Study on Spontaneous Behavior Recognition of Mice Based on Frame Stream and Feature Coordinate Matching

Biao Chen¹², Yuan Zhuang¹ and Gang Xu¹*

¹School of Software Engineering, East China Normal University, 3663 Zhongshan North Road, Shanghai, China
²National Trusted Embedded Software Engineering Technology Research Center (NTESEC), 1006 Jinshajiang Road, Shanghai, China
Email: bchen@sei.ecnu.edu.cn

Abstract. The recognition of spontaneous behavior in mice is of great significance to the biological research. It not only provides an important means for pathology, pharmacy, biological neurology, but also provides great convenience to scientific researchers. In this paper, the monitoring video of mice were taken as the research object. Based on Classical frame average method to background modeling and PBAS algorithm solved the whole detection and tracking of mice, and detection and tracking of local characteristics in mice were solved by residual neutral network (ResNet). On this basis, the spontaneous behavior recognition of mice was solved by K-Means clustering algorithm. To improve the accuracy, we proposed a method of spontaneous behavior recognition in mice based on frame stream and feature coordinate matching. The effect of recognition was intuitionistic and obvious, and met the needs of subsequent experiments on matching with large-scale neuronal spike sorting in mice.

Keywords. Moving target detection; PBAS; frame stream; feature coordinate; K-Means.

1. Introduction
Many institutions had carried out research on animal behavior recognition and analysis.
Zhao of Zhejiang University [1] proposed a behavioral recognition study based on moving trajectory. The Vibe algorithm was used to separate the foreground and background of the robot. A method of rat behavior recognition was proposed by combining probabilistic methods and curvature feature sequences. Fan of Nanjing University of Science and Technology [2] combined with three edge detection methods based on gradient operator to track and identify mouse status under the top and side monitoring angles.
Huang [3] developed a trainable, versatile, automated system for performing behavioral analysis of mice in home-cage based on the calculation model of motion processing the primate visual cortex. They provided a very large database of very good manual annotation mouse behavioral video sequences. Zhang [4], Zhejiang University, put forward two state identification methods: determining the state of the mouse according to the outline of the mouse, and a state identification method which was identified by the network classifier.
At present, there are many methods of target recognition and tracking for animals, each of which has its own advantages. For example, the background difference method provides more complete feature data to extract moving targets, but it’s too sensitive to the changes of dynamic scene caused by light or external conditions. Vibe algorithm has the advantages of small memory occupation and background adaptation, but exists some towing problems. The PBAS algorithm has advantages of both VIBE and
SACON target detection algorithms, and solves the problems caused by noise and light changes. K-Means clustering algorithm is an unsupervised algorithm, which effectively recognizes the spontaneous behavior of animals, but the recognition effect sometimes is not ideal. In some experimental environments, it’s more desirable to recognize and analyze the spontaneous behavior of animals by means of artificial participation.

2. Whole Target Detection and Background Modeling in Mice
Digital image input for target detection is generally considered to have three components: target, background, and noise. The core idea is to use the nature of the target itself to highlight and determine the position by eliminating background and noise.

The host configuration used in this experiment was as follows: Intel Core i7-7800X, 32G RAM, NVIDIA GeForce 1080TI, 250G SSD and 2T mechanical hard disk. The experimental camera used USB3.0 industrial camera of JHMU s series with a resolution of 360,000 to 10 million.

To ensure the overall behavior of the mouse could be monitored, the video capture scheme used in this experiment was mainly to fix the fixed-view camera to the center of the experimental environment directly.

According to the living habits of mice, the experimental environment was set at a temperature of 18 to 22 °C and a relative humidity of 50% to 60%. Place fixed feeding points and drinking points to avoid glare and noise, as shown in figure 1.

![Figure 1. Experimental environment for mouse behavior monitoring.](image)

2.1. Background Modeling of Mouse Video Frame Sequence
The primary task of target detection in mice was to first find the image features and modeling. The frame averaging method was a relatively easy to implement background modeling method. It took the average of the input digital image sequence as a reference template. The formula of the method is as follows:

\[ B_k = \frac{1}{N} (f_{k-N+1} + f_{k-N+2} + \cdots + f_k) \]

\(N\) is the number of frames of the input digital image sequence, \(f_k\) is the kth frame of the image sequence, \(B_k\) is the background model obtained by the frame averaging method. The value of each pixel in \(B_k\) is the average of gray values of the pixel in all input image frames.

The left image in figure 2 used a background model created from a continuous video image of one hundred frames, and the image on the right was a model created with a thousand frames of continuous video images. As the images show, target interference component on the right was much less than the left image, which met the needs of experiment.
2.2. Moving Target Detection

(1) Based on background difference method

Next, we divided the input digital image into two parts, background and foreground by the background difference method, and determined the current pixel as target or background by comparing the currently input image frame with the pixel point grayscale difference of the obtained background model. Now supposed that the background model obtained in the previous section was $B_k(x,y)$, the current image frame was $f_{k+1}(x,y)$, and $D_{k+1}(x,y)$ represented the different pixel value between background models and current image frame [5]:

$$D_{k+1}(x,y) = \begin{cases} 
1 & |f_{k+1}(x,y) - B_k(x,y)| > T \\
0 & |f_{k+1}(x,y) - B_k(x,y)| \leq T
\end{cases}$$

$T$ is a difference threshold for judging whether the pixel was a foreground or a background. Figure 3 were experimental images of target detection by background difference method [6].

(2) Based on interframe difference method

Unlike background difference method, the interframe difference method used the adjacent image frames in the video sequence to differentiate, and detected the target motion region in the image frame by difference of threshold $T$. Let $f_j(x,y)$ be the jth image frame in the input video sequence, $f_k(x,y)$ be the kth image frame in the input video sequence, and let $D_{jk}(x,y)$ denote the difference in pixel value between the jth image frame and the kth image frame:

$$D_{k+1}(x,y) = \begin{cases} 
1 & |f_j(x,y) - f_k(x,y)| > T \\
0 & |f_j(x,y) - f_k(x,y)| \leq T
\end{cases}$$
Figure 4 were experimental images of target detection by the interframe difference method.

![Figure 4](image)

**Figure 4.** Experimental images based on interframe difference method.

(3) Based on PBAS
PBAS is a highly accurate target detection algorithm that combines the advantages of both VIBE and SACON target detection algorithms. The algorithm flow is shown in figure 5 [7]:

![Figure 5](image)

**Figure 5.** PBAS algorithm flow.

$T(x_i)$ represents the adaptive update rate, $R(x_i)$ represents the adaptive threshold, and $d_{\text{min}}(x_i)$ represents the complexity function of the background.

The main steps of PBAS algorithm include initialization of background model, foreground detection, background model updated, calculation of background complexity, adaptive adjustment to judge threshold and adaptive adjustment to update rate. Figure 6 was an image of a target detection experiment based on the PBAS algorithm.

![Figure 6](image)

**Figure 6.** Experimental images based on PBAS.
2.3. Conclusion

By comparing the experimental results of above three methods, it could be intuitively concluded that PBAS algorithm had better completed the overall target detection of mice and met the needs of subsequent experiments. The method could effectively reduce the noise caused by the change of ambient light intensity and the drag of the litter by mice within the range allowed by the time complexity, so that the overall detection of mice had better effect of target extraction.

3. Mice Feature Coordinate Tracking Method Based on Residual Neutral Network (ResNet)

3.1. Mice Feature Extraction

Considering the multi-layer neural network with the increase of the number of neural network layers, the instability caused by gradient explosion or gradient disappearance, the magnitude of the gradient change would be very unstable. Therefore, we used residual neutral network (ResNet) to suppress the gradient explosion and gradient disappearance from the network structure, improve the stability of gradient changes during deep neural network training, and advance the number of network layers that could be trained to a new dimension.

Table 1 is the standard structure of existing residual neural networks with different layers [8]:

| Layer name | Output size | 18-layer | 34-layer | 50-layer | 101-layer | 152-layer |
|------------|-------------|----------|----------|----------|-----------|-----------|
| conv1      | 112×112 7×7, 64, stride 2 | 3×3 max pool, stride 2 |          |          |           |           |
| conv2.x    | 56×56       | [3×3, 64] | [3×3, 64] | [1×1, 64] | [1×1, 64] | [1×1, 64] |
|            | 3×3, 64    | 3×3, 64   | 3×3, 64  | 3×3, 64  | 3×3, 64  | 3×3, 64  |
|            | 2          | 3         | 1        | 1        | 256      | 256      |
| conv3.x    | 28×28       | [3×3, 128]| [3×3, 128]| [1×1, 128]| [1×1, 128]| [1×1, 128]|
|            | 3×3, 128   | 3×3, 128  | 3×3, 128 | 3×3, 128 | 3×3, 128 | 3×3, 128 |
|            | 2          | 4         | 1        | 1        | 512      | 512      |
| conv4.x    | 14×14       | [3×3, 256]| [3×3, 256]| [1×1, 256]| [1×1, 256]| [1×1, 256]|
|            | 3×3, 256   | 3×3, 256  | 3×3, 256 | 3×3, 256 | 3×3, 256 | 3×3, 256 |
|            | 2          | x 6       | 1        | 1        | 1024     | 1024     |
| conv5.x    | 7×7         | [3×3, 512]| [3×3, 512]| [1×1, 512]| [1×1, 512]| [1×1, 512]|
|            | 3×3, 512   | 3×3, 512  | 3×3, 512 | 3×3, 512 | 3×3, 512 | 3×3, 512 |
|            | 2          | x 3       | 1        | 1        | 2048     | 2048     |
| 1×1        | Average pool, 1000-d fc, softmax | | | | |
| FLOPs      | 1.8×10^9   | 3.6×10^9  | 3.8×10^9 | 7.6×10^9 | 11.3×10^9 | |  

The 50-layer ResNet_v1 network used in this experiment consisted of five parts: conv1, conv2.x, conv3.x, conv4.x, and conv5.x, where conv1 contained a 7×7 convolutional layer and a 3×3 maximum pooling, and didn’t contain the residual module. The latter four were composed of several residual modules. Taking conv2.x as an example, it consisted of three repetitive residual modules, each of which contained three sub-convolution layers, and the structure of three sub-convolution layers were 1×1, 64, 3×3, 64 and 1×1, 256.
3.2. Results and Analysis

(1) Setting of experimental feature parameters

The main purpose of the experiment was to extract the characteristic coordinates of mice and then analyze the behavior of mice. Therefore, the triangular regions of the mouse head, namely, the mouth, the left ear and the right ear, and the tail root were selected as features, and the characteristics of mice to be extracted were shown in figure 7.

In addition, in order to facilitate the recognition of mice behavior, the characteristics of the experimental environment were also partially extracted, including the resting area of mice, the drinking water area and mice food. The characteristics of the experimental environment to be extracted were shown in figure 8.

![Figure 7](image1)

**Figure 7.** Local features of mice to be extracted.

![Figure 8](image2)

**Figure 8.** Local characteristics of the experimental environment to be extracted.

(2) Mouse feature coordinate extraction results

After the neural network training completed, the horizontal and vertical coordinate values of the corresponding feature coordinates are output for each input image. The input images of the previous ten frames are taken as an example, and the extraction results of mice feature coordinates are as follows (table 2).

(3) Extraction results of experimental environment feature coordinates

Corresponding to the characteristic coordinates of mice, the extraction results of the experimental environment feature coordinates were as follows (table 3).

(4) Fusion of feature coordinates and input images

An example of fusing mouse feature coordinates, experimental environment feature coordinates, and raw input images is shown in figure 9.

(5) Statistical analysis of feature coordinates

In order to facilitate analysis of the statistical characteristics of neural network output coordinates, the trained digital network continuously input the digital image stream of mice monitoring video, and the obtained analysis result was shown in figures 10 and 11.
Figure 9. Fusion of local features and video frames.

### Table 2. Extraction results of mice feature coordinates.

| Indexes | Head x | Head y | LeftHear x | LeftHear y | RightHear x | RightHear y | Tail x | Tail y |
|---------|--------|--------|------------|------------|-------------|-------------|--------|--------|
| 1       | 314.408| 441.9092| 326.4526   | 424.561    | 308.4909    | 421.9213    | 323.0599| 382.8353|
| 2       | 314.2576| 442.9979 | 326.3502  | 424.2648  | 307.4346    | 422.5807    | 323.7197| 382.2207|
| 3       | 307.6303| 444.2336 | 325.0863  | 426.5477  | 306.1357    | 423.1074    | 323.3645| 383.3248|
| 4       | 307.3796| 445.1437 | 324.3453  | 426.977   | 306.2511    | 423.705     | 323.705 | 384.8968|
| 5       | 306.4357| 444.9968 | 323.9889  | 428.0336  | 305.2362    | 423.5035    | 322.5468| 383.9379|
| 6       | 305.6136| 444.3706 | 322.456   | 429.0045  | 305.9844    | 422.5453    | 324.2061| 383.3994|
| 7       | 304.9649| 445.135   | 324.353   | 426.5477  | 306.1357    | 423.1074    | 323.3645| 383.3248|
| 8       | 305.2806| 444.7756 | 323.1495  | 428.4718  | 306.8088    | 423.705     | 323.705 | 383.4847|
| 9       | 304.2801| 444.6887 | 322.411   | 429.3502  | 305.9151    | 421.6543    | 323.9103| 385.2198|
| 10      | 304.2801| 444.6887 | 322.411   | 429.3502  | 305.9151    | 421.6543    | 323.9103| 385.2198|

Table 3. Extraction results of the experimental environment feature coordinates.

| Indexes | Water x | Water y | Food x | Food y | RestPlace x | RestPlace y |
|---------|---------|---------|--------|--------|--------------|--------------|
| 1       | 204.4066| 238.8771| 297.752| 443.8398| 265.6235     | 126.3386     |
| 2       | 204.7603| 237.9688| 297.734| 445.9617| 266.2753     | 126.3889     |
| 3       | 205.021 | 238.4896| 304.7495| 447.101 | 265.8088     | 126.1508     |
| 4       | 205.6537| 238.3686| 304.9097| 447.4129| 265.5573     | 126.5915     |
| 5       | 205.6531| 238.6736| 303.9791| 447.4159| 266.1692     | 125.466      |
| 6       | 206.3801| 238.4396| 306.4974| 448.0212| 265.4633     | 126.739      |
| 7       | 204.9178| 237.9379| 307.2005| 447.8778| 266.2745     | 126.1675     |
| 8       | 205.1132| 237.1936| 304.4982| 447.2339| 266.2771     | 125.5993     |
| 9       | 205.4564| 239.3253| 306.9068| 447.9181| 265.672      | 126.3964     |
| 10      | 205.0571| 238.2096| 307.7023| 448.307  | 265.9389     | 126.2785     |

Figure 10 showed the coordinate values of different features in each frame of the video stream. In this figure, the abscissa was the frame number of each frame, and the ordinate was the coordinate value. Each color in the figure was a feature recognized by the neural network, which corresponded to two polylines, which respectively referred to the x-dimensional coordinates and the y-dimensional coordinate values of the feature. It could also be seen from figure 10 that the orange-yellow and purple six-fold lines tend to be horizontal, corresponding to the characteristics of food, drinking area and rest area, indicating that the coordinate positions of these features were relatively stable. However, the
change of fold lines corresponding to the characteristics of mouse mouth, left ear, right ear and tail were more severe, and the coordinate positions of these features varied greatly.

Figure 10. Statistical chart of feature coordinates and video frame number.

Figure 11 showed the distribution of feature coordinate points in the video stream. The abscissa was the horizontal coordinate of the input video frame and the ordinate was the vertical coordinate of the input video frame. Each color in the figure was a feature recognized by the neural network, and each point corresponds to the feature recognition result of a video frame. It could be seen from the figure that the characteristic points of the orange, yellow and purple colors corresponding to the characteristics of the food, the drinking area and the rest area were concentrated, indicating that the coordinate positions of these features were relatively stable. However, the characteristic points of mouse mouth, left ear, right ear and tail were more scattered, and the coordinate positions of these features varied greatly.

Figure 11. Spatial distribution chart of feature coordinates.

From the statistical analysis of the feature distribution, it could be concluded that in two different statistical models, the feature distribution corresponding to the body part of mice and the characteristic distribution corresponding to the experimental environment were consistent with the experimental expectation.
4. Spontaneous Behavior Recognition of Mice Based on Feature Coordinates

So far, overall detection and tracking of mice, detection and tracking of local features of mice had been achieved. Next, based on the experimental data in section 1, 2, the unsupervised K-Means clustering algorithm was used to solve the problem of spontaneous behavior of mice. In order to solve the problem of K-Means, an algorithm based on frame stream and coordinate matching was proposed to meet the subsequent needs of the experiments.

4.1. Category Definition of Spontaneous Behavior in Mice

Based on the experimental needs, we divided the spontaneous behavior in mice into four kinds of basic behaviors: mice running behavior, mice feeding behavior, mice drinking behavior and mice rest behavior. Figure 12 showed four basic types.

4.2. Based on K-Means Clustering Algorithm

The K-Means clustering algorithm is an unsupervised algorithm that divides similar data objects into the same category according to the principle of similarity, while data objects with small similarity are divided into different categories. The flow of the classic K-Means algorithm is shown in Table 4 [9].

![Figure 12. Four spontaneous behavior in mice.](image)

**Table 4. Classic K-Means Algorithm Flow Chart.**

| Step 1: Randomly select K samples from the data set as the initial cluster center $C = \{ c_1, c_2, ..., c_k \}$. |
| Step 2: For each sample $x_i$ in the dataset, calculate its distance to the K cluster centers and assign it to the class corresponding of cluster center with the smallest distance. |
| Step 3: For each category $c_i$, recalculate its cluster center $c_i$ (that is all sample centroids belonging to the class). |
| Step 4: Repeat steps 2 and 3 until the location of the cluster center isn’t changing |

(1) Sample data dimensionality reduction

Reduce the dimensionality of the sample data in Table 2. The sample of the reduced video frame of the mouse was shown in Table 5. At this time, the dimension of each sample had been reduced to two dimensions.

(2) Initialization and iterative adjustment of the behavior category center

In the K-Means clustering algorithm, the initial generation position of K coordinate points has a very important influence on the experimental results. When the initial behavior category center was selected
in this experiment, random initialization was not used. Instead, the K initial generation locations were filtered out by calculation as the category center.

| Indexes | Mouse x | Mouse y |
|---------|---------|---------|
| 1       | 318.1029 | 417.8067 |
| 2       | 317.9405 | 418.016  |
| 3       | 315.3726 | 418.621  |
| 4       | 315.3063 | 419.5309 |
| 5       | 314.5519 | 420.1179 |
| 6       | 314.5124 | 420.08   |
| 7       | 314.4893 | 419.5458 |
| 8       | 314.8974 | 419.186  |
| 9       | 314.4892 | 420.2246 |
| 10      | 314.0656 | 420.4569 |

Table 5. Dimensional reduced sample data.

In the process of adjusting the center of the behavior category, it is necessary to traverse all the sample points and confirm the traversed samples according to the similarity between the sample points and the existing behavior category centers. This paper used the method of comparing Euclidean distance to determine the similarity between sample points and existing category centers.

(3) Result analysis

The recognition result was shown in figure 13. The horizontal axis in the figure was the x-axis coordinate in the top view of the experimental scene, and the vertical axis was the y-axis coordinate. The coordinate point was recognition results of each frame in the input video. In addition, the four different colors of red, green, blue and yellow in the figure represented four different categories of recognition. The different locations of the various points represented the location of the mouse in the video frame. It could be seen that K-Means clustering algorithm could better distinguish the expected feeding behavior, drinking behavior and resting behavior. However, the classification effect on the motor behavior of mice was not particularly ideal.

Figure 13. Category result scatter plot of K-Means clustering method.

4.3. Based on Frame Stream and Feature Coordinate Matching

In order to improve the accuracy of behavioral classification judgment, we proposed a human-supervised identification method to classify and determine spontaneous behavior in mice. This identification scheme combined video frame stream and feature coordinate matching. The mice behavior recognition through the video frame stream could synthesize the sample judgment results of several previous frame samples, which could be more comprehensive and specific.
(1) Sample behavior category recognition based on single frame feature coordinate matching

Different from the unsupervised clustering of K-Means algorithm, this method used manual verification and feature coordinate matching to determine the four sample models, which were feeding behavior model, drinking behavior model, resting behavior model, running behavior model. See model's details as follows, and models of four spontaneous behavior of mice were shown in table 6:

The model of mouse feeding behavior is as follows:

$$P(E) = \begin{cases} 
    \text{True,} & \text{E} \geq 0 \\
    \text{False,} & \text{E} < 0 
\end{cases}$$

where E represents the characteristic coordinate matching value of the feeding behavior of the mouse.

$$E = f - \|x^{(m)} - x^{(f)}\|_2 = f - \sqrt{\sum_{u=1}^{n} |x_{u}^{(m)} - x_{u}^{(f)}|^2}$$

where f is the characteristic matching threshold for feeding behavior of mice. $x^{(m)}$ is the characteristic point of mouse mouth, and $x^{(f)}$ is the food feature point. $x_{u}^{(m)}$ is the coordinate value of the $u$th dimension of mouse's mouth feature point. $x_{u}^{(f)}$ is the coordinate value of the $u$th dimension of food feature point.

The model of mouse drinking behavior is as follows:

$$P(D) = \begin{cases} 
    \text{True,} & \text{D} \geq 0 \\
    \text{False,} & \text{D} < 0 
\end{cases}$$

where D represents the characteristic coordinate matching value of the drinking behavior of mice.

$$D = w - \|x^{(m)} - x^{(w)}\|_2 = w - \sqrt{\sum_{u=1}^{n} |x_{u}^{(m)} - x_{u}^{(w)}|^2}$$

where w is the characteristic matching threshold for drinking behavior of mice. $x^{(m)}$ is the characteristic point of mouse mouth, and $x^{(w)}$ is the drinking feature point. $x_{u}^{(m)}$ is the coordinate value of the $u$th dimension of mouse’s mouth feature point. $x_{u}^{(w)}$ is the coordinate value of the $u$th dimension of drinking feature point.

The model of mouse resting behavior is as follows:

$$P(R) = \begin{cases} 
    \text{True,} & \text{R} \geq 0 \\
    \text{False,} & \text{R} < 0 
\end{cases}$$

where R is the characteristic coordinate matching value of the resting behavior of mouse.

$$R = s - \|x^{(b)} - x^{(s)}\|_2 = s - \sqrt{\sum_{u=1}^{n} |x_{u}^{(b)} - x_{u}^{(s)}|^2}$$

where s is the characteristic matching threshold for resting behavior of mice. $x^{(b)}$ is the characteristic point of mouse trunk, and $x^{(s)}$ is the resting feature point. $x_{u}^{(b)}$ is the coordinate value of the $u$th dimension of mouse’s trunk feature point. $x_{u}^{(s)}$ is the coordinate value of the $u$th dimension of resting feature point.

The coordinates of the mouse torso feature points are calculated as follows:
\[ x_{(b)}^{u} = \frac{1}{4} \times \left( x_{(m)}^{u} + x_{(l)}^{u} + x_{(r)}^{u} + x_{(t)}^{u} \right) \]

where \( x_{(m)}^{u} \), \( x_{(l)}^{u} \), \( x_{(r)}^{u} \) and \( x_{(t)}^{u} \) are the characteristic coordinates of mouse mouth, left ear, right ear and tail, respectively.

The model of mouse running behavior is as follows:

\[ P(E, D, R) = \begin{cases} 
\text{True}, & E < 0 \text{ AND } D < 0 \text{ AND } R < 0 \\
\text{False}, & E \geq 0 \text{ OR } D \geq 0 \text{ OR } R \geq 0 
\end{cases} \]

Among them, \( E, D \) and \( R \) are the matching behaviors of mouse feeding behavior, mouse drinking behavior and mouse resting behavior.

(2) Improve recognition accuracy through frame stream sequences

Judgment by single frame information could lead to misjudgment situations. Because each spontaneous behavior of mice was not instantaneously completed, each of its actions last for a certain period of time. So we classified mouse behavior by video frame stream, while relying on the information in the context of the sample sequence, which could improve the accuracy of recognition and reduce the error rate.

The steps based on video frame stream classification were as follows:

Firstly, a fixed-length video frame stream sequence is established, and the video frame stream is preprocessed to obtain a sample of the current frame, that is, a coordinate value of a different feature obtained by analyzing the frame.

A preliminary classification decision was then made for each sample entered and the category results were written into the frame stream sequence. If the frame stream sequence had reached the fixed length of the initialization, the sequence was shifted backward by one by sliding the window.

The principle was shown in figure 14. When the probability of a certain behavior in the sequence of frame stream samples was the highest, it could be determined that the state was the spontaneous behavior of mouse at this time.

Table 6. Examples of four spontaneous behavior of mice.

| Spontaneous behavior | Instance | Spontaneous behavior | Instance |
|----------------------|----------|----------------------|----------|
| Feeding              |          | Resting              |          |
| Drinking             |          | Running              |          |
There were two points worth noting when establishing a sequence of frame stream samples: First, the length of the sequence of frame stream samples couldn’t be too long, so that the sequence could be kept immediacy. Second, considering the transient nature of behavior changed, in order to ensure timely and reliable effect of the behavior judgment, each sample in the sequence should be given a certain weight, and conditions for decreasing the weight value from the new sample to the old sample sequence should be satisfied.

The specific mouse behavior determination probability formula is as follows:

\[ P(Q, b) = \frac{\sum_{i=1}^{N} \omega_i \ (if \ Q[i] = b)}{\sum_{i=1}^{N} \omega_i} \]

where \( Q \) is a fixed-length frame stream sample sequence, \( b \) is the specific behavior flag value in the frame stream sample sequence, \( N \) is the length of the frame stream sample sequence, and \( \omega_i \) is the weight of the \( i \)th position in the frame stream sample sequence.

**Figure 14.** How frame stream sample sequences work.

(3) Result and analysis

Figure 15 showed results of the experiment. The color of each point in the graph represented the category result of mice for each video frame. Green represented the drinking behavior of mice. Blue represented the running behavior of mice. Red represented the feeding behavior of mice. Yellow represented the resting behavior of mice. The coordinates of each point represented the position coordinates of mice in each video frame.

**Figure 15.** Scatter plot of category results.
Compared with the K-Means clustering algorithm, the recognition results based on frame stream and feature coordinate matching were more intuitive and clearer, which could not only effectively distinguish the eating, drinking and rest behavior in mice, but also had a good recognition for the running behavior in mice. The effect met the needs of subsequent experiments.

5. Conclusion
In this paper, based on the experimental research needs of spontaneous behavior recognition in mice, through method comparison and experimental results analysis, it was determined that the whole target detection of mice was based on PBAS algorithm. The coordinate tracking and local feature recognition of mice are realized by feature extraction based on ResNet. Finally, based on the experimental results of K-Means clustering algorithm, a recognition method based on frame stream and feature coordinate matching was proposed, which could solve the four spontaneous behavior in mice of experimentally defined in a more intuitive and clear way, which laid the foundation for the matching of subsequent large-scale neuronal spike sorting in mice.

Acknowledgments
This work is supported by Shanghai STC “Brain Science and Brain-like Artificial Intelligence” Major Project (16JC1400100).

References
[1] Zhao L 2016 Rat Behavior Recognition Based on Motion Track Analysis (Zhejiang University) Master’s Thesis.
[2] Fan X 2011 Research on Image Recognition Algorithm of Body Image of Experimental Mice (Nanjing University of Science and Technology) Master’s Thesis.
[3] Huang J H, Garrote E, Yu XL, Khilnani V, Poggio T, Steele A D and Serre T 2010 Automated home-cage behavioural phenotyping of mice Nature Communications 1 (68) 1-9.
[4] Zhang M, Zhang H and Zheng Y 2006 Automatic recognition of rat posture based on contour curvature and pedigree clustering Journal of Zhejiang University (Engineering Edition) 11 (3) 524-527.
[5] Nie Y, Idaku I, Kenkichi Y, Kensuke O and Hiroshi M 2009 Real-time scratching behavior quantification system for laboratory mice using high-speed vision Journal of Real-Time Image Processing 2 181-190.
[6] Ohtsu N 1979 A threshold selection method from gray-level histograms IEEE Transactions on Systems Man & Cybernetics 9 (1) 62-66.
[7] Dodart J C, Mathis C and Ungerer A 1997 Scopolamine-induced deficits in a two-trial object recognition task in mice Neuroreport 1173-8.
[8] Cho J, Lee M, Chang H J and Oh S 2014 Robust action recognition using local motion and group sparsity Pattern Recognition 47 (5) 1813-1825.
[9] Sun L, Jia K and Chan T H 2014 IEEE Conference on Computer Vision and Pattern Recognition (Columbus: IEEE) pp 2625-2632.