2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

3D Body Parts Tracking of Mouse Based on RGB-D Video from Under an Open Field

Yoshito Tsuruda, Shingo Akita, Kotomi Yamanaka, Yuma Matsumoto,
Masataka Yamamoto, Yoshitake Sano, Teiichi Furuichi and Hiroshi Takemura

Abstract— **The mouse is a valuable animal model to address the neural mechanism of higher brain function and test the pharmacodynamics of new drugs. The development of novel behavioral analysis to detect subtleties of emotion is valuable for the evolution of neuroscience research and drug discovery. 3D pose estimation is expected to contribute significantly to them. Several methods for 3D pose estimation of the mouse using optical motion capture with markers and multiple cameras have been proposed, but these methods have problems such as preparing marker sets and the influence of the markers on mouse behavior. A low-cost and simple method for markerless 3D pose estimation of the mouse using a single RGB-D (Depth) camera is proposed. As a result, the proposed method improved the accuracy of limbs tracking compared to existing limbs tracking methods. In addition, this method could track other body parts (nose, base of tail) and the center of gravity.**

Clinical Relevance— **This study could contribute to the development of neuroscience research and drug discovery by clarifying the relationship between subtle changes in mouse behavior and emotional movements.**

I. Introduction

Behavioral analysis using the animal model is conventionally utilized in basic and applied researches, including drug discovery. In the conventional behavioral analysis, animal behaviors are recorded using the 2D camera and manually or automatically analyzed. In the manual methods by researchers, there are problems such as variation in results due to the subjective judgment criteria of the researchers and a great deal of labor and time required for analysis. Automatic analysis resolves these problems and gives consistent and objective results. For example, there is a study on detecting the scratching behavior of mice from videos taken from above [1]. This technique can detect quick scratching behaviors with a widely used camera, but it does not target other behaviors of the mouse. However, in most automatic analyses of whole-body motions of the mouse, motion tracking of animals is commonly used. So far, it is challenging to automatically define the detailed change of animal behaviors by using the movement of each body part.

The mouse is a commonly used animal model. Over its long history, many behavioral analyses have been developed, in which validity to study the physiology and disease as the human model is experimentally estimated. Additionally, utilizing the mouse has many advantages in terms of genetic manipulations, reproducibility, and so on. Open field test is one of the simple behavioral tests to analyze motility function and emotional state, in which mice allows to explore a wide-open arena. In this test, the trajectory of the mouse body and stereotyped behaviors are usually analyzed. Changes in gait patterns in response to emotional states have been studied in humans, although the neural mechanism is not understood. The mouse is useful for addressing the molecular and cellular mechanism regulating gait patterns depending on emotional state. However, an experimental technique to finely analyze the gate pattern has not been sufficiently developed.

Recently, analyses of mouse gait patterns using multiple infrared cameras and markers have been proposed [2]. However, this system has some problems with the time and cost of preparation, as well as the markers may affect mouse behavior in open-field tests. In other studies on the automation of animal experiments using a single camera for open-field tests, there is a study that tracks three points: nose, the center of gravity, and base of tail from a video of the mouse taken from above [3], and a study that tracks footprints by measuring from under the floor [4,5]. However, nose and base of tail are observable from under the floor, and the center of gravity can also be determined by images from below.

In recent years, image processing technologies based on deep learning have been attracting attention. There is a tool for tracking animal body parts using deep learning, DeepLabCut [6]. It proposed by Mathis et al. is an open-source animal pose estimation software for everyone to use (https://github.com/DeepLabCut/DeepLabCut), and this tracking tool is already used in many studies. Kinect has also made it easier to obtain 3D information, and many studies have been conducted to analyze animal behavior in 3D [7]. For example, there is a study that automatically analyzes the pecking and feeding behaviors of pigeons by estimating their posture based on RGB-D data acquired with Kinect, an RGB-D camera [8]. An RGB-D camera is a depth-sensing device that works in conjunction with an RGB camera to obtain depth information for each pixel. The proposed method is combining RGB-D video images and body part tracking by using deep learning. As for the mouse, a system that can track their limbs in 3D using Kinect has been developed [9]. However, this method cannot track the nose or base of tail, and there is room for improvement in accuracy.

In this study, we propose a simple and highly accurate method for estimating mouse posture with more measurement points, using RGB-D video images from below and body parts position estimation based on deep learning.

Yoshito Tsuruda, Shingo Akita, Kotomi Yamanaka, Yuma Matsumoto, Yoshitake Sano, Teiichi Furuichi and Hiroshi Takemura are with Tokyo University of Science, 2641, Yamazaki, Noda-shi, Chiba, Japan (corresponding author's phone: (+81)-4-7124-1501 (3931); fax: (+81)-4-7123-9814; e-mail: 7521535@ed.tus.ac.jp)
Masataka Yamamoto is with Tokyo University of Science, 2641, Yamazaki, Noda-shi, Chiba, Japan, and also with Hiroshima University, 1-3-2, Kagamiyama, Higashi-Hiroshima-shi, Hiroshima, Japan

II. METHOD

A. Recording of RGB-D Video Images

We recorded video images from under a transparent acrylic plate using an RGB-D camera. The RGB-D camera used here was Azure Kinect DK developed by Microsoft [7], and dimensions, weight, and the settings for recording are shown in Table 1. As shown in Fig. 1, a picture of the environment for capturing RGB-D video images, the background color was uniformly green to avoid disturbing the body parts tracking using color information. The camera was placed at an angle of about 27 degrees to the acrylic plate, as shown in the overview of the environment for recording RGB-D video images in Fig. 2. This angle of the RGB-D camera was set to prevent getting inaccurate depth data caused by the reflection of the acrylic plate. In this experiment, two mice were recorded for 20 minutes each. The following shows the method of body parts tracking.

Table 1. Specifications of Azure Kinect DK

Dimensions	103×39×126 mm
Weight	440 g
Color Camera Resolution (H×V)	2048×1536
Depth Camera Resolution (H×V)	640×576
Frame rate	30 fps
Color Camera Field-of-view (FOV) (H×V)	75×65 degree
Color Camera Field-of-Interest (FOI) (H×V)	90×74.3 degree

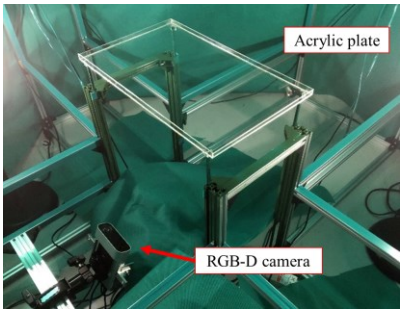


Figure 1. Picture of the environment for capturing RGB-D video images

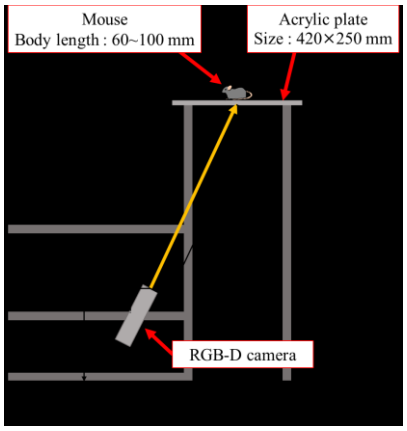


Figure 2. Overview of the environment for recording RGB-D video images

B. Tracking of the Center of Gravity of the Mouse Area

For the center of gravity of the mouse area, since there are no features in color or shape, we calculated its position from the area of the mouse extracted by image processing. It was tracked in the following procedure shown in Fig. 3.

1. From the depth image, determine threshold values for the upper and lower limits of depth and create a mask image by binarizing the area satisfying the threshold values as the area where the mouse exists (Fig. 3 (a)).
2. Take the image before the mouse is inserted as a background image. Then subtract the background image from the target image (Fig. 3 (b)).
3. Create a mask image by calculating the logical conjunction (AND) of the binarized image by the depth and the binarized image by background subtraction (Fig. 3 (c)).
4. Extract the contour with the largest area from the final mask image as the area where the mouse exists. Then calculate the center of gravity of the mouse area (Fig. 3 (d)).

C. Tracking of Limbs, Nose, and Base of Tail

For the limbs, nose, and the base of the tail, we tracked them using deep learning because their colors and shapes are distinctive. DeepLabCut, a tracking tool that uses deep learning was used to track them [6]. It is possible to perform a highly accurate estimation with less training data by using transfer learning, a method in which a model that has already been trained in one area is used for other related tasks. In addition, DeepLabCut uses a structure called a residual network to achieve accurate position estimation of body parts with a deeper network [10]. DeepLabCut has several networks with different layer depths that have been pre-trained with over a million images in ImageNet. In this study, ResNet-50, a neural network with a depth of 50 layers and the structure of a residual network, was used. The following procedure was used to track the body parts using DeepLabCut, a tracking tool that uses deep learning.

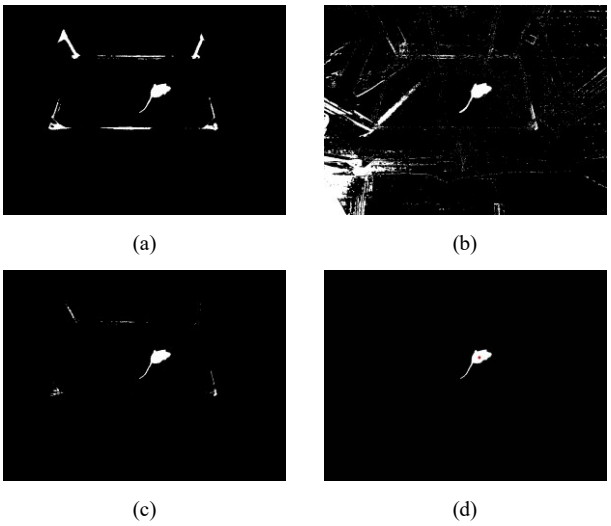


Figure 3. Tracking method of the center of gravity

Table 2. Training accuracy

	Limbs	Nose and base of tail
Training data [pixels]	2.36	2.42
Test data [pixels]	4.14	3.97

Mean Absolute Error (MAE) with training data and test data

1. Create training data (extract 1000 frames from a 20-Minute video).
2. Train the network.
3. Estimate the position of the body parts using the network.

The training was divided into two parts: the tracking of limbs and the tracking of nose and base of tail. The accuracy of the training results for limbs tracking and nose and base of tail tracking using DeepLabCut was evaluated. The error with each training data and test data was shown in Table 2. The errors with the training data were about 2.4 pixels, and the error with the test data was about 4.0 pixels. When the target body part position of the mouse was at the center of an acrylic plate with a distance of 360 mm from the camera, the error was about 0.8 mm to 1.5 mm. Considering that mouse has smaller forelimbs than hindlimbs, and the width of their forelimbs when they are on the ground is about 3 mm to 5 mm, learning was assumed to be sufficient.

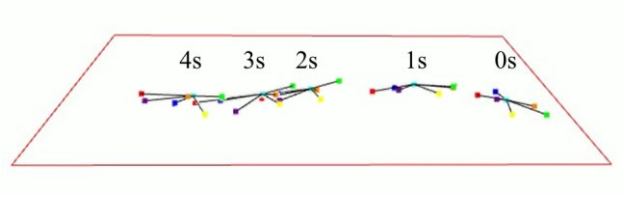
D. 3 Dimensionalization

The position of the body parts on the video image was combined with the depth information obtained from the RGB-D camera and converted to 3D by the inverse perspective transformation. The perspective transformation is represented by (1), and the inverse perspective transformation is represented by (2).

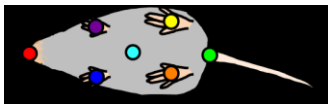
$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \quad (1)$$

$$\begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \end{bmatrix}^{-1} \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix}^{-1} \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} \quad (2)$$

where *(X, Y, Z)* is a 3D coordinate in the world coordinate space, *(u, v)* is a projection point in the image plane in pixels. The first matrix on the right-hand side of (1) is called a matrix of intrinsic parameters of the camera. The matrix of intrinsic parameters contains a coordinate of the image center (c_x, c_y) and the focal lengths expressed in pixel-related units (f_x, f_y). The elements of the matrix of intrinsic parameters are calculated from camera's FOV and color camera resolution. The joint rotation - translation matrix, the second matrix on the right-hand side of (1) is called a matrix of extrinsic parameters. r_{ij} is the element of the rotation matrix and t_i is the element of the translation matrix [11]. After that, the 3D data of 7 points were denoised using a 3-frame moving average method with one previous frame and one next frame, and all frames were updated. If either of the following two conditions are met, the data for that body part is considered inaccurate and deleted.



(a)



(b)

Figure 4. Results of the proposed method

(a) shows the results of 3 Demensionalization of body parts tracking in mouse. (b) shows the color of the points corresponding to each body part of the mouse.

A) The 3D data of the body part is outside of the range defined on the acrylic plate that would contain the body parts if they were tracked correctly.
B) The likelihood of the body part position estimation by DeepLabCut was less than threshold.

Then, the data is complemented from the 3D positions of the frames before and after the frame whose data was accurately obtained. For frames with inaccurate data, the body parts are assumed to move linearly with constant velocity.

III. RESULT

The 3D pose estimation of the mouse was obtained by the proposed methods (Fig. 4(a)). Fig. 4(b) shows the color of the corresponding point to each body part. 100 frames were randomly extracted from each of the two 20-minute videos. One was used for training (Used), and the other was not used for training (Unused). The errors between the data and ground truth are shown in Table 3. The ground truth is the data obtained by manually marking the body parts' positions on a PC by a person who is used to analyzing mice. In the video not used for training, the tracking error of the body parts other than nose was about 2.4 mm for the body part with the largest error, but the nose had a larger tracking error of 4.08 mm.

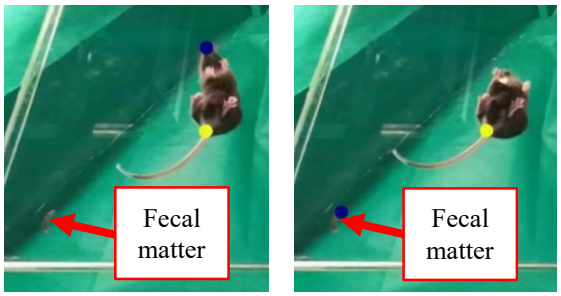
IV. DISCUSSION

As a discussion on the accuracy of tracking, compared to the limbs tracking error of 4.18 mm in the previous study [9], the limbs tracking error of the proposed method was 2.06 mm on average, which was reduced by about half. This tracking error was considered to include the human error of the manually tracked data used as the ground truth. Also, considering that the size of the mouse's front paw is about 4 mm and the size of a mouse's hind paw is about 7 mm when grounded, the accuracy was sufficient for limbs tracking. The accuracy of nose tracking in videos not used for training was worse than in other body parts but slightly improved compared to the previous study. One of the reasons for lower

Table 3. Tracking error (MAE) in distance at the center of the open field

	Limbs tracking				Nose and base of tail tracking	
Body parts	Right forelimb	Left forelimb	Right hindlimb	Left hindlimb	Nose	Base of tail
Video(Used) error [mm]	1.42	1.29	1.27	1.24	1.45	1.42
Video(Unused) error [mm]	2.43	1.54	2.24	2.05	4.08	1.40

Video (Used) is used for training, and Video (Unused) is not used for training.



(a) Success (b) Failure

Figure 5. Success and failure frames for nose tracking

The blue dots indicate where the nose was estimated and the yellow dots indicate where the base of tail was estimated. In (a), the nose was not occluded, thus nose tracking was successful. In (b), the nose was occluded by the mouse's own body, thus the fecal matter with similar color information to nose was recognized as nose.

accuracy in the nose tracking in the video not used for training (Unused) was that the color information of the nose of the mouse in the video used for training (Used) and the color information of the mouse's fecal matter during recording in the video (Unused) was similar. Therefore, when the mouse's nose was occluded by the mouse's own body, a wall, or a prop the fecal matter is identified as the nose. It is possible to improve the tracking accuracy by re-labeling and re-learning the frames where the fecal matter is recognized as a nose or by supplementing the data of other frames when the distance from the center of gravity exceeds the threshold. Fig. 6 shows a frame in the video (Unused) where fecal matter was recognized as a nose.

V. Conclusion

In this study, we proposed a low-cost and simple method for markerless 3D pose estimation of the mouse with a single RGB-D camera. The proposed method used RGB-D video images from below and body parts tracking by deep learning.

As a result, the accuracy of the proposed method was improved compared to previous studies. In addition, the proposed method can track not only 4 points of the limbs but also 3 points: nose, the center of gravity, and base of tail. However, there is still room for improvement in the accuracy of nose tracking. These results indicate that this method has the potential to contribute to the evolution of neuroscience research and drug discovery by detecting emotional movements from minute changes in mouse behavior.

In future research, we will develop a more accurate and detailed method for analyzing mouse behavior and investigate changes in mouse behavior caused by emotional movements depending on the administration of anxiolytic or antidepressant drugs.

References

[1] S. Akita, S. Tsuichihara and H. Takemura, "Detection of Rapid Mouse's Scratching Behavior Based on Shape and Motion Features," 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2019, pp. 925-928.

[2] Daniel F. Preisig, Luka Kulic, Maik Krüger, Fabian Wirth, Jordan McAfoose, Claudia Späni, Pascal Gantenbein, Rebecca Derungs, Roger M. Nitsch, Tobias Welt, "High-speed video gait analysis reveals early and characteristic locomotor phenotypes in mouse models of neurodegenerative movement disorders," Behavioural Brain Research, vol. 311, 2016, pp. 340-353.

[3] Noldus, L.P.J.J., Spink, A.J. and Tegelenbosch, R.A.J., "EthoVision: A versatile video tracking system for automation of behavioral experiments," Behavior Research Methods, Instruments, & Computers, vol. 33, 2001, pp. 398-414.

[4] Leroy, T., Silva, M., D'Hooge, R., Aerts, J.M. and Berckmans, D., "Automated gait analysis in the open-field test for laboratory mice," Behavior Research Methods, vol. 41, no. 1, 2009, pp. 148-153.

[5] Leroy, T., Stroobants, S., Aerts, J.M., D'Hooge, R. and Berckmans, D., "Automatic analysis of altered gait in arylsulphatase A-deficient mice in the open field," Behavior Research Methods, vol. 41, no. 3, 2009, pp. 787-794.

[6] Mathis, A., Mamidanna, P., Cury, K.M., Abe, T., Murthy, V.N., Mathis, M.W. and Bethge M., "DeepLabCut: markerless pose estimation of user-defined body parts with deep learning, Nature Neuroscience," vol. 21, no. 9, 2018, pp. 1281–1289.

[7] Microsoft, Azure Kinect DK hardware specifications, available from <https://docs.microsoft.com/ja-jp/azure/kinect-dk/hardware-specification>, (accessed on 14 September 2020).

[8] Lyons, D.M., MacDonall, J.S. and Cunningham, K.M., "A Kinect-based system for automatic recording of some pigeon behaviors," Behavior Research Methods, vol. 47, no. 4, 2014, pp. 1044-1054.

[9] Nakamura, A., Funaya, H., Uezono, N., Nakashima, K., Ishida, Y., Suzuki, T., Wakana, S. and Shibata, T., "Low-cost three-dimensional gait analysis system for mice with an infrared depth sensor," Neuroscience Research, vol. 100, 2015, pp. 55-62.

[10] K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition," 2016 IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770-778.

[11] OpenCV Camera Calibration and 3d Reconstruction, available from <http://opencv.jp/opencv-2svn_org/c/calib3d_camera_calibration_and_3d_reconstruction.html>, (accessed on 6 November 2020).

[12] Y. Ago, "A New Method for Assessing Behavior in Depression Model Animals and New Target Candidate Molecules for Antidepressants," Pharmacia, vol. 53, no. 7, 2017, pp. 686-690 (in Japanese).

[13] J. Deng, W. Dong, R. Socher, L. Li, Kai Li and Li Fei-Fei, "ImageNet: A large-scale hierarchical image database," 2009 IEEE Conference on Computer Vision and Pattern Recognition, 2009, pp. 248-255.

[14] T. Koide, "Series: Mysteries and Wonders of Life: Searching for Genes of Behavior and Personality - Introduction to Mouse Behavioral Genetics -," 2018, pp. 113-114 (in Japanese).

[15] M. Shinya, "Kinect-based Applications for Neuroscience Researches and Gait and Posture Researches," vol. 23, 2016, Issue 3, pp. 98-103 (in Japanese).