Computer Science and Artificial Intelligence Laboratory
Technical Report

MIT-CSAIL-TR-2009-052
CBCL-283

October 26, 2009

Automated home-cage behavioral phenotyping of mice

Thomas Serre, Hueihan Jhuang, Estibaliz Garrote, Xinlin Yu, Vinita Khilnani, Tomaso Poggio, and Andrew D. Steele

massachusetts institute of technology, cambridge, ma 02139 usa — www.csail.mit.edu

CSAIL

AUTOMATED HOME-CAGE BEHAVIORAL PHENOTYPING OF MICE

Thomas Serre*[1], Hueihan Jhuang*[1], Estibaliz Garrote[1], Xinlin Yu[2], Vinita Khilnani[2], Tomaso Poggio[1] and Andrew D. Steele[2]

* The two authors contributed equally.

[1]Department of Brain and Cognitive Sciences, McGovern Institute, Massachusetts Institute of Technology

[2]Division of Biology, California Institute of Technology

Corresponding authors: Thomas Serre serre@mit.edu or Andrew Steele steelea@caltech.edu

ABSTRACT

We describe a trainable computer vision system enabling the automated analysis of complex mouse behaviors. We provide software and a very large manually annotated video database used for training and testing the system. Our system outperforms leading commercial software and performs on par with human scoring, as measured from the ground-truth manual annotations of thousands of clips of freely behaving animals. We show that the home-cage behavior profiles provided by the system is sufficient to accurately predict the strain identity of individual animals in the case of two standard inbred and two non-standard mouse strains. Our software should complement existing sensor-based automated approaches and help develop an adaptable, comprehensive, high-throughput, fine-grained, automated analysis of rodent behavior.

INTRODUCTION

Automated quantitative analysis of mouse behavior will play a significant role in comprehensive phenotypic analyses – both on the small scale of detailed characterization of individual gene mutants and on the large scale of assigning gene function across the entire mouse genome (Auwerx, Avner et al. 2004). One of the key benefits of automating behavioral analyses 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 (Crabbe, Wahlsten et al. 1999), it allows for addressing an entirely new set of questions about mouse behavior. Indeed, the significance of alterations in home cage behavior has recently gained attention as an effective means to detect perturbations in neural circuit function – both in the context of disease detection and more generally to measure food consumption and activity parameters (Dell'Omo, Vannoni et al. 2002; Chen, Steele et al. 2005; Steele, Jackson et al. 2007; Goulding, Schenk et al. 2008; Roughan, Wright-Williams et al. 2008). Another benefit of automated analysis of behavior is the ability to conduct experiments on time scales that are orders of magnitude larger than traditionally assayed. For example, reported tests of grooming behavior span time scales of the order of minutes (Greer and Capecchi 2002; McFarlane, Kusek et al. 2008) whereas an automated analysis will allow for analysis of this behavior over hours or even days. Most previous automated systems (e.g., (Noldus, Spink et al. 2001; Dell'Omo, Vannoni et al. 2002; Jackson, Tallaksen-Greene et al. 2003; Goulding, Schenk et al. 2008), see also Supplementary Text online) have relied on the use of sensors to monitor behavior. However the physical measurements obtained from these sensor-based approaches limit the complexity of the behavior that can be measured. This problem remains even for expensive commercial systems using transponder technologies such as the IntelliCage system by NewBehavior Inc. While such systems can be effectively used to monitor the locomotion activity of an animal as well as other pre-programmed activities via operant conditioning units located in the corners of the cage, they cannot be directly used to study natural behaviors such as grooming, hanging or rearing.

Recent advances in computer vision and machine learning yielded robust computer vision systems for the recognition of objects (Viola and Jones 2001; Dalal and Triggs 2005) and human actions (see ref. (Moeslund, Hilton et al.) for review). The use of vision-based approaches is already bearing fruit for the automated tracking (Khan, Balch et al. 2005; Fry, Rohrseitz et al. 2008; Veeraraghavan, Chellappa et al. 2008) and recognition of behaviors in insects (Branson, Robie et al. 2009; Dankert, Wang et al. 2009). Several open-source and commercial computer-vision systems for the tracking of animals have been developed (e.g., ref. (Noldus, Spink et al. 2001; Spink, Tegelenbosch et al. 2001; Branson and Belongie 2005), see also Supplementary Text online). Such systems are particularly suitable for studies involving spatial measurements such as the distance covered by an animal or its speed. These tracking techniques have the same limitations as the sensor-based approaches and are not suitable for the analysis of fine animal behaviors such as micro-movements of the head, grooming or rearing.

A few computer-vision systems for the recognition of rodent behaviors have been recently described, including a commercial system (CleverSys, Inc) and two prototypes from academic groups (Dollar, Rabaud et al. 2005; Xue and Henderson 2009). They have not been tested yet in a real-world lab setting using long uninterrupted video sequences and containing potentially ambiguous behaviors or at least comprehensively evaluated against human manual annotations on large databases of video sequences using different animals and different recording sessions. In this paper, we describe a trainable, general-purpose, automated, quantitative and potentially high-throughput system for the behavioral analysis of rodents in their home-cage. We characterize its performance against human labeling and other systems. We make available ready-to-use software (under the GPL license). We also provide a very large database of manually annotated video sequences of mouse behaviors, in an effort to motivate further work and set benchmarks for evaluating progress in the field.

Our system, which is a development based on a computational model of motion processing in the primate visual cortex (Giese and Poggio 2003; Jhuang, Serre et al. 2007), consists of a few main steps: first a video input sequence is converted into a representation suitable for the accurate recognition of the underlying behavior based on the detection of space-time motion templates. After this pre-processing step a statistical classifier is trained from labeled examples with manually annotated behaviors of interest and used to analyze automatically new recordings containing hours of freely behaving animals. The full system provides an output label (behavior of interest) for every frame of a video-sequence. The resulting time sequence of labels can be further used to construct ethograms of the animal behavior and fit statistical models to characterize behavior. As a proof of concept we trained the system on eight behaviors of interest (eat, drink, groom, hang, micro-move, rear, rest and walk, see Fig. 1 for an illustration) and demonstrate that the resulting system performs on par with humans for the scoring of these behaviors. Using the resulting system, we analyze the home-cage behavior of several mouse strains, including the commonly used strains C57Bl/6J, DBA/2J, the BTBR strain that displays autistic-like behaviors, and a wild-like strain CAST/EiJ. We characterize differences in the behaviors of these strains and use these profiles to predict the strain type of an animal blindly.

Results

System overview

Our system (**Supplementary software** online) consists of three separate modules: (1) a video database, (2) a feature computation module, and (3) a classification module. We recorded a large database of video sequences of mice in their home cage using a consumer grade camcorder. These videos were then manually hand scored and used to train a computer vision system to recognize behaviors of interest. In this system, a set of about 300 distinct motion features that are based on a biological model of motion analysis in the primate cortex (Giese and Poggio 2003; Jhuang, Serre et al. 2007) are computed to convert an input video sequence into a representation, which is then used for the automated recognition of the animal's behavior by a statistical classifier.

Video database

We video recorded singly housed mice from an angle perpendicular to the side of the cage (see Fig. 1 for examples of video frames). In order to create a robust detection system we varied the camera angles as well as the lighting conditions by placing the cage in different positions with respect to the overhead lighting. In addition, we used many mice of different size, gender, and coat color. Several investigators were trained to score the mouse behavior using two different scoring techniques. The first type of annotations denoted as the *'clipped database'* included only clips scored with very high stringency, seeking to annotate only the best and most exemplary instances of particular behaviors. Through this style of annotation we created more than 9,000 short clips, each containing a unique annotation. To avoid errors, this database was then curated by a second set of human annotators who watched all 9,000 clips again, retaining only the most accurate and unambiguous assessments, leaving 4,200 clips (26,2360 frames corresponding to about 2.5 hours) from 12 distinct videos to train and tune the motion feature extraction module of our computer algorithm described below. This database is significantly larger than the currently publicly available dataset (Dollar, Rabaud et al. 2005), which contains only 5 behaviors (eating, drinking, grooming, exploring and resting) for a total of 435 clips.

The second database, called the *'full database'* involved labeling every frame (with less stringency than in the clipped database) for 12 unique videos corresponding to over 10 hours of continuously annotated video. This database was used to train and test the classification module of our computer algorithm described below. Databases such as this are not currently available. By making such database available to reliably estimate and compare the performance of vision-based systems, we hope to further motivate the development of such computer vision systems for behavioral phenotyping applications. Fig. S1 shows the distribution of labels for the clipped and the full database.

Feature computation

Motion features. The hierarchical architecture used here to pre-process raw video sequences (see Fig. 2 inset A) and extract motion features (see Fig. 2 inset C) is adapted from our previous work for the recognition of biological motion (Jhuang, Serre et al. 2007). The system is based on the organization of the dorsal stream of the visual cortex, which has been critically linked to the processing of motion information (see ref. (Born and Bradley 2005) for a recent review). Details of the implementation are described in the **Supplementary Methods** online. A hallmark of the system is its hierarchical architecture, which builds a loose collection of 3-D space-time video patches (called interchangeably "motion-features" in the following) and centered on each frame of a video sequence. The model starts with an array of spatio-temporal filters tuned to 4 directions of motion and modeled after motion-sensitive (simple) cells in the primary visual cortex (V1) (Simoncelli and Heeger 1998) (S1/C1 layers, see Fig. 2 inset D). The architecture then extracts space-time motion primitives centered at every frame of an input video sequence via a hierarchy of processing stages, whereby features become increasingly complex and invariant with respect to 2D transformations. These motion features (see Fig. 2 inset F) are obtained by combining the response of V1-like afferent motion units that are tuned to different directions of motion (see Fig. 2 inset E and **Supplementary Methods** for details).

The output of this hierarchical pre-processing module consists of a basic dictionary of about 300 space-time motion features (S2/C2 layers, see Fig. 2 inset D) that can be used by a statistical classifier to reliably classify every frame of a video sequence into a behavior of interest. This basic dictionary of motion-feature templates corresponds to discriminant motion features as learned from a training set of videos containing behaviors of interest (the 'clipped database') via a feature selection technique (see **Supplementary Methods**).

Training of the motion feature computation module. To optimize the performance of the system for the recognition of mouse behaviors, several key parameters of the model were adjusted. The parameters of the spatio-temporal filters in the first stage (e.g., their preferred speed tuning and direction of motion, the nature of the non-linear transfer function used, the video resolution, etc) were adjusted so as to maximize performance on the 'clipped database'. This was done by training and testing a multi-class Support Vector Machine (SVM) classifier on single frames via a leave-one-out procedure (see **Methods**) as performed previously (Jhuang, Serre et al. 2007).

Position- and velocity-based feature computation. In addition to the motion features described above, we computed an additional set of features derived from the instantaneous location of the animal in the cage (see Fig. 2 inset C). To derive these features, we first computed a bounding box for the animal by subtracting off the video background. For a static camera as used here, the video background can be well approximated by a median frame obtained after computing the median value across all frames (day and night frames under red lights were processed in separate videos). Position- and velocity-based measurements were estimated based on the 2D coordinates ***(x,y)*** of this bounding box for every frame. These included the position and the aspect ratio of the bounding box (indicating whether the animal is in a horizontal or vertical position), the distance of the animal from the feeder as well as the instantaneous velocity and acceleration. Fig. 2 (inset C) provides a description of the 10 position- and velocity-based features.

Frame-based evaluation of the system. In order to evaluate the quality of our motion features for the recognition of high-quality unambiguous behavior we trained and tested a linear Support Vector Machine (SVM) classifier on *single* frames from the clipped database (using the all-pair multi-class classification strategy). This approach does not rely on the temporal context of behaviors beyond the computation of low-level motion signals and classifies each frame independently. On the clipped database, we find that such a system led to 93% accuracy (motion features alone; chance level 12.5% for 8-class classification, see Supplementary Methods online for a comparison with a representation computer vision system). Performance here was estimated based on a leave-one-out procedure, whereby the clips from all except one video are used to train the system while performance is evaluated on the clips from the remaining video. The procedure is repeated n times for all videos and the average performance is reported (see **Methods** for details). This suggests that the representation provided by the dictionary of 300 motion features is suitable for the recognition of the behaviors of interest even under conditions where the temporal structure of the underlying temporal sequence is completely discarded. We also found that the addition of the position- and velocity-based features led to an improvement in recognition for behaviors that are dependent upon the location of the animal with respect to the environment (e.g., drinking occurs at the water bottle spout while eating occurs at the food bin by our definitions).

Classification module

Performing a reliable phenotyping of an animal requires more than the mere detection of stereotypical non-ambiguous behaviors. In particular, the present system aims at classifying every frame of a video sequence even for those frames that are difficult to categorize. For this challenging task, the temporal context of a specific behavior becomes an essential source of information; thus, learning an accurate temporal model for the recognition of actions becomes critical (see Supp. Fig. 3 for an illustration). Here we used a Hidden Markov Support Vector Machine (HMMSVM, see Fig. 2 inset G) (Altun, Tsochantaridis et al. 2003; Tsochantaridis, Hofmann et al. 2004; Tsochantaridis, Joachims et al. 2005; Joachims, Finley et al. 2009), which is an extension of the popular Support Vector Machine classifier for sequence tagging. This temporal model was trained on manually labeled examples extracted from about 10 hours of *continuously scored* video sequences from 12 separate videos from the 'full database' as described above.

A comparison between the resulting system and a leading commercial software (HomeCageScan 2.0, CleverSys, Inc) for mouse home cage behavior classification against human manual scoring is provided in Table I (see **Methods** for details) and Figure 3. Here we considered two sets of manual annotations: one that denoted 'set A', where every frame has been scored by one human annotator. As described above this set contains over 10 hours of videos containing different mice (different coat color, size, gender, etc) recorded at different times during day and night during 12 separate sessions. In addition, we considered a small subset of this database (denoted 'set B') corresponding to many short random segments from each of the 12 videos manually annotated by pairs of randomly selected annotators from a pool of 12 annotators (about 5-10 min in length for a total of about 1.6 hours). 'Set B' allowed us to evaluate the agreement between two independent labelers, which we estimated to be 71.6% (frame by frame agreement between both annotators). This level of agreement sets the upper bound of performance from the system since it relies completely on these manual annotations to learn to recognize behaviors. Overall we found that our system achieves 71.0% agreement with manual annotations on set B, a result significantly higher than the HomeCageScan 2.0 (56.0%) system and on par with humans (71.6%). Fig. 3 shows confusion matrices for the inter-human agreement, the proposed computer system and the HomeCageScan system.

Characterizing the home-cage behavior of diverse inbred mouse strains

To demonstrate the applicability of this vision-based approach to large-scale phenotypic analysis, we characterized the home-cage behavior of four strains of mice, including the wild-like strain CAST/EiJ, the BTBR strain, which is a potential model of autism (McFarlane, Kusek et al. 2008) as well as two of the most popular inbred mouse strains C57Bl/6J and DBA/2J. We video recorded n=7 mice of each strain during one 24-hour session, encompassing a complete light-dark cycle. An example of an ethogram obtained over a 24-hour continuous recording period for one of the CAST-EiJ (wild-like) strain is shown in Fig. 2 (inset G). One obvious feature was that the level of activity of the animal decreased significantly during the day (12-24 hr) as compared to night time (0-12hr). The mean activity peak of the CAST/EiJ mice shows a much higher night activity peak in terms of walking and hanging than any of the other strains tested (Fig. 4). As compared to the CAST/EiJ mice, DBA/2J strain showed an equally high level of walking at the beginning of video recording but this activity quickly dampened to that of the other strains C57Bl/6J and BTBR. We also found that the resting behavior of this CAST/EiJ strain differed significantly from the others: while all four strains tended to rest for the same total amount of time (except BTBR which rested significantly more than C57Bl/6J), we found that the CAST/EiJ tended to have resting bouts that lasted almost three times longer than those of any other strain (Fig. 5A-B).

As BTBR mice have been reported to hyper-groom (McFarlane, Kusek et al. 2008) we next examined the grooming behavior of BTBR mice. Following the study of McFarlane et al. (McFarlane, Kusek et al. 2008), which scored grooming manually, our system detected that the BTBR strain spent approximately 900 seconds grooming compared to the C57Bl/6J mice which spend a little more than 600 seconds grooming (Fig. 5C). These values were reproduced by a human observer scoring the same videos (Fig. 5C). Here we show that using our system we were able to reproduce the key results that the BTBR strain grooms more than the C57Bl/6J strain when placed in a novel cage environment as in McFarlane *et al*. Note that in the present study the C57Bl/6J mice were approximately 90 days old (+/- 7 days) while the BTBR mice were approximately 80 days old (+/-7 days). In the McFarlane study younger mice were used (and repeated testing was performed), but our results essentially validate their findings.

Prediction of strain-type based on behavior

To visualize the similarities/dissimilarities between patterns of behaviors exhibited by all 28 individual animals used in our behavioral study, we performed a non-metric Multidimensional Scaling (MDS). These patterns of behaviors correspond to the relative frequency of each of the 8 behaviors of interest over a 24-hour period (see **Methods** for details). Fig. 5D shows the resulting 28 data-points each corresponding to a different animal projected along the first 3 dimensions returned by the MDS analysis. Already in this relatively low dimensional space, individual animals tend to cluster by strains suggesting that different strains exhibit unique patterns of behaviors that are characteristic of their strain-type. To quantify this statement, we trained a linear SVM classifier directly on the patterns of behaviors predicted by the system (see **Methods**). Using a leave-one-animal-out procedure, we found that the resulting classifier was able to predict the strain of all animals with an accuracy of 90%. Fig. 5E shows a confusion matrix for the corresponding classifier that indicates the probability with which an input pattern of behavior (along the rows) was classified as each of the 4 strains (along the columns). The higher probabilities along the diagonal and the lower off-diagonal values indicate successful classification for all strains.

DISCUSSION

In this paper we describe the development and implementation of a trainable computer vision system capable of capturing the behavior of a single mouse in the home-cage environment. Importantly, as opposed to several proof-of-concept computer vision studies (Dollar, Rabaud et al. 2005; Jhuang, Serre et al. 2007), our system has been demonstrated with a "real-world" application, characterizing the behavior of several mouse strains and discovering strain-specific features. We provide software as well as the large database that we have collected and annotated in hope that it may further encourage the development of similar vision-based systems.

The search for "behavioral genes" requires cost effective and high-throughput methodologies to find aberrations in normal behaviors (Tecott and Nestler 2004). From the manual scoring of mouse videos (see 'full database' above), we have estimated that it requires about 22 person hours of work to manually score every frame of a one-hour video. Thus, we estimate that the 24-hour behavioral analysis conducted above with our system for the 28 animals studied would have required almost 15,000 person hours (i.e., almost 8 years of work for one person working full-time) of manual scoring. An automated computer-vision system permits behavioral analysis that would simply be impossible using manual scoring by a human experimenter. The system is currently not real-time (it takes about 10 sec to process 1 sec of video). We were able, however, to obtain real-time performance with an initial system implementation ported to a GPU (NVIDIA GTX 295) (Mutch & Poggio, in prep).

In principle, our approach can be extended to other behaviors such as dyskinetic behaviors in the study of Parkinson's disease models, seizures for the study of epilepsy, or even wheel running behavior in the context of a normal home cage repertoire. Future developments of our learning and vision approach could deal with the quantitative characterization of social behavior involving two or more freely behaving animals. In conclusion, our study shows the promise of learning-based and vision-based-techniques in complementing existing approaches towards a complete quantitative phenotyping of complex behavior.

METHODS

Mouse strains, behavioral experiment, and data collection

All experiments involving mice were approved by the MIT and Caltech committees on animal care. For generating training and testing data we used a diverse pool of hybrid and inbred mice of varying size, age, gender, and coat color (both black and agouti coat colors). In addition, varied lighting angles and conditions, using 'light' and 'dark' recording conditions where we used dim red lighting (30 Watt bulbs) to allow our cameras to detect the mice but without substantial circadian entrainment effects. JVC digital video cameras (GR-D93) were connected to a PC workstation (Dell) via a Hauppauge WinTV video card. Using this setup we collected greater than 24 distinct MPEG-2 video sequences (from one to several hours in length) used for training and testing the system. For processing by the computer vision system, all videos were down-sampled to a resolution of 320x240 pixels. This means that the throughput of our system could be increased by video recording 4 cages at a time using a two by two arrangement with a standard 640x480 pixel VGA video resolution. A separate collection of videos of the mouse strains (n=28 videos) was collected for the validation experiment performed Caltech, using different recording conditions. All mouse strains were purchased from the Jackson Laboratory (Bar Harbor, Maine), including C57Bl/6J (stock 000664), DBA/2J (000671), CAST/EiJ (000928), and BTBR *T+tf*/J (002282). Mice were singly housed for 1-3 days before being video recorded. On the recommendation of Jackson Laboratories, the CAST/EiJ mice (n=7) were segregated from our main mouse colony and housed in a quiet space where they were only disturbed for husbandry 2-3 times per week.

Data annotation

Training videos were annotated using a freeware subtitle editing tool, Subtitle Workshop freeware subtitle editing tool from UroWorks available at http://www.urusoft.net/products.php?cat=sw&lang=1. A team of twelve investigators was trained to annotate mouse home cage behaviors. Behaviors of interest included: drinking, eating, grooming (defined by a fore- or hind-limbs sweeping across the face or torso, typically the animal is reared up), hanging (defined by a 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, and standing against a wall cage), resting (defined by inactivity or nearly complete stillness), walking (defined by ambulation) and micro-movements (defined by small movements of the animal's head). For the 'full database' to be annotated, every hour of videos took about 22 hours of manual labor for a total of 230 hours of work. For the 'clipped database' it took approximately 50 hours (9 hrs/hr of video) to manually score 9,600 clips of a single behavior (corresponding to 5.4 hours of clips compiled from around 20 hours of video). We performed second screening to remove ambiguous clips (leading to 4,200 clips remaining) such that the human-to-human agreement in terms of this library is very close to 100%. This second screening took around 40 hours for the 2.5 hour long 'clipped dataset'.

Training and testing the system

The results in Table 1 were obtained using a leave-one-out cross-validation procedure. This consists in using all except one videos to train the system and the left out video to evaluate the system and repeating this procedure (n=12) times for all videos. Such procedure has been shown to provide the best estimate of the performance of a classifier and is standard in computer vision. This guarantees that the system is not just recognizing memorized examples but generalizing to previously unseen examples. The accuracy of the system was measured on a frame-by-frame basis except on the 'clipped dataset' where a single label was obtained for an entire sequence consisting of a single behavior via voting across frames.

Comparison with the commercial software

In order to compare our system with the leading commercial software HomeCageScan 2.0 (CleverSys Inc), we manually matched the 38 output labels from the HomeCageScan to the 8 behaviors used in the present work. For instance, actions such as 'slow walking', 'walking left' and 'walking right' were all re-assigned to the 'walking' label to match against our annotations. With the exception of 'jump', we matched all other HomeCageScan output behaviors to one of our 8 behaviors of interest (see Table S1 for a listing of the correspondences used between the labels of the HomeCageScan and our system). As with the UCSD system, it is possible that further fine-tuning of HomeCageScan parameters could have improved upon the accuracy of the scoring.

Statistical analysis

To detect differences among the 4 strains of mice ANOVAs were conducted for each type of behavior independently and Tukey's post-hoc test was used to test pair-wise significances. All post-hoc tests were Bonferroni corrected for multiple comparisons. For testing significance of grooming we used a Student's T test as we were only comparing between two groups (C57Bl/6J and BTBR) and we used a one-tailed test because we predicted that BTBR would groom more than C57Bl/6J.

Mouse strain comparisons

Patterns of behaviors were computed from the system output by considering 4 non-overlapping 6-hour long behavioral windows (corresponding to the first and second halves of the day and night periods) and calculating the distribution of behaviors across these windows. The resulting 8-dimensional histogram vectors were then concatenating to obtain one single 32-dimensional vector for each animal. To visualize the data, we computed a dissimilarity matrix by calculating the Euclidean distance between all pairs of points and performed an unsupervised Multidimensional Scaling (MDS) analysis on the corresponding dissimilarity matrix. This was done using the matlab command 'mdscale' with the Kruskal's normalized stress1 normalization criterion.

In addition we conducted a pattern classification analysis on the patterns of behaviors by training and testing an SVM classifier directly on the 32-dimensional vectors. This supervised procedure was conducted using a leave-one-animal out approach, whereby 27 animals were used to train a classifier to predict the strain of the animal (CAST/EiJ, BTBR, C57Bl/6J or DBA/2J). The classifier was then evaluated on the remaining animal. The procedure was repeated 28 times (once for each animal) and performance averaged across runs.

ACKNOWLEDGMENTS

This project was sponsored by the McGovern Institute for Brain Research. ADS, LY, and VK were funded by the Broad Fellows Program in Brain Circuitry at Caltech. HJ was founded by the Taiwan National Science Council (TMS-094-1-A032). We wish to thank Jessi Ambrose, Alexandra Jiang, Grant Kadokura, Xuying Li, Kelsey von Tish, Emma Tolley, Ye Yao and Eli T. Williams for their efforts in annotating the videos for this project. We are grateful to Susan Lindquist for allowing us to use her facilities and mice to generate much of the training video database and to Walker Jackson for assistance with operating HCS. We wish to thank Piotr Dollar for providing the source code of their action recognition system. The algorithm and software development and its evaluation was conducted at the Center for Biological & Computational Learning, which is in the McGovern Institute for Brain Research at MIT, as well as in the Dept. of Brain & Cognitive Sciences, and which is affiliated with the Computer Sciences & Artificial Intelligence Laboratory (CSAIL).

REFERENCES

Altun, Y., I. Tsochantaridis, et al. (2003). Hidden Markov Support Vector Machines. International Conference on Machine Learning (ICML).

Auwerx, J., P. Avner, et al. (2004). "The European dimension for the mouse genome mutagenesis program." Nat Genet **36**(9): 925-927.

Born, R. T. and D. C. Bradley (2005). "Structure and function of visual area MT." Annu. Rev. Neurosci. **28**: 157-189.

Branson, K. and S. Belongie (2005). "Tracking multiple mouse contours (without too many samples)." Proceedings of the IEEE Computer Vision and Pattern Recognition **1**: 1039-1046

Branson, K., A. A. Robie, et al. (2009). "High-throughput ethomics in large groups of Drosophila." Nat Methods **6**(6): 451-457.

Chen, D., A. D. Steele, et al. (2005). "Increase in activity during calorie restriction requires Sirt1." Science **310**(5754): 1641.

Crabbe, J. C., D. Wahlsten, et al. (1999). "Genetics of mouse behavior: interactions with laboratory environment." Science **284**(5420): 1670-1672.

Dalal, N. and B. Triggs (2005). Histograms of oriented gradients for human detection. IEEE Computer Society Conference on Computer Vision and Pattern Recognition. **1:** 886-893.

Dankert, H., L. Wang, et al. (2009). "Automated monitoring and analysis of social behavior in Drosophila." Nat Methods **6**(4): 297-303.

Dell'Omo, G., E. Vannoni, et al. (2002). "Early behavioural changes in mice infected with BSE and scrapie: automated home cage monitoring reveals prion strain differences." Eur J Neurosci **16**(4): 735-742.

Dollar, P., V. Rabaud, et al. (2005). Behavior Recognition via Sparse Spatio-Temporal Features. VS-PETS.

Fry, S. N., N. Rohrseitz, et al. (2008). "TrackFly: virtual reality for a behavioral system analysis in free-flying fruit flies." J. Neurosci. Methods **171**: 110–117.

Giese, M. A. and T. Poggio (2003). "Neural mechanisms for the recognition of biological movements." Nat Rev Neurosci **4**(3): 179-192.

Goulding, E. H., A. K. Schenk, et al. (2008). "A robust automated system elucidates mouse home cage behavioral structure." Proc Natl Acad Sci U S A **105**(52): 20575-20582.

Greer, J. M. and M. R. Capecchi (2002). "Hoxb8 is required for normal grooming behavior in mice." Neuron **33**(1): 23-34.

Jackson, W. S., S. J. Tallaksen-Greene, et al. (2003). "Nucleocytoplasmic transport signals affect the age at onset of abnormalities in knock-in mice expressing polyglutamine within an ectopic protein context." Hum Mol Genet **12**(13): 1621-1629.

Jhuang, H., T. Serre, et al. (2007). "A Biologically Inspired System for Action Recognition." Proceedings of the Eleventh IEEE International Conference on Computer Vision (ICCV).

Joachims, T., T. Finley, et al. (2009). "Cutting-Plane Training of Structural SVMs." Machine Learning **76(1)**.

Khan, Z., T. Balch, et al. (2005). "MCMC-based particle filtering for tracking a variable number of interacting targets." IEEE Trans. Pattern Anal. Mach. Intell. **27**: 1805–1819.

McFarlane, H. G., G. K. Kusek, et al. (2008). "Autism-like behavioral phenotypes in BTBR T+tf/J mice." Genes Brain Behav **7**(2): 152-163.

Moeslund, T., A. Hilton, et al. (2006). "A survey of advances in vision-based human motion capture and analysis." Computer Vision and Image Understanding **103**(2-3): 90–126.

Noldus, L. P., A. J. Spink, et al. (2001). "EthoVision: a versatile video tracking system for automation of behavioral experiments." Behav Res Methods Instrum Comput **33**(3): 398-414.

Roughan, J. V., S. L. Wright-Williams, et al. (2008). "Automated analysis of postoperative behaviour: assessment of HomeCageScan as a novel method to rapidly identify pain and analgesic effects in mice." Lab Anim.

Simoncelli, E. P. and D. J. Heeger (1998). "A model of neuronal responses in visual area MT." Vision Res **38**(5): 743-761.

Spink, A. J., R. A. Tegelenbosch, et al. (2001). "The EthoVision video tracking system--a tool for behavioral phenotyping of transgenic mice." Physiol Behav **73**(5): 731-744.

Steele, A. D., W. S. Jackson, et al. (2007). "The power of automated high-resolution behavior analysis revealed by its application to mouse models of Huntington's and prion diseases." Proc Natl Acad Sci U S A **104**(6): 1983-1988.

Tecott, L. H. and E. J. Nestler (2004). "Neurobehavioral assessment in the information age." Nat Neurosci **7**(5): 462-466.

Tsochantaridis, I., T. Hofmann, et al. (2004). "Support Vector Machine Learning for Interdependent and Structured Output Spaces." Proceedings of the 21st International Conference on Machine Learning.

Tsochantaridis, I., T. Joachims, et al. (2005). "Large Margin Methods for Structured and Interdependent Output Variables." Journal of Machine Learning Research **6**: 1453-1484.

Veeraraghavan, A., R. Chellappa, et al. (2008). " Shape-and-behavior encoded tracking of bee dances." IEEE Trans. Pattern Anal. Mach. Intell. **30**: 463–476

Viola, P. and M. Jones (2001). Robust real-time face detection. ICCV. **20(11):** 1254--1259.

Xue, X. and T. Henderson (2009). "Feature fusion for basic behavior unit segmentation from video sequences." Robotics and Autonomous Systems **57**: 239-248.

Figure Legends

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

Figure 2: Overview of the proposed system for monitoring the home-cage behavior of mice. The computer vision system consists of several modules. (A) A background subtraction procedure is first applied to an input video to compute a mask for pixels belonging to the animal vs. the cage. Two types of features are then computed: (C) Position- and velocity-based features as well as (D) space-time motion features. Position- and velocity-based features are computed directly from the segmentation mask based on the instantaneous location of the animal in the cage. In order to speed-up the computation of the motion-features (panel D) is performed on a sub-window centered on the animal and derived from the segmentation mask (panel B).). These motion features are obtained by combining the response of V1-like afferent motion units that are tuned to different directions of motion (see inset E). The output of this hierarchical pre-processing module consists of a basic dictionary of about 300 space-time motion features that is then passed to a statistical classifier (called HMMSVM, see panel F and text for details) to reliably classify every frame of a video sequence into a behavior of interest. (F) Hidden Markov Model Support Vector Machine. (G) Ethogram of a single BTBR mouse From the sequence of labels obtained from the computer software from a 24-hr continuous recording session for one of the wild-like mice an ethogram can be computed. The inset on the left provides a zoomed in version corresponding to the first 30 minutes of recording. The animal is highly active as a human experimenter just placed the mouse in a new cage prior to starting the video recording. The animal's behavior alternates between 'walking', 'rearing' and 'hanging' as it explores its new cage.

Figure 3: Confusion matrices for comparing the system to human scoring and human to human scoring. The confusion matrix for the computer system (A) was obtained by measuring the agreement between the system (row) and one arbitrarily chosen human scorer (column); the confusion matrix for human (B) was obtained by measuring the agreement between two independent scorers; and the confusion matrix for HCS ("commercial system") (C) was obtained by measuring agreement between the software and one arbitrarily chose human scorer (column).

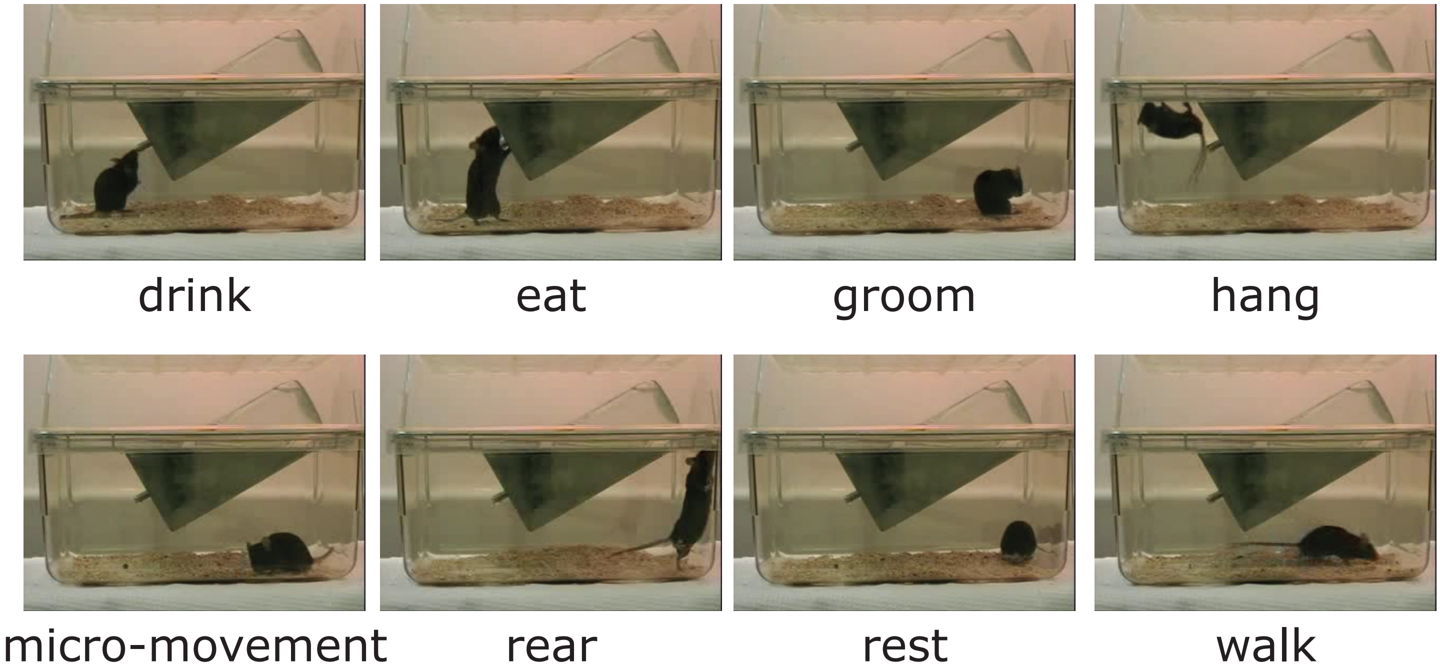
Figure 4: Patterns of behaviors obtained for 'walking' and 'hanging' behaviors for each of the four strains of mice. The CAST/EiJ (wild-like) strain is much more active than the three other strains as measured by their hanging and walking behaviors. Shaded areas correspond to 95% confidence intervals and the darker line corresponds to the mean. The intensity of the colored bars on the top corresponds to the number of strains that exhibit a statistically significant difference with the corresponding strain (indicated by the color of the bar). The intensity of one color is proportional to (N-1), where N is the number of groups whose mean is significantly different from the corresponding strain of the color. For example, CAST/EiJ at time 0 to time 7 for walking is significantly higher than the three other strains so N is 3 and the red is the highest intensity.

Figure 5: (A) Average total resting time for the four strains of mice. (B) Average duration of resting bouts. While all strains tend to spend roughly the same total amount of time sleeping, the CAST/EiJ tends to sleep fewer longer stretches. Mean +/- SEM are shown, *P<0.01 by ANOVA with Tukey's post test. (C) Average grooming duration exhibited by the BTBR strain as compared to the C57Bl/6J strain over one hour. Here we show that using the computer system we were able to match manual scoring by an experimenter and reproduce the key result from the study by McFarlane et al (McFarlane, Kusek et al. 2008) demonstrating the propensity of the BTBR strain to groom more than a control C57Bl/6J. Mean +/- SEM are shown, *P<0.05 by Student's T test, one-tailed. (P = 0.04 for System and P =0.0254 for Human). Characterizing the genotype of individual animals based on the patterns of behavior measured by the computer system. (D) Multi-Dimensional Scaling (MDS) analysis performed on the distributions of behavior types measured over 4 (6 hour-long) windows. The MDS analysis reveals that animals tend to cluster into strains (with the exception of 2 BTBR mice that tended to behave more like DBA/2J). (E) Pattern classification analysis performed on the distributions of behavior types measured over 4 (6 hour-long) windows. Using an SVM classifier on the patterns of behavior we were able to predict the genotype of individual animals with accuracy of 90% (chance level is 25% for this 4-class classification problem). The confusion matrix shown here indicates the probability with which an input pattern of behavior (along the rows) was classified as each of the 4 alternative strains (along the columns). The higher probabilities along the diagonal and the lower off-diagonal values indicate successful classification for all categories. For example, the value of 1 for C57Bl/6J means that this strain was perfectly classified.

Table 1: System evaluation on the 'full database' for the recognition of 8 behaviors. Numbers correspond to accuracy measured by % correct (chance level is 12.5% for the 8-class classification problem).

	Our system	CleverSys commercial system	Human
Frames annotated by 2 independent labelers (~ 1.6 hrs of video)	71.0 %	56.0 %	71.6 %
All frames (> 10 hrs of video)	69.5 %	57.4 %	

Figure 1



drink

eat

groom

hang

micro-movement

rear

rest

walk

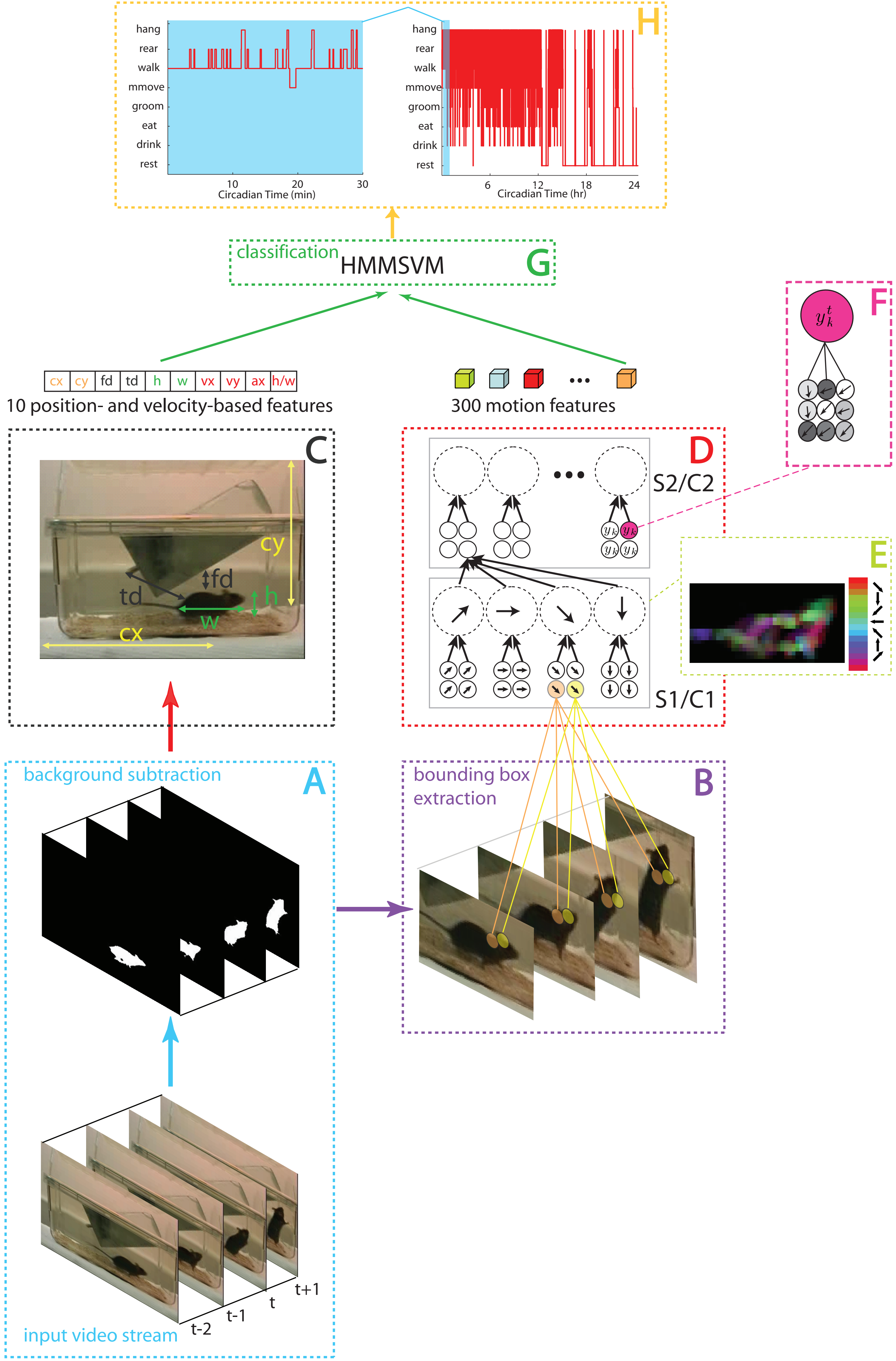
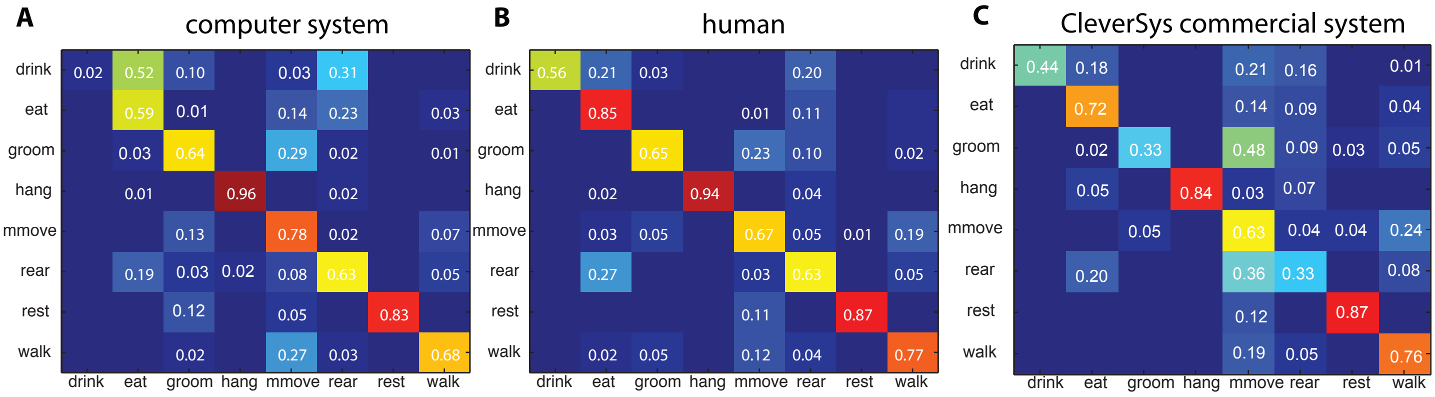

H
hang
rear
walk
mmove
groom
eat
drink
rest
10
20
30
Circadian Time (min)
hang
rear
walk
mmove
groom
eat
drink
rest
6
12
18
24
Circadian Time (hr)
classification
HMMSVM
G
F
y_k^t
cx
cy
fd
td
h
w
vx
vy
ax
h/w
10 position- and velocity-based features
300 motion features
C
cy
fd
td
h
w
cx
D
S2/C2
y_k
S1/C1
E
background subtraction
A
bounding box extraction
B
input video stream
t-2
t-1
t
t+1

Figure 3



A computer system

	drink	eat	groom	hang	mmove	rear	rest	walk
drink	0.02	0.52	0.10		0.03	0.31		
eat		0.59	0.01		0.14	0.23		0.03
groom		0.03	0.64		0.29	0.02		0.01
hang		0.01		0.96		0.02		
mmove			0.13		0.78	0.02		0.07
rear		0.19	0.03	0.02	0.08	0.63		0.05
rest			0.12		0.05		0.83	
walk			0.02		0.27	0.03		0.68

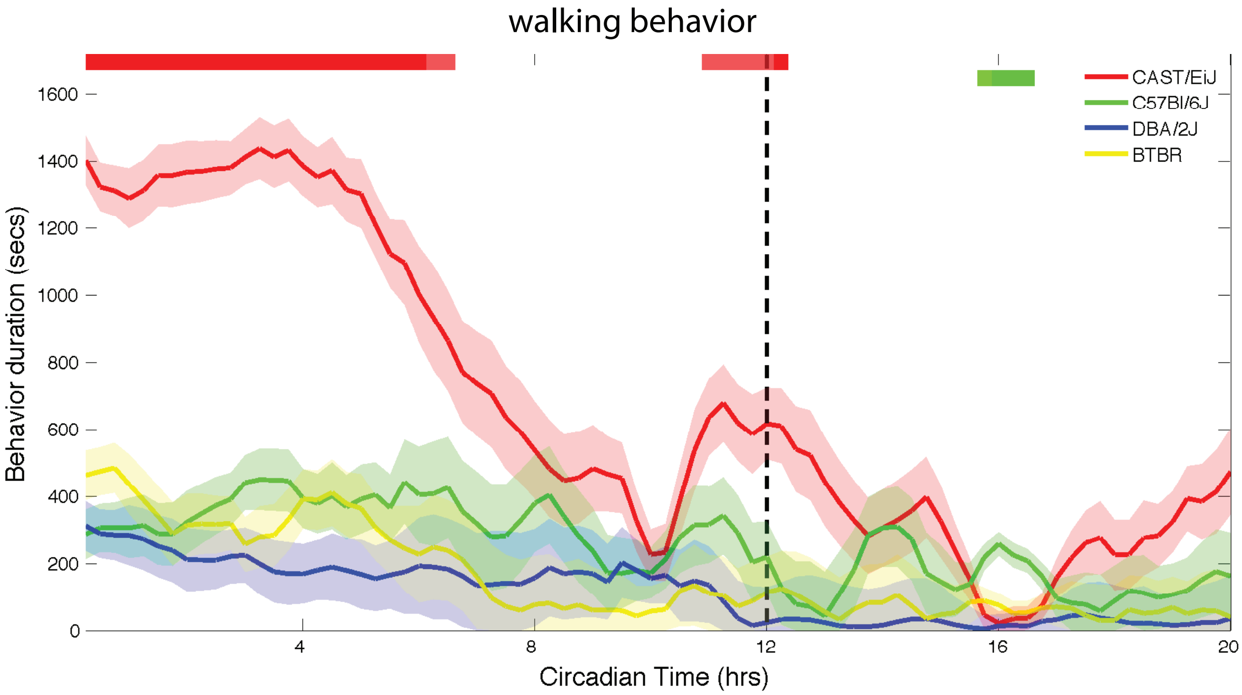
B human

	drink	eat	groom	hang	mmove	rear	rest	walk
drink	0.56	0.21	0.03			0.20		
eat		0.85			0.01	0.11		
groom			0.65		0.23	0.10		0.02
hang		0.02		0.94		0.04		
mmove		0.03	0.05		0.67	0.05	0.01	0.19
rear		0.27			0.03	0.63		0.05
rest					0.11		0.87	
walk		0.02	0.05		0.12	0.04		0.77

C CleverSys commercial system

	drink	eat	groom	hang	mmove	rear	rest	walk
drink	0.44	0.18			0.21	0.16		0.01
eat		0.72			0.14	0.09		0.04
groom		0.02	0.33		0.48	0.09	0.03	0.05
hang		0.05		0.84	0.03	0.07		
mmove			0.05		0.63	0.04	0.04	0.24
rear		0.20			0.36	0.33		0.08
rest					0.12		0.87	
walk					0.19	0.05		0.76

Figure 4


walking behavior
CAST/EiJ
C57Bl/6J
DBA/2J
BTBR
Behavior duration (secs)
0
200
400
600
800
1000
1200
1400
1600
4
8
12
16
20
Circadian Time (hrs)

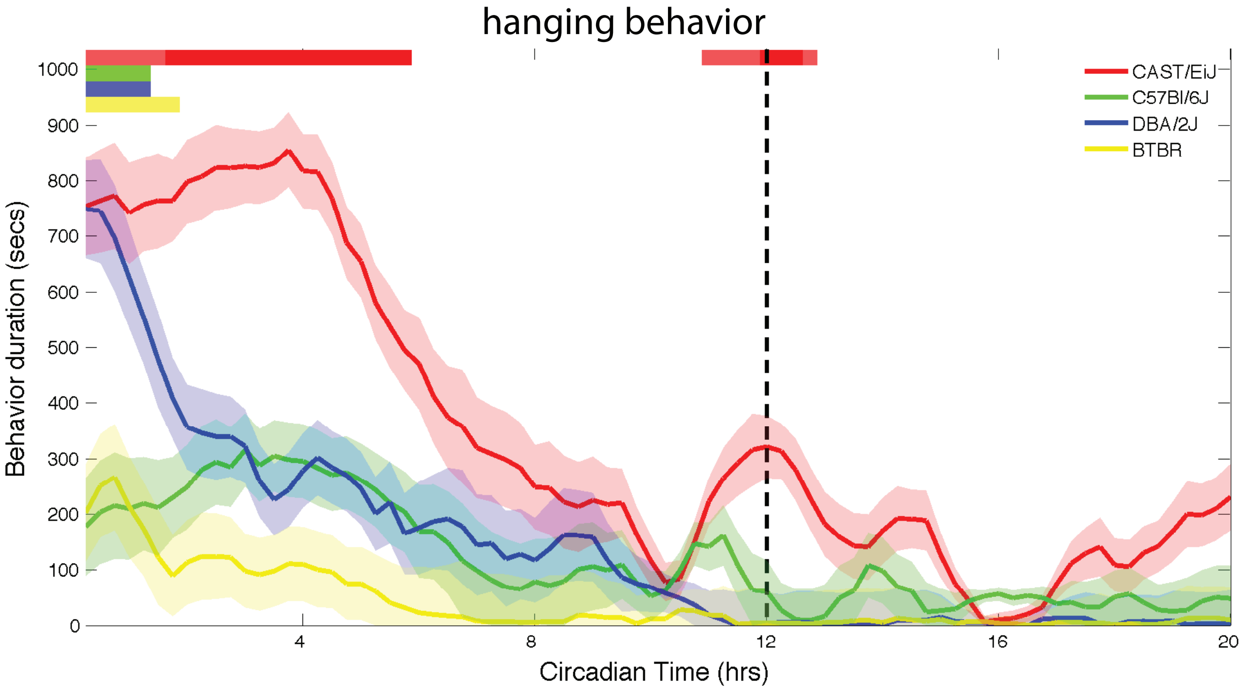
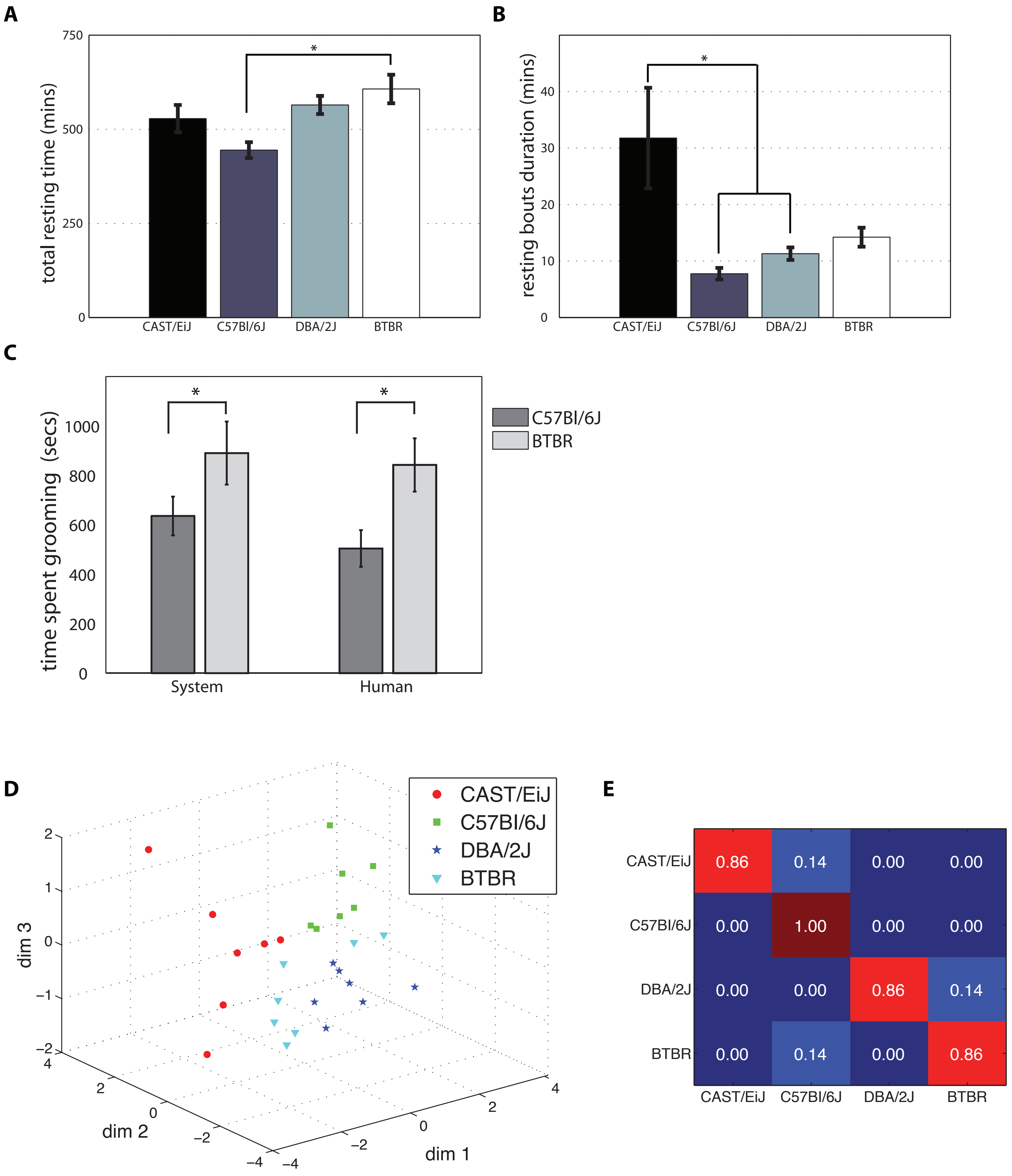

hanging behavior
CAST/EiJ
C57Bl/6J
DBA/2J
BTBR
Behavior duration (secs)
0
100
200
300
400
500
600
700
800
900
1000
4
8
12
16
20
Circadian Time (hrs)

Figure 5



	CAST/EiJ	C57Bl/6J	DBA/2J	BTBR
CAST/EiJ	0.86	0.14	0.00	0.00
C57Bl/6J	0.00	1.00	0.00	0.00
DBA/2J	0.00	0.00	0.86	0.14
BTBR	0.00	0.14	0.00	0.86

Supplementary Online Information

The Supplementary online information contains:

- Supplementary Text
- Supplementary Methods
- Supplementary References
- Supplementary Figures (S1, S2 and S3)
- Supplementary Table (S1)

SUPPLEMENTARY TEXT

Sensor-based approaches. Previous automated systems (e.g., (Noldus, Spink et al. 2001; Dell'Omo, Vannoni et al. 2002; Jackson, Tallaksen-Greene et al. 2003; Zorrilla, Inoue et al. 2005; Goulding, Schenk et al. 2008)) have relied for the most part on the use of sensors to monitor behavior. Popular sensor-based approaches include the use of PVDF sensors (Megens, Voeten et al. 1987), infrared sensors (Tamborini, Sigg et al. ; Casadesus, Shukitt-Hale et al. 2001; Dell'Omo, Vannoni et al. 2002; Tang and Sanford), RFID transponders (Lewejohann, Hoppmann et al. 2009) as well as photobeams (Goulding, Schenk et al. 2008). Such approaches have been successfully applied to the analysis of coarse locomotion activity as a proxy to measure global behavioral states such as active vs. resting. Other studies have successfully used sensors for the study of food and water intake (Gannon, Smith et al. 1992; Zorrilla, Inoue et al. 2005). However the physical measurements obtained from these sensor-based approaches limit the complexity of the behavior that can be measured. This problem remains even for commercial systems using transponder technologies such as the IntelliCage system (NewBehavior Inc). While such systems can be effectively used to monitor the locomotion activity of an animal as well as other pre-programmed activities via operant conditioning units located in the corners of the cage, such systems alone cannot be used to study natural behaviors such as grooming, sniffing, rearing, etc.

Video-based approaches. One of the possible solutions to address the problems described above is to rely on vision-based techniques. In fact such approaches are already bearing fruit for the automated tracking (Khan, Balch et al. 2005; Fry, Rohrseitz et al. 2008; Veeraraghavan, Chellappa et al. 2008) and recognition of behaviors in insects (Branson, Robie et al. 2009; Dankert, Wang et al. 2009). Several open-source and commercial computer-vision systems for the tracking of rodents have been developed (van Lochem, Buma et al. 1998; Noldus, Spink et al. 2001; Spink, Tegelenbosch et al. 2001; Twining, Taylor et al. 2001; Branson and Belongie 2005; Zurn, Jiang et al. 2007; Leroy, Stroobants et al. 2009). As for sensor-based approaches, such systems are particularly suitable for studies involving coarse locomotion activity based on spatial measurements such as the distance covered by an animal or its speed (Millecamps, Jourdan et al. 2005; de Visser, van den Bos et al. 2006; Bonasera, Schenk et al. 2008; Donohue, Medonza et al. 2008). Video-tracking based approaches tend to be more flexible and much more cost efficient. However, as in the case of sensor-based approaches, these systems alone are not suitable for the analysis of fine animal activities such as grooming, sniffing or rearing.

The first effort to build an automated computer vision system for the monitoring of rodent behavior was initiated at USC. As part of this SmartVivarium project, an initial computer-vision system was developed for both the tracking (Branson and Belongie 2005) of the animal as well as the recognition of five behaviors (eating, drinking, grooming, exploring and resting, see ref. (Dollar, Rabaud et al. 2005)). Xue & Henderson recently described an approach (Xue and Henderson 2006; Xue and Henderson 2009) for the analysis of rodent behavior however the system was only tested on synthetic data (Henderson and Xue) and a very limited number of behaviors. Overall, none of the existing systems (Dollar, Rabaud et al. 2005; Xue and Henderson 2006; Xue and Henderson 2009) have been tested in a real-world lab setting using long uninterrupted video sequences and containing potentially ambiguous behaviors or at least evaluated against human manual annotations on large databases of video sequences using different animals and different recording sessions. Recently a commercial system (HomeCageScan by CleverSys, Inc) was also introduced and the system was successfully used in several behavioral studies (Chen, Steele et al. 2005; Steele, Jackson et al. 2007; Goulding, Schenk et al. 2008; Roughan, Wright-Williams et al. 2008). Such commercial products typically rely on relatively simple heuristics such as the position of the animal in the cage to infer behavior. They thus remain limited in their scope (tracking of simple behaviors) and error-prone (see ref. (Steele, Jackson et al. 2007) and Table 1 for a comparison against our manual annotations). In addition, the software packages are proprietary: there is no simple way for the end user to improve its performance or to customize it to specific needs.

SUPPLEMENTARY METHODS

System Overview. Figure S1 provides an overview of the computer vision system used. The system takes as input a video sequence recorded from a video camera. It then converts every frame of a video sequence into a representation, which is suitable for the recognition of behaviors. This representation is based on a feature vector, where each coefficient of this vector corresponds to the degree of similarity between stored 3D space-time motion templates learned from the set of behaviors of interest and the current frame. An action label is then obtained for every frame of a video by passing this feature vector to a temporal model for classification. The temporal model used here is a hidden Markov Support Vector Machine (HMMSVM) (Altun, Tsochantaridis et al. 2003; Tsochantaridis, Hofmann et al. 2004; Tsochantaridis, Joachims et al. 2005; Joachims, Finley et al. 2009), which is an extension of the popular Support Vector Machine classifier developed by Vapnik (Vapnik 1995) in the 90's, for sequence tagging. This temporal model was trained using manually labeled examples extracted from the video database denoted 'full database' in the main text. This database involved labeling every frame for 12 videos (from different mice recorded in different conditions) for a total of 10.4 hours of annotated video. The output of the system is thus a label corresponding to a specific behavior of interest for every frame of a video sequence.

The learning of the basic dictionary of 3D space-time motion-feature templates as well as the feature computation and the temporal model are described in detail in the following sections.

The approach taken here builds directly on the work by Jhuang et al. (Jhuang, Serre et al. 2007) (which itself builds on the work by Giese & Poggio (Giese and Poggio 2003)). The system is based on the organization of the dorsal stream of the visual cortex and was shown to compete with state-of-the-art computer vision systems. The system is organized hierarchically: Low-level features are first extracted at the bottom layer and progressively transformed to become increasingly complex and invariant. This is done through successive ***S*** and ***C*** stages of processing (see (Jhuang, Serre et al. 2007) for details). In the S1 stage, feature maps are obtained by convolving an input sequence with a spatio-temporal filter bank (9 pixels x 9 pixels x 9 frames) tuned to four different directions of motion. This linear stage was followed by a non-linear contrast normalization whereby the filter response was divided by the L1 norm of the corresponding patch of image. This results in four distinct S1 maps for every input frame where each map corresponds to one direction of motion and the value of each pixel on these maps is directly proportional to the amount of motion presented in the corresponding direction. The nature of the non-linearity contrast normalization (i.e., L1 vs. L2 vs. no-normalization), the number of motion directions used and the resolution of the input video sequences were carefully optimized in a preliminary experiment carried on a subset of the clipped database.

Beyond this initial S1 stage, processing is then hierarchical: alternating between a template matching (S layers) and a max pooling operation (C layers) gradually increases feature complexity and translation invariance. That is, at the C1 stage some tolerance to small deformations is obtained via a local max operation over neighborhoods of S1 units (8x8 cells). Next, template matching is performed over the C1 maps, creating thousands of S2 maps. At each position a patch of C1 units centered at that position is compared to each of ***d*** prototype patches. Each prototype corresponds to a vector of size $\boldsymbol{4n^2}$ obtained by cropping an $\boldsymbol{n} \times \boldsymbol{n}$ ($\boldsymbol{n} = 4, 8, 12, 16$) patch of C1 units at a given location and all 4 orientations. These ***d*** prototype patches represent the intermediate-level features of the model, and are randomly sampled from the C1 layers of the training images in an initial feature-learning stage. At the top of the hierarchy, a vector of ***d*** position invariant C2 features is obtained by computing a global max for each of the ***d*** S2 feature maps.

To select a more discriminant dictionary of motion patterns (and speed up the overall system), we applied a feature selection technique called zero-norm SVM (Weston, Mukherjee et al.) on the initial set of ***d*** C2 features. This was done by computing the feature responses for the original set of ***d*** C2 features for frames randomly selected from the 'clipped' database. Selection was then done in multiple rounds: In each round, an SVM classifier was trained on the pool of C2 features and the training set was re-weighted using the weights of the trained SVM. Typically this leads to sparser SVM weights at each stage leading to a final set of ***d'***=300 features from an original set of ***d***=12,000 features.

Comparison with a benchmark computer vision system. The computer vision system used here for benchmark is the system developed by Dollar, Rabaud, Cottrell, & Belongie at the University of California (San Diego) as part of the *SmartVivarium* project (Belongie, Branson et al. 2005). The system has been shown to outperform several other computer vision systems on several standard computer vision databases and was tested for both the recognition of human and rodent behaviors (Dollar, Rabaud et al. 2005). The authors graciously provided the source code for their system. Training and testing of the corresponding system was done in the same way as for our system using a leave-one-out procedure using the 'clipped dataset' (see manuscript). Here we attempted to maximize the performance of the system by tuning some of the key parameters such as the number of features and the resolution of the videos used. Nevertheless we found that the default parameters (50 features and a 320x240 video resolution as used for our system) led to the best performance (81% vs. 93% for our system). It is possible however that further refinement of the corresponding algorithm could nevertheless improve its performance.

SUPPLEMENTARY REFERENCES

Altun, Y., I. Tsochantaridis, et al. (2003). Hidden Markov Support Vector Machines. International Conference on Machine Learning (ICML).

Belongie, S., K. Branson, et al. (2005). Monitoring Animal Behavior in the Smart Vivarium. Measuring Behavior.

Bonasera, S. J., A. K. Schenk, et al. (2008). "A novel method for automatic quantification of psychostimulant-evoked route-tracing stereotypy: Application to Mus musculus." Psychopharmacol **196**: 591–602.

Branson, K. and S. Belongie (2005). "Tracking multiple mouse contours (without too many samples)." Proceedings of the IEEE Computer Vision and Pattern Recognition **1**: 1039-1046

Branson, K., A. A. Robie, et al. (2009). "High-throughput ethomics in large groups of Drosophila." Nat Methods **6**(6): 451-457.

Casadesus, G., B. Shukitt-Hale, et al. (2001). "Automated measurement of age-related changes in the locomotor response to environmental novelty and home-cage activity." Mechan Age Develop **122**: 1887-1897.

Chen, D., A. D. Steele, et al. (2005). "Increase in activity during calorie restriction requires Sirt1." Science **310**(5754): 1641.

Dankert, H., L. Wang, et al. (2009). "Automated monitoring and analysis of social behavior in Drosophila." Nat Methods **6**(4): 297-303.

de Visser, L., R. van den Bos, et al. (2006). "Novel approach to the behavioural characterization of inbred mice: Automated home cage observations." Genes Brain Behav **5**: 458-466.

Dell'Omo, G., E. Vannoni, et al. (2002). "Early behavioural changes in mice infected with BSE and scrapie: automated home cage monitoring reveals prion strain differences." Eur J Neurosci **16**(4): 735-742.

Dollar, P., V. Rabaud, et al. (2005). Behavior Recognition via Sparse Spatio-Temporal Features. VS-PETS.

Donohue, K. D., D. C. Medonza, et al. (2008). "Assessment of a non-invasive high-throughput classifier for behaviours associated with sleep and wake in mice." Biomed Engineering Online 7:14.

Fry, S. N., N. Rohrseitz, et al. (2008). "TrackFly: virtual reality for a behavioral system analysis in free-flying fruit flies." J. Neurosci. Methods **171**: 110–117.

Gannon, K. S., J. C. Smith, et al. (1992). "A system for studying the microstructure of ingestive behavior in mice." Physiol Behav **51**: 515-521.

Giese, M. A. and T. Poggio (2003). "Neural mechanisms for the recognition of biological movements." Nat Rev Neurosci **4**(3): 179-192.

Goulding, E. H., A. K. Schenk, et al. (2008). "A robust automated system elucidates mouse home cage behavioral structure." Proc Natl Acad Sci U S A **105**(52): 20575-20582.

Henderson, T. and X. Xue "Constructing Comprehensive Behaviors: A Simulation Study."

Jackson, W. S., S. J. Tallaksen-Greene, et al. (2003). "Nucleocytoplasmic transport signals affect the age at onset of abnormalities in knock-in mice expressing polyglutamine within an ectopic protein context." Hum Mol Genet **12**(13): 1621-1629.

Jhuang, H., T. Serre, et al. (2007). "A Biologically Inspired System for Action Recognition." Proceedings of the Eleventh IEEE International Conference on Computer Vision (ICCV).

Joachims, T., T. Finley, et al. (2009). "Cutting-plane training of structural svms." Machine Learning **77**(1): 27-59.

Khan, Z., T. Balch, et al. (2005). "MCMC-based particle filtering for tracking a variable number of interacting targets." IEEE Trans. Pattern Anal. Mach. Intell. **27**: 1805–1819.

Leroy, T., S. Stroobants, et al. (2009). "Automatic analysis of altered gait in arylsulphatase A-deficient mice in the open field." Behavior Research Methods **41**(3): 787-794.

Lewejohann, L., A. M. Hoppmann, et al. (2009). "Behavioral phenotyping of a murine model of Alzheimer's disease in a seminaturalistic environment using RFID tracking." Behavior Research Methods.

Megens, A. A. H. P., J. Voeten, et al. (1987). "Behavioural activity of rats measured by a new method based on the piezo-electric principle." Psychopharmacol **93**: 382-388.

Millecamps, M., D. Jourdan, et al. (2005). "Circadian pattern of spontaneous behavior in monarthritic rats: A novel global approach to evaluation of chronic pain and treatment effectiveness." Arthritis Rheumatism **52**: 3470-3478.

Noldus, L. P., A. J. Spink, et al. (2001). "EthoVision: a versatile video tracking system for automation of behavioral experiments." Behav Res Methods Instrum Comput **33**(3): 398-414.

Roughan, J. V., S. L. Wright-Williams, et al. (2008). "Automated analysis of postoperative behaviour: assessment of HomeCageScan as a novel method to rapidly identify pain and analgesic effects in mice." Lab Anim.

Spink, A. J., R. A. Tegelenbosch, et al. (2001). "The EthoVision video tracking system--a tool for behavioral phenotyping of transgenic mice." Physiol Behav **73**(5): 731-744.

Steele, A. D., W. S. Jackson, et al. (2007). "The power of automated high-resolution behavior analysis revealed by its application to mouse models of Huntington's and prion diseases." Proc Natl Acad Sci U S A **104**(6): 1983-1988.

Tamborini, P., H. Sigg, et al. (1989). "Quantitative analysis of rat activity in the home cage by infrared monitoring. Application to the acute toxicity testing of acetanilide and phenylmercuric acetate." Arch Toxicol **63**: 85–96.

Tang, X. and L. D. Sanford (2005). "Home cage activity and activity-based measures of anxiety in 129P3/J, 129X1/SvJ and C57BL/6J mice." Physiol Behav **84**(105–115).

Tsochantaridis, I., T. Hofmann, et al. (2004). Support vector machine learning for interdependent and structured output spaces. . Proceedings of the twenty-first international conference on Machine learning.

Tsochantaridis, I., T. Joachims, et al. (2005). "Large Margin Methods for Structured and Interdependent Output Variables." Journal of Machine Learning Research **6**: 1453-1484.

Twining, C. J., C. J. Taylor, et al. (2001). "Robust tracking and posture description for laboratory rodents using active shape models." Behav Res Methods Instrum Comput **33**(3): 381-391.

van Lochem, P., M. Buma, et al. (1998). Automatic recognition of behavioral patterns of rats using video imaging and statistical classification. Measuring Behavior.

Vapnik, V. (1995). The Nature of Statistical Learning Theory. New York, Springer.

Veeraraghavan, A., R. Chellappa, et al. (2008). " Shape-and-behavior encoded tracking of bee dances." IEEE Trans. Pattern Anal. Mach. Intell. **30**: 463–476

Weston, J., S. Mukherjee, et al. (2001). Feature Selection for Support Vector Machines. Advances in Neural Information Processing Systems 13.

Xue, X. and T. Henderson (2006). "Video-based Animal Behavior Analysis From Multiple Cameras." 2006 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems: 335-340.

Xue, X. and T. Henderson (2009). "Feature fusion for basic behavior unit segmentation from video sequences." Robotics and Autonomous Systems **57**: 239-248.

Zorrilla, E. P., K. Inoue, et al. (2005). "Measuring meals: Structure of prandial food and water intake of rats." Am J Physiol **288**: R1450–R1467.

Zurn, J., X. Jiang, et al. (2007). "Video-Based Tracking and Incremental Learning Applied to Rodent Behavioral Activity Under Near-Infrared Illumination." IEEE Transactions on Instrumentation and Measurement **56**: 2804.

SUPPLEMENTARY FIGURES

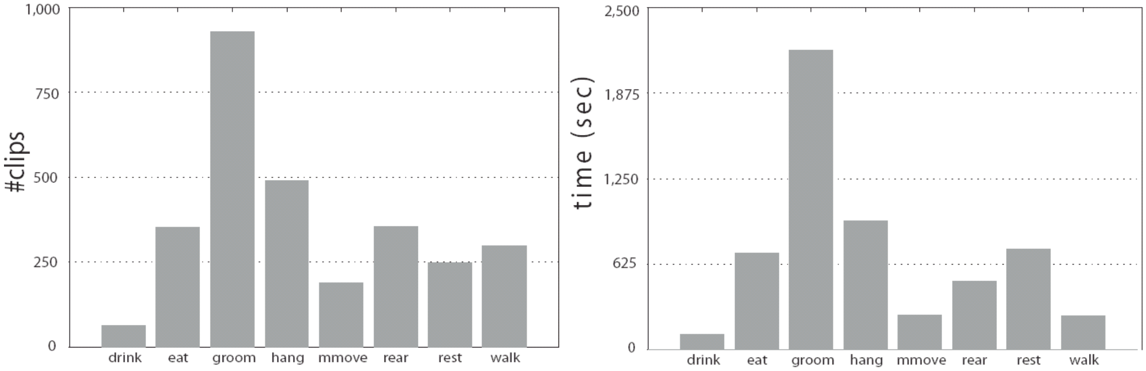


Figure S1: Distribution of behavior labels for the 'clipped database' over the number of clips (a) and total time (b).

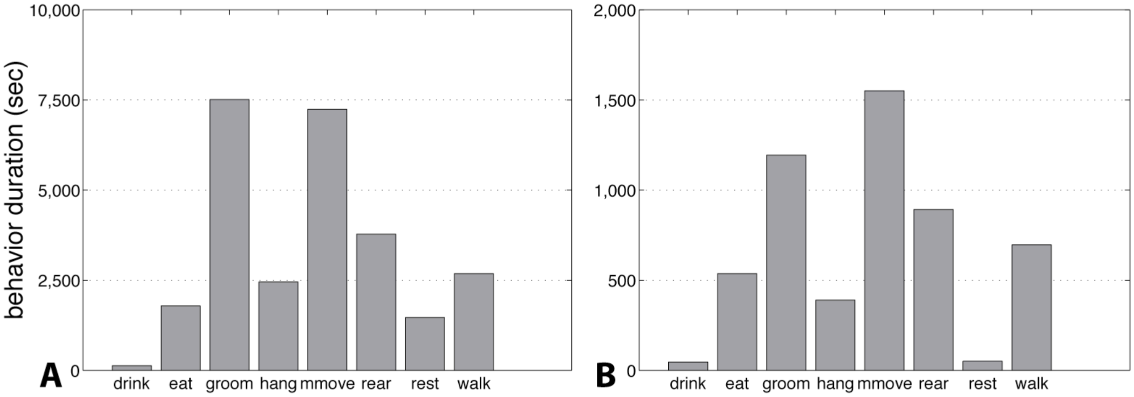


Figure S2: Distribution of behavior labels on the 'full database' annotated by one scorer (Set A) vs. set B (a subset of Set A), which was annotated by two scorers to evaluate the agreement between two independent scorers.

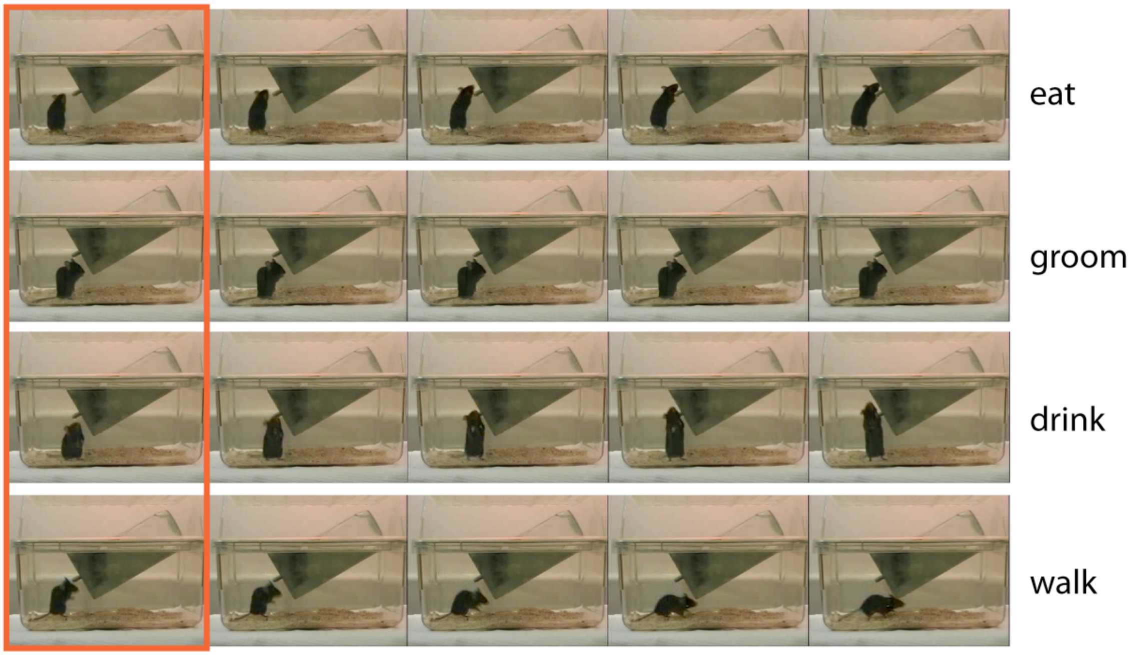


Figure S3: Single frames are ambiguous. Each row corresponds to a short video clip. While the leftmost frames (red bounding box) all look quite similar, they each correspond to a different behavior (text on the right side). Because of this ambiguity, frame-based behavior recognition is unreliable and temporal models that integrate the local temporal context over adjacent frames are needed for robust behavior recognition.

HCS label	System label
Drink	Drink
Chew	Eat
Eat	
Groom	Groom
Hang Cuddled	Hang
Hang Vertically	
Hang Vertically From Hang Cuddled Hang	
Hang Vertically From Rear Up	
Remain Hang Cuddled	
Remain Hang Vertically	
Awaken	Micro-move
Pause	
Remain Low	
Sniff	
Twitch	
Come Down	Rear
Come Down From Partially Reared	
Come Down To Partially Reared	
Stretch Body	
Land Vertically	
Rear Up	
Rear up From Partially Reared	
Rear up To Partially Reared	
Remain Partially Reared	
Remain Rear Up	
Sleep	Rest
Stationary	
Circle	Walk
Turn	
Walk Left	
Walk Right	
Walk Slowly	
Dig	Unknown Behavior
Forage	
Jump	
Repetitive Jumping	
Unknown Behavior	
Urinate	

Supplementary Table S1: HomeCage commercial system evaluation: Correspondence used for the labels.