Journal of Neuroscience Methods 219 (2013) 10–19

Contents lists available at SciVerse ScienceDirect

Journal of Neuroscience Methods

journal homepage: www.elsevier.com/locate/jneumeth

Basic Neuroscience

Automated multi-day tracking of marked mice for the analysis of social behaviour

Shay Ohayon[a,*], Ofer Avni[b], Adam L. Taylor[b], Pietro Perona[a,c], S.E. Roian Egnor[b]

[a] Computation and Neural Systems, California Institute of Technology, 1200 East California Boulevard, Pasadena, CA 91125, USA
[b] Howard Hughes Medical Institute, Janelia Farm Research Campus, 19700 Helix Drive, Ashburn, VA 20147, USA
[c] Division of Engineering and Applied Science, California Institute of Technology, 1200 East California Boulevard, Pasadena, CA 91125, USA

HIGHLIGHTS

- A fully automated system to track multiple animals in a large arena without losing their identities is presented.
- The system learns unique bleach patterns on the mice's fur and tracks them during both dark and light cycles.
- Identification of six mice in the experimental setup was 97% correct during non-sleep intervals.
- As a proof of principle, we tracked groups of four mice and report social trends that develop across hours and days.

ARTICLE INFO

Article history:
Received 29 March 2013
Received in revised form 29 May 2013
Accepted 30 May 2013

Keywords:
Multiple mice tracking
Long term
Social behaviour
Automated

ABSTRACT

A quantitative description of animal social behaviour is informative for behavioural biologists and clinicians developing drugs to treat social disorders. Social interaction in a group of animals has been difficult to measure because behaviour develops over long periods of time and requires tedious manual scoring, which is subjective and often non-reproducible. Computer-vision systems with the ability to measure complex social behaviour automatically would have a transformative impact on biology. Here, we present a method for tracking group-housed mice individually as they freely interact over multiple days. Each mouse is bleach-marked with a unique fur pattern. The patterns are automatically learned by the tracking software and used to infer identities. Trajectories are analysed to measure behaviour as it develops over days, beyond the range of acute experiments. We demonstrate how our system may be used to study the development of place preferences, associations and social relationships by tracking four mice continuously for five days. Our system enables accurate and reproducible characterisation of wild-type mouse social behaviour and paves the way for high-throughput long-term observation of the effects of genetic, pharmacological and environmental manipulations.

Published by Elsevier B.V.

1. Introduction

Mouse models have been recently developed to study the cognitive and social deficits observed in autism (Jamain et al., 2008; Penagarikano et al., 2011), schizophrenia (Hikida et al., 2007; Tremolizzo et al., 2002), Down syndrome (Olson et al., 2004; Reeves et al., 1995) and fragile X syndrome (Kooy et al., 1996; Zang et al., 2009). Social relationships in mice develop and evolve over the course of many days (Hurst et al., 1993; Poole and Morgan, 1975). The ability to carry out thorough, quantitative, long-term observations would likely have transformative effects on understanding and measuring social behaviour and its pathologies. However, widely used assays are often performed for short durations that can miss persistent durable traits (Fonio et al., 2012). A key challenge in performing long-term assays is the ability to obtain reliable annotation. However, it is not practical to have these assays done by human experts because they are tedious, expensive and not easily reproducible (de Chaumont et al., 2012; Spencer et al., 2008). Computer vision systems that are able to analyse animal behaviour automatically hold much promise (Reiser, 2009). Despite recent progress, state-of-the art computer vision systems are limited to the observation of two mice sharing an unfamiliar enclosure for a period of 10–20 min, often in partition cages, which limit social interaction (de Chaumont et al., 2012; Spencer et al., 2008). Significant progress in the classification of actions, once animal trajectories have been computed, has recently been reported (Burgos-Artizzu et al., 2012; de Chaumont et al., 2012; Jhuang et al., 2010). However, reliable tracking and the identification of individual mice when multiple mice share the same enclosure for several days remains an open problem.

Automatically tracking the identities of multiple animals in a video sequence is difficult. Current approaches are based on the

* Corresponding author. Tel.: +1 626 395 6672.
E-mail address: shay.ohayon@gmail.com (S. Ohayon).

0165-0270/$ – see front matter. Published by Elsevier B.V.
http://dx.doi.org/10.1016/j.jneumeth.2013.05.013

assumptions that the animals are always visible, do not overlap, and do not move too quickly, or employ heuristics, such as size differences across animals (Dankert et al., 2009), constrained environments (Branson et al., 2009) or artificially coloured markers (EthoVision, Noldus) to resolve animal identities. Attached coloured markers are easily groomed out and are not discriminable in infrared lighting, which is required for observation during dark cycles. All of the above approaches can fail and require human verification and correction of the results (de Chaumont et al., 2012). Furthermore, mice have flexible bodies, are highly interactive (cuddling, chasing, jumping on top of each other, mounting, etc.), and live in fairly complex environments (e.g., environments involving nests and bedding into which the mice burrow, which makes them invisible to the camera for periods of time). These factors make tracking and identification challenging, particularly when prolonged observation of social behaviour is desired.

We present a method that is capable of tracking individual mice interacting socially in a group over days without confusing identities; identities are maintained even when individuals hide and burrow in the bedding. The method consists of a single-camera computer vision system that automatically learns the appearance of each mouse and uses that appearance to infer each animal's identity throughout the experiment. We developed a set of uniquely discriminable patterns for marking the back of each animal. These patterns are produced by applying harmless hair bleach to the fur, cannot be groomed out, and can be tracked under infrared illumination during both dark and light cycles. The trajectories computed by our system may be used to detect and quantify mouse social behaviour (courtship, aggression, dominance, etc.) and to study its evolution over days. The system is easily reproducible, inexpensive, does not use any specialized hardware, user-friendly, and scalable to allow high throughput (the system and installation instructions are available at http://motr.janelia.org). Using our system, we characterised how social interaction developed in groups of four wild-type mice (two males and two females) over a five-day period.

2. Results

2.1. Method overview

Recognising individual mice from overhead pictures is difficult for both human observers and machines. To overcome this limitation, we developed a method to apply a distinct pattern to the back of each mouse using hair bleach (see Fig. 1a, Section 4). After patterning, each mouse is filmed alone for 5–10 min to collect diverse samples of its appearance during normal behaviour (Fig. 1b and c). The samples are then used to train image classifiers (one per mouse). All mice are then placed together in the same enclosure, where they are video-recorded continuously for five days under infrared lighting for the actual study. A purpose-built computer vision system tracked the positions of the mice and computed their trajectories (Fig. 1d). In the final step, the system computed mouse identities for each trajectory using Bayesian inference (Fig. 1e). On a single CPU, the processing of each video frame is ~300 ms (10× slower than real time). Processing can be done on a computer cluster to improve performance. Processing a five-day video (at 30 FPS) takes approximately 12 h on a cluster of one hundred 2.66 GHz four-core processors. Short sequences (1–2 h) can be easily analysed on a single computer overnight.

2.2. Mouse patterns

Inspired by naturally occurring patterns from the animal world (Gordon, 1985) and by patterns used in error-correcting codes (Blahut, 2003), we designed and tested more than a dozen different patterns, ten of which are presented in Fig. 1a. The patterns included large spots and thick stripes at different orientations and positions. Many more patterns can be generated using the same dyes. Our goal was to design patterns that could easily, quickly, and reproducibly be drawn on the backs of mice and that were highly discriminable from each other. The fur patterns slowly fade due to dark hair regrowth but remain visible for almost three weeks.

To train our computer vision system to identify the mice, we filmed each patterned mouse alone for several minutes (5–10 min) as the mouse was exploring the arena. Our tracking algorithm detected the position and orientation of the mouse in each frame and extracted a small image patch that was centred and aligned on the mouse (http://motr.janelia.org). Dense histogram of gradient (Dalal and Triggs, 2005) (HOG) features were extracted from each image patch and used to train a classifier to discriminate each mouse pattern from all other mouse patterns (1 vs. all, see Supplementary Fig. 2, Supplementary Text).

The performance of each pattern classifier was then evaluated in a cross validation procedure ($k = 4$) that tested it against the patterns from all ten mice ($10k$ samples per mouse) to discover which patterns were maximally discriminable.

We found that most patterns could be discriminated with high accuracy. The average true positive rate (TPR) was 0.9 ± 0.04, and the average false positive rate (FPR) was 0.01 ± 0.06 (see the confusion matrix in Fig. 1f). However, we found that some patterns were more easily confused than others. For example, pattern five (two vertical stripes) was likely to be confused with pattern eight (three vertical stripes). Manual inspection of misclassified samples revealed that errors occurred when patterns were heavily deformed (due to the flexible nature of the mouse body), partially obscured or completely occluded. This phenomenon typically occurred when mice sat or reared.

To find the optimal set of four patterns, we tested all possible pattern quadruplets and computed the error frequency (average false positive + false negative) for each quadruplet (Supplementary Fig. 3a and b). We found that many quadruplets of patterns produced roughly similar performance levels (the top ten combinations are given in Supplementary Fig. 3c), indicating that the method is relatively robust to the particular patterns used. For all of our experiments, we chose patterns 1–4 (Fig. 1a).

Small image patches obtained from videos showing only one mouse in the imaging setup ("solo samples", Fig. 1g) contained less variability than samples obtained from videos with four mice in the imaging setup ("group samples", Fig. 1i). Classifiers were trained on solo samples and required no human annotation. Classifiers performed well on solo samples (Fig. 1h, average TPR 0.96 ± 0.01), but their performance dropped when tested on group samples (average TPR 0.88 ± 0.13, Fig. 1j). Thus, frame-by-frame classification was not always reliable due to occlusion and large variations in appearance (Fig. 1i), suggesting that integration of the information from multiple frames was needed to accurately recover identities.

2.3. Detection and tracking

The function of the tracker in our system is to detect and track the poses (position and orientation, modelled by an ellipse) of multiple mice without concern for identity (Supplementary Text, section 3). The tracker works incrementally from the beginning to the end of the video. For each new frame, the poses of the mice from the previous frame are extrapolated and perturbed randomly to generate multiple hypotheses regarding mice positions in the current frame. Multiple instances of the expectation maximisation (Bishop, 2006) (EM) algorithm are initialised with these random hypotheses to estimate the most likely poses in the current frame. The best fitting hypothesis is then selected as the current pose, and

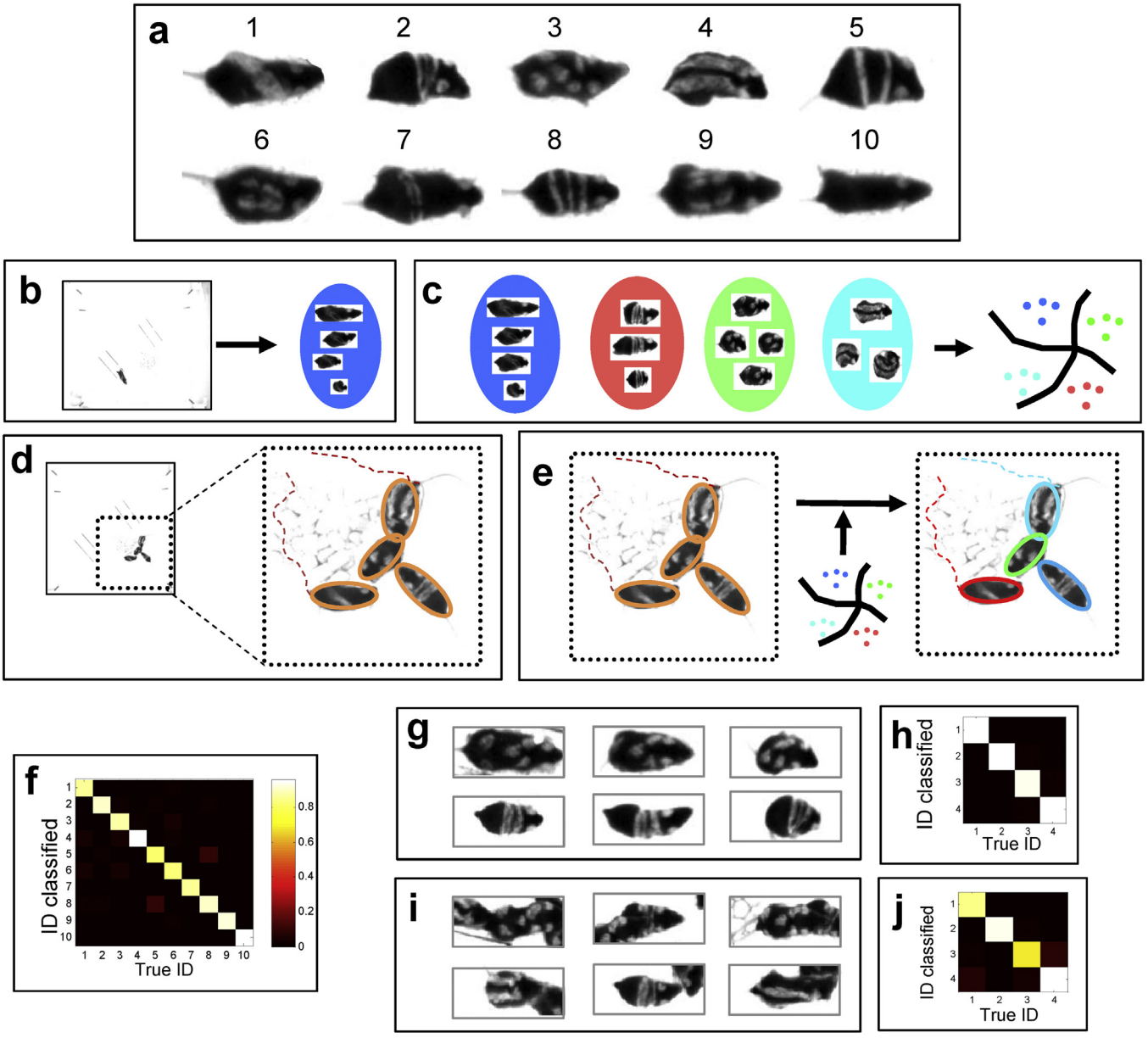

Fig. 1. System framework. (a) Ten patterns dyed onto the backs of the mice. Each pattern was created by bleaching the fur for several minutes (see Online Methods). (b) A single mouse was placed in the imaging setup and filmed for 5–10 min. Multiple images of the mouse were collected. (c) The process was repeated, and images were collected for each individual in the group. Using the four mice's images as a training set, a classifier was trained to distinguish individual mice. (d) The mice were tracked in the video of the experiment without using identity information, which generated trajectories with possible identity swaps. (e) Information from trajectories and the learned classifiers was combined to generate correct, identity-preserving trajectories. (f) Performance of mouse identity classifiers on images collected when each mouse was filmed alone. Each column represents the performance of a classifier trained to identify a single mouse. Entries on the main diagonal represent the true positive rate (correct identification of the mouse in a test set). Off-diagonal entries correspond to false alarm rates (incorrectly assigned mouse identity). (g) Examples of the variations in appearance observed when the mice were filmed alone. (h) Performance of four classifiers trained to identify the patterns [1–4] from (a) and tested on images obtained from single mouse videos. (i) Examples of the variations in appearance from a 30 min video of four mice. (j) Performance of the same classifiers in (d) on images obtained from a 30 min video of four mice simultaneously present in the enclosure. Rather than directly using the classifier's output, our method identified mice by combining the classifier's identity with trajectory information (see Fig. 2).

that hypothesis is associated with the corresponding pose in the previous frame. The tracking of a mouse stops when not enough pixels are available (e.g., when the mouse burrows in the bedding) and reinitialises when new unassigned pixels appear (e.g., when the mouse emerges from the bedding). Multiple mice disappearing and reappearing (e.g., due to burrowing) do not pose a problem because their identities are resolved in a later step (see Supplementary Text, Supplementary Fig. 4). The process is repeated for all frames in the video in a single pass from the beginning to the end, resulting in four trajectories. To reduce processing time, the video is automatically split into shorter segments that are processed in parallel on different computers (see Supplementary Text, section 3.1, 3.3, Supplementary Fig. 5).

Each trajectory obtained from the tracker may track different mice at different times because when two mice interact in close proximity, their identities may be swapped. These identity errors are resolved in the next step using the patterns on each mouse.

2.4. Propagating identity information

Once trajectories are obtained (in the previous step), the mouse identity classifiers are used to assign identities to the mice that

are associated with each trajectory in each frame. Good identity assignments result in each mouse's identity being consistent with its appearance in each frame and in each mouse's trajectory being smooth.

Our system uses a hidden Markov model (HMM) to associate the most likely mouse identities with each trajectory in each frame. The model is defined over all possible assignments of trackers to identities. For example, given a frame with four mice, there are 24 (4!) possible ways to assign identities to the four detected ellipses (two possible assignments are shown in Fig. 2a, each identity is colour-coded). The identity classifiers assign probabilities to each identity assignment. The probability of transitioning from one identity assignment to another is low when the mice are well separated in space and high when the mice are very close to each other (Fig. 2a, Supplementary Text, section 4.3). The probabilities of each identity assignment, which are purely based on frame-by-frame appearance-based identity classification, for a short (15 min long) sequence are shown in Fig. 2b. Each row corresponds to an identity assignment, and each column represents a frame. States with high identity probabilities are denoted in red.

Selection of the most probable identity (ID) assignment in each frame that is purely based on mouse appearance results in a jagged solution (see pink outline in Fig. 2c) because the most probable identity of each mouse in each trajectory changes frequently when visual classification is ambiguous. Comparison to ground truth identities showed that frame-by-frame selection of the most likely assignment had an error rate of approximately 10%. The HMM uses the additional constraint that cross-trajectory swaps are only likely when two trajectories come very close (i.e., see the example in Fig. 2d) and thus computed better assignments of identities and yielded 100% correct identification (Fig. 2e).

2.5. Validation

To evaluate our system's performance, we classified each mouse as huddled when it was in close contact with another mouse and non-huddled otherwise (see Section 4 and Fig. 3b). Huddled mice are typically clustered together sleeping and are difficult to tell apart, which poses a difficult problem for both correct segmentation and identification for both human and automatic annotators. This problem has little effect on behavioural analysis because the huddled mice are most often sleeping, and their behaviour is easily classified even when identification is uncertain. By contrast, correct mouse identification during non-huddled events is crucial for the study of individual and social behaviour. Huddling events were abundant and accounted for 55% of video frames. Huddling events were much more frequent during the light cycle (when the mice were less active) than during the dark cycle and increased in number over the course of the five-day experiment (Fig. 3a).

We quantified the performance of our system in estimating mouse pose and found that it performed comparably to human annotators regardless of whether the mice were huddling. To perform this quantification, we trained two human observers to draw tight ellipses around the bodies of the mice in 470 frames randomly sampled from our video recordings. We found that the average discrepancy in determining the position of each mouse between the two human annotators was 1.6 ± 0.8 mm, while the discrepancy between a human annotator and the machine was 1.8 ± 2.8 mm (see Supplementary Fig. 6 and Supplementary Text sections 6 and 7).

We also measured the accuracy of our system in classifying mouse identities over long periods of time. A human annotator manually labelled mouse identities in hour-long sections of the recording during the dark and light cycles over five days (Fig. 3b). We compared the annotator-determined identities with those computed by our algorithm for one frame every 5 s during the annotated sections. Overall, 34,416 mouse images were manually annotated, which amounted to 12 h of annotated video (out of the total of 120).

Mice were correctly identified during non-huddling in 97.3% (19,649/20,193) of the images. Performance was approximately constant across the five days of the experiment. Identification errors (2.7%) were in part due to segmentation errors (Fig. 3c). Huddled events posed a much harder problem for our system; we found that 58% (8262/14,223) of those frames contained correct segmentation and correct identities, while 28% of the mouse images were poorly segmented, and 13% were properly segmented but were assigned incorrect identities. Thus, our system was capable of maintaining correct identities during active behaviour over days during both the dark and light cycles, and errors were almost entirely limited to mice that were huddled together and motionless.

To further evaluate the performance and generalisation of our system, we recorded 12 continuous hours of video of six mice in the imaging setup during a dark cycle. We ground-truthed the video by manually annotating mouse identities every 30 s regardless of huddling condition. Out of 8400 annotated mouse images, 99.4% were properly segmented and correctly identified, 0.3% were assigned incorrect identities and 0.3% were segmentation errors (Fig. 3d).

Fighting behaviour can often involve rapid movements, as mice jump and wrestle with each other. We identified several fighting bouts in one of our 5-day sequences by thresholding mouse velocity. Out of 10 randomly selected fighting bouts (four are shown in Supplementary Fig. 12), only 5% of the frames contained incorrect identities of the fighting mice. In all cases, identities were correct just before and just after the fight. Fights typically lasted 15–60 frames (0.5–2 s).

2.6. Development of social behaviour in wild-type mice

We characterised the behaviour of six sets of four C57BL/6J wild-type mice (two brothers and two sisters) over five days. Males and females had been housed separately prior to the experiment, which allowed us to observe how social hierarchies develop when mice are grouped together for the first time. At the beginning of the recording, the mice were added to a large (.6 m × .6 m × .6 m) home cage equipped with food, water, and two tube shelters (see Fig. 1a, Supplementary Fig. 7).

After capturing video for five days (12,960,000 frames), we used our system to compute the trajectories of each individual over the entire period. We analysed the trajectories by calculating statistics (places visited, velocity, and distance between mice) and detecting actions. For the latter task, we employed JAABA, a freeware software tool for detecting behaviours in animal trajectories (Kabra et al., 2013).

Fig. 4a shows how much time the mice in the first set spent at any given location in the enclosure. The four corners, the entrances to the tubes and inside the tubes were preferred locations (Fig. 4b). A similar pattern was observed across multiple experiments (Supplementary Fig. 8). Fig. 4c shows a histogram of time spent at these locations. We found that mice switch, as a group, between the two tubes during the light cycle (events marked by white arrows in Fig. 4c). We observed this phenomenon in all groups, and it appeared to be spontaneous and not associated with human presence or disturbance. Additionally, over days, the mice tended to spend more time at one of the corners (in this case, the bottom left corner, see Fig. 4c).

Overall, the mice spent less time at the corners compared to the tubes and tube entrances ($p < 0.0001$, U-test, Supplementary Fig. 9a). This was true for all mice in all experiments except one male in Experiment 5 (Supplementary Fig. 9a, fourth experiment column). Mice spent more time at the corners on the last day compared to the previous days ($p < 0.05$, U-test, Supplementary Fig. 9b).

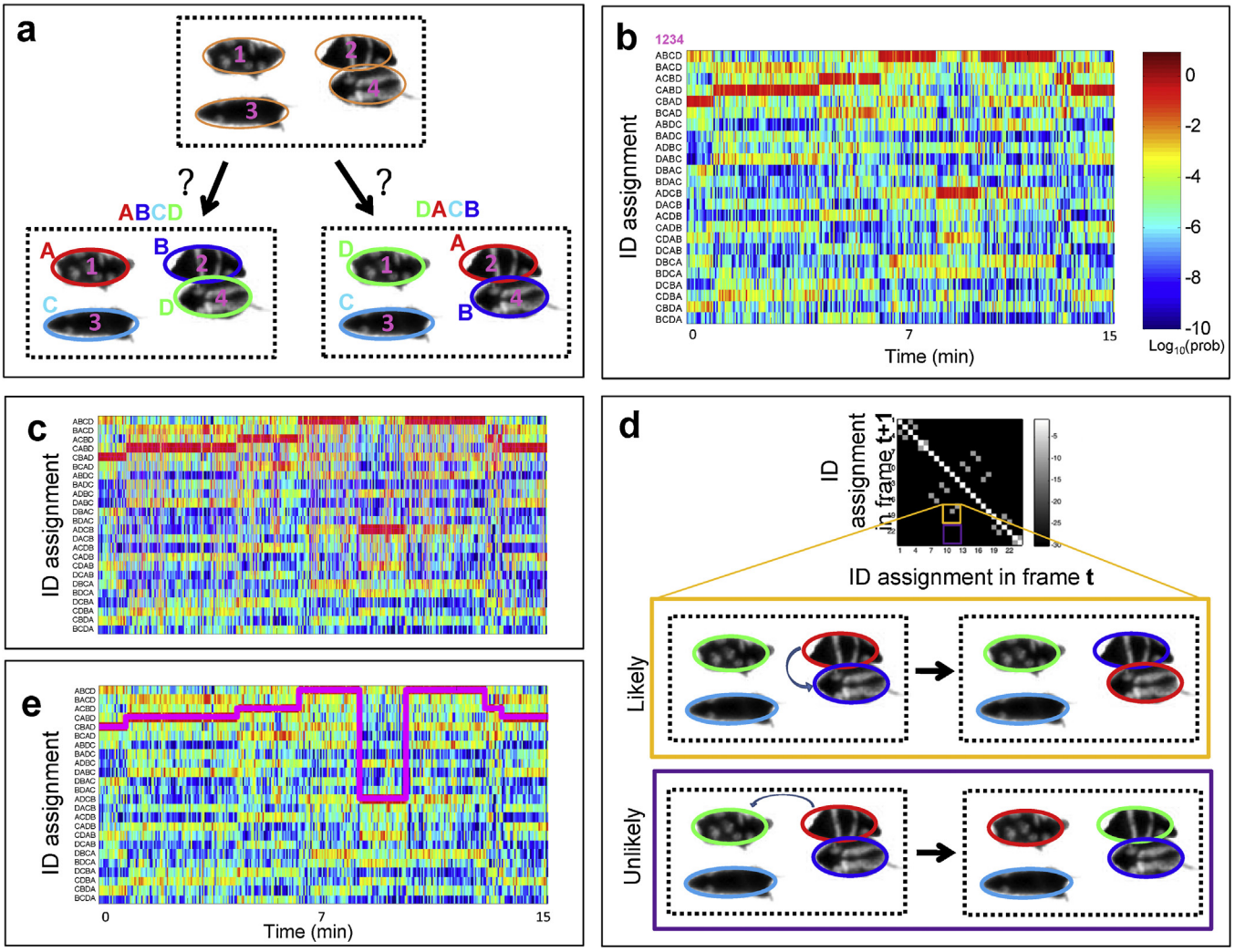

Fig. 2. Propagating identity information. (a) In each frame, the identities (identified by letters and corresponding colours) of the four tracked mice (identified by numbers) are unknown. Twenty-four identity assignments were possible (two are depicted). (b) Identity assignment probability matrix for a 15 min video (red denotes high likelihood). Each row represents a fixed identity assignment for each of the four tracked mice. Each column corresponds to a video frame. (c) Identities selected according to the maximum likelihood found in each frame, i.e., by the classifiers shown in Fig. 1. Notice the jagged solution, which suggests that the assignments were switched frequently (incorrectly) between different trajectories. (d) Identities can only change when mice are in close proximity. Some identity swaps were more likely than others, given the current identity assignment. For example, swapping of the red and blue identities was more likely (due to their proximity) than swapping of the red and green identities. (e) Identity likelihood computed by mouse classifiers was combined with mouse proximity using a hidden Markov model (HMM) to produce correct identity assignments over the entire video sequence (piecewise-constant pink trace).

To quantify how groups are formed and which groups formed most frequently, we counted all possible mice group configurations. We considered two mice to be in the same group if the minimal distance between their ellipses was smaller than half their body width. Given four mice, 15 group configurations that range from all mice forming a single group (Fig. 5a, first row, group configuration #1) to every mouse being in isolation are possible (Fig. 5a, last row, group configuration #15). We found that mice spent the majority of their time during the first dark cycle in isolation (Fig. 5b, top). However, this behaviour gradually changed, and mice spent less and less time in isolation over the next days. We found this trend to be significant ($p < 0.01$ one-way ANOVA). Two-way ANOVAs for each experiment with husbandry condition as a factor (standard or enriched) did not reveal any significant effect of rearing conditions on this behaviour ($p < 0.001$ for day, $p > 0.5$ for husbandry). We also observed a significant increase in the fraction of time the mice spent all together, and again, there was no difference between husbandry conditions (Fig. 5b, $p < 0.001$ for day, $p > 0.1$ for husbandry, two factor-ANOVA, Fig. 5b bottom). These changes in group composition suggest that the social relationships of the mice were developing continuously throughout the five-day experiment.

Preferred location and preferred associates in a group are passive proxies of social preference. To investigate active behaviours, we quantified social interaction by focusing on male following behaviour (e.g., both male-following-male and male-following-female; see Supplementary Text for further classifier details). An example of male following is shown in Fig. 6a. In both standard and enriched conditions, following behaviour was strongly circadian, with the vast majority of follows occurring during the dark cycle (Fig. 6b, $p < 0.006$). In all cases, the largest number of follow events occurred in the first dark cycle. In the enriched condition cages (Exp 4, 5 and 6) intermediate levels of following were maintained over the five days, while in two of the three standard condition cages (Exp 1 and 2), follow rates dropped to low levels after the first dark cycle, suggesting a reduction in social interaction in these cages. Follow durations and speed distributions were similar across experiments (see Supplementary Fig. 10a and b).

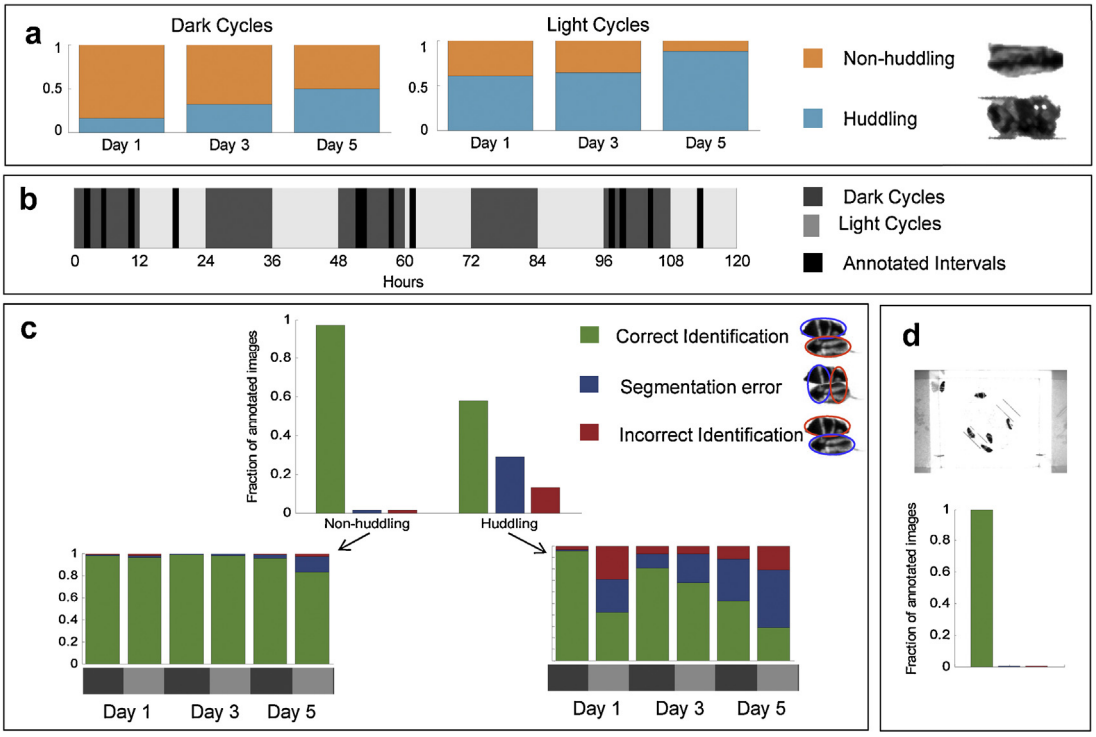

Fig. 3. Validation of identity assignments. (a) Fraction of time spent each day in a huddling configuration (orange) compared to a non-huddled configuration (cyan). Left: during dark cycles, right: during light cycles. (b) Intervals of the 5-day test sequence for which mouse identities were established by a trained human observer in Experiment 5 (black). Dark cycles are represented in dark grey, and light cycles are represented in light grey. (c) Performance of the tracking system measured in terms of correct identification (green), incorrect identification with correct segmentation (red) and incorrect segmentation (blue). The upper plot denotes the performance averaged across the entire five-day experiment broken down into huddling and non-huddling events. The bottom plot depicts performance as a function of day. (d) The performance of the system tracking six mice for 12 continuous hours during a dark cycle. Conventions are the same as (c). A frame from the video is shown.

It has been shown that male mice develop dominance relationships in which one male is both successful in agonistic interactions and has more mating opportunities (Dewsbury, 1981) and higher reproductive success (D'amato, 1988; Hurst et al., 1993). We wondered whether following behaviour would display a similar asymmetry between males and made the prediction that one male would do the majority of the following (i.e., following both the other male and the females). To explore this possibility, we developed the two following indices: the first was based on male–male following behaviour, and the second was based on male–female following behaviour (see Section 4). The male–male index was based on the amount of time each male spent following the other male such that a value of +1 indicates that all of the male–male follows were performed by male 1 following male 2, while a value of −1 indicates that all of the male–male follows were performed by male 2 following male 1. An example of the male–male index as a function of time is shown in Fig. 6c (open circles, data from Exp 1). The males began by following each other equally (index close to zero), but as time progressed, male 1 spent more time following male 2. The male–female index was computed similarly using the amount of time each male spent following the females (see Section 4). We also observed a gradual increase in the female follow index of male 1 over the first 12 h (Fig. 6c, filled circles).

We then plotted the male and female follow indices against each other for every hour to produce a follow index graph (see Fig. 6d). To simplify comparison across cages, we designated the male with the higher male–male index in the first 12 h as male 1 and the other as male 2. If the male–male and male–female indices are correlated and stable, all values of male and female follow indices should be greater than 0 and should result in points in the upper right-hand corner of the follow index graph (as in Fig. 6d, first dark cycle of Exp 1). The follow index graph for all six cages is shown in Fig. 6e. In all enriched cages (Exp 4–6), the male–male and male–female follow indices were greater than zero from the first block, indicating that a single male was responsible for the majority of both the male–male follows and the male–female follows, while all standard cages had values outside the upper right-hand corner in the first dark cycle, indicating that male–male behaviour and male–female behaviour were not completely correlated at first. By the end of the first dark cycle (12 h), however, all six cages had male and female follow indices in the upper right-hand corner.

The previous analysis focuses on the use of following behaviour, detected using the output of our tracker, to train a behavioural classifier. It is important to note that many different behaviours could easily be quantified using this system. For example, the system can also be used to detect simple behaviours such as walking (Kabra et al., 2013) or more complex behaviours such as mating events (see Supplementary materials).

3. Discussion

We developed a method for tracking multiple socially interacting, individually identified mice across multiple days that does not confuse their identities. Our system is fully automated and requires minimal human intervention. The software is open source and freely available at http://motr.janelia.org. Our method integrates information over time and reliably computes the identity of each mouse, even in video frames in which instantaneous identity is difficult to discriminate due to pattern occlusion or deformation. We demonstrated the applicability of our system by tracking several groups of four mice over a five-day period and observing how behaviour evolved over hours and days. To verify the applicability of our method to different numbers of mice, we computed

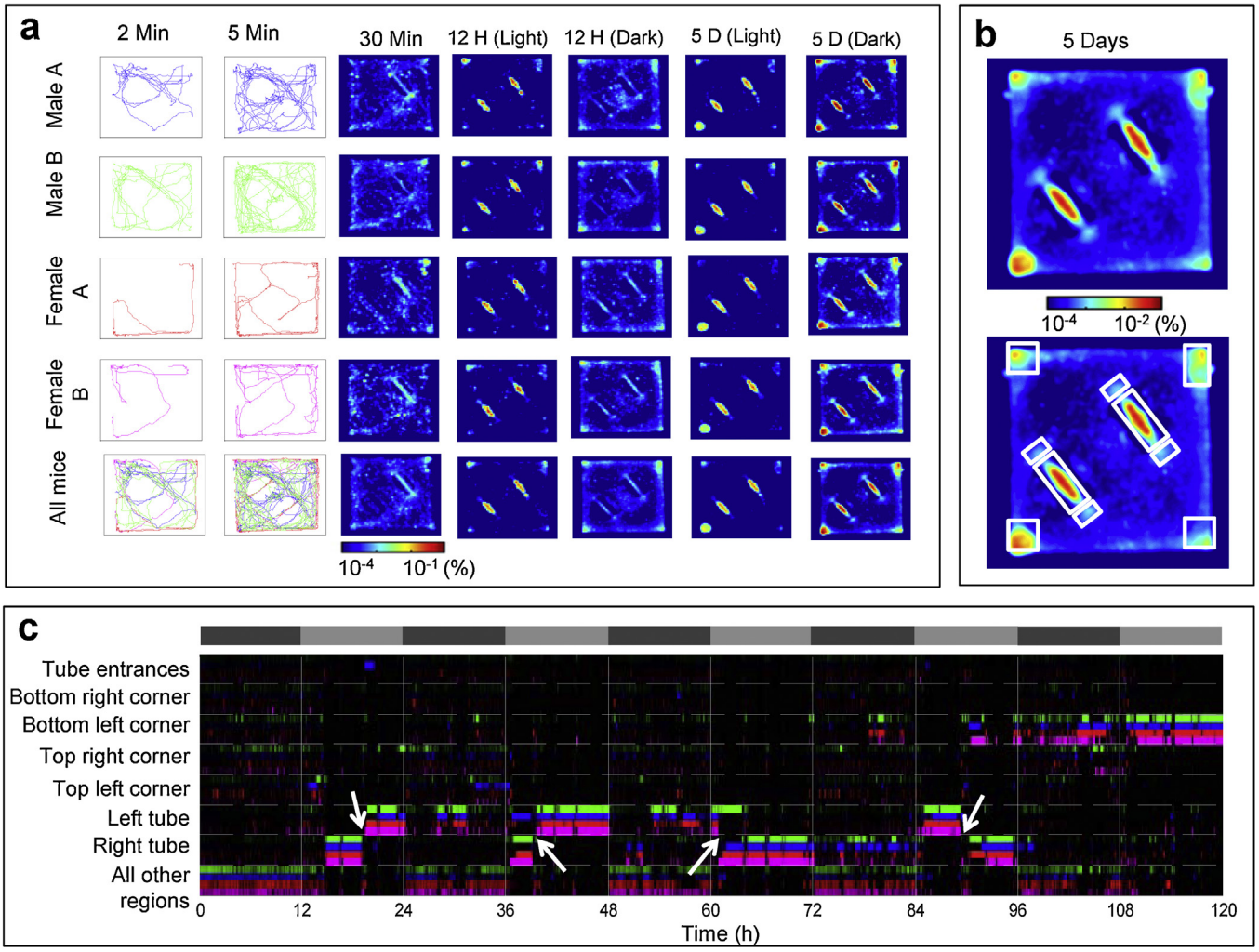

Fig. 4. Mouse trajectories and dwelling places for Experiment 5. (a) Example trajectories and position histograms for each individual mouse and for the entire group. Data are presented for 2 min, 5 min, 30 min, 12 h (first light cycle), 12 h (first dark cycle), all light cycles (5 days), and all dark cycles (5 days). Each coloured histogram was constructed by computing the percentage of time spent in a given pixel. Data were smoothed and are presented on a log scale for improved visualisation. (b) Two dimensional 2D position histogram for all mice (top) and selected monitored regions (bottom, highlighted in white). (c) Ethogram summarising the fraction of time each mouse spent in each of the monitored regions. Colour codes denote mice identities, similar to (a). White arrows denote events in which mice changed their sleeping place from one tube to the other.

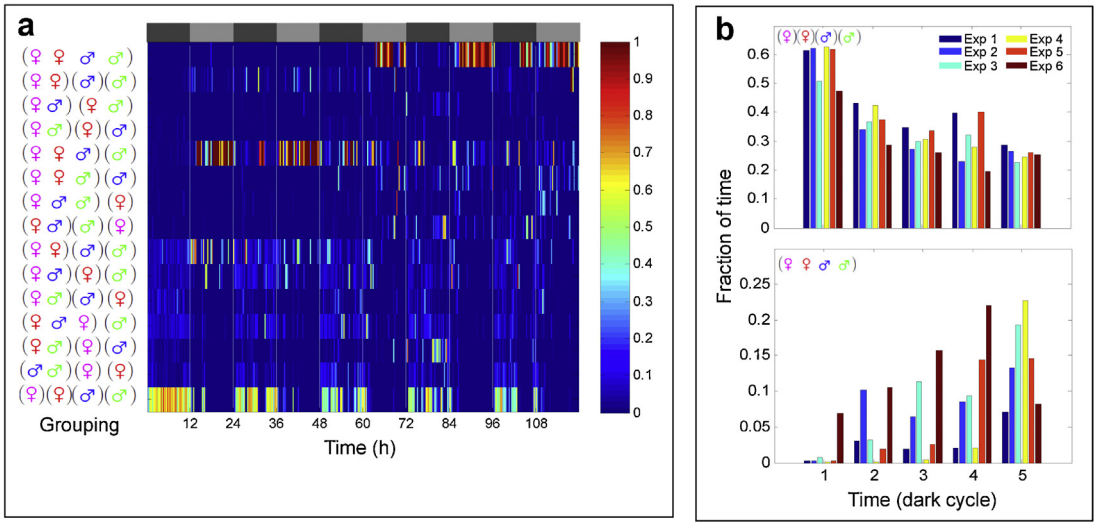

Fig. 5. Group configuration analysis. (a) Ethogram denoting the percentages of time spent in one of 15 possible group configurations. Group is denoted by the colour-coded male and female symbols on the left. Dark and light cycles are denoted by the grey bars on top. (b) Top: fraction of time spent during dark cycles in group configuration 15 (every mouse on its own). Each colour denotes a different five-day experiment. Bottom: fraction of time spent during dark cycles in group configuration 1 (all mice in a single group). (c) Difference in the fractions of time males 1 and 2 spent in a group with females. Each colour denotes a different experiment.

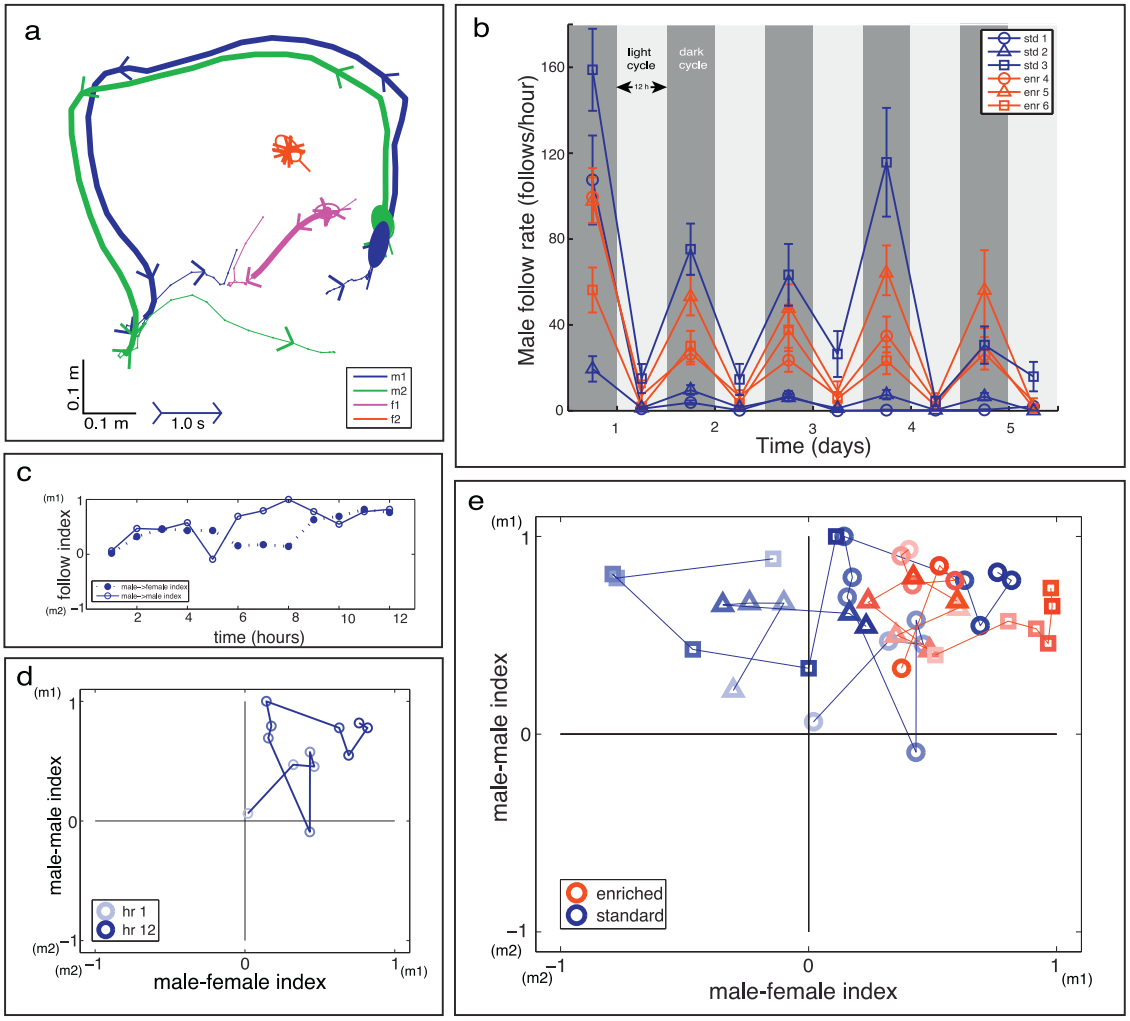

Fig. 6. Male following behaviour. (a) Example of male 1 (____) following male 2 (____). The trajectory line is thick during the following event and becomes thin at the end of the event, and the time between arrows is 1 s. Ellipses indicate the position of the four mice at the beginning of the follow event, and the sticks indicate the tails. The positions and movements of the female mice are indicated by the pink and red symbols. (b) Following rate as a function of time for all six experiments. (c) Example male–male follow (open circles) and male–female follow indices (filled circles) for the first dark cycle of Experiment 1. (d) Data from the first dark cycle of Experiment 1 (standard rearing conditions). Each hour of observation is represented by an open circle. The male follow index is plotted as a function of the female follow index; time is indicated by colour saturation, with more saturated colours representing later times. (e) Following data for all six experimental cages. Standard cages are in blue, and enriched cages are in red. All enriched cages were fully contained in the upper right-hand quadrant, while each standard cage produced data points that spilled into the other quadrants, indicating a more complex evolution of male–female and male–male social interaction patterns.

trajectories in a six-mouse cage and achieved excellent identification performance.

We measured proxies of social behaviour (preferred location, group setting, following) and found that they changed across days. Additionally, we found no differences between standard-reared and enriched-reared mice in simple social metrics, such as group association, but we found differences in more complex metrics, such as male and female following behaviour. The lack of differences between standard and enriched cages in simple association metrics may be due to the mice's tendency to associate with each other even across dominance relationships (Uhrich, 1938). This observation underscores the importance of quantitative and detailed behavioural descriptions in untangling social deficits. Such behaviour would be difficult to assess in a short-term experiment. Additionally, our method was able to demonstrate that animals that experienced enriched rearing environments more quickly adopted consistent social roles, an observation that has been previously made using labour-intensive manual scoring (Branchi et al., 2006).

Our method was designed with cost and reproducibility in mind. It is based on a single overhead camera to reduce the need to store and process multiple video feeds. Processing long videos (days) is fast on a large computer cluster, and shorter experiments (spanning a few hours) may be analysed on a single CPU.

The ability to correctly keep track of identities over long periods of time opens up a wide range of possibilities for developing new assays for the study of aggression and courtship. We expect that our system will be a valuable tool for genetic screening because it enables the examination of the effects of genetic, pharmacological and environmental manipulations on long-term social behaviour.

4. Materials and methods

4.1. Animals

Male and female C57Bl/6J mice (Jackson Labs) aged 6–17 weeks were used. Prior to recording, two female mice (sisters) and two male mice (brothers) were housed in separate cages. Mice were

raised in either standard or enriched conditions. Standard-reared mice were acquired from Jackson Labs at 3 weeks of age and housed in same-sex pairs (siblings) in large mouse cages until the recording session. Enriched-reared mice were born as the second of three litters into a large (0.61 m × 0.61 m × 0.61 m) population cage with two adult males and two adult females. Enriched-reared mice were removed from the population cage at 3 weeks of age and housed in same-sex pairs (siblings) in large mouse cages until the recording session.

We exposed the female mice in the study to bedding from the males to be used in the study at least 7 days prior to recording to ensure that the females were cycling regularly (Whitten, 1959). Vaginal smears from both females of each pair were then collected and used to determine their oestrus states. Recordings began when both females were in proestrus. Mice always had *ad libitum* access to food and water.

4.2. Fur patterns

Individually distinctive patterns were bleached into the fur of the mice. Mice were anaesthetised with isoflurane (2%) in an induction chamber. Lab tape was used to mask out a chosen pattern on the back of each anaesthetised mouse. Human hair bleach (Clairol Nice 'N Easy Born Blond Maxi) was mixed using the manufacturer's instructions. Bleach was applied only to the top of the fur to avoid irritating the skin. The tape was removed, and the mice were maintained under anaesthesia (1.5–2% isoflurane) for 20 min. The bleach was then rinsed thoroughly using warm water, the fur was dried and the mice were placed in a heated cage to recover from anaesthesia.

4.2.1. Mouse enclosure and recording equipment

Mice were housed in a 0.61 m × 0.61 m × 0.61 m polycarbonate population cage. Bedding was composed of a 25%/75% mix of corn cob and Alpha-Dri (Shepherd). Shelters for the mice were custom-made square-section tunnels made of IR-transparent acrylic (cylindrical-section tunnels distorted the image of the mice within the tunnel and degraded tracking performance). Video was recorded using an overhead Basler A622f monochrome 1394 camera (16 mm fixed focal length lens with a manual focus and iris, C-mount, 2/3″ format, F-stop: 1.4, filter: 25.5 mm, pitch: 0.5, graftek.com; part # HF16HA-1B). The camera was placed centrally, facing downwards, approximately 120 cm above the cage floor (see Supplementary Fig. 7). Illumination was provided by four infrared LED light sources placed adjacent to the camera (IR-LT30, 850 nm, 30° beam, Reytec Imaging). Because the mice were filmed continuously across multiple days and were on a 12 h day/night cycle, an infrared-pass filter (Hoya RM72 Infrared filter, B&H Photo; OIR7252) was used to minimise the effect of changes in ambient illumination on the recordings as the room lights were turned on and off. Video recording was monitored from an adjacent control room. Video (30 Hz, 1024 × 768 pixel image) was streamed continuously to an external hard drive using StreamPix 5 software (Norpix). Camera gains and black levels were adjusted prior to the experiments to obtain good contrast between the mice and the background without saturating the mice.

We recorded the groups of four mice for five days and then recorded the single-mouse videos used to train the mouse classifiers so that all mice would be new to the enclosure at the beginning of the experiment.

4.3. Huddled mice

We define an image of a mouse as "huddled" if the minimal distance between the mouse ellipse and the closest other ellipse was smaller than a pre-defined threshold, which was 6 mm, and if the mouse's velocity was smaller than 3 pixels/frame (7.2 cm/s).

4.4. Follow index

We define the male and female follow indices as follows:

$$\text{male} - \text{male follow index} = \frac{m1m2 - m2m1}{m1m2 + m2m1}$$

$$\text{male} - \text{female follow index} = \frac{m1f - m2f}{m1f + m2f}$$

where $m1m2$ is the amount of time male 1 spent following male 2, $m2m1$ is the amount of time male 2 spent following male 1, $m1f$ is the time male 1 spent following females and $m2f$ is the time male 2 spent following females.

4.5. Statistical methods

The duration and speed distributions of the follow events were compared using paired Kolmogorov–Smirnov tests with Bonferroni corrections for multiple comparisons. Comparisons of follow numbers were made with two-factor repeated measures ANOVAs.

Funding

This work is funded by NIH and the Howard Hughes Medical Institute.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jneumeth.2013.05.013.

References

Bishop CM. Pattern recognition and machine learning (Information Science and Statistics). Secaucus, NJ, USA: Springer-Verlag New York, Inc; 2006, ISBN 0387310738.

Blahut RE. Algebraic codes for data transmission. Cambridge University Press; 2003.

Branchi I, D'Andrea I, Fiore M, Di Fausto V, Aloe L, Alleva E. Early social enrichment shapes social behavior and nerve growth factor and brain-derived neurotrophic factor levels in the adult mouse brain. Biol Psychiatry 2006;60:690–6.

Branson K, Robie AA, Bender J, Perona P, Dickinson MH. High-throughput ethomics in large groups of *Drosophila*. Nat Methods 2009;6:451–7.

Burgos-Artizzu XP, Dollar P, Lin D, Anderson DJ, Perona P. Social behavior recognition in continuous video. In: Paper presented at: IEEE computer society conference on computer vision and pattern recognition; 2012.

D'amato FR. Effects of male social status on reproductive success and on behavior in mice (*Mus musculus*). J Comp Psychol 1988;102:146–51.

Dalal N, Triggs B. Histograms of oriented gradients for human detection. In: Paper presented at: International conference on computer vision & pattern recognition. San Diego: IEEE Computer Society; 2005.

Dankert H, Wang L, Hoopfer ED, Anderson DJ, Perona P. Automated monitoring and analysis of social behavior in *Drosophila*. Nat Methods 2009;6:297–303.

de Chaumont F, Coura RD, Serreau P, Cressant A, Chabout J, Granon S, et al. Computerized video analysis of social interactions in mice. Nat Methods 2012;9:410–7.

Dewsbury D. Social dominance, copulatory behavior, and differential reproduction in deer mice (*Peromyscus maniculatus*). J Comp Physiol Psychol 1981;95:880–95.

Fonio E, Golani I, Benjamini Y. Measuring behavior of animal models: faults and remedies. Nat Methods 2012;9:1167–70, http://dx.doi.org/10.1038/nmeth.2252.

Gordon RD. The Coccinellidae (Coleoptera) of America north of Mexico. J NY Entomol Soc 1985;93:1–912.

Hikida T, Jaaro-Peled H, Seshadri S, Oishi K, Hookway C, Kong S, et al. Dominant-negative DISC1 transgenic mice display schizophrenia-associated phenotypes detected by measures translatable to humans. Proc Natl Acad Sci USA 2007;104:14501–6.

Hurst J, Fang J, Barnard C. The role of substrate odours in maintaining social tolerance between male house mice *Mus musculus domesticus*. Anim Behav 1993;45:997–1006.

Jamain S, Radyushkin K, Hammerschmidt K, Granon S, Boretius S, Varoqueaux F, et al. Reduced social interaction and ultrasonic communication in a mouse model of monogenic heritable autism. Proc Natl Acad Sci USA 2008;105:1710–5.

Jhuang H, Garrote E, Mutch J, Yu X, Khilnani V, Poggio T, et al. Automated home-cage behavioural phenotyping of mice. Nat Commun 2010;1:68.
Kabra M, Robie AA, Rivera-Alba M, Branson S, Branson K. JAABA: interactive machine learning for automatic annotation of animal behavior. Nat Methods 2013;10:64–7.
Kooy RF, D'Hooge R, Reyniers E, Bakker CE, Nagels G, De Boulle K, et al. Transgenic mouse model for the fragile X syndrome. Am J Med Genet 1996;64:241–5.
Olson LE, Roper RJ, Baxter LL, Carlson EJ, Epstein CJ, Reeves RH. Down syndrome mouse models Ts65Dn, Ts1Cje, and Ms1Cje/Ts65Dn exhibit variable severity of cerebellar phenotypes. Dev Dyn 2004;230:581–9.
Penagarikano O, Abrahams BS, Herman EI, Winden KD, Gdalyahu A, Dong H, et al. Absence of CNTNAP2 leads to epilepsy, neuronal migration abnormalities, and core autism-related deficits. Cell 2011;147:235–46.
Poole T, Morgan D. Aggressive behaviour of male mice (*Mus musculus*) toward familiar and unfamiliar opponents. Anim Behav 1975;23:470–9.
Reeves RH, Irving NG, Moran TH, Wohn A, Kitt C, Sisodia SS, et al. A mouse model for Down syndrome exhibits learning and behaviour deficits. Nat Genet 1995;11:177–84.
Reiser M. The ethomics era? Nat Methods 2009;6:413–4.
Spencer CM, Graham DF, Yuva-Paylor LA, Nelson DL, Paylor R. Social behavior in Fmr1 knockout mice carrying a human FMR1 transgene. Behav Neurosci 2008;122:710–5.
Tremolizzo L, Carboni G, Ruzicka WB, Mitchell CP, Sugaya I, Tueting P, et al. An epigenetic mouse model for molecular and behavioral neuropathologies related to schizophrenia vulnerability. Proc Natl Acad Sci USA 2002;99:17095–100.
Uhrich J. The social hierarchy in albino mice. J Comp Psychol 1938;25:373–413.
Whitten WK. Occurrence of anoestrus in mice caged in groups. J Endocrinol 1959;18:102–7.
Zang JB, Nosyreva ED, Spencer CM, Volk LJ, Musunuru K, Zhong R, et al. A mouse model of the human Fragile X syndrome I304N mutation. PLoS Genet 2009;5:e1000758.