

Trainable, Vision-Based Automated Home Cage Behavioral Phenotyping

Hueihan Jhuang
hueihan@mit.edu

Estibaliz Garrote
estibaliz.garrote@gmail.com

Nicholas Edelman
edelmann@mit.edu

Tomaso Poggio
tp@ai.mit.edu

McGovern Institute of Brain Research,
Massachusetts Institute of Technology, Cambridge, MA, USA

Andrew Steele

Broad Fellows in Brain Circuitry Program,
Division of Biology, California Institute of
Technology, Pasadena, CA, USA
steelea@caltech.edu

Thomas Serre

McGovern Institute of Brain Research,
Massachusetts Institute of Technology,
Cambridge, MA, USA
serre@mit.edu

ABSTRACT

We describe a fully trainable computer vision system enabling the automated analysis of complex mouse behaviors. Our system computes a sequence of feature descriptors for each video sequence and a classifier is used to learn a mapping from these features to behaviors of interest. We collected a very large manually annotated video database of mouse behaviors for training and testing the system. Our system performs on par with human scoring, as measured from the ground-truth manual annotations of thousands of clips of freely behaving mice. As a validation of the system, we characterized the home cage behaviors of two standard inbred and two non-standard mouse strains. From this data, we were able to predict the strain identity of individual mice with high accuracy.

Author Keywords

Computer vision, behavior recognition, rodent, mouse, phenotyping.

ACM Classification Keywords

I.5.4 Pattern Recognition: Computer Vision

INTRODUCTION

Automated quantitative analysis of mouse behavior will play a significant role in comprehensive phenotypic

analysis – both on the small scale of detailed characterization of individual gene mutants and on the large scale of assigning gene functions across the entire mouse genome [1]. One key benefit of automating behavioral analysis arises from inherent limitations of human assessment: namely cost, time, and reproducibility. Although automation in and of itself is not a panacea for neurobehavioral experiments, it allows for addressing an entirely new set of questions about mouse behavior, such as conducting experiments on time scales that are orders of magnitude longer than traditionally assayed. For example, reported tests of grooming behavior span time scales of minutes whereas an automated analysis will allow for analysis of grooming behavior over hours or even days.

Most previous automated systems [3, 6] rely on the use of non-visual sensors (i.e. infrared beam) or video tracking techniques to monitor behavior. Such systems are particularly suitable for studies involving spatial measurements such as the distance covered by an animal or its speed. The physical measurements obtained from these sensor-based and tracking-based approaches limit the complexity of the behavior that can be measured. In particular, these approaches are not suitable for the analysis of fine animal behaviors such as grooming or micro-movements of the head. A few computer-vision systems for the recognition of mice behaviors have recently been described, including a commercial system (CleverSys, Inc) and two prototypes from academic groups [2, 9]. These computer-vision systems have not yet been tested in a real-world lab setting using long, uninterrupted video sequences containing potentially ambiguous behaviors. In addition, the systems have not been comprehensively evaluated against large, human annotated video databases containing different animals and different recording sessions.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.

In this paper, we describe a trainable, general-purpose, automated, and potentially high-throughput system for the behavioral analysis of mice in their home cage. Developed from a computational model of motion processing in the primate visual cortex [4], our system computes a sequence of feature descriptors for each input video based on the motion and position of the mouse. In the learning stage, a classifier is trained from manually annotated labels (behaviors of interest) and used to predict an output label for every frame of the video sequence. We compare the resulting system against human labeling and existing commercial software. We also discuss a range of applications demonstrating the flexibility of this approach.

EXPERIMENTS

All experiments involving mice were approved by the MIT and Caltech committees on animal care.

Behaviors of Interest and Definition

We annotate 8 types of common behaviors of inbred mice: drinking (defined by the mouse's mouth being juxtaposed to the tip of the drinking spout), eating (defined by the mouse reaching and acquiring food from the food bin), grooming (defined by the fore- or hind-limbs sweeping across the face or torso, typically as the animal is reared up), hanging (defined by grasping of the wire bars with the fore-limbs and/or hind-limbs with at least two limbs off the ground), rearing (defined by an upright posture and forelimbs off the ground), resting (defined by inactivity or nearly complete stillness), walking (defined by ambulation) and micro-movements (defined by small movements of the animal's head or limbs). Figure 1 illustrates these typical behaviors.

Video Datasets

In order to train a set of motion templates that are useful for discriminating between behavior categories, we manually collected a dataset (*clipped dataset*) consisting of 4,200 clips with the best and most exemplary instances of each behavior (each clip contains one single behavior). This set contains different mice (differing in coat color, size, gender,

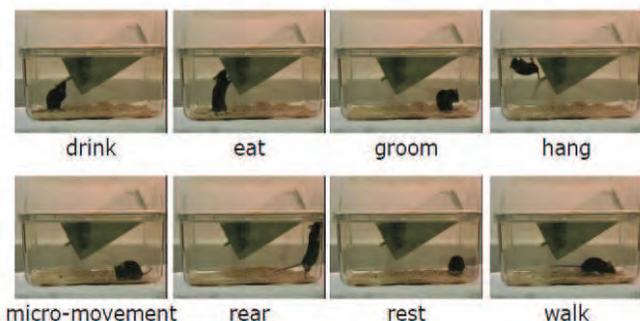


Figure 1. Snapshots taken from representative videos for the eight home cage behaviors of interest.

etc.) recorded at different times during day and night over 12 separate sessions.

Currently, the only public dataset for mice behaviors is limited in the scope [2]. In order to train and test our system on a real-world lab setting where mice behaviors are continuously observed and scored over hours or even days, we collected a second dataset (*full dataset*). This set contains 12 continuous labeled videos, in which each frame is assigned a behavior of interest. Each video is 30-60 minutes in length, resulting in a total of over 10 hours of continuously annotated videos. As in the *clipped dataset*, these videos are chosen from different mice at different times to maximize generalization of the dataset.

A team of 8 trained investigators ('Annotators group 1') manually annotated the videos. Two annotators of the 'Annotator group 1' performed a secondary screening on these annotations to correct mistakes and ensure the annotation style is consistent throughout the whole database. In order to measure the agreement between human labelers, we asked 4 of the original 8 investigators ('Annotators group 2') to label a subset of the already labeled videos (*doubly annotated dataset*). The *doubly annotated dataset* consists of many short video segments, which are randomly selected from the *full dataset*. Each segment in the *doubly annotated dataset* is 5-10 minutes long for a total of about 1.6 hours of video.

Training and Testing the System

The system computes two types of features for recognizing behaviors: the motion features developed by Jhuang et al. [4], as well as position and velocity features. Combining these two feature sets, the system learns a classifier that maps these features to the behaviors of interest.

Training based on the *clipped dataset* and the *full dataset* is done in two stages. In the first stage, we compute a set of 12,000 motion features on the *clipped dataset*. To reduce these 12,000 features to a more computationally tractable subset, we applied a feature selection technique called a zero-norm SVM [8] to select a subset (approximately 300) of the features that are most useful for discriminating between the behaviors categories. In the second stage, we compute the approximately 300 motion features, and the position and velocity features for the *full dataset*. The performance on the *full dataset* is evaluated using a leave-one-video-out cross-validation procedure: use all but one of the videos to train a classifier and the video not used in the training to evaluate the system. This process is repeated 12 times, once for each video. The system predictions for all the videos are then used to compute the accuracy as the percentage of frames correctly predicted by the system. Here a prediction or a label is 'correct' if it matches ground truth made by 'Annotators group 1'.

| | Our System | CleverSys Commercial System | Human ('Annotator Group 2') |
|---|------------|-----------------------------|-----------------------------|
| <i>Doubly annotated dataset</i> (1.6 hours of video) | 76.6% | 60.9% | 71.6% |
| <i>Full dataset</i> (10 hours of video) | 77.6% | 61.0% | N/A |

Table 1. Accuracy of our system, human annotators and HomeCageScan 2.0 CleverSys system evaluated on the doubly annotated dataset and the whole set of full dataset for the recognition of 8 behaviors.

For the system's classifier, we used the Hidden Markov Support Vector Machine (SVMHMM) [5]. The SVMHMM models the temporal dependence between behaviors (i.e. drinking is unlikely to occur directly following hanging), and thus performed much better than a simple SVM classifier that evaluates each frame in isolation.

Comparison with Commercial Software and Humans

Here we evaluate the system performance on the *doubly annotated dataset* and the *full dataset*. The system is compared against commercial software (HomeCageScan 2.0, CleverSys, Inc) for mouse home cage behavior classification and against human manual scoring. The *doubly annotated dataset* benchmarks the agreement between human annotators. Table 1 shows the comparison. Overall, we found that our system achieves 76.6% agreement with human labelers ('Annotators group 1') on the *doubly annotated dataset*. The agreement is significantly higher than the 60.9% agreement between HomeCageScan 2.0 system and 'Annotators group 1'. Our system agreement is on par with the 71.6% agreement between human labelers, defined as the agreement between 'Annotators group 1' and 'Annotators group 2' on the *doubly annotated dataset*. Two online videos demonstrating the automatic scoring of the system are available at <http://techtv.mit.edu/videos/5561> and <http://techtv.mit.edu/videos/5562>.

Running Time of the System

For performance reasons, much of the system is implemented on a graphical processing unit (GPU). After preprocessing, feature computation and classification run in nearly real time (30 frames per second).

Identifying Strain Based on Behavior

To demonstrate the applicability of this vision-based approach to large-scale phenotypic analysis, we characterized the home cage behavior of 4 strains of mice, including the wild-derived strain CAST/EiJ, the BTBR strain (a potential model of autism [7]), and two of the most popular inbred mouse strains, C57BL/6J and DBA/2J. We

recorded video of seven mice of each strain during one 24-hour session, encompassing a complete light-dark cycle.

From these videos, we computed patterns of behaviors for each mouse. We segmented the predictions for each 24-hour video into four non-overlapping 6-hour long segments (corresponding to the first and second halves of the night, and the first and second halves of the day). For each segment, we calculated the histogram of each behaviors type (walking, hanging, etc.). The resulting 8-dimensional (one for each behavior) vectors of the four segments were then concatenated to obtain one single 32-dimensional vector (8 dimensions x 4 segments) as the pattern of behavior for each animal. The pattern of behavior corresponds to the relative frequency of each of the eight behaviors of interest, as predicted by the system, over a 24-hour period. Using a leave-one-animal-out cross-validation procedure, we found that the resulting support vector machine (SVM) classifier predicted the strain of all animals with 90% accuracy.

Training the System to Handle More Complex Behaviors

To train and evaluate the performance of the system we chose the eight behaviors described above to capture standard home cage behaviors. We next asked if the system could be extended to other, more complex behaviors based on the same motion features.

We collected a new set of videos for an entirely new set of behaviors corresponding to animals interacting with "low profile" running wheels. This wheel-interaction set contains 13 fully annotated one-hour videos taken from six C57BL/6J mice. The four actions of interest are as follows: running on the wheel (defined as all 4 paws on the wheel and the wheel to be rotating), interacting with the wheel but not running (any other behavior on the wheel), awake but not interacting with wheel, and rest outside the wheel. These actions are shown in the video available at <http://techtv.mit.edu/videos/5567>. Using the leave-one-video-out cross-validation procedure as in the *full dataset*, the system achieves 92.8% of accuracy.

Although we certainly cannot generalize to all types of behaviors, the wheel results demonstrate that for many typical mouse behaviors no additional features need to be designed: the system learns new actions from annotated examples of new behaviors.

We have on-going work in the monitoring and analysis of abnormal behaviors in the context of neurological disorders. In particular, we are currently training the system to automatically detect and rate the severity of dyskinetic movements in the context of Parkinson's disease. Results will be presented at the meeting.

CONCLUSION

We have applied a biological model of motion processing to the recognition of mice behaviors. For common behaviors of interest, the system achieves performance on

par with human observers. The system demonstrates the promise of learning-based and vision-based techniques in complementing existing approaches towards a complete quantitative phenotyping of complex behavior.

ACKNOWLEDGMENTS

This research would not have been possible without the dedicated work of our data collectors and annotators: Jessi Ambrose, Andrea Farbe, Cynthia Hsu, Alexandra Jiang, Grant Kadokura, Xuying Li, Anjali Patel, Kelsey Von Tish, Emma Tolley, Ye Yao and Eli T Williams.

This research was sponsored by grants from DARPA (IPTO and DSO), NSF-0640097, NSF-0827427, IIT, and the McGovern Institute for Brain Research. Andrew Steele was funded by the Broad Fellows Program in Brain Circuitry at Caltech. Hueihan Jhuang was funded by the Taiwan National Science Council (TMS-094-1-A032).

REFERENCES

1. J. Auwerx and A. El. The European dimension for the mouse genome mutagenesis program. *Nature Genetics*, 36, 9 (2004), 925-927.
2. P. Dollar, V. Rabaud, G. Cottrell, and S. Belongie. Behavior recognition via sparse spatio-temporal features. *Visual Surveillance and Performance Evaluation of Tracking and Surveillance*, 2005.
3. E. H. Goulding, A. K. Schenk, P. Juneja, A. W. Mackay, J. M. Wade, and L. H. Tecott. A robust automated system elucidates mouse home cage behavioral structure. *Proceedings of the National Academy of Sciences USA*, 105, 22 (2008) 20575–20582.
4. H. Jhuang, T. Serre, L. Wolf, and T. Poggio. A biologically inspired system for action recognition. *IEEE International Conference on Computer Vision*, 2007.
5. T. Joachims. http://www.cs.cornell.edu/People/tj/svm_light/svm_hmm.html. Version V3.10, 2008.
6. L. P. Noldus, A. J. Spink, and R. A. Tegelenbosch. EthoVision: a versatile video tracking system for automation of behavioral experiments. *Behavior Research Methods, Instruments, Computers*, 33, 3 (2001), 398-414.
7. H. G. Mcfarlane, G. K. Kusek, M. Yang, J. L. Phoenix, V. J. Bolivar, and J. N. Crawley. Autism-like behavioral phenotypes in BTBR T1tf/J mice. *Genes, Brain and Behavior*, 7 (2008), 152-163,.
8. J. Weston, S. Mukhrejee, O. Chapelle, M. Pontil, T. Poggio, V. Vapnik. Feature Selection for SVMs. *Advances in Neural Information Processing*, 2001.
9. X. Xue and T. C. Henderson. Feature fusion for basic behavior unit segmentation from video sequences. *Robotics and Autonomous Systems*, 57, 3 (2009), 239-248.