Detection of Rarely Occurring Behaviors Based on Human Trajectories and Their Associated Physical Parameters

Hesham M. Shehata[1](✉), Nam Do[2], Shunl Inaoka[2], and Trung Tran Quang[2]

[1] Asilla, Inc., 1 Chome-4-2 Nakamachi, Machida, Tokyo, Japan
hesham@asilla.jp

[2] Asilla Vietnam, 3F Hoang Linh Building, No. 6, Lane 82, Duy Tan, Dich Vong Hau Ward, Cau Giay District, Ha Noi City, Vietnam
{namdo,inaoka,trungtq}@asilla.net

Abstract. The complexity of detecting rarely occurring behaviors through human trajectories is closely related to a lack of data, unclear behavioral characteristics, and complex variations in their related physical parameters (e.g., velocity and orientation angles, etc.). In this context, we propose a methodology to maximize the detection performance of rarely occurring behaviors in public places by investigating the data collection process, trajectory representation based on detected skeleton poses from videos, and the use of 2D (X, Y) trajectory positional data only versus its combination with their associated physical parameters as the input for trajectory learning models. In order to evaluate the proposed method, we studied a rare Japanese behavior in public places called UroKyoro, which is a combination of the two Japanese words Urouro and Kyorokyoro. This behavior includes aimlessly moving while frequently looking in both directions. Since there is a lack of related data from real-life cases, we hired professional actors to role-play the behavior alone or with normal pedestrians moving around. The learning system was trained using limited and augmented data. The trajectory learning system, trained with combined human trajectories and orientation angles following the proposed method, succeeds in detecting the studied behavior with an accuracy of 91.33%, outperforming the accuracy of the trained model using only human 2D (X, Y) trajectories by 4.33%. The results show the effectiveness of the proposed method to detect complex, rarely occurring human behaviors by training the LSTM classifier with a combination of human trajectories and physical parameters. However, the effectiveness of physical parameters on training performance may differ from one case study to another based on behavioral characteristics.

Keywords: Rarely Occurring Behaviors · Trajectory-Related Human Behaviors · Trajectory Learning · Trajectory Associated Physical Parameters · UroKyoro Behavior

A. Al. Ali et al. (Eds.): ICSR 2023, LNAI 14453, pp. 276–293, 2024.
https://doi.org/10.1007/978-981-99-8715-3_24

1 Introduction

In recent years, there has been a need for the use of security robots and autonomous security systems fitted with assistive devices (e.g., cameras and sensors), which are becoming increasingly popular to ensure high security in public places such as shopping malls. However, it is critical to thoroughly assess these systems' performance in recognizing anomalous behavior and maintaining public safety. A study conducted by [1,2] investigated the deployment of robots in public places and discovered that the level of autonomy granted to the robots can have a substantial impact on their capacity to recognize and respond to possible deviant actions. The authors suggested that the level of autonomy be carefully studied and adjusted in order for the robots to make real-time judgments while prioritizing public safety. Furthermore, it is crucial to consider the ethical and legal implications of using autonomous security systems in public spaces. A study by [3] investigated the ethical considerations of using security robots and concluded that it is essential to ensure that the robots are programmed to respect individual privacy and civil liberties. In conclusion, the use of robotic bodyguards in public spaces is a promising development for enhancing security measures. However, it is important to thoroughly evaluate and consider the level of autonomy, movement and action strategies, and ethical and legal implications to ensure that these systems provide a high level of security while also respecting public safety and individual rights.

On the other hand, the contribution of machine learning in the field of social behavior detection is promising and has increased rapidly. The basic information, including human trajectories and their associated physical parameters, should be precisely studied to detect the behaviors successfully. Recent studies rely on videos or sensors in order to extract or formulate trajectory information [4–6]. Usually, trajectory-related normal behaviors in public places can be detected using basic movement information (e.g., a person who is loitering while looking for a shop or waiting for a friend, etc.). However, for abnormal behaviors (e.g., shoplifting cases, etc.), it is difficult to capture and collect their data with high accuracy since they are rare to occur and require special permissions to acquire related information in public places. In addition, the detection process for such behaviors is challenging due to unclear characteristics, complex variations in their related physical parameters, and a lack of high-quality trajectories. In this context, we propose a method to successfully detect rarely occurring human behaviors using an appropriate representation of human trajectories and their associated physical parameters processed from video data. These parameters include human 2D position (X, Y), movement velocities, orientation angles, and other possible physical parameters (e.g., accelerations, deviations, etc.).

Therefore, we investigate the data collection process, trajectory representation based on detected skeleton poses from videos, and the use of trajectory positional data only versus combined positional data with physical parameters for trajectory-based learning approaches. The purpose of this research is to maximize machine learning detection performance of rarely occurring behaviors and to investigate the following research question: "How to achieve high-quality trajectories of rarely occurring behaviors from video data and maximize their detection performance through machine learning?". To answer this question, this study proposes a method to detect rarely occurring behaviors in real life. To implement the proposed method, we study the case of a rarely occurring Japanese behavior in public places called UroKyoro. A combination of Kyorokyoro (to move around restlessly) and Urouro (aimless wandering). Typically, shoplifter behavior is characterized by Kyorokyoro [7]. Security cameras should be equipped with AI software that automatically detects such suspicious activities. The results confirmed that training with combined human trajectories and their associated physical parameters can optimize the detection performance of rarely occurring behaviors.

The contributions of this work are:

- We propose effective analysis of human skeleton poses to precisely formulate human trajectories and confirm the accurate distribution of the physical parameters based on the overall body movements.
- We propose a simple yet effective method for detecting rarely occurring behaviors in public places and achieved a promising detection performance for a case study (UroKyoro behavior) by following a trajectory learning approach and considering the combined input of 2D (X, Y) positional data and their associated physical parameters.
- We apply an effective data augmentation, utilizing the data size through overlapping between samples, to enhance the overall detection performance.

The remaining sections are organized as follows: Sect. 2 discusses the background. Section 3 proposes the methodology for a successful trajectory learning approach for rarely occurring behaviors based on human trajectories and their associated physical parameters. Section 4 presents a case study of rarely occurring "UroKyoro" Japanese behavior. Section 5 shows the results following the proposed methodology. Section 6 discusses the findings of the study based on the results. Finally, Sect. 7 shows our conclusions.

2 Background

2.1 Human Trajectory and Machine Learning

Recent advancements in the field of detection and recognition of trajectory-related human behaviors have showcased the effectiveness of Long Short Term Memories (LSTMs) [8–14], Variational Recurrent Neural Networks (VRNNs) [15], and Gated Recurrent Units (GRUs) [16]. These neural network architectures have proven highly capable of sequence-to-sequence prediction tasks. For instance, in [4] Lee et al. employed RNNs to forecast future motion positions based on scene context and agent interactions. Su et al. [5] introduced a methodology utilizing LSTMs in conjunction with recurrent Gaussian processes to characterize crowd transitions and uncertainties in human trajectory prediction. Nevertheless, it's worth noting that this approach doesn't distinguish between pedestrians and exclusively considers the presence of surrounding pedestrians. These methodologies deviate from the traditional social force model, where social forces are calculated based on standard physical parameters [17–19]. Furthermore, they predominantly address normal behaviors, failing to account for abnormal behaviors or the potential benefits of integrating physical parameters with trajectory data. Only a limited number of research papers, such as App-LSTM [6], have explored models that generate an agent's trajectory towards a group of agents while considering orientation angles. Given this landscape, our research aims to fill the gap by developing models that not only excel at detecting rarely occurring behaviors based on movement trajectories but also leverage related physical parameters to enhance their predictive capabilities.

2.2 Social Behaviors and Machine Learning via Visual Approaches

In recent years, computer vision techniques, including those discussed in [20] by Wu et al., and supervised machine learning models, which leverage stand-alone Convolutional Neural Networks (CNNs) as presented in the work of Zamboni et al. [21] or combine them with Long Short-Term Memory networks (LSTMs) as demonstrated by Quan et al. [22] and Zhong et al. [23], have been extensively employed for learning social behaviors from human trajectories. These methods focus on detecting human trajectories effectively, building upon computer vision methodologies described in the works of Alahi et al. [8], Yi et al. [24], and Su et al. [5], who incorporated CNN architectures following the principles laid out in the computer vision literature [25,26]. For instance, Yi et al. [24] utilized a CNN-based architecture to model pedestrian behavior, predicting their walking patterns and goals. Furthermore, various approaches have addressed the detection of abnormal behaviors in public spaces, such as fighting or kicking, by employing image processing from videos in conjunction with CNNs, autoencoders, and LSTM networks for behavior detection, as described in the works of Tay et al. [27], Ribeiro et al. [28], Ko et al. [29], Xu et al. [30], and Pennisi et al. [31]. However, these approaches do not delve into the strategies used to formulate trajectories, which is a challenging aspect due to the necessity of minimizing detection errors and enhancing trajectory representations based on overall body movement.

In addition to the above-mentioned methods, alternative approaches have been explored for behavior detection. Nater et al. [32] utilized the tracker tree method to specify actions at higher levels, while Lv et al. [33] employed the Pyramid Match Kernel algorithm for feature matching. Du et al. [34] proposed a recognition technique for low-moral behaviors, such as smoking or using mobile phones in public spaces, based on depth skeleton data obtained from Kinect sensors, achieving a maximum accuracy of 90%. Their approach focused on recognizing low-moral behaviors in public spaces by using depth data and extracted skeletons from Microsoft Kinect v2, conducting experiments with a group of 20 individuals aged between 22 and 54. Ko et al. [29] developed a CNN framework incorporating a Kalman filter to classify various behaviors. They fed images into the framework and transferred the output to another LSTM structure, primarily aiming to enable instant detection of risky behavior in video surveillance systems, specifically for socially disadvantaged groups like the elderly, using standard RGB images. Their models were trained on the "UT-Interaction-Data" dataset, containing video clips with multiple moving human subjects engaged in six different activities: "hand shaking,"' "hugging," "kicking," "pointing," "punching," and "pushing." The maximum reported recall and precision values were 0.95 and 0.97 for kicking behavior, respectively. Although these approaches studied the most common abnormal behaviors in public places, they did not investigate the quality of the skeleton data or the influence of their distributions on the physical parameters.

3 Methodology

Detection of rarely occurring behaviors in crowded public places is challenging due to complex variations in their trajectory-related physical parameters, a lack of high-quality trajectories, and unclear behavioral characteristics, e.g., loitering without clear intentions, shoplifting, etc. To achieve successful detection of these behaviors, the possible shortest detection duration should be considered for each sample. Hence, the tracking process should consider the results based on the continuously detected samples. A detection threshold should be tuned to confirm the abnormal behavior based on the detected number of samples within a specific duration. This helps to precisely identify and differentiate the abnormal occurrence from deviated normal behaviors on certain occasions or suspicious individuals who interfere with or align with normal people to use the same pattern of motion.

Human trajectory and their related physical parameters (e.g., positional noise, velocity, etc.) play an important role in achieving successful detection performance for rarely occurring human behaviors. Recent research studies have failed to propose efficient processing approaches to strengthen trajectory representation and minimize detection errors in video data. Hence, there is a need to investigate the effect of these parameters on the learning process. To solve these problems, we propose below steps and subsections to follow towards successful detection of rarely occurring behaviors based on the appropriate representation of human trajectories from video data and consideration of their associated physical parameters.

1. Collection of rarely occurring trajectory-related human behavior data from real life However, if it's difficult to collect related behavioral data, a role-playing experiment should be followed based on behavioral observations from real-life cases.
2. Strengthen trajectory representation and minimize detection errors by following proper processing of the collected data.
3. Deciding on an appropriate time series learning structure using 2D trajectory positional data along with the minimum possible sample size.
4. Maximize the detection performance of the rarely occurring behavior to be distinguished from normal behavior by considering combined input features

for training (2D trajectory positional data along with potential physical parameters).

5. Data augmentation by overlapping between samples to achieve the best possible performance of the trained model.

3.1 Collecting Data

Collection of rare human behavioral data should take place in real-world public areas using genuine cases. This is determined by a number of elements, including the likelihood of behavior occurring in public places, the ability to acquire reasonably large data sets, and the licenses to collect data in specified locations. If these conditions cannot be met, a potential alternate solution can include a role-playing experiment. The collecting method necessitates a proper choice of data collection location based on the desired behavior, an effective collection system, and precise observations for behavioral features in real life.

In a role-playing experiment, participants should be guided to act out the behavior based on the observed behavioral characteristics from real life while interacting freely with other pedestrians and obstacles in the environment (e.g., changing movement directions, slower or faster walking) based on the scenario or situation. In addition, to ensure that the data are as diverse as possible, the participants should be asked to change the starting point for each trial.

Human trajectories in two dimensions (X and Y) can be acquired using vision or non-vision tracking technologies. Non-vision-based tracking approaches address social privacy problems. However, in public settings, the related collection system using sensors, such as Li-DARs, might be complex and expensive to install. To overcome this problem, vision-based systems using cameras are cost-effective, and recent related identification algorithms can disguise the faces of each individual in the scene. The discovered data mostly consists of skeleton poses of people in the scene, which should be processed effectively in order to improve human trajectory representation. Figure 1 shows an example of detected skeleton poses from a moving pedestrian using Asilla product.

3.2 Strengthen Trajectory Representation

Reasonable human trajectory formulation based on collected videos via cameras is challenging due to detection errors, hidden poses on certain occasions (e.g., lower body poses because of sitting on chairs), etc. The detection performance via videos depends on the deployed algorithm, which detects and extracts human body poses. Hence, there is a need to strengthen the trajectory representation by considering the best possible distribution of associated physical parameters (e.g., positional noise, velocity, etc.) to reflect the actual overall body movement through clearly visible poses and minimize detection errors. In this context, there is a need to analyze formulated trajectories and their associated physical parameters based on different types of skeleton poses to confirm the best key point to formulate human trajectories.

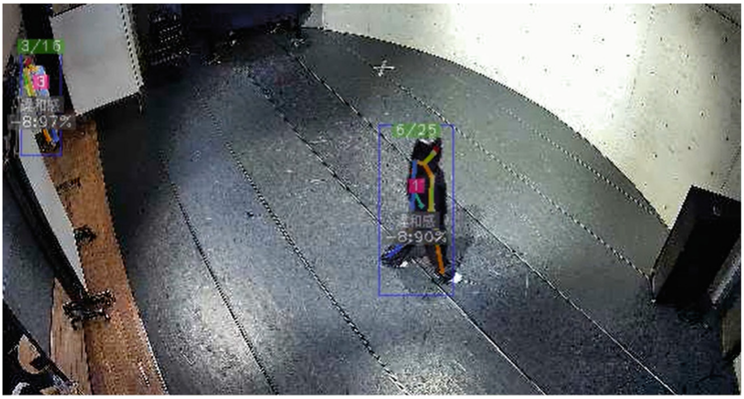

Fig. 1. Example of detected skeleton poses using Asilla product

3.3 Learning Framework

The learning framework should be a fast, light model that can successfully learn sequence-to-sequence data. In this context, recurrent neural network architectures have proven highly capable in related prediction tasks. For instance, Long Short Term Memories (LSTMs) [8–14], Variational Recurrent Neural Networks (VRNNs) [15], and Gated Recurrent Units (GRUs) [16] have showcased the effectiveness in the field of detection and recognition of trajectories data. Then, the learning framework should be selected from these recent approaches. In addition, the number of layers and neurons should be minimized as much as possible while still showing promising detection performance.

3.4 Maximizing Detection Performance

To maximize the detection performance of rarely occurring behaviors, there is a need to learn behavioral-related features. Then, training data should include potential trajectory-related features closely relate to their physical parameters (e.g., velocity, orientation angle, etc.). These parameters should be fed to the learning framework along with basic trajectory information (2D of X and Y poses) to successfully detect complex hidden features (e.g., a person changing walking direction inappropriately while walking in crowded areas or continuously changing the moving pattern with unknown intentions). In addition, data augmentation by overlapping between samples can be a final step to achieving the best possible detection performance.

4 A Case Study of UroKyoro Behavior in Public Places

There is always a need to enhance public security by detecting abnormal behaviors at an early stage. To do so, we selected a rarely occurring Japanese behavior in public places called UroKyoro. The word itself is a combination of the other two Japanese words, Urouro (aimless wandering) and Kyorokyoro (to move around restlessly). The behavior involves a person moving around without a specific target or objective while behaving suspiciously by always changing the angle of view to different directions (i.e., looking in the right and left directions) and exploring their surroundings (i.e., loitering). It's considered an early stage of other suspicious behaviors (e.g., shoplifting) [7]. The studied behavior mostly occurs in public places (e.g., shopping malls), including normal pedestrians moving around the person who is behaving with the targeted behavior.

4.1 Data Collection

Due to the lack of UroKyoro behavior data from real-life cases, we conducted a field experiment in a closed environment as an alternative to a public place (e.g., a shopping mall) (as shown in Figs. 2 and 3) by hiring professional actors to role-play the behavior alone and with normal pedestrians moving around. Since the Asilla product is adapted for implementation on security cameras in public places, the collected data included recorded videos via similar setting for each scenario, with several trials lasting a total of twenty-four minutes, while every trial lasted for around two minutes. Different camera angles (e.g., 25, 30, and 40°) are considered so that we can ensure diverse trials as much as possible. Two professional actors were asked to play the role of UroKyoro behavior freely from different locations for every trial based on the behavior definition and their imagination about it, e.g., moving around while continuously changing the angle of view, stopping for a while rotating around the body, and loitering. Since our target is to distinguish UroKyoro behavior from normal pedestrians moving around, we also collected normal data for real pedestrians in a shopping mall (Tokyo, Japan) with the same size as the other class (normal class data is collected in two corridors with people moving in two opposite directions and includes straight-line movement, loitering, changing movement direction, slowing down, entering or exiting shops, etc.). The environment included familiar obstacles in a shopping mall (e.g., separating stands, public chairs, etc.). Unfortunately, we could not share pictures from the public environment due to policies and privacy concerns.

4.2 Trajectory Representation

We used the Asilla product to detect and extract the skeleton poses for humans in the collected videos. It include the following 10 poses: nose, neck, average ankles (right and left), average shoulders, average elbows, average wrists, average hips, average knees, average eyes, and average ears. To achieve a reasonable formulation of human trajectory using the detected poses from the videos, we analyzed trajectories in the collected data using each pose value (X, Y) separately, then compared the distribution of absolute positional noises (in terms of pixels) and absolute velocities (in terms of pixels/second) versus the overall average values for the whole body movement. The concept that we followed to

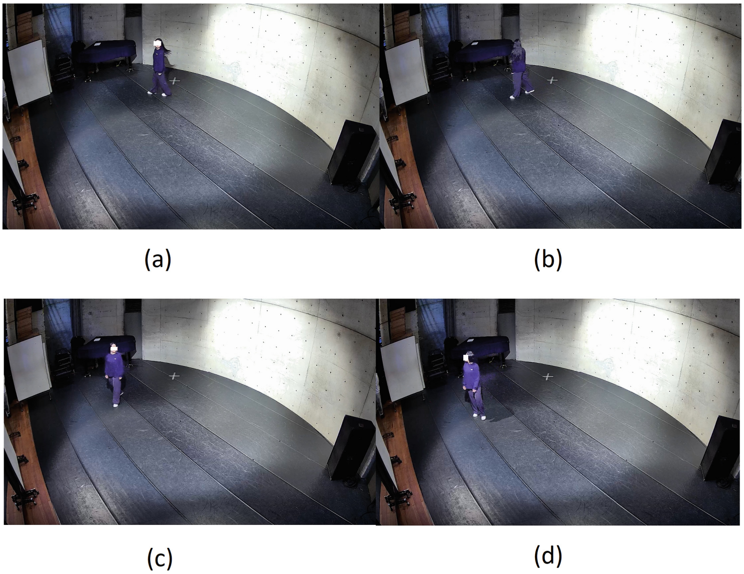

Fig. 2. Field experiment from different scenes (a, b, c, and d) in a closed environment where an actor is role-playing the UroKyoro behavior alone

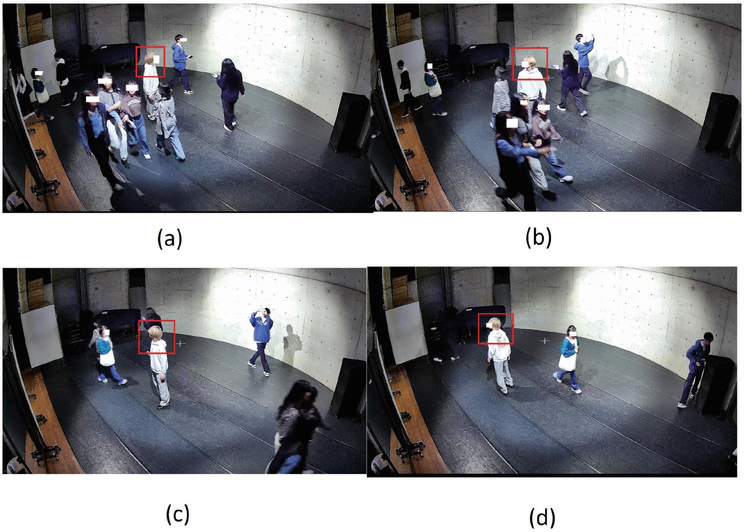

Fig. 3. Field experiment from different scenes (a, b, c, and d) in a closed environment where an actor (labelled in red) is role-playing the UroKyoro behavior with normal pedestrians moving around (Color figure online)

estimate trajectory noise (as shown in Fig. 4) is to calculate the relative distance between the middle points in the actual trajectory and the fake trajectory (straight line between the first and last point) every specific duration (e.g., 5 s).

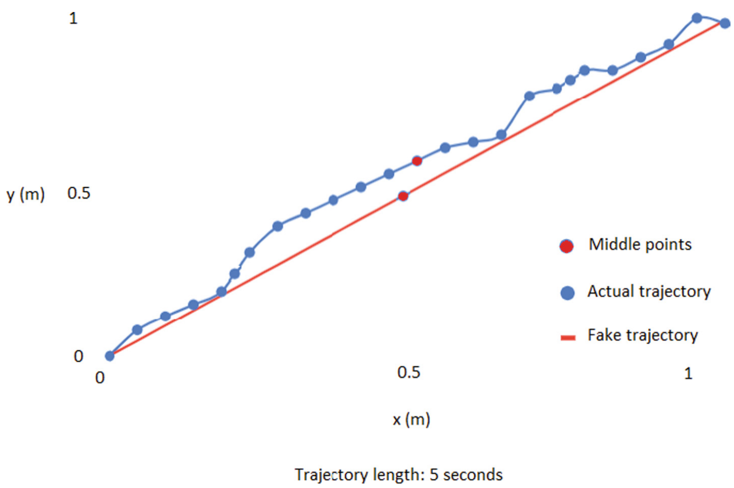

Fig. 4. Noise estimation process

Tables 1 and 2 show the average values for positional noise and velocities. The average noise ranges from neck poses show the closest value (29.86 pixels) to the overall average noise from all poses (31.26 pixels), while the absolute average velocities from hip poses (27.06 pixels/s) seem to show the closest value to the overall average velocity from all poses (27 pixels/s). However, the quality of the detected neck poses is better than that of hip poses due to the fact that humans upper body sections are almost visible and clear in most cases. In addition, the average absolute velocity ranges vary within a too small range [26.42 to 27.78 pixels per second], which makes it difficult to consider specific poses to represent the best velocity distribution. Based on that, we decided to use neck poses to formulate the human trajectory from our data. Figures 5 and 6 show a couple of 2D trajectory plots using neck poses for UroKyoro and normal behaviors (5 s per sample).

5 Results

We include in this section the training results following the LSTM training structure (explained in the next subsection). Based on our behavioral observations on UroKyoro and several training trials using different sample sizes, we found that 5 s is the minimum possible duration to include most of the related characteristics, and it showed acceptable results. Based on that, a series of five-second windows (26 frames per window) of the UroKyoro trajectory and the normal

Table 1. Average Positional Noise for All Poses, Ordered from Maximum to Minimum

Pose Name	Average Positional Noise (in pixels)
Eyes	37.07
Nose	36.68
Ears	35.99
Neck	29.86
Shoulders	29.52
Wrists	29.44
Elbows	29.29
Hips	28.57
Ankles	28.34
Knees	27.82
Average for All	31.26

Table 2. Average Velocity for All Poses Ordered from Maximum to Minimum

Pose Name	Average Velocity (in pixels/second)
Eyes	27.78
Nose	27.64
Ears	27.55
Hips	27.06
Knees	26.93
Wrists	26.77
Elbows	26.77
Ankles	26.67
Neck	26.44
Shoulders	26.42
Average for All	27

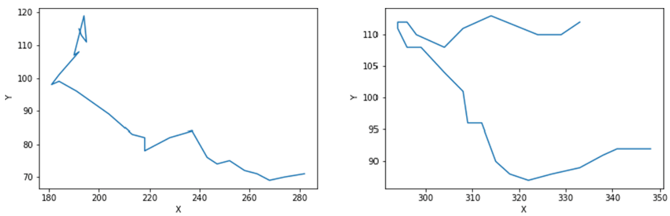

Fig. 5. a couple of 2D (X, Y in pixels) trajectory plots using neck poses for UroKyoro behavior (5 s per sample)

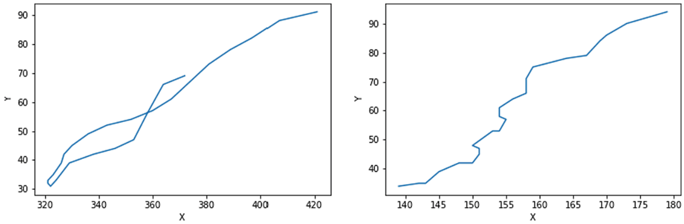

Fig. 6. a couple of 2D (X, Y in pixels) trajectory plots using neck poses for normal behavior (5 s per sample)

trajectory are fed to the network. For the UroKyoro class, we split the processed data from the videos with the single UroKyoro person into 90% for training and 10% for validation, while for evaluation (testing), we used the processed data from the videos, which include normal pedestrians moving around our targeted behavior. On the other hand, all normal class data is collected from a shopping mall and split for training, validation, and evaluation with the same size as the other UroKyoro class. In addition, to avoid over-fitting, a dropout of 0.1 is used after each LSTM layer. We also monitored the validation loss every 50 epochs to stop the training when there was no enhancement in performance.

5.1 Training Details

LSTMs demonstrated promising performance in various applications when trained with time series data [5,6,8]. Based on that, we followed a LSTM-supervised training structure as shown in Fig. 7. The training is performed using Python and PyTorch, where a series of 2D windowed trajectories (X, Y) are fed to the network with scaled samples of five seconds (26 frames per window) as a single window, where each window consists of the UroKyoro or normal pedestrian trajectory. The size of the data used for training is balanced (50% normal, 50% UroKyoro). The applied learning rate is 0.0001, while the optimizer is RMS. Two fully connected LSTM layers are used (512 units), while the classifier layer activation function is "RelU" (256 units) to finally distinguish between UroKyoro and normal behaviors (2 classes).

5.2 Baseline

The baseline is trained using only 2D (X, Y) positional data (406 windows) by following the presented LSTM network structure to be compared with the combined (i.e., considering additional training inputs, "physical parameters") and augmented (i.e., increasing data size) models, which are explained in the following subsections. The resultant testing accuracy is 87% (testing data in 300 windows).

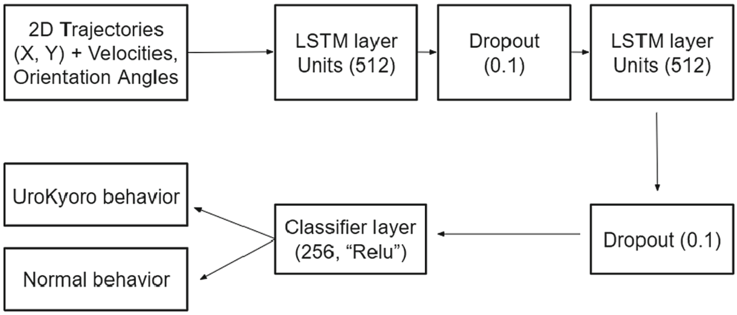

Fig. 7. The LSTM network structure distinguishes UroKyoro from normal behavior

5.3 Combined Features

To successfully detect complex behavioral characteristics, we considered training using combined input features by feeding absolute velocities (in pixels/second) and orientation angles (in degrees) along with 2D (X, Y) trajectory positional data to the training structure. Velocities and orientation angles are calculated based on the absolute position between every frame and the following one, so that finally every frame has its own 2D (X, Y) absolute position along with the absolute velocity and orientation angle. The results showed that orientation angle is a powerful factor to enhance the detection of the UroKyoro class. Also, combining all features together showed better evaluation performance, up to 90.33%. Table 3 shows the overall results achieved.

Table 3. Training Results Using Combined Features

Input	Training	Validation	Testing
X, Y	100%	97.50%	87%
X, Y, Velocity	100%	95%	88.66%
X, Y, Theta	98.09%	95%	89.66%
X, Y, Velocity, Theta	98.91%	92.50%	90.33%

5.4 Data Augmentation

We augmented the data by following five rounds of overlapping between the five-second samples with a step of 1 s for every round to check the effectiveness of increasing the data size on the detection performance. The total size of the augmented data is 2380 windows. In this case, the improvement in the evaluation (testing) performance is shown when considering training inputs as (X, Y, Theta) up to 91.33%. Table 4 shows the corresponding results.

Table 4. Training Results Using Augmented Data

Data Size	Training	Validation	Testing
406	98.09%	95%	89.66%
2380	99.35%	96.22%	91.33%

The lowest number of failure cases to detect UroKyoro behavior (7 samples) is shown from the model, which is trained using 2D positional data along with orientation angles (X, Y, and Theta). The confusion matrix for evaluated (tested) samples is shown in Fig. 8. It is obvious from the matrix results that the model detects UroKyoro behavior with good performance. The majority of failure cases are caused by normal pedestrians' confusing trajectories, which are similar to UroKyoro trajectory shapes on certain occasions, while a minority of failure cases are caused by UroKyoro trajectories, which are similar to those of normal pedestrians. Figure 9 shows a couple of trials of the failure cases.

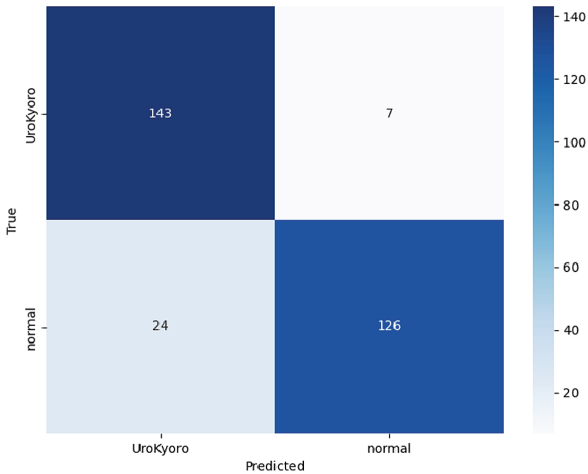

Fig. 8. Confusion matrix for evaluated data

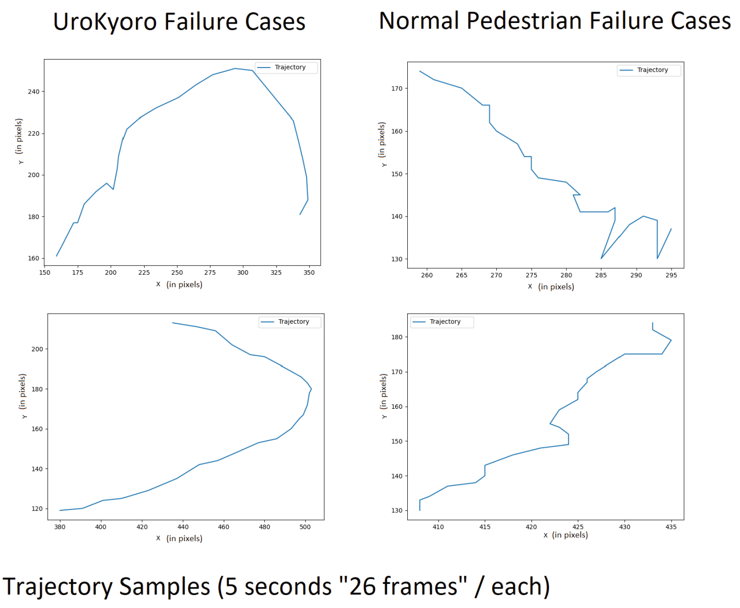

Fig. 9. Samples from failure cases

6 Discussions

Based on research findings and our UroKyoro case study, we confirmed that the proposed approach resulted in promising detection performance for rarely occurring behaviors by developing a trajectory learning model using combined training features:

- The proposed analysis of human skeleton poses resulted in precise trajectory formulation and confirmed the accurate distribution of the physical parameters based on the overall body movements. This resulted in a realistic representation of human trajectories from the collected video data.
- Velocity and orientation angle have a promising effect on enhancing the detection of rarely occurring behaviors. However, combining all parameters together for training could not be an effective approach to enhancing detection performance. This is due to the fact that related characteristics for every behavior differ from other ones. For instance, orientation angle seems to be the most powerful parameter to enhance the detection of UroKyoro behavior.

This is the first research study to focus on analyzing human skeleton poses to formulate precise trajectories with accurate distribution of the physical parameters and investigating the effect of trajectory-related physical parameters (e.g., velocities and orientation angles, etc.) on the detection performance of rarely occurring behaviors by following different combinations of training inputs along with 2D (X, Y) positional data. In this context, the research methodology and findings should generalize to other complex, rarely occurring behaviors if their characteristics are closely related to human movement. In addition, since the data are collected from different angles of view, the approach should generalize in the case of implementing security robots fitted with cameras in public places.

The limitations of the proposed method include:

- The detection accuracy of trajectory-learned models can be affected by the quality of the detected poses from videos.
- Behavioral trajectories may lack diversity in the data due to the collection of data in specific environments. Data collection from several environments may be needed based on the focused behavior.

Although the accuracy reported in [34] or [29] to detect abnormal actions is rather high (around and over 90%), most of the investigated behavioral characteristics do not relate to trajectory shape and are closely related to the whole skeleton (e.g., kicking, smoking, talking on the phone, etc.). Up to our knowledge, this is the first focused research on trajectory-related, rarely occurring behaviors. The accuracy reported from our case study is promising and reaches 91.33%. It shows the effect of trajectory-related physical parameters (velocity and orientation angle) on the detection performance of the studied case of UroKyoro behavior. Future work includes studying the effect of other physical parameters (e.g., acceleration, deceleration, etc.) on the detection performance and validating the research findings based on real-life case studies. Also, the Asilla team will consider collecting datasets using different scenarios of rarely occurring behavior, collecting data under similar conditions with different environmental shapes and sizes, training the network using these different datasets, and comparing the results.

7 Conclusions

We propose a new approach towards successful detection of rarely occurring behaviors in public places by strengthening human trajectory formulation from videos and developing trajectory learning models by following combined training inputs of 2D positional data and their associated physical parameters (e.g., velocity and orientation angle). To evaluate the proposed method, we studied a rare Japanese behavior in public places called UroKyoro. The behavior involves moving around while repeatedly looking in the right and left directions without clear intention or purpose. Since there is a lack of related data from real-life cases, we hired actors to role-play the behavior. The best detection performance of the studied behavior showed an accuracy of 91.33%. Also, we compared the detection performance by training using different inputs of physical parameters along with 2D (X, Y) positional data and confirmed that the performance may differ based on the focused behavioral characteristics.

Data Availability Statement. The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Conflict of Interest. The authors declare that they have no conflicts of interest.

References

1. Beer, J.M., Fisk, A.D., Rogers, W.A.: Toward a framework for levels of robot autonomy in human-robot interaction. J. Hum.-Robot Interact. **3** (2014)
2. Nahavandi, S.: Trusted autonomy between humans and robots. IEEE Syst. Man Cybern. Mag. (2017)
3. Lin, P., Abney, K., Bekey, G.: Robot ethics: mapping the issues for a mechanized world. Artif. Intelli. J. **175** (2011)
4. Lee, N., Choi, W., Vernaza, P., Choy, C.B., Torr, P.H.S., Chandraker, M.: Desire: distant future prediction in dynamic scenes with interacting agents. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 336–345 (2017)
5. Su, H., Zhu, J., Dong, Y., Zhang, B.: Forecast the plausible paths in crowd scenes. In: IJCAI, vol. 1 (2017)
6. Yang, F., Peters, C.: App-LSTM: data-driven generation of socially acceptable trajectories for approaching small groups of agents. In: Proceedings of the 7th International Conference on Human-Agent Interaction, pp. 144–152 (2019)
7. AI Guardman' system aims to cut shoplifting losses, JAPANTODAY. https://japantoday.com/category/features/kuchikomi/vigilant-%27ai-guardman%27-system-aims-to-cut-shoplifting-losses. Accessed 20 June 2018
8. Alahi, A., Goel, K., Ramanathan, V., Robicquet, A., Fei-Fei, L., Savarese, S.: Social LSTM: human trajectory prediction in crowded spaces. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 961–971 (2016)
9. Sun, L., Yan, Z., Mellado, S.M., Hanheide, M., Duckett, T.: 3DOF pedestrian trajectory prediction learned from long-term autonomous mobile robot deployment data. In: IEEE International Conference on Robotics and Automation (ICRA), pp. 5942–5948 (2018)
10. Shi, K., Zhu, Y., Pan, H.: A novel model based on deep learning for Pedestrian detection and Trajectory prediction. In: IEEE 8th Joint International Information Technology and Artificial Intelligence Conference (ITAIC), pp. 592–598 (2019)
11. Xu, Y., Piao, Z., Gao, S.: Encoding crowd interaction with deep neural network for pedestrian trajectory prediction. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 5275–5284 (2018)
12. Crivellari, A., Beinat, E.: LSTM-based deep learning model for predicting individual mobility traces of short-term foreign tourists. Sustain. J. **12**, 349 (2021)
13. Ono, T., Kanamaru, T.: Prediction of pedestrian trajectory based on long short-term memory of data. In: 21st IEEE International Conference on Control, Automation and Systems (ICCAS), pp. 1676–1679 (2021)
14. de Freitas, N.C.A., da Silva, T.L.C., de Macêdo, J.A.F., Junior, L.M., Cordeiro, M.G.: Using deep learning for trajectory classification. In: Proceedings of the 13th International Conference on Agents and Artificial Intelligence (ICAART 2021) (2021)
15. Brito, B., Zhu, H., Pan, W., Alonso-Mora, J.: Social-VRNN: one-shot multi-modal trajectory prediction for interacting pedestrians. arXiv preprint arXiv:2010.09056 (2020)
16. Zhang, Y., Zheng, L.: Pedestrian trajectory prediction with MLP-social-GRU. In: 13th International Conference on Machine Learning and Computing, pp. 368–372 (2021)

17. Helbing, D., Molnar, P.: Social force model for pedestrian dynamics. Phys. Rev. E J. **51** (1995)
18. Luber, M., Stork, J.A., Tipaldi, G.D., Arras, K.O.: People tracking with human motion predictions from social forces. In: IEEE International Conference on Robotics and Automation, pp. 464–469 (2010)
19. Mehran, R., Oyama, A., Shah, M.: Abnormal crowd behavior detection using social force model. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 935–942 (2009)
20. Wu, S.: Approach to auto-recognition of human trajectory in squares using machine learning-based methods-an application of the Yolo-v3 and the DeepSORT algorithm. CUMINCAD, CAAD research (2021)
21. Zamboni, S., Kefato, Z.T., Girdzijauskas, S., Norén, C., Dal Col, L.: Pedestrian trajectory prediction with convolutional neural networks. Pattern Recogn. J. **121** (2022)
22. Quan, R., Zhu, L., Wu, Y., Yang, Y.: Holistic LSTM for pedestrian trajectory prediction. IEEE Trans. Image Process. J. **30**, 3229–3239 (2021)
23. Zhong, Jianqi and Sun, Hao and Cao, Wenming and He, Zhihai: Pedestrian motion trajectory prediction with stereo-based 3D deep pose estimation and trajectory learning. IEEE Access **8**, 23480–23486 (2020)
24. Yi, S., Li, H., Wang, X.: Pedestrian behavior understanding and prediction with deep neural networks. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016. LNCS, vol. 9905, pp. 263–279. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46448-0_16
25. Ren, S., He, K., Girshick, R., Sun, J.: Faster R-CNN: towards real-time object detection with region proposal networks. Adv. Neural Inf. Process. Syst. J. **28** (2015)
26. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. Adv. Neural Inf. Process. Syst. J. **25** (2012)
27. Tay, N.C., Connie, T., Ong, T.S., Goh, K.O.M., Teh, P.S.: A robust abnormal behavior detection method using convolutional neural network. In: Computational Science and Technology. LNEE, vol. 481, pp. 37–47. Springer, Singapore (2019). https://doi.org/10.1007/978-981-13-2622-6_4
28. Ribeiro, M., Lazzaretti, A.E., Lopes, H.S.: A study of deep convolutional autoencoders for anomaly detection in videos. Pattern Recogn. Lett. J. **105**, 13–22 (2018)
29. Ko, K.-E., Sim, K.-B.: Deep convolutional framework for abnormal behavior detection in a smart surveillance system. Eng. Appl. Artif. Intell. J. **67**, 226–234 (2018)
30. Xu, D., Yan, Y., Ricci, E., Sebe, N.: Detecting anomalous events in videos by learning deep representations of appearance and motion. Comput. Vision Image Underst. J. **156**, 117–127 (2017)
31. Pennisi, A., Bloisi, D.D., Iocchi, L.: Online real-time crowd behavior detection in video sequences. Comput. Vision Image Underst. J. **144**, 166–176 (2016)
32. Nater, F., Grabner, H., Van Gool, L.: Exploiting simple hierarchies for unsupervised human behavior analysis. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 2014–2021 (2010)
33. Lv, F., Nevatia, R.: Single view human action recognition using key pose matching and Viterbi path searching. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–8 (2007)
34. Du, K., Kaczmarek, T., Brščić, D., Kanda, T.: Recognition of rare low-moral actions using depth data. Sens. J. **20** (2020)