu, M. (2010). On-site measurement of delamination and surface crack in concrete structure by visualized NDT. *Constr. Build. Mater.* **24**, 2381-2387.

Metz, C. E. (1978). Basic principles of ROC analysis. *Semin. Nucl. Med.* **8**, 283-298.

Otsu, N. (1975). A threshold selection method from gray-level histograms. *Automatica* **11**, 23-27.

Oufiero, C. E., Holzman, R. A., Young, F. A. and Wainwright, P. C. (2012). New insights from serranid fishes on the role of trade-offs in suction-feeding diversification. *J. Exp. Biol.* **215**, 3845-3855.

Reis, P. M., Jung, S., Aristoff, J. M. and Stocker, R. (2010). How cats lap: water uptake by Felis catus. *Science* **330**, 1231-1234.

Ribak, G. and Swallow, J. G. (2007). Free flight maneuvers of stalk-eyed flies: do eye-stalks affect aerial turning behavior? *J. Comp. Physiol. A* **193**, 1065-1079.

Shnapp, R. and Liberzon, A. (2015). A comparative study and a mechanistic picture of resuspension of large particles from rough and smooth surfaces in vortex-like fluid flows. *Chem. Eng. Sci.* **131**, 129-137.

Toro, E., Herrel, A. and Irschick, D. (2004). The evolution of jumping performance in Caribbean *Anolis* lizards: solutions to biomechanical trade-offs. *Am. Nat.* **163**, 844-856.

Traugott, H., Hayse, T. and Liberzon, A. (2011). Resuspension of particles in an oscillating grid turbulent flow using PIV and 3D-PTV. *J. Phys. Conf. Ser.* 318, 052021.

Wainwright, P. C. and Bellwood, D. R. (2002). Ecomorphology of feeding in coral reef fishes. In *Coral Reef Fishes. Dynamics and Diversity in a Complex Ecosystem* (ed. P. F. Sale), pp. 33-55. San Diego, CA: Academic Press.

Wainwright, P. C., Ferry-Graham, L. A., Waltzek, T. B., Carroll, A. M., Hulsey, C. D. and Grubich, J. R. (2001). Evaluating the use of ram and suction during prey capture by cichlid fishes. *J. Exp. Biol.* **204**, 3039-3051.

Wainwright, P. C., Carroll, A. M., Collar, D. C., Day, S. W., Higham, T. E. and Holzman, R. (2007). Suction feeding mechanics, performance, and diversity in fishes. *Integr. Comp. Biol.* **47**, 96-106.

Wang, H., Kläser, A., Schmid, C. and Liu, C.-L. (2011). Action recognition by dense trajectories. In *2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3169-3176. IEEE.

Westneat, M. W. (2006). Skull biomechanics and suction feeding in fishes. In *Fish Biomechanics* (ed. G. V. Lauder and R. E. Shadwick), pp. 29-75. San Diego, CA: Elsevier Academic Press.

Westphal, R. E. and O'Malley, D. M. (2013). Fusion of locomotor maneuvers, and improving sensory capabilities, give rise to the flexible homing strikes of juvenile zebrafish. *Front. Neural Circuits* **7**, 108.

Wolpert, D. H. (1992). Stacked generalization. *Neural Netw.* **5**, 241-259.