IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received October 20, 2019, accepted November 2, 2019, date of publication November 11, 2019,
date of current version November 21, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2952598

Tracking Multiple Indistinguishable and Deformable Objects Based on Multi-Anchor Flow With Annular Sector Model

BIAO GUO AND YUESHENG ZHU, (Senior Member, IEEE)
Laboratory of Communication and Information Security, Peking University Shenzhen Graduate School, Shenzhen 518055, China

Corresponding author: Yuesheng Zhu (zhuys@pku.edu.cn)

This work was supported in part by the Grants from the NSFC-Shenzhen Robot Jointed Founding under Grant U1613215, in part by the Shenzhen Municipal Development and Reform Commission Disciplinary Development Program for Data Science and Intelligent Computing, and in part by the Shenzhen International Cooperative Research Projects under Grant GJHZ20170313150021171.

ABSTRACT Most of current multi-object tracking (MOT) methods solve the task of identity assignment mainly by using the distinguishable features and predictable motions of nearly rigid objects. However, when the objects are indistinguishable, non-rigidly deformable and erratically motorial, it is challenging to differentiate the identities and regions of multiple objects in occlusions. In this paper, a novel method based on multi-anchor flow with annular sector model (ASM) is proposed to handle occlusions and complex motions for tracking multiple of these kinds of objects. The multi-anchor flow is built on anchor points which are extracted from the union contours of deformable objects. By optimizing the multi-anchor flow over frames, the proposed method solves the task of identity assignment when multiple objects are in clustered. Meanwhile, the annular sector model is also combined into the tracking method not only to rectify the results of identity assignment, but also to depict the poses or regions of deformable objects in occlusions. In our experiments, the proposed method is evaluated on two publicly challenging dataset for tracking multiple *Drosophila* Larvae. The results demonstrate that the proposed method has better performance on both of the accuracy of identity assignment and the computational time when compared to other multi-object tracking algorithms.

INDEX TERMS Multi-object tracking, larvae tracking, annular sector model, multi-anchor flow, indistinguishable objects tracking.

I. INTRODUCTION

Multi-object tracking (MOT) plays an important and fundamental role in wide range of computer vision applications such as monitoring pedestrians [1], [2] or vehicles [3], acquiring action data of human body [4], and analyzing the behaviors of laboratorial animals [5]–[7], etc. The general MOT problem is to provide the trajectories and identities of multiple objects in image sequences. Most of current tracking-by-detection approaches focus on solving the data association (the identity assignment of observations to object tracks) problem by using the distinguishable features [8]–[10] and/or the predictable trajectories [11], [12] to identify the labels of nearly-rigidly tracked objects, for example disambiguating pedestrians or cars in different colors or shapes, and/or supposing that they moving by linear motion models.

However, when the tracked objects are tiny animals, such as *Drosophila* larvae [13], *Zebrafish* [14], or *Caenorhabditis elegans* [15], the objects are always poorly distinguishable from each other, non-rigidly deformable, and can exhibit an erratic motion. In this situation, it is challenging to track and disambiguate the identity of multiple objects, especially when they can heavily mutually occlude for several frames.

Also, estimating the state for each object is important for some applications, such as recognizing social behavior of animals [7]. In this case, getting the poses and positions of animals is helpful to extract features for analyzing social behavior when they interact with each other. As these animals are visually indistinguishable and deformable, tracking the states of those in occlusions is a difficult problem, even in controlled laboratory conditions [5].

The associate editor coordinating the review of this manuscript and approving it for publication was Li He.

In this paper, inspired by the work [13], a novel method based on multi-anchor flow is proposed to handle the occluding fragments for multiple objects. The principle difference between our approach and the work [13] is that the multi-anchor flow proposed in this paper is quite different from the latent mass flow used in [13]. The multi-anchor flow is built based on the anchor points which are sampled from contours of objects, and can describe the rough region of each object when multiple deformable objects move in cluster for several frames. The cost of multi-anchor flow in this paper is acting on limited number of points and is minimized by Hungarian algorithm, when the cost of mass flow in the work [13] is built on pixel-level features and is formulated to a convex energy function model. Also, instead of using multiple identity interpretations [13], the joint states of anchors are applied to indicate the identity of each single object when it moves out from clusters. Therefore, the proposed method can vastly accelerate the tracking speed.

Moreover, in order to depict the state of each deformable object in occlusions and improve the accuracy of identity assignment, the annular sector model (ASM) which is proposed in our previous work [19] is combined into our method. The difference between our approach and the work [19] is that the state of ASM for each object is fitted by the distribution of anchor points in this proposed method, when the parameter of ASM in the work [19] is optimized by minimizing the formulated energy function. Besides, the labels of anchor points in each frame can be refined by the fitted ASM and it can finally reduce identity switches for the whole tracking task.

In this paper, the developed approach is used in the task for tracking multiple *Drosophila* larvae, a popular model organism in biology, by requiring visual data from a single camera. The larvae move on a well plate, and may touch or occlude others over several frames. In these controlled laboratory settings, detection and tracking of isolated individuals is easily to obtain by extracting the foreground of larvae. The main task is the identification and tracking of the larvae during they are touching or occluding.

The main contributions in this paper are:

- The multi-anchor flow is proposed to solve the identity assignment for tracking multiple indistinguishable and deformable objects.
- The ASM is combined to depict the state of each deformable object in occlusions and improve the accuracy of identity assignment.
- The proposed method can track the locations and poses of multiple indistinguishable and deformable objects in occlusions simultaneously in near real time. Experiments on two publicly *Drosophila* larvae datasets demonstrated the robustness and good performance of the proposed method on both of identity assignment and computational time.

The content of this paper is organized as follows. Section II presents the related work that applies to multiple animals tracking. Section III describes the proposed method. The results and discussions of experiments are performed in Section IV. Finally, a conclusion is given in Section V.

II. RELATED WORK

In the past several decades, most of works for MOT in computer vision focus on pedestrian and vehicle tracking. State of the art approaches often solve the identity assignment of multiple objects in occlusions by using the factors that the target has distinguishable features from each other (such as different colors or shapes [8]–[10], including deep features [20], [22], [23]), and/or supposing that they moving in predictable trajectories [11], [12], [21]. Meanwhile, in most of works the objects are often regarded as nearly rigid targets, and are commonly tracked with the locations when they are in clustered, ignoring the region or the pose of each object.

Tracking multiple animals from a single view is also an essential task in the field of complex animal behavior studies. However, there are several challenging factors that current works are not enough for this task. Firstly, the tracked animals in common species, especially the tiny animals such as the *Drosophila* larvae [13], the embryos of *Zebrafish* [14], or the *Caenorhabditis elegans* [15], are always difficult to be disambiguated from each other by the appearance features. Then, the shapes of some kinds of animals always vary more than affine deformations and the animals always exhibit erratic motions, thus the state of object is not easily to describe and the trajectories are hard to be predicted. Meanwhile, when animals are in contact with each other, obtaining their regions or poses is also important for analyzing the social behavior of animals. Finally, it demands more computational efficiency for the tracking method because of the vast volume of vision data for animals in practical applications.

In biology domain, in order to obtain reliable animal tracking results, current works usually leverage manual or natural measures to mark individuals in cluster to distinguish different objects. Mersch *et al.* [24] apply barcode-based identifiers for long tracking ants, Benjamin *et al.* [18] use fluorescent proteins to mark the cells of animals, and Hong *et al.* [7] obtain poses of mice in different coat color. Shemesh *et al.* [25] and Lorbach *et al.* [26] build their dataset by dyeing the fur of mice with different colors. However, these methods are limited in the number of unique markings individuals, and these expensive or complex measures are not robust for automatically analyzing the social behavior of animals.

Unmarked methods are also developed for tracking multiple animals. A recent review of open-source worm like object tracker is given in the work [15], but the reference trackers simply terminate on collisions and reinitialize afterwards. The idTracker [27] method is proposed to provide a long object tracker and solve the data association problem by extracting a characteristic fingerprint from each animal in a video recording of a group. But this approach is easily influenced by the illumination variation and also does not handle the occluded objects. The improved version

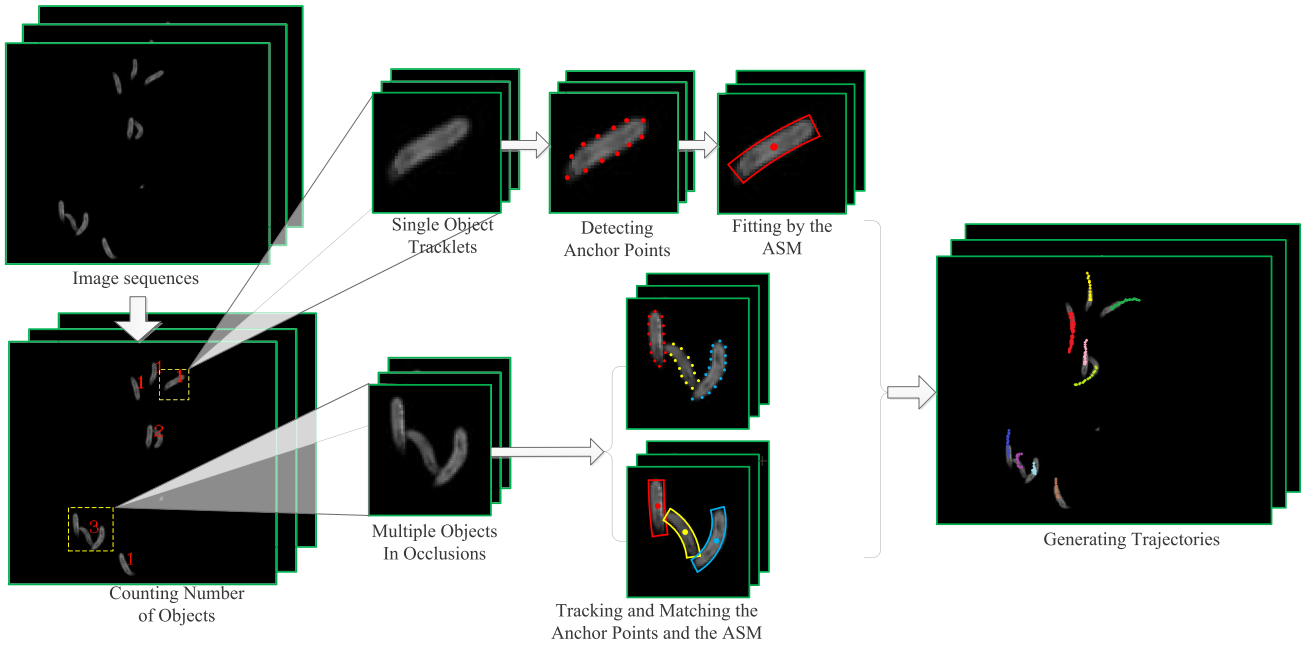


FIGURE 1. The framework of the tracking method: the tracking approach breaks the entire video into non-occluded parts and occlusion parts, and then processes all parts respectively.

idTracker.ai [28] is proposed to eliminate the illumination influence by using two convolutional networks to detect touch animals and to identify the animals. However, this method is prone to handle occlusions for nearly rigid animals or part of animals with time-consuming learning from single objects. Wählby *et al.* [29] provide a method for segmenting overlapped worms, but it depends on the skeletons of objects which are easily changed by the noises.

There are also some works that explicitly tx rack and disambiguate the identity and region/shape of animals in occlusions simultaneously. In the work [30], the contours of mice are tracked by using particle filters. But this approach is limited to matching the affine transformations of some manually designed shape templates. Particle filter is also used in approximating inference of ant tracking [14] and in this approach the objects are fitted by ellipse models. Simpler and independent constant velocity models are implemented in Ctrax [31], although this approach limits to handle slight occlusion. For tracking multiple *Drosophila* larvae, the work [13] provides to leverage a structured supervised learning [32] formulation to automatically find the optimized parameters of their energy function. But the complexity of this approach explodes with the number of objects and frames because of the latent variables are combined with the intensity flows of pixels. Michels *et al.* [33] propose an elaborate multi-circle model to track multiple *Drosophila* larvae based on a reversible jump Markov Chain Monte Carlo (RJMCMC) method, when the parameters of the model should be designed carefully. In front work [19], we propose an annular sector model to describe the states of larvae in each frame, and solving the identity assignment by optimizing a formulated energy function. However, the computational efficiency of this work should be further improved and the drifting problem in some cases should be solved.

In this paper, our tracking method is inspired by the work [13] that optimizing the flows to solve the identity assignment of *Drosophila* larvae in occlusions. Instead of using the pixel-level latent mass flow [13], this method proposes multi-anchor flows which are formed from multiple extracted anchor points to track multiple larvae frame-by-frame. Also, unlike the work [13] that focuses only on the identity assignment task, this method combined the ASM which is proposed in [19] to describe the state of larva and track the region and pose of each larva in cluster simultaneously.

III. METHOD

In this section, we firstly introduce the framework of our tracking method. Fig.1 illustrates the ideas of the tracking framework for multiple larvae in image sequences. The image sequences are firstly preprocessed by subtracting the background and counting the number of larvae for each connect component. Then the components over frames are associated as tracklets by their relations on spatial and temporal context. The state of single object is described by using the ASM which fits the detected anchor points. In order to generate their trajectories, the challenging problem is how to track multiple larvae and obtain their regions or poses when they move to touch or overlap each other, especially that they are visually indistinguishable and deformable.

In this paper, the process that two or more larvae move into clusters and split again is called an encounter. The primary task in our approach is to solve the identity assignment of multiple larvae and track their states in encounters. The main

of our thought to solve this task is based on the assumption that the larva moves at a low velocity and the pose of each one changes continuously frame-by-frame. Thus, the observed image intensities during mutual occlusion can be imaged to flow smoothly like water through the ''channel'' which constructed by the regions of occluded larvae in adjacent frames. Fiaschi *et al.* [13] modeled this process as the movement of mass flows inspired by the literature on Earth Mover Distance (EMD) [34]. However, as the used flows were constructed on the pixel-level features, the complexity of their work is high although their formulated energy can be optimized by linear programming.

In fact, the larvae are always tracked by human-eye only through several specific contour points, e.g. the head/tail point, some side points, etc. Therefore, choosing these specific points to represent the characteristics of the observed larvae can be used to replace the pixel-level features. These specific points are called ''anchor points'' in this paper. In intuition, if the anchor points are initialized to be assigned labels as the identities of their related larvae in the first frame, determining the identities of anchor points frame by frame can finally solve the larval identity assignment. This process can be simplified as building relationships of anchor points between two adjacent frames, and can be regarded as the flows of anchor points. The optimized multi-anchor flow can be obtained through minimizing cost energy function of a bipartite graph.

When extracting the anchor points and establishing the cost function of anchor flows, we consider it on four aspects. Firstly, the anchor points should be able to be extracted from the region of larvae as much easily as possible. Secondly, the anchor points can be differentiated from each other by some features despite of the indistinguishability of larvae. Thirdly, the cost function should be sufficiently expressive to allow assigning energy as much low as possible when the anchor points are correctly connected between two adjacent frames. Finally, the process should also consider the fact that the anchor flows may be terminated or emerged in certain frames.

A. EXTRACTING ANCHOR POINTS

When two or more larvae move into clustered, the human-eye always track these indistinguishable objects by primarily focusing on the larval head or tail continuously over time, or by gazing the larval side edges when the head and tail are covered. Therefore, in our method, we extract the head and tail points and sample several side edge points as the anchor points.

Similar to the work [35], the larval head and tail points are extracted by calculating the curvature values for each point p on the contour of foreground region using the first pass of the IPAN algorithm [36]. In our method, the difference from the work [35] is that the larval head and tail points don't need to be distinguished from each other, both of which are treated as the sharp points. Also, the sharp point is limited to be extracted from the convex part of the larval contour,

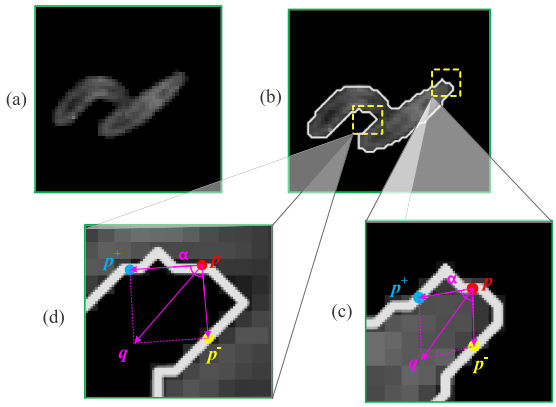


FIGURE 2. Illustration of an example for calculating the curvature values of sharp points on the contour of two occluded larvae: (a) raw images of two occluded larvae, (b) contour of larvae, (c) convex part (head or tail) of larvae, (d) concave part (cross section) of larvae.

excluding the concave part. Because that when two or more larvae move into clustered, the contour points near the cross section would be concave and sharp despite that they are not the larval head or tail points. We illustrate an example that calculating the curvature values of sharp points on the contour of two occluded larvae in Fig.2. In Fig.2(b), the two yellow dotted boxes indicate a concave contour section and a convex contour section respectively. The curvature for point p is defined as:

$$\angle p = \begin{cases} \alpha + \pi, & \text{if } I(\overrightarrow{pq}) < 0.5 \\ \alpha, & \text{otherwise,} \end{cases} \tag{1}$$

where α is the angle between two vectors $\overrightarrow{pp^-}$ and $\overrightarrow{pp^+}$, and $\overrightarrow{pq} = \overrightarrow{pp^-} + \overrightarrow{pp^+}$. The points p^- and p^+ locate on the contour near the point p by using a sliding window approach, as shown in Fig.2(c) and Fig.2(d). And $I(\overrightarrow{pq})$ is the proportion of foreground pixels on the line segment $\overrightarrow{pq}$. Fig.2(c) shows the convex contour section that the point q is located in the foreground area, while the point q for the convex contour section locates outside the foreground area, as shown in Fig.2(d). In our method, the center points of contour sections whose curvatures $\angle p$ are less than the threshold value τ are defined as sharp points.

After extracting the sharp points from the larval region, we sample several edge points from the contour between each two sharp points by fixed interval. These edge points and the sharp points are combined to compose the anchor points.

B. MULTI-ANCHOR FLOW

The multi-anchor flow is built on the anchor points which are extracted from the contour of larval region. After detecting the anchor points, we track them across frames by optimizing the multi-anchor flows. This process is realized by matching a bipartite graph formulated by a set of label nodes and a set of observation nodes. We assume that there are N targets in an encounter, where N is fixed, and that the continuous movement time is T. The labels of anchor points for single

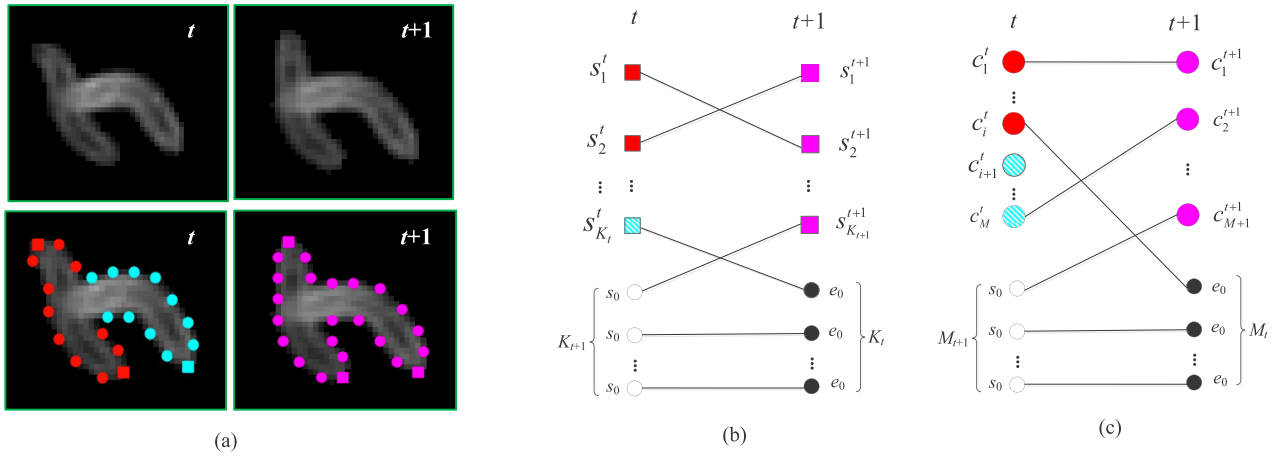


FIGURE 3. Extracting anchor points and generating multi-anchor flows in two adjacent frames: (a) the extracted anchor points in frame t and $t+1$ (the labels of anchor points in frame t are assigned), (b) the bipartite graph of sharp points, and (c) the bipartite graph of edge anchor points between frame t and $t+1$.

larva are initialized in the first frame of encounters. The task is to match the anchor points in frame t with the anchor points in frame $t+1$.

In Fig.3, we illustrate an example for tracking two larvae in clustered between two adjacent frames. The anchor points are extracted from the larval regions in both of frames t and $t+1$, which are labeled by larval number in frame t. As shown in Fig.3(a), the red and cyan points have different labels and it represents that they belong to different larvae. And the magenta points in frame $t+1$ have not been assigned labels yet. We assume that there are K_t sharp points and M_t edge anchor points extracted in frame t. Then, the anchor points in frame t and frame $t+1$ are matched by two stages: i) matching the sharp points by using their distances, and ii) matching the rest of anchor points by using the spatial distances and their relative orders.

Extracting the sharp points set $\mathbf{s}^t = \{s_1^t, s_2^t, \ldots, s_{K_t}^t\}$ from the frame t and set $\mathbf{s}^{t+1} = \{s_1^{t+1}, s_2^{t+1}, \ldots, s_{K_{t+1}}^{t+1}\}$, the sharp bipartite graph is formed as Fig.3(b). We add K_{t+1} source nodes s_0 and K_t terminal nodes e_0 into the bipartite graph. And the cost energy function of sharp points is formulated as follows:

$$E_{sharp}(\Phi) = \sum_{ij} \phi_{ij} D_{ij},$$
$$\text{where} \quad i \in \mathbf{s}^t \cup \{s_0\}^{K_{t+1}}, \quad j \in \mathbf{s}^{t+1} \cup \{e_0\}^{K_t}, \tag{2}$$

in which ϕ is the anchor flow and is defined as:

$$\phi_{ij} = \begin{cases} 1 & \text{if point } i \text{ is linked by point } j, \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

In addition, the added constraints that $\sum_j \phi_{ij} = 1$ and $\sum_i \phi_{ij} = 1$. D_{ij} is the pairwise transition score between point i and point j, which denotes the negative log-likelihood that point j should be linked with point i, and is defined as follows:

$$D_{ij} = \begin{cases} C & \text{if } i = s_0 \text{ or } j = e_0, \\ -\lambda \log \mathcal{N}(d_{ij}; u_1, \sigma_1) & \text{otherwise,} \end{cases} \tag{4}$$

where $\mathcal{N}$ is the Guassian function with the expected value u_1 and the standard deviation σ_1, and d_{ij} is the Euclid distance between point i and point j. λ is a scale number for the negative log-likelihood, and the range of its value is (0, 1). C is a constant number for the situations that the flow is started or terminated. We use the Hungarian algorithm to minimize the energy cost function E_{sharp} in (2), and the sharp anchor flow set can be estimated as $\hat{\Phi}_{sharp}$. If the frame rate is not too low and the pairwise transition score of the same sharp point in two adjacent frames is not more than C, the sharp bipartite graph can be matched well and the anchor flows can work well.

Similarity, the edge anchor bipartite graph is formulated by the extracted edge points set $\mathbf{c}^t = \{c_1^t, c_2^t, \ldots, c_{M_t}^t\}$ in frame t and set $\mathbf{c}^{t+1} = \{c_1^{t+1}, c_2^{t+1}, \ldots, c_{M_{t+1}}^{t+1}\}$ in frame $t+1$, as shown in Fig.3(c). Also, M_{t+1} source nodes s_0 and M_t terminal nodes e_0 are also added into the graph. Then the cost energy function of edge anchor points is defined as:

$$E_{edge}(\Phi) = \sum_{ij} \phi_{ij} F_{ij},$$
$$\text{where } i \in \mathbf{c}^t \cup \{s_0\}^{M_{t+1}}, j \in \mathbf{c}^{t+1} \cup \{e_0\}^{M_t}, \tag{5}$$

in which F_{ij} is pairwise transition score similar to D_{ij}, such that:

$$F_{ij} = \begin{cases} C & \text{if } i = s_0 or j = e_0, \\ -\lambda \log \mathcal{N}(d_{ij}; u_2, \sigma_2) & \text{if } RO(i,j) \geq 1 \\ -\log \mathcal{N}(d_{ij}; u_2, \sigma_2) & \text{otherwise,} \end{cases} \tag{6}$$

where the $RO(i, j)$ is the relative order relationship function between the point i and point j, and is formulated as:

$$RO(i, j) = G(l_i, l_j) + G(r_i, r_j),$$

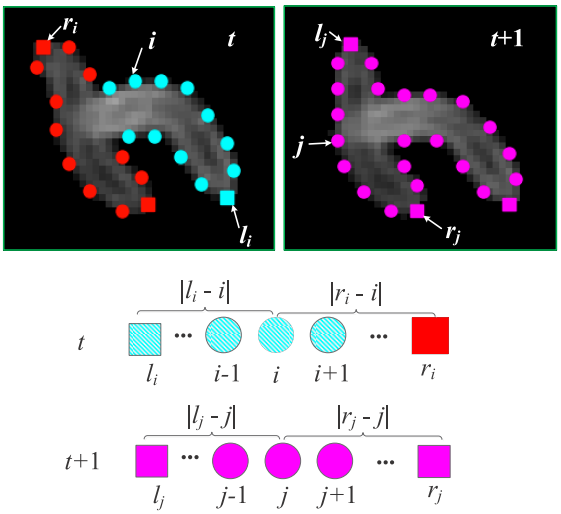


FIGURE 4. An example for calculating the relative order distance.

$$\text{where } G(p_i, q_j) = \begin{cases} 0 & \text{if } |p_i - i| \neq |q_j - j|, \\ \hat{\Phi}_{sharp}(p_i, q_j) & \text{otherwise,} \end{cases} \tag{7}$$

in which l_i and r_i are the left and right sharp points adjacent to point i, because that all of the anchor points for one component can be linked as a circle. Here, $|\bullet|$ is the relative order distance, which is defined as the number of interval anchor points. We illustrate an example in Fig.4 for calculating the relative order distance. In this example, although the relative order distance $|r_i - i|$ equals to $|r_j - j|$ (both of them are 4), the $RO(i, j)$ should be 0 because that the flow $\hat{\Phi}_{sharp}(r_i, r_j)$ is 0. This means that the relative positions of point i and the point j are different. Thus, their pairwise cost value should be more than which have same relative positions.

To match the edge bipartite graph, we also use the Hungarian algorithm to minimize the cost energy function E_{edge} in (5). And the edge anchor flow set can be obtained as $\hat{\Phi}_{edge}$. The multi-anchor flow set is composed by the sharp anchor flow set and the edge anchor flow set, which is:

$$\hat{\Phi} = \hat{\Phi}_{sharp} \cup \hat{\Phi}_{edge}. \tag{8}$$

By estimating the multi-anchor flows frame-by-frame, the identity assignment can be ultimately solved by jointly analyzing the labels of anchor points in the last frame of an encounter. However, errors may occur and accumulate over time because of this locally optimizing process. Thus, the assigning labels need to be revised in due course to ensure the correction of ultimate larval identities matching. Moreover, the location and pose for each larva also should be tracked when two or more get in clustered. For solving these two problems, the ASM is used to describe the state of larva and rectify the labels of anchor points.

C. ANNULAR SECTOR MODEL (ASM)

When tracking larvae by using multi-anchor flows, the junction of two different labels may deviate for one or more points, such that the shape that composed of anchor points with the same label would be abnormal. Therefore, using the larval shape model to fit the anchors allows adjusting the junction of labels and reducing deviation. By considering the shape characteristics of the *Drosophila* larva, we use the former proposed annular sector model (ASM) [19] as the larval shape model.

As shown in Fig.5(a), the annular sector is generated from the subtraction of two sectors which have the same circle center and included angle, but have different radiuses. In this paper, in order to conveniently infer the larval pose and location from the anchor points, the vector $\mathbf{v}$ of ASM is modified as follows:

$$\mathbf{v} = (c_x, c_y, r, l, e, \beta), \tag{9}$$

where

- (c_x, c_y) is the coordinate of the center point on larval spine line,
- r is the middle radius, which is calculated as:

$$r = (r_{max} + r_{min})/2, \tag{10}$$

  in which r_{max} is the radius of the bigger sector mentioned above and r_{min} is the radius of the other,
- l is the length of the arc related to the middle radius r,
- e is the width of the annular sector, and e equals to $r_{max} - r_{min}$,
- $\beta \in (-\pi, \pi]$ is the orientation of the axes vector of the annular sector.

As the relationship of the arc length, the radius, and the included angle, the other variables (like the included angle θ, the radius r_{max} and r_{min}) of the annular sector model can be calculated from these six variables. Also, when the shape of the larva stretches along a line, it should be fit with a rectangle, as shown in Fig.5(b). And this situation can be also regarded as a special annular sector, whose $r \to \infty$, θ closes or equals to 0.

The variables of the ASM for single larva can be calculated directly from the larval spine line. Fig.5(c) illustrates examples for using the ASM to fit a bendy larva and a straighter larva respectively. By obtaining the spine points from the extracted anchor points, the spine line is fit with an arc or a line by using the least square method. Then the variables of the ASM for the larva can be estimated from the arc or the line.

However, when two or more larvae moving into clusters, it is difficult to obtain the larval spine line and the corresponding ASM directly from the occluded area, due to the absence of some of the anchor points. Instead, as shown in Fig.6, we split the distribution of the anchor points with the same label into four situations:

i) If there are two sharp points in the extracted anchor points, the arc or line is estimated by the anchors on the edge that maintains integrity, as shown in Fig.6(a);
ii) If there is only one sharp point and there are more than ω anchor points on both sides of the sharp point, the arc or

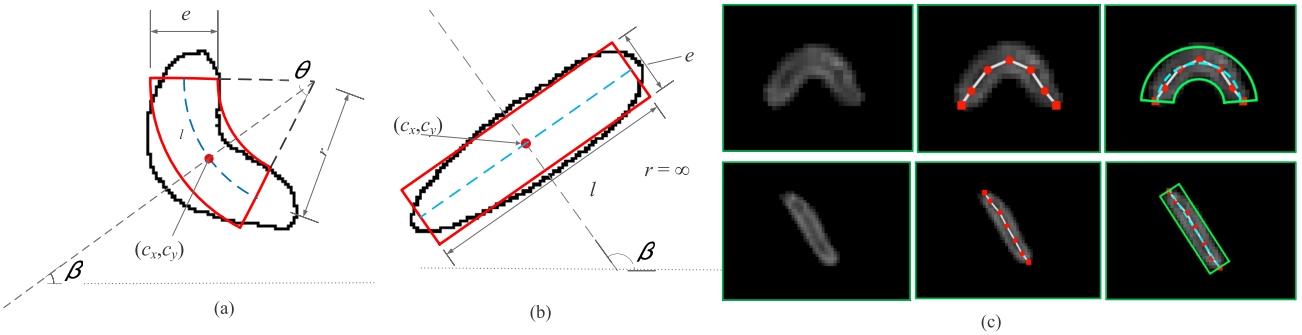


FIGURE 5. The sketch of ASM and two cases of larvae fitted by the ASM: (a) the annular sector (red contour) fits the bendy larva (black contour) and (b) the particular case of the ASM – the rectangle – (red contour) fits the stretched larva (black contour) and (c) two larvae fitted by the ASM, the left column is the raw image, the center column is the spine line of larva, and the right column is the illustration of fitting ASM.

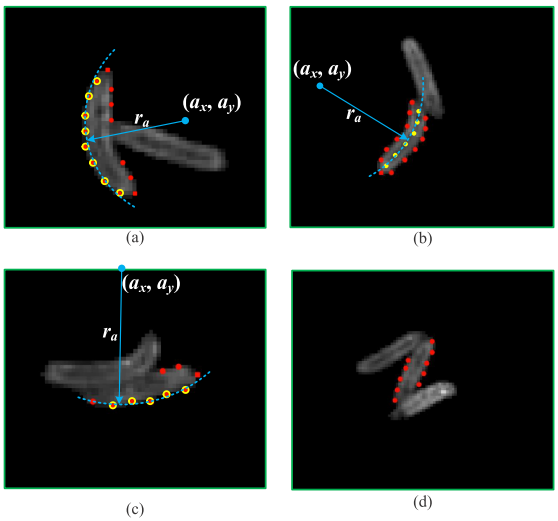


FIGURE 6. Four situations for the distribution of the anchor points with the same label and generating fitting arcs from the anchor points.

line is estimated by the spine points that can be inferred by the edge anchor points, as shown in Fig.6(b);

iii) If there is only one sharp point and the number of anchor points on only one side of the sharp point is more than ω, the arc or line is fit by the corresponding anchors, as shown in Fig.6(c);

iv) The others are classified to the situation that don't have enough anchors, as shown in Fig.6(d). The variables of the ASM are initialized as the values for the same larva in the former frame, and then sampling on the near space of the variables.

By estimating the larval ASM in occlusions, the labels of junction points are revised by the minimal distance between the anchor point and the larval spine line. Also, in order to accelerate the processing of multiple frames, only the anchor points on the junctions are considered by the revising process.

IV. EXPERIMENTS

To evaluate the performance of our approach, the experiments were performed on two publicly *Drosophila* larvae datasets which have different spatial and temporal resolutions. In this section, we firstly introduce the details of datasets and present the implementation of the proposed method respectively. Then, we evaluate the performance of our method quantitatively and qualitatively.

A. DATASETS

We evaluate our approach on two publicly available challenging larval collision datasets from the HCI laboratory [13] and from the work [18]. Both of the datasets provide videos of multiple *Drosopila* larvae in laboratory condition from a single camera on the top view, but they have different resolutions and characteristics for the larvae collisions.

The HCI dataset [13] includes 33 high resolution movies with 1000 frames each and a temporal resolution of 3.3 frames per second. Each movie provides the holistic view for average 20 larvae that moving from a cluster to all directions. The larvae are small in this dataset and the average number of pixels for one larva is less than 250 in this dataset. Furthermore, the dataset also provides the encounters which are extracted from the original videos, and each encounter is composed of subimages that contain the region of two or more larvae. The ground truth is that the larval identities for each encounter are labeled manually in the frames before and after overlapping. In this dataset, there are 1478 encounters for two larvae, 96 for three larvae and 28 for four or more larvae.

The dataset from the work [18] is composed of 2262 image sequences and each sequence provides local view for two or more larvae involved in collisions with a temporal resolution of 20 frames per second. In this dataset, the average size of one larva is about 1500 pixels, and the appearances of larvae are more diversified with different lightning conditions. There are 4524 encounters for two larvae and 33 for three larvae that are extracted from the collision sequences. As the same with the prior dataset, to evaluate the proposed method, we label these encounters with the larval identities manually in the frames before and after overlapping.

B. IMPLEMENTATION DETAILS

For each image sequence or video, a simple background image is learned as the per-pixel median of 10 images which are sampled uniformly from all frames [37].

The segmentation is performed by subtracting the background and taking morphological operations. Connected components of the segmented image are regarded as detections. Then we use the context of connected components to count the larval number for each cluster, which also can obtain a precision of 99.9% on both of these datasets.

As the larval resolutions are diverse in different datasets and different cases, and even the larval size varies in different frames, sliding window with adaptive length is adopted to extract sharp points. In this paper, the adaptive length is in proportion to the number of contour points. Also, the sampling interval for individual larva between sharp points is also in proportion to the number of contour points. When occlusion occurs, the fixed interval for sampling edge points between sharp points is the average of sample intervals for individual larvae before they move into cluster. And the sharp threshold value τ is set as 0.5π.

Besides, in order to adapt the proposed method to different datasets with various larval sizes and diverse resolutions, all of the distances in cost energy function in (4) and (6) are in proportion to the average length of the related larva. The parameters in Guassian functions for pairwise transition score are obtained statistically from the isolated individual larval movements. In this paper, the sharp expected value u_1 and standard deviation σ_1 are respectively set as 0.03 and 0.09, and the edge expected value u_2 and standard deviation σ_2 are set as 0.025 and 0.07. And the scale number λ in (4) and (6) is set as 0.33. The number ω which is the threshold value for fitting the ASM in cluster is set as 5. It needs to be noted that when evaluating the proposed method on the above two larval datasets, we use the same parameter values in our method, despite that these two datasets have different spatial and temporal resolutions.

C. RESULTS AND DISCUSSIONS

First, we evaluate the performance of the proposed method on the HCI dataset. The comparison between our method and other approaches on the accuracy rates to solve identity assignment for encounters with different number of larvae is presented in Table 1. Ctrax [28] is apt to handle the object that moves regularly, thus its tracking results are prone to be relatively poor for multiple larvae. CT[35] has a higher accurate rate than Ctrax for the clusters of two larvae but has difficulties to disambiguate the occlusions of more than two entangled targets. The weakly supervised structured learning method proposed by Fiaschi *et al.* [13] has relative more stable performance than the above approaches, but its performance of accuracy rate declines quickly with the number of larvae raising up. The ASM [20] method and the approach proposed by Branson and Belongie [30] have relative higher accuracies for this dataset. Here, for the method proposed by Branson and Belongie [30], the results are under the situation which is trade-off of accuracy rate and computation time (with 1000 samples). We also compare the results of our method with and without ASM. Among all of the approaches, our method with ASM achieves the highest accuracy rate on identity assignment in the encounters with the number of larvae less than or equal to 3 on this dataset, and it still has very good performance when the larval number is more than 3.

TABLE 1. The identity assignment results on HCI Dataset [13] with different number larvae in encounters.

Method	Avg.	$N = 2$	$N = 3$	$N \geq 4$
Ctrax [31]	86.8%	88.5%	73.7%	42.9%
CT [38]	90.6%	93.3%	65.3%	35.7%
Fiaschi *et al.* [13]	94.7%	95.8%	85.3%	67.9%
ASM [19]	98.2%	98.6%	92.1%	**84.0**%
Michels *et al.* [33]	97.6%	98.4%	92.8%	75.0%
Ours(without ASM)	94.6%	96.8%	82.1%	67.9%
Ours(with ASM)	**98.6**%	**99.3**%	**93.7**%	78.6%

TABLE 2. The testing computation time on HCI Dataset [13].

Method	Computation Time
Fiaschi *et al.* [13]	42.9h
ASM [19]	3.1h
Michels et al. [33]	48min
Ours(without ASM)	4.8min
Ours(with ASM)	8.7min

TABLE 3. The identity assignment results on the Dataset [18] with different number larvae in encounters.

Method	Avg.	$N = 2$	$N = 3$
ASM [19]	97.86%	97.90%	93.94%
Michels *et al.* [33]	97.57%	97.62%	90.91%
Ours(without ASM)	95.43%	95.72%	76.32%
Ours(with ASM)	**99.04**%	**99.05**%	**98.99**%

The testing computation time on this dataset for some of these methods is listed in Table 2. We can see that the computation costs of our method (with and without ASM) and the work [33] are much lower than that of the ASM method [19] and the work [13]. In the table, results show that our method without ASM is the fastest among all of these approaches. However, it is obvious that the proposed method with ASM can achieve the best performance on accuracy with a bit more computational cost than that without ASM.

We also compare our method with two approaches [19], [33] on the dataset from the work [18] and the results of accuracy rates for encounters with different number of larvae are shown in Table 3. Also for the method proposed by Michels *et al.* [33], the number of samples is set as 500, which is the trade-off of accuracy rate and computation time. From this table, we can see that our method with ASM also has the best performance for accuracy rate in encounters with both 2 and 3 larvae on this dataset. In addition, Table 4 shows the computation time of these three methods on this dataset. The ASM [19] method has the lowest speed because that the convergence and interpretation mechanism spend more time. And the result also shows that the proposed method is faster than the literature [33].

All of the above quantitative results demonstrate that our proposed method is robust when evaluating on different datasets with different spatial and temporal resolutions, and that our method can improve the accuracy rate of the identity

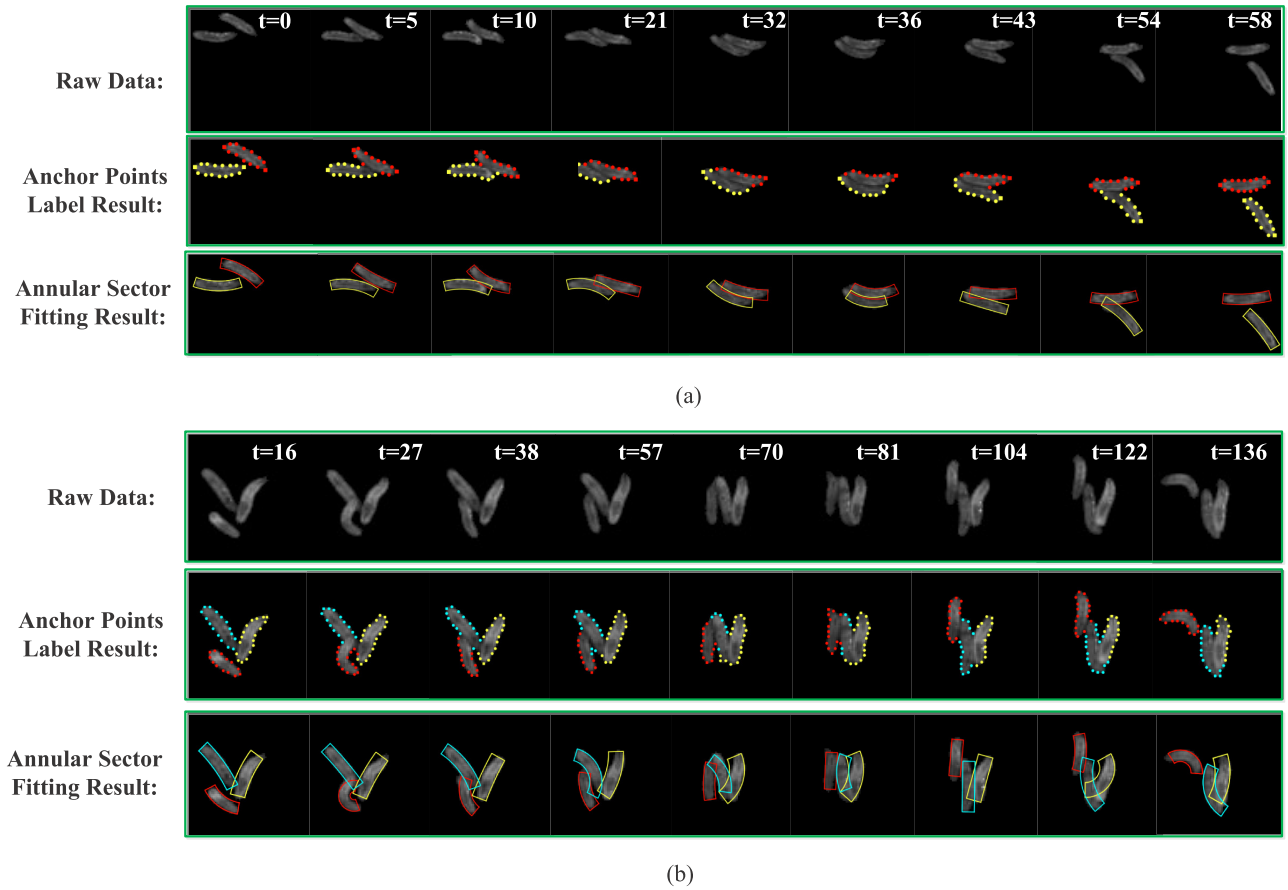


FIGURE 7. Tracking results of two cases: (a) two larvae in occlusions and (b) three larvae in occlusions.

TABLE 4. The testing computation time on the Dataset [18].

Method	Computation Time
ASM [19]	20h
Michels *et al.* [33]	89min
Ours(without ASM)	22min
Ours(with ASM)	42min

assignment and accelerate the tracking process simultaneously. Moreover, in order to further demonstrate the robustness of our method, we also evaluate the proposed method (with ASM) on different temporal and spatial scales of the HCI dataset [13] and the dataset from the work [18]. And the experimental results have shown in the Supple. Table 1 and Supple. Table 2.

To qualitatively evaluate the proposed method, Fig.7 illustrates our results for tracking multiple larvae with two challenging cases. The first case presents the tracking result of the encounter of two larvae, as shown in Fig.7(a). The tracking result shows that our method not only disambiguates the identity of each object before and after encounters through the labels of anchor points, but also can still well indicate the poses of larvae in each frame even during they are heavily tangled together. Our tracking algorithm uses the ASM not only to rectify the labels of anchor points, but also to fit

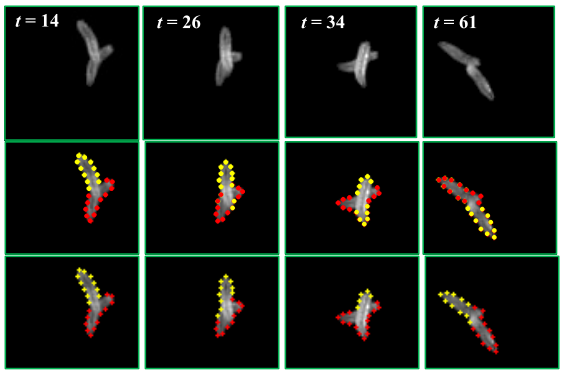


FIGURE 8. Illustration of a failure example. **Top row:** raw data, **Center row:** manually labeled ground truth, **Bottom Row:** tracking result with color labeled anchor points. The tracking fails since the yellow marked larva crawls over the red marked.

the poses of larvae in occlusions. In Fig.7(b), the second case indicate part of the encounter of three larvae, in which the larvae move together and their entanglement lasts for a long time. In this case, the larva labeled with blue anchor points and blue annular sector is sandwiched between two other larvae and in appearance heavily merged with others. Despite that the results of labels of anchor points occurs some errors in several frames, the annular sector fitting result still

can recognize the poses and rectify the identity assignment. Thus, the proposed tracking method still performs well on the encounters that are in heavily occlusions.

Fig.8 shows a failure case that the predicted result for the proposed tracking method is incorrect. From this case, we find that the proposed method is difficult to assign correct labels to anchor points for the situation that one larva crawls over the middle of another larva. The reason of these kinds of errors occurring is that in the proposed method the labels of anchor points are mainly decided by the identities of neighboring anchor points in previous frame, but the truth label of anchor points would be abrupt changed when the above kind of situation appears. In addition, it is rarely happens for this kind of case in the practical movements of multiple larvae, that there are less than 10 cases for the above two datasets.

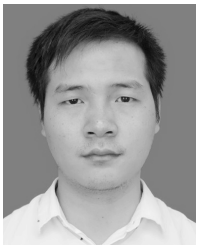
V. CONCLUSION

We have proposed a novel method for tracking multiple indistinguishable and deformable objects in occlusions and tested it on two publicly challenging larvae datasets. In our method, the approach of extracting anchor points has been presented and the bipartite graphs of sharp points and edge anchor points respectively have been established to measure the relationships of anchor points between adjacent frames. By using Hungarian algorithm to minimize the energy function, the multi-anchor flows were obtained to solve the identity assignment of multiple objects before and after clusters. Meanwhile, the ASM has been combined with the multi-anchor flows to rectify the labels of anchor points and describe the poses of larvae in occlusions. The experiment results have shown that our proposed method not only achieves highest accuracy rate of identity assignment among state-of-the-art approaches on both of the datasets, but also improves the computational speed when compared with current methods. The results also demonstrate the efficiency of our method for tracking multiple larval poses, even though they are entangled heavily lasting for a long time. Moreover, as the multi-anchor flow ignores the erratic motions and deformable shapes, it can be used for tracking other animals.

REFERENCES

[1] A. Maksai and P. Fua, "Eliminating exposure bias and metric mismatch in multiple object tracking," in *Proc. CVPR*, Long Beach, CA, USA, Jun. 2019, pp. 4639–4648.
[2] P. Voigtlaender, M. Krause, A. Osep, J. Luiten, B. B. G. Sekar, A. Geiger, and B. Leibe, "MOTS: Multi-object tracking and segmentation," in *Proc. CVPR*, Long Beach, CA, USA, Jun. 2019, pp. 7942–7951.
[3] Z. Tang, M. Naphade, M.-Y. Liu, X. Yang, S. Birchfield, S. Wang, R. Kumar, D. Anastasiu, and J.-N. Hwang, "CityFlow: A city-scale benchmark for multi-target multi-camera vehicle tracking and re-identification," in *Proc. CVPR*, Long Beach, CA, USA, Jun. 2019, pp. 8797–8806.
[4] S. Jin, W. Liu, W. Ouyang, and C. Qian, "Multi-person articulated tracking with spatial and temporal embeddings," in *Proc. CVPR*, Long Beach, CA, USA, Jun. 2019, pp. 5664–5673.
[5] A. A. Robie, K. M. Seagraves, S. E. R. Egnor, and K. Branson, "Machine vision methods for analyzing social interactions," *J. Exp. Biol.*, vol. 220, no. 1, pp. 25–34, Jan. 2017.
[6] L. Giancardo, D. Sona, H. Huang, S. Sannino, F. Managò, D. Scheggia, F. Papaleo, and V. Murino, "Automatic visual tracking and social behaviour analysis with multiple mice," *PLoS ONE*, vol. 8, no. 9, Sep. 2013, Art. no. e74557.
[7] W. Hong, A. Kennedy, X. P. Burgos-Artizzu, M. Zelikowsky, S. G. Navonne, P. Perona, and D. J. Anderson, "Automated measurement of mouse social behaviors using depth sensing, video tracking, and machine learning," *Proc. Nat. Acad. Sci. USA*, vol. 112, no. 38, pp. E5351–E5360, 2015.
[8] S. Schulter, P. Vernaza, W. Choi, and M. Chandraker, "Deep network flow for multi-object tracking," in *Proc. CVPR*, Honolulu, HI, USA, Jul. 2017, pp. 6951–6960.
[9] Y. Xiang, A. Alahi, and S. Savarese, "Learning to track: Online multi-object tracking by decision making," in *Proc. ICCV*, Santiago, Chile, Dec. 2015, pp. 4705–4713.
[10] X. Wang, B. Fan, S. Chang, Z. Wang, X. Liu, D. Tao, and T. S. Huang, "Greedy batch-based minimum-cost flows for tracking multiple objects," *IEEE Trans. Image Process.*, vol. 26, no. 10, pp. 4765–4776, Oct. 2017.
[11] R. T. Collins, "Multitarget data association with higher-order motion models," in *Proc. CVPR*, Providence, RI, USA, Jun. 2012, pp. 1744–1751.
[12] A. Milan, K. Schindler, and S. Roth, "Multi-target tracking by discrete-continuous energy minimization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 10, pp. 2054–2068, Oct. 2016.
[13] L. Fiaschi, F. Diego, K. Gregor, M. Schiegg, U. Koethe, M. Zlatic, and F. A. Hamprecht, "Tracking indistinguishable translucent objects over time using weakly supervised structured learning," in *Proc. CVPR*, Columbus, OH, USA, Jun. 2014, pp. 2736–2743.
[14] Z. Khan, T. Balch, and F. Dellaert, "MCMC data association and sparse factorization updating for real time multitarget tracking with merged and multiple measurements," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 12, pp. 1960–1972, Dec. 2006.
[15] S. J. Husson, W. S. Costa, C. Schmitt, and A. Gottschalk, "Keeping track of worm trackers," in *WormBook: The Online Review of C. Elegans Biology*. WormBook, Sep. 2012. [Online]. Available: http://www.wormbook.org/chapters/www_tracking/tracking.html, doi: 10.1895/wormbook.1.156.1.
[16] P. Ramdya, T. Schaffter, D. Floreano, and R. Benton, "Fluorescence behavioral imaging (FBI) tracks identity in heterogeneous groups of *Drosophila*," *PLoS ONE*, vol. 7, no. 11, Nov. 2012, Art. no. e48381.
[17] S. Ohayon, O. Avni, A. L. Taylor, P. Perona, and S. E. R. Egnor, "Automated multi-day tracking of marked mice for the analysis of social behaviour," *J. Neurosci. Methods*, vol. 219, no. 1, pp. 10–19, Sep. 2013.
[18] B. Risse, N. Otto, D. Berh, X. Jiang, M. Kiel, and C. Klämbt, "FIM^{2c}: Multicolor, multipurpose imaging system to manipulate and analyze animal behavior," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 3, pp. 610–620, Mar. 2017.
[19] B. Guo, G. Luo, Z. Weng, and Y. Zhu, "Annular Sector Model for tracking multiple indistinguishable and deformable objects in occlusions," *Neurocomputing*, vol. 333, pp. 419–428, Mar. 2019.
[20] W. Feng, Z. Hu, W. Wu, J. Yan, and W. Ouyang, "Multi-object tracking with multiple cues and switcher-aware classification," Jan. 2019, *arXiv:1901.06129*. [Online]. Available: https://arxiv.org/abs/1901.06129
[21] P. Bergmann, T. Meinhardt, and L. Leal-Taixe, "Tracking without bells and whistles," Mar. 2019, *arXiv:1903.05625*. [Online]. Available: https://arxiv.org/abs/1903.05625
[22] P. Chu and H. Ling, "FAMNet: Joint learning of feature, affinity and multi-dimensional assignment for online multiple object tracking," in *Proc. CVPR Workshops*, Long Beach, CA, USA, Jun. 2019, pp. 1–10.
[23] R. Henschel, Y. Zou, and B. Rosenhahn, "Multiple people tracking using body and joint detections," in *Proc. CVPR Workshops*, Long Beach, CA, USA, Jun. 2019, pp. 1–10.
[24] D. P. Mersch, A. Crespi, and L. Keller, "Tracking individuals shows spatial fidelity is a key regulator of ant social organization," *Science*, vol. 340, no. 6136, pp. 1090–1093, May 2013.
[25] Y. Shemesh, Y. Sztainberg, O. Forkosh, T. Shlapobersky, A. Chen, and E. Schneidman, "High-order social interactions in groups of mice," *Elife*, vol. 2, Sep. 2013, Art. no. e00759.
[26] M. Lorbach, E. I. Kyriakou, R. Poppe, E. A. van Dam, L. P. J. J. Noldus, and R. C. Veltkamp, "Learning to recognize rat social behavior: Novel dataset and cross-dataset application," *J. Neurosci. Methods*, vol. 300, pp. 166–172, May 2018.
[27] A. Pérez-Escudero, J. Vicente-Page, R. C. Hinz, S. Arganda, and G. G. De Polavieja, "idTracker: Tracking individuals in a group by automatic identification of unmarked animals," *Nat. Methods*, vol. 11, no. 7, pp. 743–748, Jun. 2014.

[28] F. Romero-Ferrero, M. G. Bergomi, R. C. Hinz, F. J. H. Heras, and G. G. de Polavieja, "idtracker.ai: Tracking all individuals in small or large collectives of unmarked animals," *Nat. Methods*, vol. 16, no. 2, pp. 179–182, Jan. 2019.
[29] C. Wählby, T. Riklin-Raviv, V. Ljosa, A. L. Conery, P. Golland, F. M. Ausubel, and A. E. Carpenter, "Resolving clustered worms via probabilistic shape models," in *Proc. ISBI*, Rotterdam, The Netherlands, Apr. 2010, pp. 552–555.
[30] K. Branson and S. Belongie, "Tracking multiple mouse contours (without too many samples)," in *Proc. CVPR*, Jun. 2005, pp. 1039–1046.
[31] K. Branson, A. A. Robie, J. Bender, P. Perona, and M. H. Dickinson, "High-throughput ethomics in large groups of *Drosophila*," *Nat. Methods*, vol. 6, no. 6, pp. 451–457, May 2009.
[32] C.-N. J. Yu and T. Joachims, "Learning structural SVMs with latent variables," in *Proc. ICML*, Montreal, QC, Canada, Jun. 2009, pp. 1169–1176.
[33] T. Michels, D. Berh, and X. Jiang, "An RJMCMC-based method for tracking and resolving collisions of drosophila larvae," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 16, no. 2, pp. 465–474, Mar./Apr. 2017.
[34] Y. Rubner, C. Tomasi, and L. J. Guibas, "The earth mover's distance as a metric for image retrieval," *Int. J. Comput. Vis.*, vol. 40, no. 2, pp. 99–121, Nov. 2000.
[35] B. Risse, D. Berh, N. Otto, C. Klämbt, and X. Jiang, "FIMTrack: An open source tracking and locomotion analysis software for small animals," *PLoS Comput. Biol.*, vol. 13, no. 5, May 2017, Art. no. e1005530.
[36] D. Chetverikov, "A simple and efficient algorithm for detection of high curvature points in planar curves," in *Proc. Conf. CAIP*, Salerno, Italy, Sep. 2003, pp. 746–753.
[37] R. Cucchiara, C. Grana, M. Piccardi, and A. Prati, "Detecting moving objects, ghosts, and shadows in video streams," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 10, pp. 1337–1342, Oct. 2003.
[38] M. Schiegg, P. Hanslovsky, B. X. Kausler, L. Hufnagel, and F. A. Hamprecht, "Conservation tracking," in *Proc. ICCV*. Sydney, NSW, Australia: Darling Harbour, Dec. 2013, pp. 2928–2935.



BIAO GUO received the B.S. degree in computer science from Lanzhou University, in 2012. He is currently pursuing the Ph.D. degree with the Laboratory of Communication and Information Security, Peking University Shenzhen Graduate School. His current research interests include computer vision, machine learning, and biomedical information processing.

YUESHENG ZHU received the B.Eng. degree in radio engineering, in 1982, the M.Eng. degree in circuits and systems, in 1989, and Ph.D. degree in electronics engineering, in 1996. He is currently a Professor with the Laboratory of Communication and Information Security, Peking University Shenzhen Graduate School. His current research interests include digital signal processing, multimedia technology, communication, and information security. He is a Fellow of the China Institute of Electronics and a Senior Member of the China Institute of Communications.