



# Improved 3D tracking and automated classification of rodents' behavioral activity using depth-sensing cameras

Ana Gerós<sup>1,2,3</sup> · Ana Magalhães<sup>1,4</sup> · Paulo Aguiar<sup>1,2</sup>

Published online: 30 March 2020  
© The Psychonomic Society, Inc. 2020

## Abstract

Analysis of rodents' behavior/activity is of fundamental importance in many research fields. However, many behavioral experiments still rely on manual scoring, with obvious problems in reproducibility. Despite important advances in video-analysis systems and computational ethology, automated behavior quantification is still a challenge. The need for large training datasets, background stability requirements, and reduction to two-dimensional analysis (impairing full posture characterization), limit their use. Here we present a novel integrated solution for behavioral analysis of individual rats, combining video segmentation, tracking of body parts, and automated classification of behaviors, using machine learning and computer vision methods. Low-cost depth cameras (RGB-D) are used to enable three-dimensional tracking and classification in dark conditions and absence of color contrast. Our solution automatically tracks five anatomical landmarks in dynamic environments and recognizes seven distinct behaviors, within the accuracy range of human annotations. The developed free software was validated in experiments where behavioral differences between Wistar Kyoto and Wistar rats were automatically quantified. The results reveal the capability for effective automated phenotyping. An extended annotated RGB-D dataset is also made publicly available. The proposed solution is an easy-to-use tool, with low-cost setup and powerful 3D segmentation methods (in static/dynamic environments). The ability to work in dark conditions means that natural animal behavior is not affected by recording lights. Furthermore, automated classification is possible with only ~30 minutes of annotated videos. By creating conditions for high-throughput analysis and reproducible quantitative measurements of animal behavior experiments, we believe this contribution can greatly improve behavioral analysis research.

**Keywords** Animal tracking in 3D · Automated behavior classification · Automated phenotyping · Depth sensors · Dynamic background segmentation · Free and user-friendly software · Public RGB-D Dataset · Wistar Kyoto model

## Introduction

Analysis of how animals interact with, respond to, and control their environment is a fundamental methodological

**Electronic supplementary material** The online version of this article (<https://doi.org/10.3758/s13428-020-01381-9>) contains supplementary material, which is available to authorized users.

✉ Paulo Aguiar  
pauloaguiar@ineb.up.pt

- <sup>1</sup> Instituto de Investigação e Inovação em Saúde, Universidade do Porto, Porto, Portugal
- <sup>2</sup> INEB – Instituto de Engenharia Biomédica, Universidade do Porto, Porto, Portugal
- <sup>3</sup> FEUP – Faculdade de Engenharia da Universidade do Porto, Porto, Portugal
- <sup>4</sup> IBMC – Instituto de Biologia Molecular e Celular, Universidade do Porto, Porto, Portugal

approach in many research fields (Anderson & Perona, 2014; Berman, 2018). This is particularly relevant in behavioral neuroscience and in the challenge to understand brain function (Dickinson et al., 2000; Hong et al., 2015; Krakauer, Ghazanfar, Gomez-Marin, MacIver, & Poeppel, 2017). Besides being a pillar in the health sciences, supporting research translation to human clinical trials (Richardson, 2015; Unger et al., 2017), animal behavior analysis is an increasingly important tool in industry, namely in essential animal welfare monitoring in food production (Ahrendt, Gregersen, & Karstoft, 2011; Hong et al., 2015; Stavrakakis et al., 2015).

A full characterization of phenotypic domains in behavioral analysis requires screening test batteries, with different degrees of coverage and validation, implemented in a non-subjective and standardized way. Computerized video-analysis systems have thus emerged as potential tools for automatically assessing behavior, combining two-dimensional (2D) video recordings with image

processing (Robie, Seagraves, Egnor, & Branson, 2017; Valletta, Torney, Kings, Thornton, & Madden, 2017) and machine learning (ML) methods (Aguiar, Mendonca, & Galhardo, 2007; de Chaumont et al., 2012; Jhuang et al., 2010; Preisig et al., 2016). Most published solutions rely on standard background subtraction methods (Aguiar et al., 2007; Jhuang et al., 2010; Twining, Taylor, & Courtney, 2001) for animal segmentation, with dynamic background conditions still under active development. Body-part classification can be addressed using algorithms for learning/computing the individual's pose (Mathis et al., 2018; Pereira et al., 2019). In turn, trajectory-based features (Burgos-Artizzu, Dollár, Lin, Anderson, & Perona, 2012; Kabra, Robie, Rivera-Alba, Branson, & Branson, 2013) can be extracted from video sequences (Dollár, Rabaud, Cottrell, & Belongie, 2005; Jhuang et al., 2010) to describe low-level representations of behavior. These features can then be used for automated behavior classification by applying rule-based classifiers (de Chaumont et al., 2012), or supervised (Burgos-Artizzu et al., 2012; Kabra et al., 2013) and unsupervised (Berman, Choi, Bialek, & Shaevitz, 2014; Schwarz, Branicky, Grundy, Schafer, & Brown, 2015) ML methods to train classifiers. Alternatively, semi-supervised and weakly supervised learning may be introduced in this context, although only modest progress has been made here (Egnor & Branson, 2016; Lorbach, Poppe, & Veltkamp, 2019; Robie et al., 2017).

Nevertheless, as expected, the estimation of an animal's pose in 2D is unsatisfactory in most cases. Therefore, some studies have begun to address the problem in three dimensions (3D), using multiple conventional cameras, or cameras capable of combining color and depth sensing (RGB-D cameras) (Hong et al., 2015; Matsumoto et al., 2013; Wang, Mirbozorgi, & Ghovanloo, 2018).

The present study describes a novel computational solution for automated, markerless 3D segmentation and tracking (in static and dynamic environments) of both whole-body and body parts in experiments with a single freely behaving rodent. This tool uses low-cost RGB-D sensors and machine learning/computer vision techniques to precisely quantify behavioral features in 3D space. Given its focus on automated classification and tracking in depth ( $z$ -axis), our computational tool is called CaT-z. The tool is tested and validated in controlled experiments to assess its performance and precision. It is made freely available to the research community in order to foster reproducible and reliable quantitative behavioral analysis in labs with limited resources.

The CaT-z software is publicly available for download at GitHub: [https://github.com/CaT-zTools/CaT-z\\_Software](https://github.com/CaT-zTools/CaT-z_Software). The open-access dataset (41 GB) is also publicly available for download at Zenodo: <https://doi.org/10.5281/zenodo.3636136>.

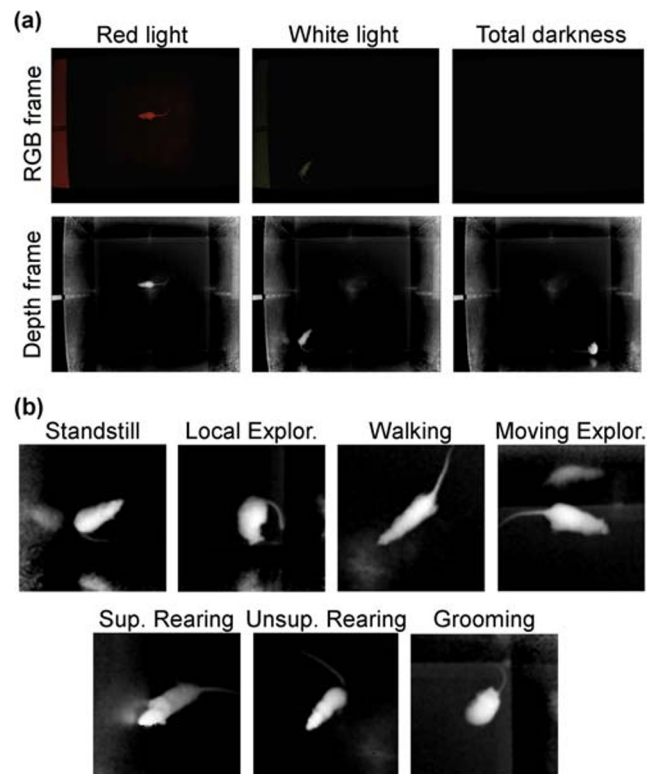
## Materials and methods

### Behavioral protocol

Behavioral experiments for dataset construction and system validation were conducted during three consecutive weeks for each animal ( $n = 2$ ). Inside the experimental environment (an opaque acrylic open-field cage,  $1 \times 1 \times 0.5$  m, made in-house), three types of light conditions were alternatively used: dim red light, dim white light, and total darkness (Fig. 1a). Animals were recorded while moving freely for 15 minutes, using CaT-z software. For behavioral phenotyping studies, Wistar Kyoto rats (WKY;  $n = 10$ ) and wild-type rats ( $n = 10$ ) were subjected to the elevated plus maze (EPM) test (standard apparatus). Animals were allowed to freely explore the maze for 5 minutes. The following measurements were taken: percentage of time spent in the open arms, percentage of time spent in center arena, and total distance, as well as automated classification of seven behaviors (see below).

### Video acquisition

RGB-D videos were recorded using a Microsoft Kinect v2 camera, with  $1920 \times 1080$  color and  $512 \times 424$  depth pixel resolution. It records at a maximum of 30 frames per second (fps), but in low light conditions, this value drops to 15 fps



**Fig. 1** RGB-D behavioral dataset. (a) RGB and depth frames under three different lighting conditions: dim red light, dim white light, and total darkness. (b) Depth frames for the seven types of rodent behaviors

(typically). The operating range is from 0.5 to 4.5 m, with a spatial resolution of  $\approx 2$  mm. The camera was placed centrally above the open field (OF) and the EPM (at a height of 1.20 m, to fully include setup dimensions) and connected to a computer. A pre-heating time of 30 minutes for the camera was adopted for stabilization of the depth sensor (Lachat, Macher, Landes, & Grussenmeyer, 2015).

### Manual annotation of rodents' behaviors

The RGB-D dataset containing frames for supervised classification (ground truth) was fully annotated by researchers with experience in ethology, with one of seven mutually exclusive behavioral labels: standstill, local exploration, moving exploration, walking, supported and unsupported rearing, and grooming (Table 1 in Appendix 2; see Fig. 1b for examples). An extended list of classes is sometimes not necessary, or advisable (increased subjectivity), and consequently a simplified list was also considered: Standstill+ (standstill and local exploration), Walking+ (walking and moving exploration), Rearing (unsupported and supported rearing), and Grooming. The CaT-z software also includes an interface for manual annotation, which was used for the manually annotated dataset for the supervised classification algorithms ("ground truth"). Regarding the observation method, the annotation interface enables the construction of the animal's ethogram based on focal-animal annotations, and all actions of one animal are annotated for a specified time period (all video frames are annotated).

The level of agreement between observers for the annotated dataset was calculated using two different metrics. In the frame-based approach, a one-frame tolerance was allowed in the transitions. In the quality-based approach, the number of matching (overlapping) behavior periods between observers was used.

For the WKY/Wistar EPM experiments, seven mutually exclusive behaviors were also defined: standstill+ (local exploration and standstill), walking+ (walking and moving exploration), rearing (supported and unsupported rearing), head dipping (snout sloping down from the EPM and body standing in the same place with the four legs in the open arms), protective head dipping (snout sloping down from the EPM and body standing in the same place with at least one limb in the closed arms), stretch-attend posture (SAP; hind legs remaining stationary and body extending forward with head raised), and grooming (see Table 1, Appendix 2, for definitions).

### Tracking and classification algorithms

Four computational components are addressed in our method (Fig. 1, Appendix 2): animal segmentation, tracking, feature detection, and classification. All algorithms were

implemented in C++ language, for computational performance, and using the Qt Creator (The Qt Company, Finland) environment to integrate the algorithms in the user-friendly CaT-z software. Three graphical user interfaces (GUIs) were developed to support video acquisition, annotation, and processing (segmentation, tracking, and classification of behavioral data).

**Animal detection and tracking** Animal segmentation was performed using three different background modeling methods. The static median difference method sets a static background model using the median of the pixels over a set of initial frames. A 2D median filter ( $5 \times 5$  size) was also applied. Along the frames, the foreground detection was performed by computing the difference between the current frame and the background model.

In order to cope with dynamically changing environments (e.g. bedding material, small objects moving/(dis)appearing), two other algorithms were developed. Both methods are initialized with a background model similar to the static method.

The landscape change detection (LCD) method uses the background subtraction technique but updates the background model continuously. The updating algorithm uses the assumption that local environment modifications are smaller than the animal's area. The background model is updated using information from the current frame to incorporate possible objects that (dis)appeared/moved in the frame. Finally, the probabilistic Gaussian mixture model (GMM) was adapted from (Stauffer & Grimson, 1999) to incorporate 16-bit depth images in the processing algorithm and improve background estimation.

The validation of these methods under dynamic environments was performed using a controlled synthetic dataset. This dataset consisted of 1000 depth frames whose intensity values followed a normal distribution of mean 1000 mm and standard deviation 5 mm (experimental precision value of this depth sensor). A dynamic environment was simulated by synthetically creating well-defined dips or rises in the depth map. The validation was performed by comparing background models and ground truth.

**Body part detection and tracking** From the 3D segmented animal, five anatomical points were tracked: nose, head, body center (centroid), tail base, and tail end. Importantly, these landmarks were estimated using scale-free geometric constraints/properties. For example, after determining the rodent body contours, the tail end is defined as the furthest contour point from the centroid (independent of animal size). Simple heuristics were implemented to check the validity of the detected body part locations (for example, discrepancy between the positions in consecutive frames). Frames with uncertain body part detection are flagged, and this information is later used for the frame classification (see Table 2 in Appendix):

not only is this flag important for signaling tracking anomalies, but also, interestingly, the absence of particular body parts (e.g. by occlusion) itself can help in detecting certain behaviors (for example, during grooming events, the nose is frequently not detected).

The performance of the body part detection algorithm (which relies on scale-free geometric rules) was evaluated by comparing the automated tracking results with manually annotated locations of body parts in a set of 600+ frames.

**Feature extraction** For the automatic classifiers, low-level representations of behavior were organized into “features” describing trajectory-based aspects and shape-based information (Table 2, Appendix 2). In order to add information from previous frames (temporal memory) and to help distinguish between behaviors with different temporal dynamics, the feature sets for each frame were combined with the features from ~1 second in the past, obtaining a final set of 22 features for each time point. The features were normalized using Z-score transformation.

**Automated behavior classification** The support vector machine (SVM) classifier was selected for supervised and multiclass behavior recognition (Boser, Guyon, & Vapnik, 1992). A nonlinear classifier with a radial basis function kernel was used. The best combination of SVM parameters was selected by grid search, and the parameters with higher cross-validation accuracy were selected, using a  $k$ -fold cross-validation approach ( $k = 5$ ) on the training set.

Performance was estimated using the leave-one-video-out technique, where all but one video of a pool of  $n$  videos were used to train the model, and the performance was evaluated on the remaining video. This procedure was repeated  $n$  times for all videos. Learning curves were constructed to show the classification performance as a function of the training dataset size, and to determine the minimum  $n$  size to construct this pool of videos.

Model predictions for all the testing frames were filtered (with a  $5 \times 5$  median filter) to reduce erroneous classification of isolated frames, and then concatenated to compute the overall accuracy (ratio of correct frames) and performance per class using confusion matrices and the F1-score. The F1-score is the harmonic average of the precision and recall, ranging from 0, with no correct predictions, to 1 for perfect precision and recall, calculated as follows:

$$F1 \text{ score} = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}}$$

where  $\text{precision} = \frac{\text{true positive}}{(\text{true positive} + \text{false positive})}$  and  $\text{recall} = \frac{\text{true positive}}{\text{true positive} + \text{false negative}}$

This metric is better suited for datasets with behaviors that occur with different frequencies (Lorbach et al., 2018). This leave-one-video-out approach provides the

best estimate of the future performance of a classifier, and was also applied to avoid testing bias due to the consecutive frames effect and “double-dipping” (Kriegeskorte, Simmons, Bellgowan, & Baker, 2009).

When studying the activity of WKY rats inside the EPM, only RGB-D data from Wistar rats was used to train the classifier, but both Wistar and WKY data was used as testing sets.

## Behavioral phenotyping

The ability to detect behavioral differences (phenotyping) between different strains was assessed using a  $k$ -nearest neighbor algorithm ( $k$ -NN). This choice served the purpose of demonstrating that even a simple classifier can be used for this step. The model’s accuracy and posterior probabilities of belonging to the control class were calculated for both the Wistar and WKY strains in order to select a reduced set of metrics and to construct a behavioral profile for phenotyping of the strains.

The extended methodology is presented in Appendix 1.

## Results

### An RGB-D behavioral dataset to promote advances in computational ethology

As in other fields, important contributions to computational ethology can arise from ML researchers not directly engaged in behavioral experiments. The availability of large, public, annotated datasets is therefore of fundamental importance to empower these contributions. With this in mind, instead of producing a specific dataset for developing CaT-z, we have compiled a general-purpose dataset, which is made public to catalyze new developments in computational ethology and automated classification of rat behavior activity.

The produced RGB-D dataset was compiled from videos and respective behavior annotations that capture freely walking Wistar rats in an OF arena. The OF was chosen since it is a standard setup commonly used in ethology studies to measure behavioral and locomotor activity in animal models (Belzung, 1999; Cryan & Holmes, 2005; Overstreet, 2012). The dataset is composed of several ~10/15-minute RGB-D video sequences of individual rat behavior, where the animal is allowed to move freely inside the OF cage (Fig. 1). Three different lighting conditions were used (Fig. 1a) to recreate the typical light setups used in behavioral recordings. Total darkness is the ideal lighting condition for the animals’ active phase, but it is usually replaced by dim red light or dim white light due to limitations of the standard recording systems. The full dataset consists of 24 videos, with a total of 6 hours: 4 hours of fully annotated sequences (for

supervised ML methods; ~180,000 annotated frames) and 2 additional hours of raw behavioral sequences (adding data for unsupervised ML methods).

Every RGB-D video frame in the annotated dataset was manually labeled with one of the seven mutually exclusive rat behavioral labels (Fig. 1b), by researchers with experience in ethology. These specific behaviors were selected as they are commonly used in manual scorings in neurobehavioral research. Information regarding the frequency of each behavioral event within the annotated dataset, which ranges from 2.5% for walking events to 37.9% for local exploration events, is described in Table 1 in Appendix 2.

In the manual annotation of animal behavior, reliability between human observers is typically limited to 70–80% (Anderson & Perona, 2014; Spruijt & DeVisser, 2006). This limitation was, in fact, a core motivation for this work. In the annotated dataset, the average level of agreement between the observers was  $83.3\% \pm 5.7$  in a frame-based approach ( $n = 21,988$  frames), and  $81\% \pm 0.8$  in a quality-based approach (agreement on behavioral type; please see Materials and Methods). Taken together, these results reveal that both agreement scores for the annotation of this dataset are consistent with the reported range. The full open-access dataset is freely available for download at Zenodo: <https://doi.org/10.5281/zenodo.3636135>.

### Depth information improves whole-body segmentation in both static and dynamic background conditions

Animal segmentation, a challenging problem in RGB video sequences, is significantly improved and facilitated using depth information combined with the implemented static/dynamic background algorithms (Fig. 2). In the segmented images, it is possible to visually distinguish specific body parts such as tail, snout, and upper and lower limbs (Fig. 2a). For different lighting conditions, there were no differences in detection performance, which means that animal detection is independent of ambient lighting.

The performance of the three background segmentation algorithms (standard static, modified GMM, and the new LCD algorithm) was quantified in controlled dynamic background landscapes (Fig. 2, Appendix 2). The results showed that the LCD method is more effective at dealing with background changes, incorporating them quickly into its depth profile: as the background changes, the pixel depth values change instantaneously, allowing a more accurate estimate of the background. In turn, the modified GMM method also incorporates pixel modifications in the estimated background, but much more slowly than the LCD method, which is consistent with the defined learning rate. As expected, the widely used static median difference method has very limited performance in dynamic environments.

### Tracking multiple anatomical landmarks in 3D

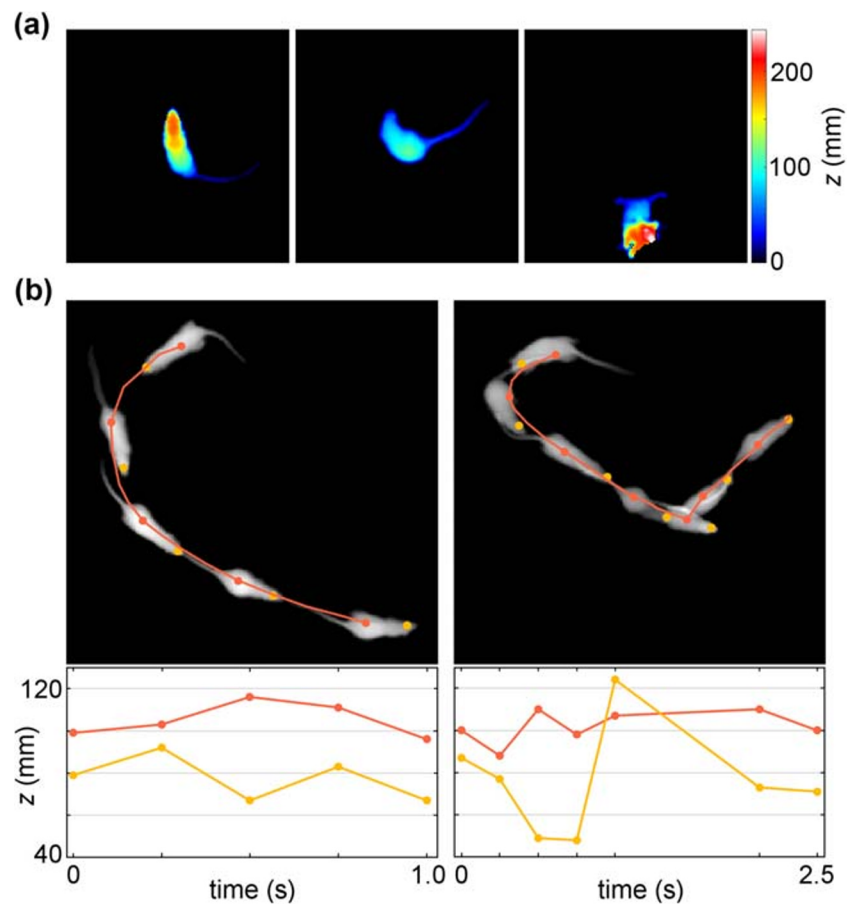
Geometric methods for the detection of body parts greatly benefit from depth information, enabling the detection of the 3D trajectories of each anatomical landmark. Using these representations, it is possible to identify subtle fluctuations in depth which would not be discernible by visual inspection (Fig. 2b).

Overall tracking performance was assessed by comparing automatically predicted coordinates with the manually labeled ones (Fig. 3). In particular, automatically detected positions of the animal's body center are in very high agreement with the carefully manually traced trajectories (Fig. 3a). The trajectories overlap along the frames, with a 5% error in the estimation of the distance traveled by the animal. The errors in estimating the traveled distance can be driven by differences between the visual estimate of the animal's body center and the centroid mathematical estimate, which is affected by other segmented body parts (e.g., tail). For each labeled frame, the  $x$ - and  $y$ -coordinate differences between predicted and manually defined body center coordinates were computed for error quantification (Fig. 3b). The differences in most cases (median) were less than 2 pixels (Fig. 3b). In fact, a difference of 2 pixels between the predicted and manually labeled body center coordinates is barely noticeable and is within the variability of human annotation (Fig. 3c).

The system is also able to automatically locate the position of landmarks for a variety of animal postures (Fig. 3d, e). Nevertheless, when the animal is in ambiguous poses, the performance is reduced (Fig. 3f). Globally, the performance of the system is very high, with the majority (median) of the landmark detection errors being below 2 pixels for the nose and tail-base anatomical points, 4 pixels for the head estimate, and 1 pixel for the tail-end point detection (Fig. 3g–j). The geometric algorithms defined to determine the anatomical points are scale-free, making the tracking system robust to changes in animal size.

### Automated behavior classification using depth information

The proposed automated classification system, based on multiclass SVMs, shows the capability to attain high performance levels even if trained with only 30 minutes of annotated video recording (Fig. 4). As the number of training examples increases, the mean gap between the validation and training scores narrows, and from a training set size of 30,000 examples (~30-min video), both scores stabilize. This level of performance is observed using either simplified or extended annotations, corresponding to either four or seven different types of behaviors (Fig. 4a and b, respectively). The performance levels were assessed using a 5-fold cross-validation approach and avoiding testing bias problems (see Methods). The 30-minute



**Fig. 2** Depth information improves whole-body segmentation. Segmented depth frames, using the static median difference method for background removal, capturing (a) three different behaviors: unsupported rearing (left), local exploration (middle), and supported rearing (right). Background pixels in black. Maximum depth values (240 mm) in white.

Depth colomap in mm. (b) Body part tracking (centroid and nose) and their depth profile. Top: Two sequences of segmented depth frames with identification of some body parts: centroid (in orange) and head (in yellow). Bottom: Corresponding depth profile (in mm) for the centroid and head points in the depth frames sequences

figure is important, as compared with the very large training datasets required by other approaches, particularly deep learning. It means that the manual annotation effort may be markedly reduced in supervised training approaches. For consistency, the results presented from here on were all obtained with training datasets with roughly 30 minutes of video.

Standard methods for automated behavior analysis (EthoVision XT, Noldus, Netherlands; SMART, Panlab, Spain; Kabra et al., 2013) are not fully functional under total dark conditions, which is an important limiting factor for recording natural rodent behavior. Our methods are independent of ambient light conditions (dim red, dim white, and total darkness) as shown by automated classification accuracy and F1-scores (Table 3, Appendix 2). Moreover, the system generalizes among different lighting conditions; for example, dim red light videos can be used for training and total darkness for testing (Table 3, Appendix 2).

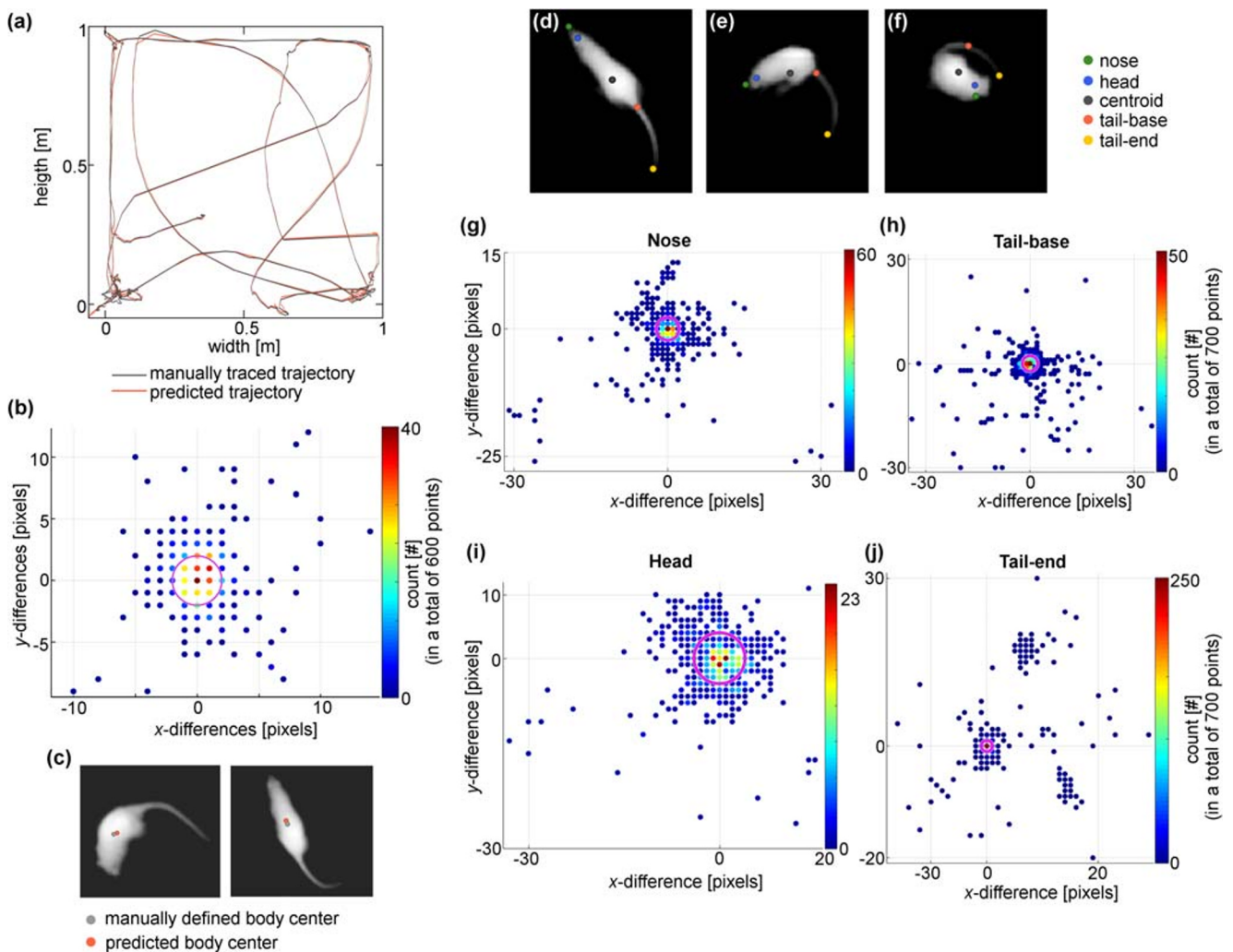
For a detailed analysis of the classification errors for each type of behavior, we constructed confusion matrices, showing the combinations of predicted and real/annotated values (examples in Fig. 4c, d). For the simplified annotations (four

classes), the average accuracy was 84.9%, with high F1-score values for all behaviors (Fig. 4c), whereas in the extended annotations (seven classes) the average accuracy was 76.9%. In both conditions, the presently defined features for the SVM classifier enable the system to correctly recognize most behaviors (Movie 1, Appendix 2). In the extended annotations, the current system shows some limitations. Walking periods belong to the most misclassified behaviors, occasionally classified as moving exploration, leading to low F1-scores. Also, F1-scores for standstill are very low, or not possible to calculate due to lack of representativeness in the training set.

The automated classification methods presented here allow the direct generation of ethograms to describe the behavioral data, and the time spent on each behavior (Fig. 4e).

### CaT-z: a user-friendly computational solution for quantifying animal behavior

Acknowledging the paramount importance of encapsulating all algorithms in a user-friendly application suited for



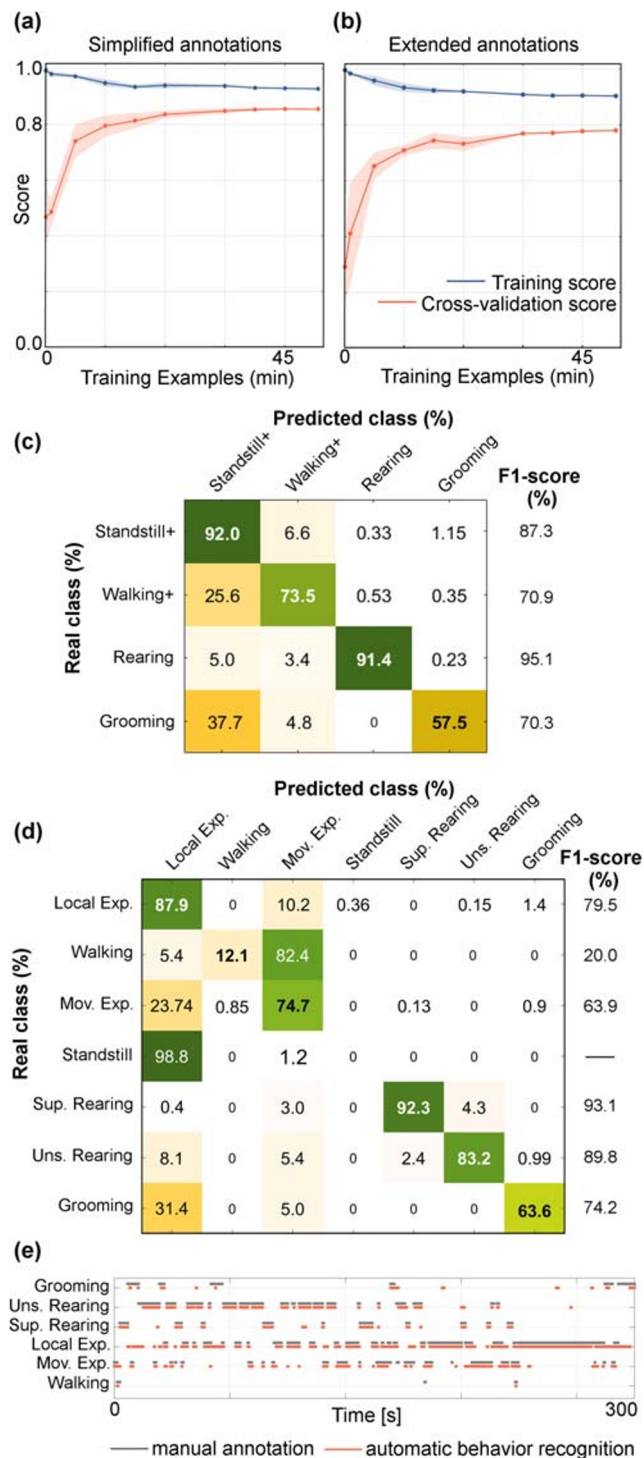
**Fig. 3** Multiple anatomical landmarks can be accurately tracked in 3D. (a–c) Comparison between manually defined body center and automatically predicted coordinates for a 40-second frame sequence. (a) Manually traced (gray) and predicted (orange) trajectories inside the open-field cage. (b)  $x$ - and  $y$ -coordinate differences, in pixels, between manually defined and predicted centroid coordinates. Colorbar indicates  $x$ - and  $y$ -coordinate difference occurrences. The circle in magenta (2-pixel radius) represents 50% of the results. (c) Example images with manually defined body center (gray) and predicted (orange) coordinates, where the

distance is equal to the median value (2 pixels). (d–f) Examples of body part detection in several frames of a single video. (f) An example of incorrect detection of tail-base and nose body parts. (g–j) Histograms of coordinate differences, in pixels, between manually defined and predicted body part coordinates, for a 46-second frame sequence. Colorbar indicates  $x$ - and  $y$ -coordinate difference occurrences. The circle, in magenta, represents 50% of the results, whose distance radius is 1.0, 2.1, 2.2, and 4.0 pixels for (g–j) histograms, respectively. Scale factor calculated using open-field setup dimensions (scale factor = 3.2 mm/pixel)

laboratory environments, an effort was made to create an integrated, easy-to-use, and freely available software that works off the shelf—CaT-z. This computational tool contains three different modules to support annotation and recording of RGB-D frames, and automated tracking and classification of rodent behavior (Fig. 5). The graphical user interface (GUI) for RGB-D data visualization and annotation (Fig. 5a) allows the manual scoring of color and depth frame, simultaneously into user-defined behaviors. Depth frames can be displayed in three different visualizations, and RGB-D videos can be played using media controls (in different velocities). During annotation, a behavioral ethogram is automatically updated to

give color feedback on the behaviors previously identified. It is possible to resume an unfinished annotation, and finally, the data is saved in .csv format for later use in automated behavior analysis. To the best of our knowledge, such RGB-D data annotation tools are not presently available.

New RGB-D data can be acquired using the data acquisition GUI (Fig. 5b), and later annotated or analyzed by the tracking/behavior classification GUI (Fig. 5c). Segmentation and tracking are performed using different available methods, and a particular region of interest can be selected. Body-part tracking information ( $x$ -,  $y$ -, and  $z$ -coordinates) can be exported to a user-defined



**Fig. 4** Automated behavioral recognition performance. **(a)**, **(b)** Learning curves of trained model for the recognition of four (simplified annotations) or seven (extended annotations) behaviors, respectively. Results represented as mean (filled line) and SD (colored shadow) for training (blue) and cross-validation (orange) scores. **(c)**, **(d)** Examples of normalized confusion matrix of automated behavioral recognition and corresponding F1-scores, for four or seven classes, respectively. **(e)** Example of ethogram for manual annotations (gray) and automated behavioral recognition labels (orange), over 300 seconds of testing video

directory. Finally, using previous tracking information and annotated data, the classifier can be trained, tested, or applied for the recognition of new behavioral data. The GUI also allows the classifier to be trained with multiple videos simultaneously, without the need for multiple launches. Importantly, CaT-z is made available to the community with a detailed user manual and tutorial/walkthrough videos ([https://github.com/CaT-zTools/CaT-z\\_Software](https://github.com/CaT-zTools/CaT-z_Software)).

### Ability to distinguish between strains: automated behavioral phenotyping

The behavioral profile of WKY rats was quantitatively compared with that of Wistar rats using CaT-z. The system was capable of automatically detecting behavioral differences between strains (behavioral phenotyping) (Fig. 6). Specific ethology metrics were calculated from the tracking data to assess the degree of activity within EPM: percentage of time in open arms, total distance traveled, and percentage of time in the EPM center. In most cases, no significant differences were found between genders within the same strain (Fig. 3, Appendix 2), and therefore the gender variable was dropped.

As expected, WKYs generally spend less time in the open arms of the EPM ( $p < 0.05$ ), since they are a strain characterized by high levels of anxiety and depression, as well as less time in the center of the EPM ( $p < 0.05$ ) (Fig. 6a). There also appears to be a decrease in the traveled distance in WKY when compared to Wistar rats (but not reaching statistical significance). These results are consistent with the fact that WKY animal behavior is generally less exploratory (D'Souza & Sadananda, 2017; Langen & Dost, 2011).

The specific set of types of behaviors for the EPM were quantified and compared between strains, and, as before, no differences were found between genders within the same strain (Fig. 3, Appendix 2). When comparing the two strains (Fig. 6b), WKY animals spent less time in rearing periods than the Wistar rats ( $p < 0.01$ ), whereas there were no statistically significant differences between groups in the other behaviors.

The combination of the metrics %time walking, %time rearing, and %time in the open arms enable a high discrimination power when comparing strains using a  $k$ -NN classifier (Fig. 6c): accuracy of 79% and average posterior probabilities of  $96\% \pm 12.6$  and  $25\% \pm 15.4$  for a control or WKY sample, respectively, belonging to the control class. In addition, and according to the confusion matrix using these metrics, two in every ten WKY rats were misclassified as belonging to the control class (20% false-positives rate), while 22% of the controls were misclassified as not belonging to the Wistar class (false-negative rate).



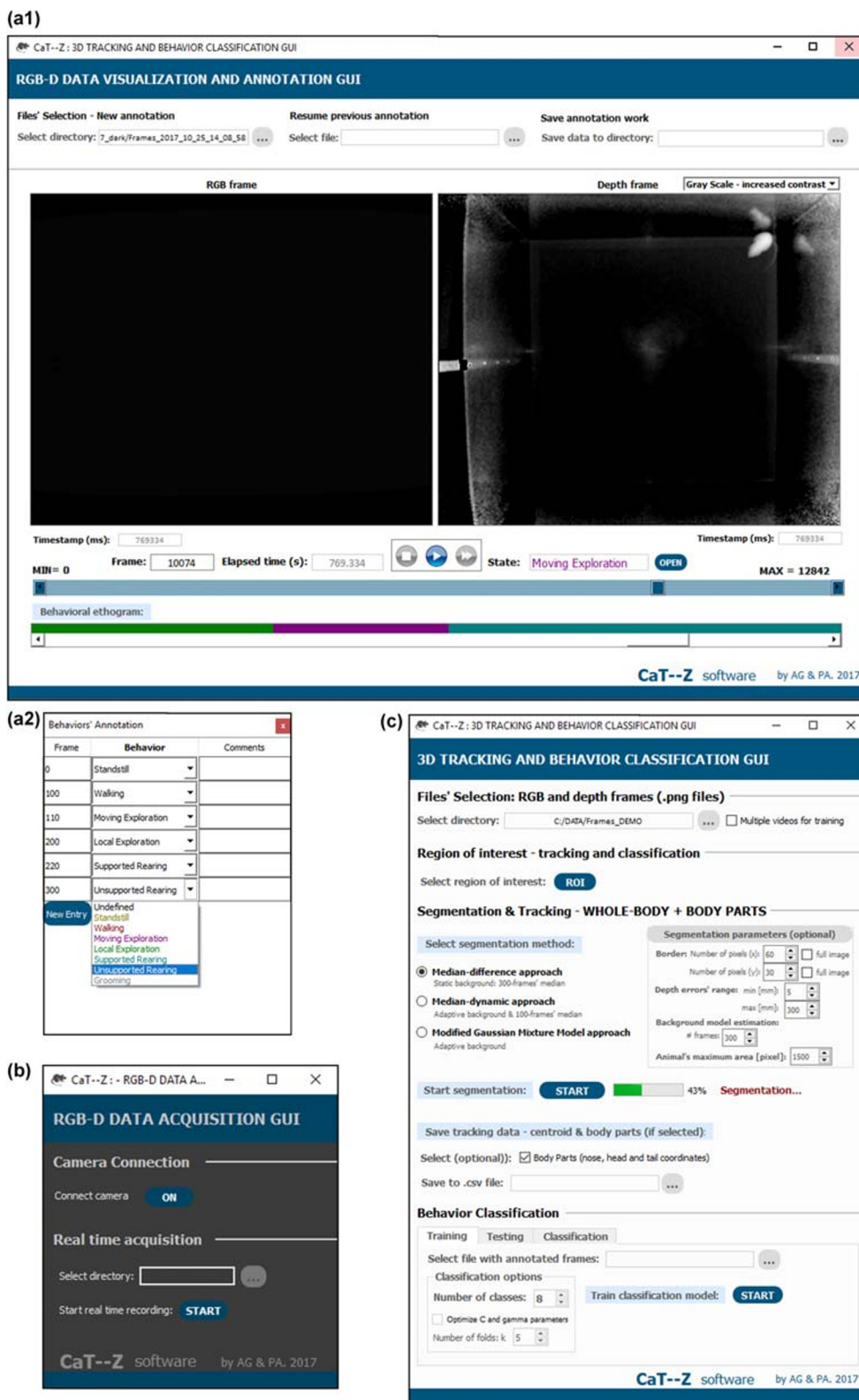
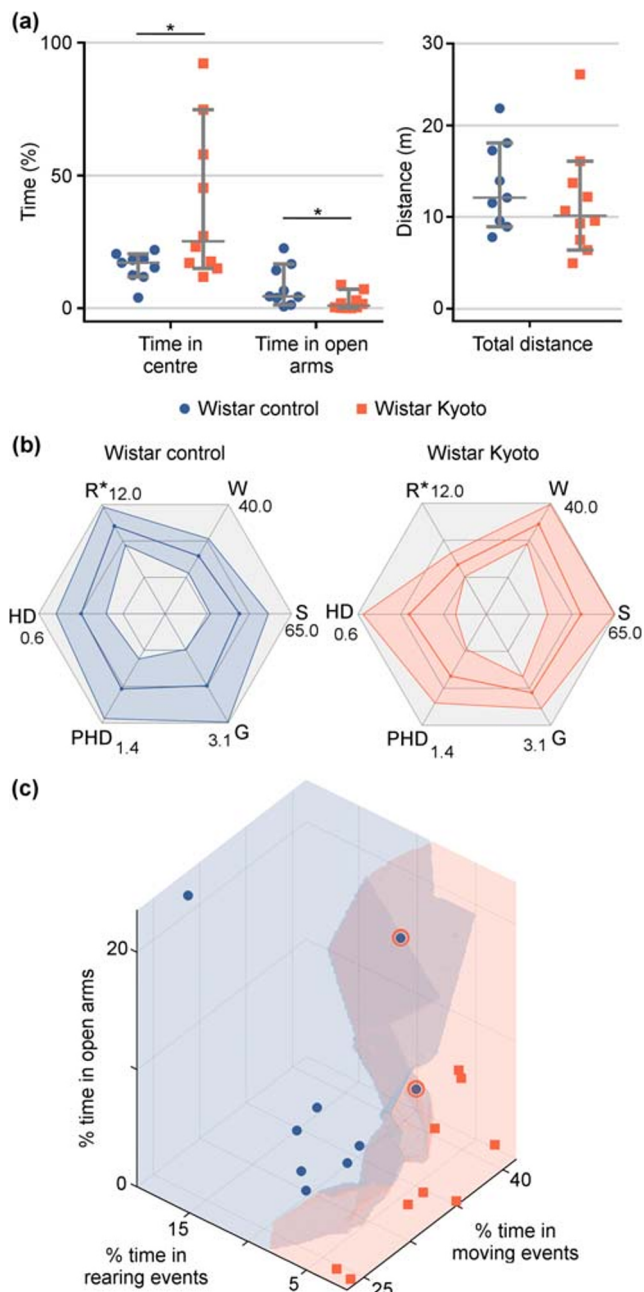


Fig. 5 CaT-z: a free computational solution for quantifying animal behavioral features in depth (z). Graphical user interface (GUI) of the applications developed for (a1) RGB-D frames visualization and

annotation (main window); (a2) dock window for the annotation; (b) RGB-D data acquisition (dark mode for animal facility environments); (c) 3D segmentation, tracking and behavior classification



**Fig. 6** Distinction between Wistar Kyoto strains (behavioral phenotyping) made easy using RGB-D information. **(a)** Motor activity measurements inside the elevated plus maze, for Wistar control (blue circles) and Wistar Kyoto (orange squares) rats. Data represented as median  $\pm$  95% confidence interval. \* $p < 0.05$ . **(b)** Radar plot of automated classification of behaviors for Wistar control (blue) and Wistar Kyoto (orange) rats. Solid lines (both blue and orange) represent median values. Shaded areas (both blue and orange) represent  $\pm$  95% confidence interval. \* $p < 0.05$ . **(c)** Three-dimensional representation of clustering results, for Wistar control (blue area) and Wistar Kyoto (orange area) rats, regarding three features: % of time moving, rearing, and in open arms. Blue circles and orange squares represent well-classified points, for Wistar control and Wistar Kyoto, respectively. Blue circles with orange margin represent misclassified points, regarding the decision region of the clustering algorithm: both points should belong to the Wistar control area but were misclassified as Wistar Kyoto points. Standstill (S); Walking (W); Rearing (R); Head dipping (HD); Protective head dipping (PHD); Grooming (G)

Thus, the results show that, although statistically significant differences are not found in isolated metrics, when they are combined, it is possible to distinguish the two strains with a 79% degree of accuracy. Furthermore, it is possible to construct behavioral profiles characteristic of each strain, with 20% false positives.

## Discussion

The core goal of this work was to develop a free and fully integrated system for 3D segmentation, tracking, and classification to automatically detect and quantify behaviors in rodents. With the developed algorithms, the CaT-z tool is capable of performing segmentation of a single animal's whole body in complex backgrounds, tracking multiple body parts, and detecting different behaviors. These methods are embedded in a user-friendly software package, supported by a publicly available manual. The outputs of this tool include 3D coordinates of body parts, automatically predicted behaviors, and, if applicable, corresponding performance metrics. From the 3D coordinates, one can construct trajectories and extract other motor parameters, such as distance traveled, average velocities, and periods of active movement.

Importantly, this work also introduces the first publicly available RGB-D rat behavioral dataset that is suitable for training automated behavior recognition in rodents, catalyzing new ML developments.

From the results, it was shown that 30 minutes of annotated video of freely walking movement is already sufficient to train our multiclass SVM classifier and attain accuracy levels comparable to the level of agreement in human observers (70–80%). The 30-minute figure is worth emphasizing, since other methods, namely deep learning, typically require many hours of annotated videos to reach high accuracy levels (but see Mathis et al., 2018; Pereira et al., 2019). The ability to generalize is also fundamental in ML systems and, as demonstrated with the phenotyping experiments, CaT-z is able to cope not only with different setups but also with new types of behavior (without the need to redefine the features).

The use of depth sensors in analyzing animal behavior includes advantages that go well beyond simply adding a third dimension. Several research groups have considered its potential application for segmentation and tracking of rodents (Ou-Yang, Tsai, Yen, & Lin, 2011; Paulino Fernandez, van Dam, Noldus, & Veltkamp, 2014), as well as to estimate their pose, and social and nonsocial interactions (Hong et al., 2015; Matsumoto et al., 2013; Wang et al., 2018). However, limitations such as marker imposition, basic pose/behavior recognition, manual intervention, integration in a user-friendly public software, or insufficient classifier performance have limited their use. In addition to presenting important advantages over other

approaches, CaT-z can be used to compare behavioral profiles (“behavioral fingerprints”) of different strains. Previous studies have shown that WKY rats exhibit a combination of anxiety- and depressive-like behaviors, as well as hypoactivity and decreased locomotion and social interaction levels (Burke et al., 2016; D'Souza & Sadananda, 2017; Langen & Dost, 2011). With our system, we were able to automatically quantify several behavioral differences that confirm these findings. More importantly, it was possible to automatically predict the strain of individual animals (with low false-positive and false-negative rates). While automated behavioral phenotyping can be achieved in some conditions using home-cage 2D video data (EthoVision XT, Noldus, Netherlands; Jhuang et al., 2010), this process can be greatly facilitated and improved when 3D information is available. Currently available solutions for automated behavioral phenotyping are often very expensive and are limited to constrained/controlled environments (HomeCageScan, CleverSys Inc, USA; LABORAS, Metris, Netherlands; PhenoCube, PsychoGenics, USA) or require the use of radio-frequency identification (RFID) implants which may affect animal behavior itself (IntelliCage, TSE, Germany; Weissbrod et al., 2013). For all of these reasons, we are convinced that CaT-z has an important role to play in the computational ethology landscape.

The CaT-z software is freely available for download at GitHub ([https://github.com/CaT-zTools/CaT-z\\_Software](https://github.com/CaT-zTools/CaT-z_Software)). The open-access dataset (41 GB) is also available at Zenodo (<https://doi.org/10.5281/zenodo.3636136>).

**Acknowledgments** AG and PA would like to thank Jaime Cardoso for a helpful discussion on potential strategies for performance improvement of SVMs. This work was funded in part by National Funds through FCT – Fundação para a Ciência e a Tecnologia, under the project EXPL-AMAGALHÃES-IF/00753/2014/CP1241/CT0005, that also funds AM. Funding also by FEDER – Fundo Europeu de Desenvolvimento Regional – funds through the COMPETE 2020 Operational Programme for Competitiveness and Internationalisation (POCI), Portugal 2020, and by Portuguese funds through FCT – Fundação para a Ciência e a Tecnologia/Ministério da Ciência, Tecnologia e Ensino Superior in the framework of the project “Institute for Research and Innovation in Health Sciences” (POCI-01-3870145-FEDER-007274). PA was supported by Programa Ciência – Programa Operacional Potencial Humano (POPH) – Promotion of Scientific Employment, ESF. AG is funded by FCT – Fundação para a Ciência e a Tecnologia, grant contract SFRH/BD/137385/2018.

**Author contributions** AG implemented the algorithms, performed the experiments, and was responsible for acquiring all the data. AM and AG annotated the datasets. AG and PA developed the algorithms, analyzed and interpreted the data, and wrote the main manuscript. PA devised the project and main conceptual ideas. All authors discussed the results and contributed to the final manuscript.

**Data accessibility** The following dataset was generated: Gerós A, Magalhães A, Aguiar P, 2019, RGB-D rat dataset. The open-access dataset (41 GB) is available for download at Zenodo (<https://doi.org/10.5281/zenodo.3636135>). The CaT-z software is also available for download at GitHub: [https://github.com/CaT-zTools/CaT-z\\_Software](https://github.com/CaT-zTools/CaT-z_Software).

## References

- Aguiar, P., Mendonca, L., & Galhardo, V. (2007). OpenControl: a free opensource software for video tracking and automated control of behavioral mazes. *J Neurosci Methods*, 166(1), 66–72. doi:<https://doi.org/10.1016/j.jneumeth.2007.06.020>
- Ahrendt, P., Gregersen, T., & Karstoft, H. (2011). Development of a real-time computer vision system for tracking loose-housed pigs. *Computers and Electronics in Agriculture*, 76(2), 169–174. doi:<https://doi.org/10.1016/j.compag.2011.01.011>
- Anderson, D. J., & Perona, P. (2014). Toward a science of computational ethology. *Neuron*, 84(1), 18–31. doi:<https://doi.org/10.1016/j.neuron.2014.09.005>
- Belzung, C. (1999). .11 Measuring rodent exploratory behavior. In *Techniques in the behavioral and neural sciences* (Vol. 13, pp. 738–749): Elsevier.
- Berman, G. J. (2018). Measuring behavior across scales. *BMC biology*, 16(1), 23. doi:<https://doi.org/10.1186/s12915-018-0494-7>
- Berman, G. J., Choi, D. M., Bialek, W., & Shaevitz, J. W. (2014). Mapping the stereotyped behaviour of freely moving fruit flies. *J R Soc Interface*, 11(99). doi:<https://doi.org/10.1098/rsif.2014.0672>
- Boser, B. E., Guyon, I. M., & Vapnik, V. N. (1992). *A training algorithm for optimal margin classifiers*. Paper presented at the Proceedings of the fifth annual workshop on Computational learning theory.
- Burgos-Artizzu, X. P., Dollár, P., Lin, D., Anderson, D. J., & Perona, P. (2012). *Social behavior recognition in continuous video*. Paper presented at the Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on.
- Burke, N. N., Coppinger, J., Deaver, D. R., Roche, M., Finn, D. P., & Kelly, J. (2016). Sex differences and similarities in depressive- and anxiety-like behaviour in the Wistar-Kyoto rat. *Physiol Behav*, 167, 28–34. doi:<https://doi.org/10.1016/j.physbeh.2016.08.031>
- Cryan, J. F., & Holmes, A. (2005). Model organisms: the ascent of mouse: advances in modelling human depression and anxiety. *Nature reviews Drug discovery*, 4(9), 775.
- de Chaumont, F., Coura, R. D., Serreau, P., Cressant, A., Chabout, J., Granon, S., & Olivo-Marin, J. C. (2012). Computerized video analysis of social interactions in mice. *Nat Methods*, 9(4), 410–417. doi:<https://doi.org/10.1038/nmeth.1924>
- Dickinson, M. H., Farley, C. T., Full, R. J., Koehl, M., Kram, R., & Lehman, S. (2000). How animals move: an integrative view. *Science*, 288(5463), 100–106.
- Dollár, P., Rabaud, V., Cottrell, G., & Belongie, S. (2005). *Behavior recognition via sparse spatio-temporal features*. Paper presented at the Visual Surveillance and Performance Evaluation of Tracking and Surveillance, 2005. 2nd Joint IEEE International Workshop on.
- D'Souza, D., & Sadananda, M. (2017). Anxiety- and depressive-like profiles during early- and mid-adolescence in the female Wistar Kyoto rat. *Int J Dev Neurosci*, 56, 18–26. doi:<https://doi.org/10.1016/j.ijdevneu.2016.11.003>
- Egnor, S. E., & Branson, K. (2016). Computational Analysis of Behavior. *Annu Rev Neurosci*, 39, 217–236. doi:<https://doi.org/10.1146/annurev-neuro-070815-013845>
- Hong, W., Kennedy, A., Burgos-Artizzu, X. P., Zelikowsky, M., Navonne, S. G., Perona, P., & Anderson, D. J. (2015). Automated measurement of mouse social behaviors using depth sensing, video tracking, and machine learning. *Proc Natl Acad Sci U S A*, 112(38), E5351–E5360. doi:<https://doi.org/10.1073/pnas.1515982112>
- Jhuang, H., Garrote, E., Mutch, J., Yu, X., Khilnani, V., Poggio, T., ... Serre, T. (2010). Automated home-cage behavioural phenotyping of mice. *Nat Commun*, 1, 68. doi:<https://doi.org/10.1038/ncomms1064>
- Kabra, M., Robie, A. A., Rivera-Alba, M., Branson, S., & Branson, K. (2013). JAABA: interactive machine learning for automatic annotation of animal behavior. *Nat Methods*, 10(1), 64–67. doi:<https://doi.org/10.1038/nmeth.2281>

- Krakauer, J. W., Ghazanfar, A. A., Gomez-Marín, A., MacIver, M. A., & Poeppel, D. (2017). Neuroscience Needs Behavior: Correcting a Reductionist Bias. *Neuron*, *93*(3), 480–490. doi:<https://doi.org/10.1016/j.neuron.2016.12.041>
- Kriegeskorte, N., Simmons, W. K., Bellgowan, P. S. F., & Baker, C. I. (2009). Circular analysis in systems neuroscience: the dangers of double dipping. *Nature Neuroscience*, *12*, 535. doi:<https://doi.org/10.1038/nn.2303>, <https://www.nature.com/articles/nn.2303#supplementary-information>
- Lachat, E., Macher, H., Landes, T., & Grussenmeyer, P. (2015). Assessment and Calibration of a RGB-D Camera (Kinect v2 Sensor) Towards a Potential Use for Close-Range 3D Modeling. *Remote Sensing*, *7*(10), 13070–13097. doi:<https://doi.org/10.3390/rs71013070>
- Langen, B., & Dost, R. (2011). Comparison of SHR, WKY and Wistar rats in different behavioural animal models: effect of dopamine D1 and alpha2 agonists. *Atten Defic Hyperact Disord*, *3*(1), 1–12. doi:<https://doi.org/10.1007/s12402-010-0034-y>
- Lorbach, M., Kyriakou, E. I., Poppe, R., van Dam, E. A., Noldus, L., & Veltkamp, R. C. (2018). Learning to recognize rat social behavior: Novel dataset and cross-dataset application. *J Neurosci Methods*, *300*, 166–172. doi:<https://doi.org/10.1016/j.jneumeth.2017.05.006>
- Lorbach, M., Poppe, R., & Veltkamp, R. C. (2019). Interactive rodent behavior annotation in video using active learning. *Multimedia Tools and Applications*, 1–20.
- Mathis, A., Mamidanna, P., Cury, K. M., Abe, T., Murthy, V. N., Mathis, M. W., & Bethge, M. (2018). DeepLabCut: markerless pose estimation of user-defined body parts with deep learning. *Nat Neurosci*, *21*(9), 1281–1289. doi:<https://doi.org/10.1038/s41593-018-0209-y>
- Matsumoto, J., Urakawa, S., Takamura, Y., Malcher-Lopes, R., Hori, E., Tomaz, C., ... Nishijo, H. (2013). A 3D-video-based computerized analysis of social and sexual interactions in rats. *PLoS One*, *8*(10), e78460. doi:<https://doi.org/10.1371/journal.pone.0078460>
- Ou-Yang, T. H., Tsai, M. L., Yen, C. T., & Lin, T. T. (2011). An infrared range camera-based approach for three-dimensional locomotion tracking and pose reconstruction in a rodent. *J Neurosci Methods*, *201*(1), 116–123. doi:<https://doi.org/10.1016/j.jneumeth.2011.07.019>
- Overstreet, D. H. (2012). Modeling depression in animal models. In *Psychiatric Disorders* (pp. 125–144): Springer.
- Paulino Fernandez, O., van Dam, E. A., Noldus, L., & Veltkamp, R. (2014). *Robust Point Cloud Segmentation of Rodents using Close Range Depth Cameras in Controlled Environments*. Paper presented at the proceedings ICPR workshop on Visual observation and analysis of Vertebrate And Insect Behavior.
- Pereira, T. D., Aldarondo, D. E., Willmore, L., Kislin, M., Wang, S. S.-H., Murthy, M., & Shaevitz, J. W. (2019). Fast animal pose estimation using deep neural networks. *Nature Methods*, *16*(1), 117.
- Preisig, D. F., Kulic, L., Kruger, M., Wirth, F., McAfoose, J., Spani, C., ... Welt, T. (2016). High-speed video gait analysis reveals early and characteristic locomotor phenotypes in mouse models of neurodegenerative movement disorders. *Behav Brain Res*, *311*, 340–353. doi:<https://doi.org/10.1016/j.bbr.2016.04.044>
- Richardson, C. A. (2015). The power of automated behavioural homecage technologies in characterizing disease progression in laboratory mice: A review. *Applied Animal Behaviour Science*, *163*, 19–27. doi:<https://doi.org/10.1016/j.applanim.2014.11.018>
- Robie, A. A., Seagraves, K. M., Egnor, S. E., & Branson, K. (2017). Machine vision methods for analyzing social interactions. *J Exp Biol*, *220*(Pt 1), 25–34. doi:<https://doi.org/10.1242/jeb.142281>
- Schwarz, R. F., Branicky, R., Grundy, L. J., Schafer, W. R., & Brown, A. E. (2015). Changes in Postural Syntax Characterize Sensory Modulation and Natural Variation of *C. elegans* Locomotion. *PLoS Comput Biol*, *11*(8), e1004322. doi:<https://doi.org/10.1371/journal.pcbi.1004322>
- Spruijt, B. M., & DeVisser, L. (2006). Advanced behavioural screening: automated home cage ethology. *Drug Discov Today Technol*, *3*(2), 231–237. doi:<https://doi.org/10.1016/j.ddtec.2006.06.010>
- Stauffer, C., & Grimson, W. E. L. (1999). *Adaptive background mixture models for real-time tracking*. Paper presented at the Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on.
- Stavrakakis, S., Li, W., Guy, J. H., Morgan, G., Ushaw, G., Johnson, G. R., & Edwards, S. A. (2015). Validity of the Microsoft Kinect sensor for assessment of normal walking patterns in pigs. *Computers and Electronics in Agriculture*, *117*, 1–7. doi:<https://doi.org/10.1016/j.compag.2015.07.003>
- Twining, C., Taylor, C., & Courtney, P. (2001). Robust tracking and posture description for laboratory rodents using active shape models. *Behavior Research Methods, Instruments, & Computers*, *33*(3), 381–391.
- Unger, J., Mansour, M., Kopaczka, M., Gronloh, N., Spehr, M., & Merhof, D. (2017). An unsupervised learning approach for tracking mice in an enclosed area. *BMC Bioinformatics*, *18*(1), 272. doi:<https://doi.org/10.1186/s12859-017-1681-1>
- Valletta, J. J., Torney, C., Kings, M., Thornton, A., & Madden, J. (2017). Applications of machine learning in animal behaviour studies. *Animal Behaviour*, *124*, 203–220. doi:<https://doi.org/10.1016/j.anbehav.2016.12.005>
- Wang, Z., Mirbozorgi, S. A., & Ghovanloo, M. (2018). An automated behavior analysis system for freely moving rodents using depth image. *Med Biol Eng Comput*, *56*(10), 1807–1821. doi:<https://doi.org/10.1007/s11517-018-1816-1>
- Weissbrod, A., Shapiro, A., Vasserman, G., Edry, L., Dayan, M., Yitzhaky, A., ... Kimchi, T. (2013). Automated long-term tracking and social behavioural phenotyping of animal colonies within a semi-natural environment. *Nature communications*, *4*, 2018. doi:ARTN 2018. <https://doi.org/10.1038/ncomms3018>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.