LONG-TERM NON-CONTACT TRACKING OF CAGED RODENTS

Behnaz Rezaei[1], *Xiaofei Huang*[1], *Jason R. Yee*[2], *Sarah Ostadabbas*[1]

[1]Electrical and Computer Engineering Department, Northeastern University, Boston, USA.
[2]Department of Psychology, Northeastern University, Boston, USA.

ABSTRACT

Automatic tracking of rodents' behaviors over time in their home cages is of great interest in psycho-physiological studies. The commercially-available animal monitoring systems use RGB videos or bio-potential signals to monitor behaviors of animals when exploring their surroundings. The based models of these devices starts from several thousands of dollars and the cost would increase if extra analysis features were added. In this study, we present a low-cost, non-contact animal tracking system which records depth data from the caged rodent to detect the animal's location and pose over time. An adaptive Gaussian Mixture Model (GMM) algorithm is employed to detect animal's center of mass and extract its movement trajectory over an extended period of time. The animal's pose is determined by applying Principle Component Analysis (PCA) on 3D depth data of the located animal. In conjunction with our previously-introduced respiratory detection algorithm, this system can be utilized as an automatic long-term and unobtrusive monitoring system for animal experiments. We validated the tracking accuracy of our system by monitoring two different caged voles. The voles' locations were correctly detected in 80% of times, while the poses were detected correctly in 100% of times confirmed by visually inspecting the color-coded depth videos.

Index Terms— Animal behavior monitoring, Gaussian mixture model, Kinect depth sensor, Principle component analysis.

1. INTRODUCTION

1.1. Animal Behavior Monitoring

Pervasive use of rodents as animal models in biological and psychological studies have generated a growing interest in developing automated laboratory apparatus for long-term monitoring of animal behaviors [1, 2]. Classically, the animal's behavioral patterns are watched (or taped) by researchers during the experiments, especially when a certain stimulus is induced. The real-time inspection by human observers is usually performed in short time intervals immediately after the stimulation. However, to acquire a comprehensive evaluation of animal behaviors before/during/after the stimulation, long-term monitoring of freely behaving animals in their cages seems necessary [1, 3, 4]. Automated behavioral monitoring systems are able to record hours or days of animal behaviors during the experiment and make the long-term in-cage screening feasible. If equipped with further data processing units, these systems may even highlight the behaviors of interest or detect behavioral abnormalities, which will significantly increase the researchers' time efficiency in animal studies.

1.2. Commercial Systems

Due to the high impact of such laboratory machinery in providing high throughput analysis, different companies and research groups are competing to fully automate the long-term animal behavioral monitoring and recognition. Among them, there are several commercial video and bio-potential monitoring devices [5, 6, 7, 8]. The main focus of the video tracking systems is on observing the behaviors of the rodents while the animal explores the test objects and its surrounding environment. Telemetry bio-potential monitoring systems use implanted electrodes to record potential body signals. The four leading companies working on the rodent behavioral monitoring in the cage are ViewPoint Behavioral Technology [5], Clever Sys Inc. [6], Noldus [7], and emka Technology [8].

ViewPoint has commercialized a neural headset that allows continuous and simultaneous telemetry monitoring of caged rodents [5]. CleverSys has designed four separate RGB scanning system in order to provide the best angle for collecting the most information from animal in each category of behavioral studies: side-view, top-view, ventral-view, and dual or stereo-view [6]. Noldus has introduced EthoVision XT10 system for video tracking of rodents as a tool for behavioral analysis of rodents [7]. Emka Technologies provides implanted bio-potential measurement devices for freely moving animals 50g and upward. The telemetry system is head mounted or backpack with the ability to measure EEG, ECG, EMG and acceleration [8]. Currently, the cost of these systems based on their functionality starts from $\sim$ \$7K and could increase to a few hundred thousands of dollars after adding different features such as pose detection or multiple animals tracking.

Corresponding author's e-mail: ostadabbas@ece.neu.edu

1.3. Our Animal Tracking System

In this paper, we present our vision-based rodent monitoring system designed for long-term behavioral tracking of freely moving voles in a home-cage environment. Our system locates the vole in its cage and extracts its movement trajectory and the speed of the movement over time by capturing the depth video of the animal using Microsoft Kinect depth sensor. We model the vole as the foreground object and extract its location by employing an adaptive Gaussian Mixture Model (GMM) algorithm to subtract the background. According to the calculated traveling distance in a specific time interval, our algorithm detects the time periods during the experiment when the animal is at rest. At these intervals, the animal's pose is recognized by applying principle component analysis (PCA) on the 3D cloud of the located animal and the trunk part of the vole as the region of interest for the respiration estimation is selected. The details of respiration calculation using depth data was previously presented in [9, 10].

Our system provides two key advantages compared to the state-of-the-art automated animal monitoring systems: (i) being completely non-contact which preserves the natural behaviors of the animal under study, and (ii) it is developed based on a low-cost hardware (Microsoft Kinect as one of the inexpensive but accurate depth sensors) and an open-source software (developed in C# and MATLAB). In addition, unlike the implanted bio-potential measurement instruments, it is thoroughly non-invasive for the respiration monitoring, which is one of the most engaging physiological indicators in psycho-physiological studies in rodents [11, 12, 13].

2. METHODOLOGY

We used a Microsoft Kinect v2 to record depth data for tracking and respiration monitoring of rodent voles. The advantage of using depth sensor data instead of RGB camera in detection is that the system can perform tracking in different lighting conditions without performance degradation.

2.1. Animal Mask Detection

Object tracking in dynamic scenes, by modeling the objects of interest as the foreground scenes, is an active research topic in computer vision especially in surveillance applications [14]. In our animal tracking application, although the dynamics of visual field is relatively limited due to voles being caged, there still exist some situations hindering the successful tracking of animal over an extended period of time. These include slow moving, even motionless animal during some periods of times, Kinect depth sensor recording noise, and moving elements of the scene such as cage's bedding. We employed an adaptive background mixture models to conquer these problems [15, 16].

2.1.1. Pre-processing

The depth sensor measurements are fed as images to our object detection algorithm, after an outlier elimination and data normalization on the raw data were applied. Due to both inherent low signal-to-noise-ratio of Kinect depth sensor in measuring length smaller than 5mm and existence of areas with low infrared reflectance, there are some false measurements with values much greater than the maximum distance from the Kinect camera. To deal with this issue, we clipped these pixel values to the average of their eight neighboring pixels and then normalized all pixel values according to the distance of the Kinect from the floor. In addition, to speed up the object detection and decrease the probable false detection outside the cage, we cropped each video frame based on the cage boarders. Since the setup is fixed during the experiment, we just need to detect the cage within the first few frames and use its coordinates for all of the frames during the monitoring period.

2.1.2. Background Subtraction

One of the significant attributes of mixture of Gaussian as a multimodal density function is its capability to create smooth approximations of any arbitrary shaped distribution. Considering this fact, we first characterized the distribution of n pixel values of a recently observed depth image (color-coded image extracted from original depth data), $\mathcal{D}_t$, at time instance t with a weighted mixture of Gaussian distributions as:

$$P(\mathcal{D}_t) = \sum_{j=1}^{M} \omega_{t,j}\, \mathcal{N}(\mathcal{D}_t, \mu_{t,j}, \Sigma_{t,j}) \tag{1}$$

where M is the number of Gaussian distributions, $\omega_{t,j}$ determines the portion of the data accounted for by jth distribution, $\mu_{t,j}$ and $\Sigma_{t,j}$ are the mean value and covariance matrix of the related distribution, and $\mathcal{N}$ represents a Gaussian probability density function as:

$$\mathcal{N}(\mathcal{D}_t, \mu_t, \Sigma_t) = \frac{exp(-\frac{1}{2}(\mathcal{D}_t - \mu_t)^T \Sigma_t^{-1} (\mathcal{D}_t - \mu_t))}{(2\pi)^{\frac{n}{2}} |\Sigma_t|^{\frac{1}{2}}} \tag{2}$$

In general, approximation of GMM is done by representing each pixel with a Gaussian distribution from the mixture model. The standard method for likelihood maximization in this process is Expected Maximization (EM), which maximizes the expected value of the likelihood function. However, implementing EM for each frame is a computationally complex procedure, and it is not applicable for the real-time tracking applications. Instead, we used a sample mode K-means approximation to learn the model parameters. Moreover, we assumed a diagonal matrix for the covariance matrices in Gaussian distributions, such that $\Sigma_{t,j} = \sigma_{t,j}^2 I$, where $\sigma_{t,j}$ is a scalar value for the covariance of jth distribution at time t, and I is the matching size identity matrix.

After learning the parameters of the model, probability of each pixel in the current depth video frame is compared to a background model threshold to determine whether it is part of the background or the foreground. In order to model each frame accurate enough for foreground detection while make it computationally fast, we employed empirically three Gaussian density functions to model each pixel value. To have a long-term tracking with accurate detection, parameters of the model are updated every three minutes after the initial learning, according to the ten recent coming frames. This updating step especially for the cages covered with bedding is crucial, since the background changes as the animal moves around.

By comparing the contribution of each pixel to the background against model threshold, we achieve a binary image with foreground regions. Morphological opening and closing operations are applied on the detected binary image to remove noisy detections and fill the holes in the connected components. Finally, we choose the two largest connected components as the candidates for our target object. When the animal is near the cage walls, the second detected object is its reflection on the wall, so we remove this unwanted object based on the fact that its center of the mass is outside the cage. In other cases, the second object is the noise (due to bedding pile-up in some corners of the cage) which is removed by calculating the area of each object and discarding the smaller one. The largest connected component in the foreground image in a given frame then will be labeled as the animal mask at time t, $\mathcal{M}(t)$. We consider the smallest rectangle that encloses the extracted $\mathcal{M}(t)$ as the bounding box of detected object, $\mathcal{B}(t)$.

2.2. Movement Trajectory Extraction

Using the detected animal mask, $\mathcal{M}(t)$, and the original depth values, $\mathcal{D}_t$, we compute a weighted average of the depth measurements inside the animal body mask to determine its center of the mass coordinates at time t:

$$\mathcal{C}^{X(\text{or}\,Y)}(t) = \frac{\sum_{x=b_x}^{b_x+b_w} \sum_{y=b_y}^{b_y+b_l} x(\text{or}\,y)\mathcal{D}(x,y,t)\mathcal{M}(x,y,t)}{\sum_{x=b_x}^{b_x+b_w} \sum_{y=b_y}^{b_y+b_l} \mathcal{D}(x,y,t)\mathcal{M}(x,y,t)} \tag{3}$$

where $\mathcal{C}^X$ and $\mathcal{C}^Y$ are the row and column of the mass center, respectively. b_x and b_y are the row and column of the diagonal end point of bounding box, $\mathcal{B}(t)$, and b_w and b_l are the width and length of the $\mathcal{B}(t)$, respectively.

We use the location of the mass center, $\mathcal{C}^X(t)$ and $\mathcal{C}^Y(t)$, in each frame to find the movement trajectory and speed of the animal over the time. Based on the distance traveled during different time intervals, we can determine whether or not the animal was stationary during a particular interval.

2.3. Animal Pose Detection

We create a 3D cloud using the depth measurements inside the animal body mask and apply the classical principal component analysis (PCA) approach on the cloud data to compute the direction of the highest variance of this cloud to get the pose direction of the animal in each frame. Before execution of PCA, we scale the pixel values of each frame to the metric pixel dimensions to enable PCA to estimate the actual body direction, this scaling is done using the setup and Kinect camera configuration:

$$S^{x(\text{or}\,y)} = \frac{2}{R^{x(\text{or}\,y)}} \times \tan\left(\frac{F_{view}^{x(\text{or}\,y)}}{2}\right) \times L \tag{4}$$

where S^x and S^y are scaling factors for row and column values, R^x and R^y are horizontal and vertical space resolution of the Kinect camera, F_{view}^x and F_{view}^y are horizontal and vertical camera's field of view, and L is the distance of the camera from floor. Field of view of camera in each direction gives us the largest angle of side view with respect to the vertical view. By calculating the tangent of this angle and scale it with respect to the camera distance and space resolution, we can find the scaling factor for pixel values.

2.4. Respiratory Rate Estimation

After determining the trunk part of the vole according to its position in stationary time intervals, the depth data is ready to be presented to our respiratory rate measurement algorithm, described in details in [9].

3. EXPERIMENTAL RESULTS

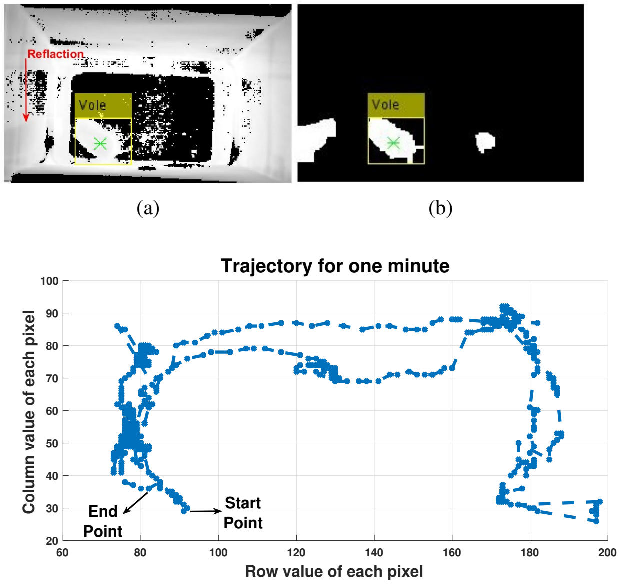
We conducted two sets of experiments: (1) preliminary depth data recording of restrained voles in order to obtain distance with the highest movement tracking resolution, and (2) primary (main) depth data recording from freely moving voles in their home cages for the tracking purpose.

3.1. Finding Distance with Highest Resolution

In the preliminary experiment, in order to find the distance with highest depth resolution, we restrained one vole and collected depth video in different distances, while the animal was breathing. We changed the distance from 50cm to 90cm with 10cm increments and recorded four minutes of depth video in each distance. We also simultaneously collected side-view RGB videos and performed visual inspection on each video to obtain respiration rate ground truth for calculating the respiratory estimation error in each distance. Using the respiration estimation algorithm developed in our previous work [9], the most accurate estimation was obtained at 80cm.

3.2. Animal Movement Trajectory and Pose Direction

In the primary experiment, the Kinect camera was setup on top of a cage in 80cm and 10 minutes of depth video recordings from two different voles in the cage were collected to validate the tracking performance of our approach. Our algorithm detected the animal in each video frame and showed the



(a) (b) (c)

Fig. 1: (a) Original depth video frame with superimposed center of mass and bounding box; (b) extracted foreground mask with detected information. The left hand component is the reflection of the vole in cage wall which is discarded by the algorithm; (c) movement trajectory of the vole inside the cage for one minute of tracking.

foreground mask along with the original depth videos with the detection information, including center of mass and bounding box, superimposed to both frame images as shown in Fig. 1(a) and (b). After a series of morphological image processing, the binary image in Fig. 1(b) shows three connected components. Considering the two largest components, the left hand one is the reflection of the vole in cage wall, which its center of the mass is outside the cage. The remaining component is the detected vole in the cage.

We have verified our detection system by visually inspecting the color-coded depth videos which were recorded from two different voles in their cage for 10 minutes. It detected the vole correctly in 80% of times and there were no detection in other times. Assuming that consecutive miss-detections are less than one second and the animal does not have any sudden movement at this time interval, we compensated the miss-detections by applying a linear extrapolation on most recent extracted locations from recorded history and used the same bounding box for the frames with no animal detection.

One of the voles' movement trajectory inside the cage during one minute monitoring is illustrated in Fig. 1(c). After analyzing each vole's trajectory, we found that the voles tend to move along the boarder of the cage and repeat that trail. Our algorithm also calculates the travelling distance and average speed of each vole's over time. For instance, in one minute of high activity period, the traveling distance of one of the voles was 434cm and the average speed of that time intervals was around 7.2cm/s.

After extracting the foreground mask of each depth frame

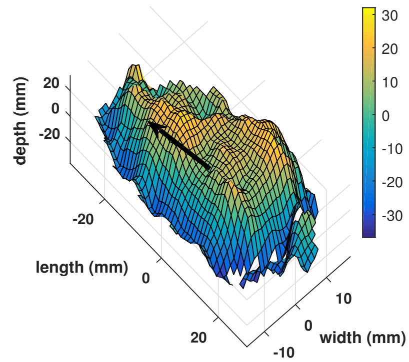


Fig. 2: 3D representation of the detected vole in the bounding box superimposed with its first principle component showing the animal's pose direction.

in the recorded video, the detection information was used to detect animal's pose in that frame. All of the pixel positions were scaled using the Equation (4) with the space resolution of 514×424 pixels and field-of-view of 70.6×60 degrees for the Microsoft Kinect V2. Pose direction of each vole was detected by applying PCA on the scaled 3D depth data and choosing the first principal component as the body main direction. Our pose estimation algorithm could estimates the direction of the animal in 3D space correctly for all of the video frames. Fig. 2 displays representation of one of the voles in a bounding box, centered in animal's mass center, and its main body direction. The angle between the cage's Cartesian coordinates and the first principal component is satisfactory for classifying the animal position in two classes of standing and on-all-fours as well as its head direction in on-all-four pose.

4. CONCLUSION AND FUTURE WORK

In this paper, we introduced an automatic non-contact tracking system designed for unobtrusive behavior monitoring of in-caged rodents. By integrating our previously-developed respiratory monitoring algorithm [9], the presented system provides an accurate but low-cost means to researchers for studies requiring long-term animal behavior monitoring. We tested our system on two sets of data recorded from two different voles in a cage, using Microsoft Kinect depth sensor. The animal's movement trajectory were extracted using an adaptive GMM algorithm for background estimation and online object tracking. In addition, we applied PCA on 3D representation of animal's depth data to detect pose direction of the vole in the cage. To the best of our knowledge the only work on automatic behavior assessment of animals using depth sensor is [17], in which the Microsoft Kinect camera was employed for behavioral tracking of caged mice, which are quite larger than voles, hence easier to detect/track. As part of a future work, we aim to perform tracking as well as behavioral analysis on multiple voles interacting in a single cage, as these laboratory models are valued for their similar social behavior to humans. One of our candidate models for tracking multiple objects will be recursive Bayesian state estimation method.

5. REFERENCES

[1] Thierry F Vandamme et al., "Use of rodents as models of human diseases," *Journal of Pharmacy And Bioallied Sciences*, vol. 6, no. 1, pp. 2, 2014.

[2] Neal R Swerdlow and Mark A Geyer, "Using an animal model of deficient sensorimotor gating to study the pathophysiology and new treatments of schizophrenia," *Schizophrenia bulletin*, vol. 24, no. 2, pp. 285–301, 1998.

[3] AA Deeny, M Fumanelli, and B Illgen-Wilcke, "Recommendations for the health monitoring of rodent and rabbit colonies in breeding and experimental units," *Management*, 2001.

[4] Tams Fzesi, Nuria Daviu, Jaclyn I. Wamsteeker Cusulin, Robert P. Bonin, and Jaideep S. Bains, "Hypothalamic crh neurons orchestrate complex behaviours after stress," *Nature Communications*, vol. 7, pp. 11937, 2016.

[5] ViewPoint Behavior Technology, "http://www.viewpoint.fr/en/home," 2016.

[6] CleverSys Behavior Recognition, "http://cleversysinc.com," 2016.

[7] Noldus Innovation, "http://www.noldus.com," 2016.

[8] emka Technologies, "http://www.emka.fr," 2016.

[9] Behnaz Rezaei, Jared Lowe, Jason R Yee, Stephen Porges, and Sarah Ostadabbas, "Non-contact automatic respiration monitoring in restrained rodents," pp. 4946–4950, 2016.

[10] Sarah Ostadabbas, Christoph Bulach, David N Ku, Larry J Anderson, and Maysam Ghovanloo, "A passive quantitative measurement of airway resistance using depth data," *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 5743–5747, 2014.

[11] Angela J Grippo, Damon G Lamb, C Sue Carter, and Stephen W Porges, "Cardiac regulation in the socially monogamous prairie vole," *Physiology & behavior*, vol. 90, no. 2, pp. 386–393, 2007.

[12] Luca Carnevali, Eugene Nalivaiko, and Andrea Sgoifo, "Respiratory patterns reflect different levels of aggressiveness and emotionality in wild-type groningen rats," *Respiratory physiology & neurobiology*, vol. 204, pp. 28–35, 2014.

[13] Muammar M Kabir, Mirza I Beig, Mathias Baumert, Mimosa Trombini, Francesca Mastorci, Andrea Sgoifo, Frederick R Walker, Trevor A Day, and Eugene Nalivaiko, "Respiratory pattern in awake rats: effects of motor activity and of alerting stimuli," *Physiology & behavior*, vol. 101, no. 1, pp. 22–31, 2010.

[14] Weiming Hu, Tieniu Tan, Liang Wang, and Steve Maybank, "A survey on visual surveillance of object motion and behaviors," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 34, no. 3, pp. 334–352, 2004.

[15] Chris Stauffer and W Eric L Grimson, "Adaptive background mixture models for real-time tracking," *Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on.*, vol. 2, 1999.

[16] Abhishek Kumar Chauhan and Prashant Krishan, "Moving object tracking using gaussian mixture model and optical flow," *International Journal of Advanced Research in Computer Science and Software Engineering*, vol. 3, no. 4, 2013.

[17] Alexander B Wiltschko, Matthew J Johnson, Giuliano Iurilli, Ralph E Peterson, Jesse M Katon, Stan L Pashkovski, Victoria E Abraira, Ryan P Adams, and Sandeep Robert Datta, "Mapping sub-second structure in mouse behavior," *Neuron*, vol. 88, no. 6, pp. 1121–1135, 2015.