Tracking and Re-Identification System for Multiple Laboratory Animals

Filip Naiser, Matěj Šmíd, Jiří Matas

Center for Machine Perception, Czech Technical University in Prague, Czech Republic
filip@naiser.cz, {smidm,matas}@cmp.felk.cvut.cz

Abstract—We present a tracking system for ecology and biology researchers suitable for movement and interaction analysis of multiple animals in laboratory conditions. On the input is a single video with multiple animals and the outputs are animal trajectories. The system is agnostic with regard to animal species. It can adapt to a new animal appearance automatically without annotation. For animal re-identification we use discriminatively trained CNN embedding. The system was tested on sequences with multiple ants, zebrafish and sowbugs.

1. Introduction

Image-based animal tracking recently enabled high-throughput methods in biology and ecology. Numerous insights were achieved using automated tracking [2]. There are already several freely available tracking systems for end users [10, 12–14, 16].

As stated in [2], the range of possible tracking problem difficulty and output quality is broad in ecology research. The tracking task ranges from straightforward laboratory environment with few animals to a complex landscape with many individuals directly in a field. The basic tracking systems provide output without individual identities and infer only animal positions. Advanced systems maintain identities [6, 10] and also recover animal poses [1, 15].

The presented system is suitable for laboratory environment. The arena where animals move should have uniform background and constant lighting. The shape of the environment can be arbitrary.

We expect the camera to be stationary. The objects to be tracked are often nearly indistinguishable animals of one species. The camera is typically observing the scene from above. The animals move on a plane or in shallow water. This setup ensures that the animals are viewed from a single direction and the scale changes are negligible.

We based our work on [8]. We evaluated the system on four videos of ants, zebrafish and sowbugs. We compare our results with the established baseline idTracker [10].

There are several state of the art tracking systems for laboratory animals suitable for end users and a range of publications dealing with the topic without accompanying software package. The mentioned idTracker proposed a re-identification method that compares fingerprints in a unique histogram form (colour correlogram) that encodes spatial distribution of pixel pairs along with their intensities. ABCtracker [14] recently originated in the group of Min C. Shin. The research team has a string of publications on laboratory animals tracking [3, 4, 9, 11]. Toxtrack [12] uses background subtraction, thresholding, mathematical morphology and Kalman filter for basic tracking. It also includes re-identification module with texture features and intensity histogram. Attractive Toxtrack feature is tracking speed. Toxid [13] is re-identification method currently used in Toxtrack based on intensity histograms and Hu-moments. The work [16] uses CNNs for identity matching of zebrafish tracklets. The network is trained online with automatically extracted zebrafish head images.

2. Methods

In input video frames are first segmented regions containing tracked animals. An initial tracking graph is constructed out of regions in spatiotemporal space. Nodes represent regions and edges possible transitions between them in consecutive frames. The graph edges are further pruned in such way that only the most probable transitions between regions remain. The paths in the graph where no branching occurs are joined into tracklets. The user then annotates few regions with categories single id, multi id and other (further called cardinality classification). A classifier trained on the annotated regions is able to classify tracklets cardinality. The re-identification module is learned on data automatically extracted from single id tracklets. The re-identification module computes a probability that two tracklets belong to the same individual. Tracklets of the same individual are joined into tracks and the identity information is propagated in the tracking graph. The overview is illustrated in Figure 1.

2.1. Segmentation

In every video frame, the animals have to be segmented from the background. The arenas in typical laboratory experiments have mostly uniform colours and are not cluttered. We were able to avoid complex object detectors and still achieve satisfactory results using maximally stable extremal regions (MSERs) [7] algorithm. MSERs are also superior to simple thresholding and background subtracting algorithms often used in laboratory tracking algorithms. The mentioned two segmentation methods often produce not compact regions and in the case of the background subtraction, the animals that stay still for
longer time are blended into the background. The MSERs
are further filtered by multiple criteria: MSER margin,
nested regions removal and suppression of bright regions.

2.2. Tracking Graph
Once an image in the frame $t$ is segmented, the region
set $R_t$ is defined. Between $R_t$ and $R_{t+1}$ a fully connected
bipartite graph is established. Edges represent possible
transitions between two regions. Unambiguous fragments
of tracks called tracklets are found in the graph. Track-
lets are typically parts of object trajectories where the
object is separated from other objects. The tracklets are
constructed from isolated paths in the graph where no
branching occurs. More formally a tracklet is a sequence
of regions corresponding to the nodes on an isolated path.

The edge cost is then an inverse probability of a valid
transition. The changes in appearance and movement
features of the two regions are checked for anomalies with
isolation forest algorithm. The anomaly score is converted
to a probability using logistic regression. The edges with
probability $< \theta$ are removed.

2.3. Region Cardinality Classification
To distinguish regions with a single animal, more animals
and other regions, we train a nearest neighbour region car-
dinality classifier. First, $k$ regions are randomly sampled,
and then clustered into $\frac{k}{10}$ groups. The clustering is done
in space of high-level region descriptors (e.g. area, ma-
ajor/minor axis, pixel density). Each cluster is represented
by a region that is labelled by a user as single id, multi
id or other. The unsupervised clustering groups similar
regions and so reduces the number of user annotations.
Example regions along with labels are shown in Figure 2.

Figure 1. Tracking system overview. The nodes in an asymmetric shape are inputs or outputs of the system. The
rectangular nodes represent actions and the texts with grey background over the arrows describe the type of data
flowing between action nodes. The parts of the system are described in the Section 2.
A classifier is trained and tracklet cardinalities are decided by majority vote over cardinality of all tracklet regions.

### 2.4. Re-identification

Every region is described by a low-dimensional descriptor computed by a simple CNN with eight convolutional layers followed by a single fully connected layer. We trained the CNN in a Siamese architecture with a triplet loss [5]. The training examples were randomly sampled from two time intersecting single id tracklets. This guarantees two different classes.

Each track appearance is represented by \( k \) prototypes \( \psi = \{ \vec{\mu}, \sigma, w \} \). A prototype is representing \( w \) descriptors \( D = \{ \vec{d}_1, \vec{d}_2, \ldots, \vec{d}_w \} \) with mean \( \vec{\mu} = \overline{D} \), \( \sigma^2 = \frac{1}{|D|} \sum_{\vec{d} \in D} ||\vec{\mu} - \vec{d}||^2 \). Prototypes are defined as results of agglomerative clustering of tracklet descriptors into \( k \) clusters. When \( \Gamma(t_i) \) is an id-set of track \( t_i \), probability of tracks having the same id, based on appearance, is computed as:

\[
P(\Gamma(t_1) = \Gamma(t_2)) = \frac{f(t_1,t_2) + f(t_2,t_1)}{2} \cdot P_s(t_1,t_2),
\]

where \( P_s \) is a probability term based on a spatio-temporal distance. It is zero in prohibited cases (e.g. \( t_2 \) begins sooner than \( t_1 \) ends). It is switched off for big temporal distances thus the decision is done using only appearance. The probability \( f(t_1,t_2) \) represents \( t_2 \) drawn from \( t_1 \) distribution and \( f(t_2,t_1) \) vice versa.

Id assignment is done as follows: first, each tracklet is initialized with a unique id. Then ids are propagated when \( P > C \) - certainty threshold. The decision is not done independently. Instead, two sets of concurrent single id tracklets are solved together using maximum weighted matching. This guarantees id consistency (e.g. an id cannot be assigned twice in the same frame).

### 3. Experiments

An insight on the performance of re-identification module is visualized on Figure 3. The matrices show probabilities of tracklet pairs (represented as rows and columns) belonging to the same identity. We can see the distinctive patterns in three out of four datasets. The end-to-end tracking performance was evaluated and compared to the established idTracker. As we are not solving animal close encounters yet, we compared with the idTracker results where the encounters are missing. We used four datasets: Ants1, Ants2, Zebrafish and Sowbug. For description see [8]. All evaluated metrics are percentages of all animals in all frames that: were detected correctly, were left undecided, or were wrongly detected (includes missing detections). The used metric is suitable for measuring identity preservation and is motivated by biology and ecology research objectives. The results are summarized in Table 1 and visualized in Figure 4. The presented tracker, marked ours performs better than idTracker in all metrics on three out of four datasets.

### 4. Conclusion

We presented a tracking system for multiple laboratory animals. Although it is still in a work in progress state, it already performs slightly better than idTracker on three out of four datasets we evaluated. The tracklets with multiple animals are currently not handled, but preliminary solver already exist and we expect to include it soon.

We plan to test the tracker on more datasets and compare the results to ABCTracker and Toxtrack. We will make an open source release.

### Acknowledgements

The authors acknowledge the support of the OP VVV funded project CZ.02.1.01/0.0/0.0/16_019/0000765 “Research Center for Informatics”, Technology Agency of the Czech Republic funded project TH03010191 and CTU funded project SGS17/185/OHK3/3T/13.

### References

**Figure 3.** Distance matrices of all tracklets sorted by true identities based on re-identification descriptors. Values close to one mean that two tracklets belong to the same identity, values close to zero mean otherwise. The identities are separated by the dashed lines. The capability of re-identification descriptors to match respective tracklets is visible on three out of four datasets. On the Sowbug3 dataset are the descriptors failing probably due to low resolution and compression artefacts.


**Figure 4.** Tracking results on four video sequences. For description see Table 1. The tracking system described in this paper performs slightly better than idTracker in three out of four sequences.


