"Smart Mouse" Tracking Performance Report

19 June 2010–06–19

Introduction

The aim of this report is to quantify the current performance level of the mouse tracking software.

Ground Truth Labeling Procedure

A five day sequence was selected. Since it was not feasible to go over every frame of the data, it was decided to test only a subset, which is supposed to be representative. The subset contains frames from the first, third and fifth day. Four hours were selected from each day (see table 1). Mice were on a reversed cycle. Lights turned off at 10:00 and back on at 22:00.

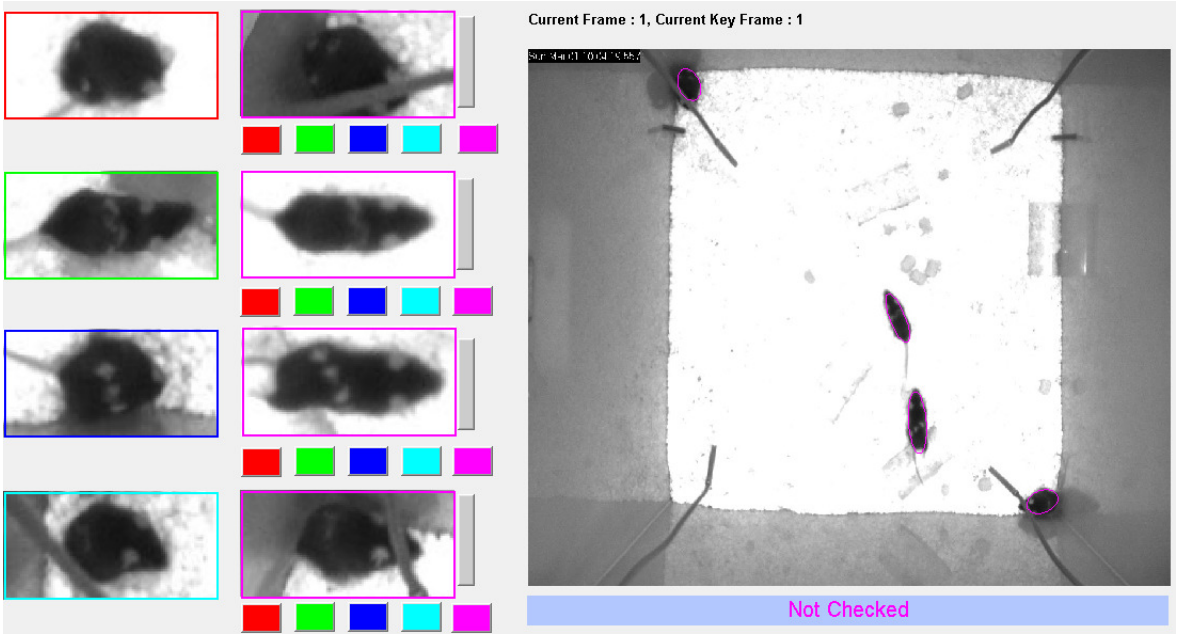
To reduce the manual labor further, only a single frame every 150 frames was actually labeled. This corresponds to checking the tracking results every 5 seconds ~~within the hour~~.

Table 1: Subset of the 5 day sequence used in the ground truth procedure.

Day	First Hour	Second Hour	Third Hour	Forth Four
1	12:00-13:00	15:00-16:00	20:00-21:00	01:00-02:00
3	13:00-14:00	14:00-15:00	19:00-20:00	04:00-05:00
5	17:00-18:00	19:00-20:00	20:00-21:00	05:00-06:00

To reduce possible bias from the human labeler, frames were shuffled and tracking results were displayed without the identities decided by the automatic algorithm. Thus, in each frame, the human labeler saw four ellipses, colored uniformly (magenta). The small image patch centered at each ellipse was magnified ~~at displayed~~ on the right of the screen. The task of the human labeler was to match the small image patch to image patches representing the true identities (i.e., these images were fixed throughout the entire ground truth procedure). A snapshot from the ground truth application is shown in figure 1. The human labeler matched the set of magnified images by clicking the small color rectangles below each image.

Figure 1 – Snapshot from the ground truth appliation



Definitions

1. "Fully Annotated Frame" – a frame in which the human labeler annotated all mice. In the rare cases in which the human labeler forgot to mark all mice the frame will be denoted "Partially annotated"
2. "Failed Segmentation" – a frame in which the human labeler decided that tracking output does not accurately portray mice position. A single mice that was not accurately segmented will cause the entire frame to be declared as failed segmentation.
3. "Incorrect" – A key frame with at least one identity swap
4. "Correct" – A key frame where all identities are correct

Results

For some reason, the last hour has extremely poor results. Further investigation is needed to understand exactly why.

When excluding the last hour, the final statistics are:

7838 Frames (10.89 hours) were checked (fully annotated)

113 (**1.44 %)** Frames were labeled as Failed Segmentation

The statistics for the remaining 7725 Frames:

6602 (**85.46 %**) were marked as Correct

1123 Incorrect frames (**14.54 %)**

Out of which, **338** **(30%)** were when all mice were close together

40.87% of the incorrect key frames occurred when the flipped identity was near one of the microphones.

The 1123 incorrect frames can be further decomposed into:

2 Identity Swaps : 935 (**83.26 %)**

3 Identity Swaps : 140 (**12.47 %)**

4 Identity Swaps : 36 (**3.21 %)**

The correct percentages per hour are:

Day	First Hour	Second Hour	Third Hour	Forth Four
1	78.36%	97.94%	77.25%	68.79%
3	69.94%	90.43%	97.23%	90.54%
5	76.58%	81.78%	96.39%	Excluded