Unsupervised Abnormal Behaviour Detection with Overhead Crowd Video

Shoujiang Xu[1,2], Edmond S. L. Ho[1], Nauman Aslam[1], and Hubert P. H. Shum[1]

[1]Northumbria University, UK, Email: {shoujiang.xu, e.ho, nauman.aslam, hubert.shum}@northumbria.ac.uk
[2]Jiangsu Food and Pharmaceutical Science College, China

***Abstract*—Due to the increasing threat of terrorism, it has become more and more important to detect abnormal behaviour in public areas. In this paper, we introduce a system to identify pedestrians with abnormal movement trajectories in a scene using a data-driven approach. Our system includes two parts. The first part is an interactive tool that takes an overhead video as an input and tracks the pedestrians in a semi-automatic manner. The second part is a data-driven abnormal trajectories detection algorithm, which applies iterative k-means clustering to find out possible paths in the scene and thereby identifies those that do not fit well in any paths. Since the system requires only RGB video, it is compatible with most of the closed-circuit television (CCTV) systems used for security monitoring. Furthermore, the training of the abnormal trajectories detection algorithm is unsupervised and fully automatic. It means that the system can be deployed into a new location without manual parameter tuning and training data annotations. The system can be applied in indoor and outdoor environments and is best for automatic security monitoring.**

I. Introduction

In recent years, the increasing threat of terrorism has caused more and more concern for public security. In particular, lone-wolf terrorist attacks in public areas have caused a significant amount of damage to society. It is, however, difficult to apply traditional human-based security monitoring systems to tackle these threats. On the one hand, it is unfeasible to monitor a large number of security video streams manually considering the cost of manpower. On the other hand, monitoring systems that involve facial identification or close-up video raise concerns about privacy.

Previous research [1], [2] has shown that security monitoring can be presented as the problem of abnormal pedestrian trajectories detection, in which the behaviour of a person is expressed as a temporal sequence of 2D movement trajectories in the scene. The main advantage is that while being able to automatically identify people with abnormal behaviour effectively, these systems do not require personal information such facial features, which eases the privacy concerns. However, due to the difficulties in accurate pedestrian trajectories tracking, there are very few public databases based on real-world scenarios. While [1] make their dataset available to the public, the 2D trajectories of the pedestrian required tremendous effort to annotate manually. Also, it is a challenge to identify whether a trajectory is normal or not (i.e. the ground-truth) as it involves some subjective evaluation, making it inflexible to apply supervised machine learning algorithms such as the deep learning or supported vector machines to identify abnormal behaviour.

In this paper, we introduce a system for effectively detecting abnormal trajectories in a public area. The system consists of two parts. First, we apply a computer vision algorithm to track the trajectories of pedestrians in the street. We propose an effective, semi-automatic interface such that we can obtain high-quality results with a reasonable amount of labour. Second, we propose an unsupervised algorithm based on iterative k-means clustering to identify common paths in the scenario, thereby identifying pedestrian trajectories that do not fit into any of these paths as abnormal.

There are several key advantages of our proposed system. First, it takes an overhead video as the input, and therefore does not require personal information to be captured. Second, it does not require manually annotated class labels of trajectories for machine learning, saving a huge cost for annotating training data when setting up the system in a new location. Third, the system is adaptive to changes. Even when the general behaviour of the crowd has changed, for example, due to the closure of a path, our system can adapt to the changes automatically by retraining on the newly captured data. These features enable our system to be adaptable to most of the CCTV systems in the world.

Experimental results are highly positive. We create a database by capturing long sequences of video in a real-world public area. Our proposed system can effectively identify possible paths that connect different incoming and outgoing routes, as well as some abnormal trajectories that deviate from the norm of the data.

There are two main contributions of this work:

- We propose a semi-automatic system that involves a two-pass interface to track pedestrian trajectories in a captured overhead video. This facilitates the effective creation of a pedestrian trajectory database.
- We propose an automatic, unsupervised framework for abnormal behaviour detection. It involves an iterative k-means clustering algorithm to detect possible paths in the scene, and thereby identifies the trajectories that do not fit well with these paths.

The rest of the paper is organized as follows. In Section II, we briefly review researches that are related to abnormal behaviour detection. In Section III, we give an overview of the proposed system. Section IV explains how we create the database using a pedestrian tracking algorithm and a two-pass tracking interface. Section V details the abnormal behaviour detection system we proposed. Section VI presents the experimental results. Section VII concludes and discusses the work.

II. Related Work

In this section, we will focus on the related research in detecting abnormal behavior in crowd and pedestrians behavior analysis. Mehran et al. [2] proposed to detect abnormal behavior in crowd by modeling the movement of individuals using a Social Force model. Each individual is treated as a particle and their movement is controlled by the interaction force estimated from the social force model. However, since the model is driven by simple equations, the accuracy drops when the configurations of the environment are slightly modified [3] which makes it difficult to apply in different environments. Xiang et al. [4] proposed to automatically track moving objects in the scene and detect abnormal behaviors based on spatial-temporal features such as movement orientation over time. Again, manual annotations are required on each frame. In addition, only a limited number of pedestrians are included in the scenes in the experiments. It is unclear if the proposed method will work well in crowds.

A recent work by Yi et al. [1] proposed to understand pedestrian behaviors by taking into account the stationary crowd groups in the scene and inferencing the interactions between those stationary groups and pedestrians. they demonstrated such a model can be used for abnormal pedestrian behavior detection. However, a supervised training process is required and a ground-truth label is required for each individual pedestrian. Sun et al. [5] proposed using a Temporal-Spatial Coherence model to detect local abnormal behavior. Image blocks instead of pixels are extracted for training normal and abnormal models of image blocks. Using this method, the accuracy and speed of abnormal detection can be improved. However, abnormal image blocks have to be manually annotated in the training stage. Rasheed et al. [6] proposed to track and detect abnormal behavior in real-time for video surveillance. The object of interest is tracked by the optical flow model and a classifier for abnormal detection is trained using a Neural Network. While the end-to-end approach can directly detect abnormal behavior from an input video, the results heavily rely on whether an appropriate tracking model and noise removal technique are used.

In summary, while abnormal behavior detection is an active research area and encouraging outcomes have been presented over the years, most of the existing techniques require either 1) manually work to annotate abnormal situations, 2) tracking the object of interest or pedestrians which can be difficult, or 3) image appearance features which can be difficult when low-resolution video or surveillance video are being used. To tackle these problems, we propose a semi-automatic method to extract low-dimensional pedestrian trajectories from video and an unsupervised approach to detect abnormal behavior from the trajectories.

III. System Overview

Fig. 1 shows the overview of the proposed system. We first capture overhead video in a real-world scenario. Second, we implement a semi-automatic tool to track the pedestrian trajectories in the video. We propose a two-pass tracking algorithm for effective tracking. The first pass is a rapid tracking with one manually identified position per pedestrian. The second pass is a correction cycle with multiple identified positions for each incorrectly tracked trajectory in the first pass. Third, we design an iterative k-means clustering algorithm that automatically finds out possible paths in the scenario. Those trajectories that do not fit well into the found paths are considered to be abnormal. Finally, we implement a visualization system to visualize these evaluated paths and abnormal trajectories.

IV. Database Creation

In this section, we explain how we capture the crowd video in a real-world scenario. We then explain the computer vision based semi-automatic tracking tool we have adapted for pedestrian tracking. Finally, we introduce the two-pass interface we have designed for more effective tracking.

A. Crowd Video Capture

To construct a pedestrian trajectory database, we capture overhead video in a real-world scenario as shown in Fig. 2. This was done by setting up a camera at the top floor of a near-by building. Multiple long video sequences are captured. The sequences are all in a high-resolution of 1920×1080 at 50 frames per second. This ensures that we obtain enough detail for successful tracking.

We have obtained approval to capture the video within a university area. Ethical approval has been sought and approved by the university ethics committee. During the capture, we setup notification boards to declare the capturing process and our contacts. If a person wishes his/her data to be erased for any reason, this will be done immediately. We enforce two strategies to minimize the risk that people are being identified in order to protect their privacy. First, we employ overhead cameras such that we do not capture the front and back of the people. Therefore, in the images, we can only see from the top of the people, including the top of the head (i.e. hair) as well as a small part of the shoulder. The information we captured should not lead to someone being identified. Second, in case we accidentally capture the faces of some pedestrian (e.g. when someone looks upwards), our system detects and masks these faces such that they become unidentifiable.

B. Pedestrian Tracking

We adapt the tracking method in [7], which was recently enhanced in [8] for better computational efficiency, to track the pedestrians in the video. A major advantage of the tracking method is that it handles significant scale variations very well compared to existing approaches, making it highly suitable to track overhead video of pedestrians in which the scale of the pedestrians change a lot across frames due to camera angles and lens distortions.

The tracking method takes a video and a number of manually defined tracking patches that indicate the tracked objects as the input. It extracts HOG features for the translation filter and concatenates it with the usual image intensity features. Tracking is done by minimizing a cost function that evaluates the correlation of the tracked object across frames, through estimating the translation and scale independently.

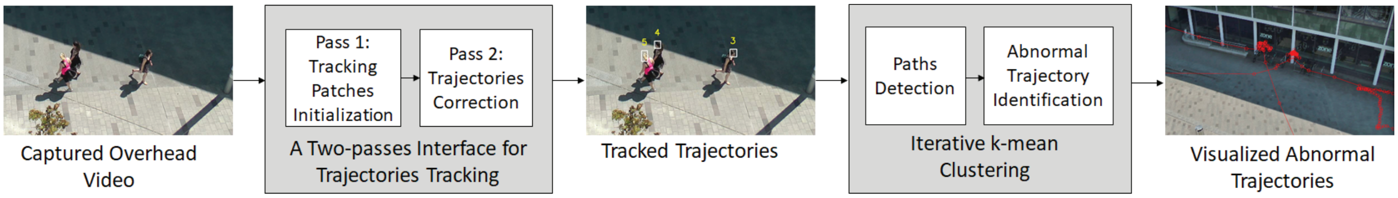


Fig. 1. The overview of the proposed system.

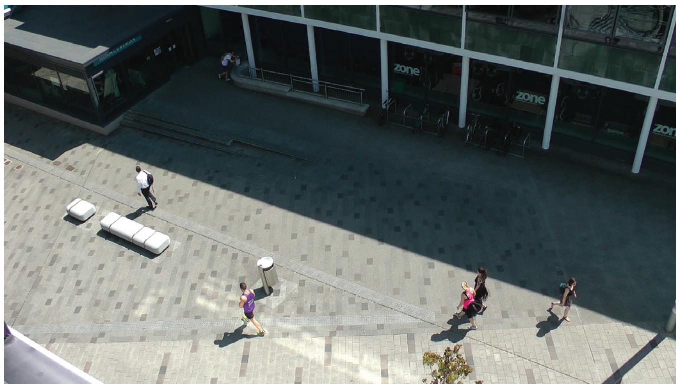


Fig. 2. An example frame of the captured video.

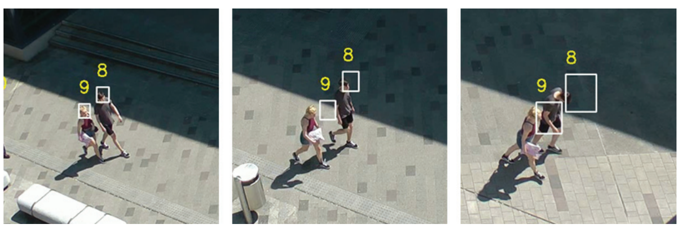


Fig. 3. Left: A user indicated tracking location. Middle and Right: Incorrect tracking locations obtained in later on frames after the first pass tracking.

Fig. 4. The final tracked results after the second pass tracking.

The result of the tracking is a sequence of tracked patches represented by four 2D positions in each frame, indicating the four corners of the tracked patch in the frame. The trajectory position in a frame is calculated as the center point of the tracked patch.

It is possible to run the tracking system without human input by integrating a human shape recognition algorithm [9] to initialize the tracker locations. However, in our system, we prefer to manually initialize the trackers as this can produce more accurate results.

C. A two-pass Interface for Effective Tracking

Like all other trackers, the tracking system explained in Section IV-B may generate inaccurate results due to occlusions and changing lighting conditions. A naive solution would be manually updating the tracking position during run-time whenever an error occur, which results in a large overhead in labour cost. Here, we propose a two-pass interface for effective tracking.

The first pass of the system requires the user to initialize one position in each pedestrian trajectory. The system then tracks the pedestrian both forwards and backwards in time, until the trajectory either (1) reaches the boundary of the captured video, or (2) reaches the predefined polygons that represent the entrances of the buildings. The user is therefore advised to indicate the mentioned trajectory position in a frame where the pedestrian has a clear, non-occluded appearance. Since the computation cost of the tracker system is not consistent and depends on the number of tracking objects, for more efficient user interaction, we first obtain the user input for all the pedestrian trajectories. We then run the tracker system as an offline process based on the user input. This minimizes user's waiting time and allows the user to playback the video in the preferred speed. Once the tracking process is finished, we obtain a set of resultant trajectories. Notice that some of the tracked results may be incorrect, as visualized in Fig. 3.

The second pass of the system aims at correcting all the mis-tracked trajectories. Based on the tracked results, the user first identifies which pedestrian trajectories are incorrect. Then, for each of these trajectories, the user indicate a number of trajectory positions in different frames. The tracker system iteratively start tracking at each of this position until it reaches the frame when the next position is indicated. In this case, even if the tracker gradually runs into error from one position, the error will be rectified at the next indicated position. We empirically found that one position in every 60 frames produces good results. Again, this process is implemented as an offline process for better user interaction. After this, we will obtain a set of highly accurate pedestrian trajectories. Examples of tracked results are shown in Fig. 4.

V. Abnormal Behaviour Detection

In this section, we explain our abnormal behaviour detection system. We first explain how we represent the trajectories. Then, we detail how to apply iterative k-means clustering to find out the possible paths in the scenario, with which we evaluate the abnormal trajectories.

A. Trajectories Representation

Since different people have different walking speeds, the tracked trajectories come with different lengths. To avoid the influence of walking speed and to focus on the topology of the trajectories, we uniformly subsample all trajectories into a predefined length. Each trajectory is represented as $P(t) \quad \forall \quad t \in [1, t']$, where $P(t)$ is the t^{th} 2D position in the trajectory, t' is the predefined number of samples to represent one trajectory. We empirically found that setting $t' = 40$ produces good results, but other values work as long as they allow the retention of the geometrical features of the trajectories.

We define a distance function between two trajectories to be:

$$D(P_1, P_2) = \sum_{t=1}^{t'} \sqrt{(P_1(t) - P_2(t))^2}, \tag{1}$$

where P_1 and P_2 are two trajectories, $P_1(t)$ and $P_2(t)$ are the corresponding t^{th} 2D positions.

B. Path Detection

In order to detect possible paths in the scene, we apply k-means clustering on the set of trajectories, using Equation 1 as the distance function. A major disadvantage of k-means clustering is that the number of clusters, k, has to be defined manually. We solve this problem by implementing an iterative approach to find out the best k in a given scenario.

To do so, we first start with $k = 1$ and iteratively increase the value of k by 1. In each iteration, we evaluate the Davies-Bouldin index [10], which evaluates the quality of the clusters. A small value indicates a good separation of the clusters and the tightness inside the clusters. The system iteratively increments the value of k until the Davies-Bouldin index reaches a local minimum. The advantage of analyzing the Davies-Bouldin value instead of defining the cluster size k is that the system can be adapted to other scenarios with different number of paths.

The result of the k-means clustering process is a set of k clusters. We calculate the mean trajectory for each cluster. This results in k mean trajectories denoted as $P'_i \quad \forall \quad i \in [1, k]$.

As suggested by previous works [11], the distribution of pedestrian trajectories is time dependent. For example, in a university environment, there are significantly more incoming pedestrians in the early morning, and more outgoing ones in the evening. Also, the pedestrians walk faster during peak hours, and walk slower otherwise. This implies that abnormality is also a time-varying function.

To cope with the temporal aspect of the data, we utilize a rolling window method such that the system can adapt to changes of pedestrian distribution and behaviour. We implement a rolling window to consider the past m minutes of data when extracting the paths with k-means clustering. The clustering procedure is performed every n minutes to reflect the most up-to-date pedestrian behaviour. Empirically, we found that setting $m = 30$ and $n = 5$ produces good results.

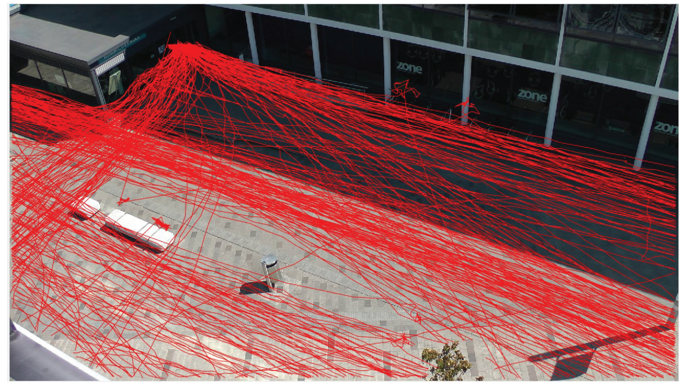


Fig. 5. An example of the 350 pedestrian trajectories extracted from a rolling window.

C. Abnormal Trajectories Identification

After finding out the possible paths in the scenario using k-means clustering, we evaluate individual trajectories to decide if they are abnormal. To do so, we define an evaluation function that calculates how the considered trajectory, P is different from each of the detected paths:

$$E(P) = \min_{\forall i \in [1,k]} \left(D(P, P'_i) \right), \tag{2}$$

where D is the distance function in Equation 1, $P'_i \quad \forall \quad i \in [1, k]$ are the k mean trajectories representing the paths in the scene detected by the k-means clustering process. If the evaluation $E(P)$ from Equation 2 is larger than a predefined threshold, the trajectory P is considered to be abnormal.

The proposed method of abnormal trajectories identification is unsupervised, as we do not require annotated class labels of whether a trajectory is normal or abnormal. An unsupervised method is essential for this problem as accurate ground-truth data is difficult, if not impossible, to obtain, due to the subjective aspect of the label annotators, as well as the limited information we have about the intentions of the pedestrians.

VI. Experimental Results

In this section, we present our experimental results on pedestrian tracking, possible paths extraction and abnormal trajectory detection. The video clips used in all experiments are captured by ourselves and the details are explained in Section IV.

A. Pedestrian Tracking Statistics

We present the pedestrian tracking results obtained using the method proposed in Section IV-C. We captured 105 minutes of full high definition (1920×1080) videos in a university campus with a *near*-bird's-eye view. An example of the 350 pedestrian trajectories extracted from a rolling window is shown in Figure 5. The results demonstrated that our proposed method can assist the user to extract high-quality pedestrian trajectories efficiently.

k	Davies-Bouldin index
2	1.044
3	0.897
4	0.797
5	0.854
6	0.739
7	0.733
8	**0.725**
9	0.792
10	0.748
11	0.806

TABLE I. AN EXAMPLE OF THE DAVIES-BOULDIN INDEX COMPUTED FROM THE PEDESTRIAN TRAJECTORIES IN A ROLLING WINDOW.

B. Path Visualization

Having extracted the pedestrian trajectories in each rolling window, iterative k-means clustering is performed and the optimal value of k is selected by computing the Davies-Bouldin index as explained in Section V-A. Table I shows the Davies-Bouldin index with different k-values computed from the pedestrian trajectories in a rolling window.

In our experiments, we found that the optimal k-value is 8 in all rolling windows in our dataset.

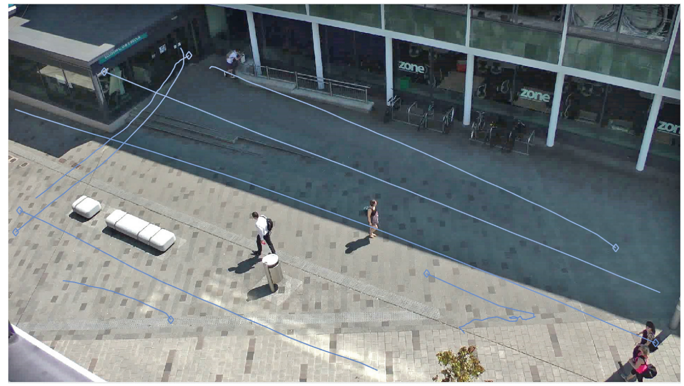
Next, the possible paths in the scene can be found by computing the centroid of each cluster from the k-means clustering results (with the selected optimal k-value. Example of the 8 possible paths extracted from different rolling windows are shown in Figure 6. While some paths are geographically close to and similar to others, such as the two entering/exiting the entrance of the building in Figure 6, the pedestrians are moving in different directions. Each possible path is ended with a diamond symbol to indicate the direction of movement in Figure 6. The results show that our proposed method can successfully extract the possible paths from a large amount of tracked pedestrian trajectories.

C. Abnormal Trajectories Detected

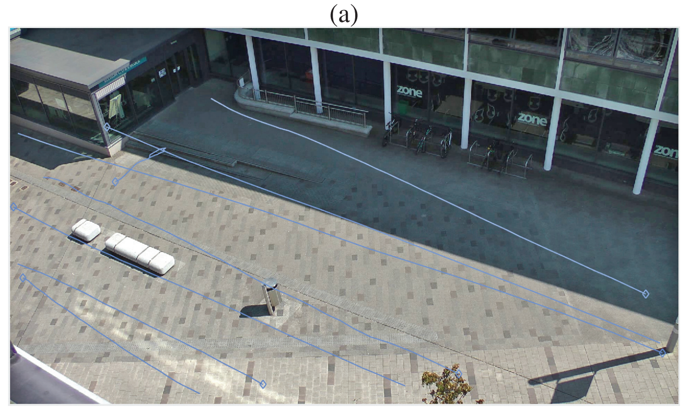
To detect abnormal trajectories, we calculate the distance between each tracked trajectory and the possible paths using Equation 2. We empirically selected the $threshold = 3200$ in the experiments. An example of the 3 abnormal paths detected from a rolling window is shown in Figure 7. From the results, it can be seen that the detected abnormal paths (in Figure 7) are very different from the possible paths (in Figure 6 (a)). This highlights the effectiveness of our proposed method. Each of the detected abnormal paths are illustrated separately with further explanation in Figure 8.

VII. DISCUSSIONS AND CONCLUSION

In this paper, we propose an unsupervised method system for abnormal behaviour detection using overhead video. We first present a two-pass interface for efficient tracking of pedestrians in a given overhead video. Then, we propose an iterative k-means clustering framework to identify possible paths in the scenario, as well as to evaluate pedestrians with abnormal behaviour. The system does not need annotated labels and can be adaptive to the temporal aspect of the data. Experiment results shows that the system can successfully detect and visualize abnormal trajectories.



(a)



(b)

Fig. 6. Examples of the 8 possible paths extracted from different rolling windows.

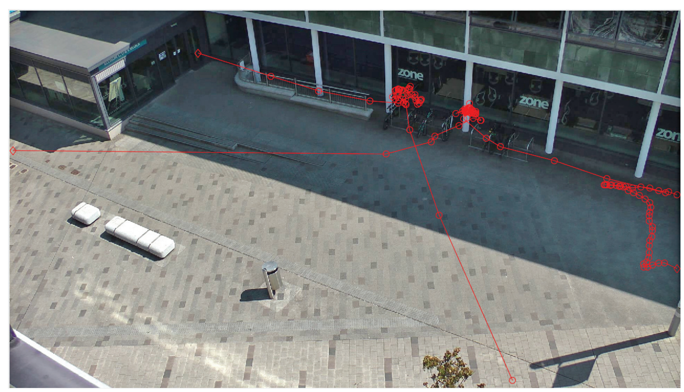
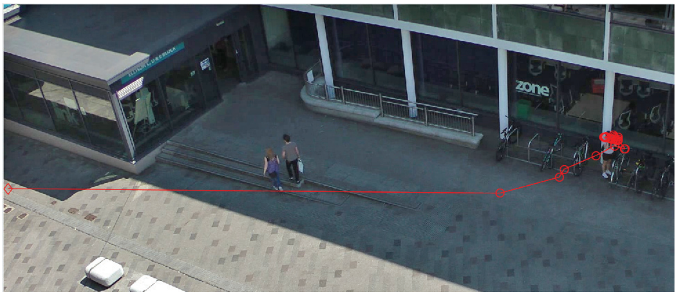


Fig. 7. An example of the 3 abnormal paths detected from a rolling window.

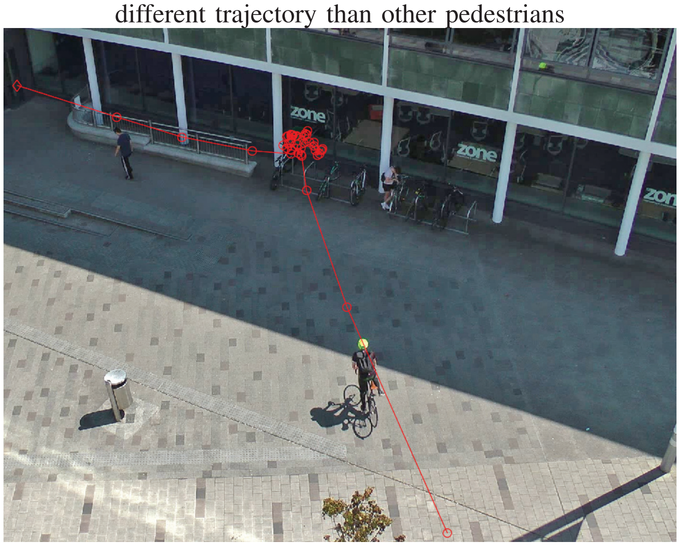
While the proposed system can be adapted to the temporal variation of the data using a data-driven approach, it does not explicitly model such temporal variation. We are interested in introducing regression models such that we can effectively understand the variation of data, such as the increase/decrease of pedestrians and the change of movement behaviour, over time. This can further improve identification accuracy, especially in scenarios with fewer pedestrians.

We are also interested in exploring complex nonlinear systems such as deep neural networks in abnormal behaviour detection. A Convolution neural network with an autoencoder setup can effectively extract the low-dimensional features of the high-dimensional pedestrian trajectories, and such low-



(a) The person stayed at the parking area for a long period of time before leaving

(b) The person was cleaning the street and results in a very different trajectory than other pedestrians



(c) The person was cycling and then stayed at the parking area for a long period of time before entering the building

Fig. 8. Showing each of the detected abnormal paths separately.

dimensional features can be used for identifying trajectories with abnormal patterns. A major challenge in adapting deep learning algorithms is the need of a large amount of training data.

Person re-identification is another potential future work direction. Due to the limited capture range of CCTV cameras and their incomplete coverage, there may be situations in which a person exits and enters the same monitored areas repeatedly. In order to identify potentially abnormal behaviour in these situations, it is important for the system to be able to uniquely identify people who enter the scene multiple times.

While we use pedestrian trajectories in this research, the proposed methodology can be used for other types of trajectories for different purposes. For example, it can be applied in a car park scenario to detect abnormal car movement.

Acknowledgment

This work was supported in part by the Engineering and Physical Sciences Research Council (EPSRC) (Ref: EP/M002632/1), the Royal Society (Ref: IE160609), and the Jiangsu Overseas Research & Training Program for University Prominent Young & Middle-aged Teachers and Presidents.

References

[1] S. Yi, H. Li, and X. Wang, "Pedestrian behavior modeling from stationary crowds with applications to intelligent surveillance," *IEEE Transactions on Image Processing*, vol. 25, no. 9, pp. 4354–4368, Sept 2016.

[2] R. Mehran, A. Oyama, and M. Shah, "Abnormal crowd behavior detection using social force model," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, June 2009, pp. 935–942.

[3] D. R. Parisi, M. Gilman, and H. Moldovan, "A modification of the social force model can reproduce experimental data of pedestrian flows in normal conditions," *Physica A: Statistical Mechanics and its Applications*, vol. 388, no. 17, pp. 3600 – 3608, 2009. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0378437109004075

[4] J. Xiang, H. Fan, and J. Xu, "Abnormal behavior detection based on spatial-temporal features," in *2013 International Conference on Machine Learning and Cybernetics*, vol. 02, July 2013, pp. 871–876.

[5] X. Sun, S. Zhu, and Y. Cheng, "Temporal-spatial coherence based abnormal behavior detection," in *2017 29th Chinese Control And Decision Conference (CCDC)*, May 2017, pp. 1997–2001.

[6] N. Rasheed, S. A. Khan, and A. Khalid, "Tracking and abnormal behavior detection in video surveillance using optical flow and neural networks," in *2014 28th International Conference on Advanced Information Networking and Applications Workshops*, May 2014, pp. 61–66.

[7] M. Danelljan, G. Hger, F. Shahbaz Khan, and M. Felsberg, "Accurate scale estimation for robust visual tracking," in *Proceedings of the British Machine Vision Conference*. BMVA Press, 2014.

[8] M. Danelljan, G. Hger, F. S. Khan, and M. Felsberg, "Discriminative scale space tracking," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 8, pp. 1561–1575, Aug 2017.

[9] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proceedings of the 25th International Conference on Neural Information Processing Systems*, ser. NIPS'12. USA: Curran Associates Inc., 2012, pp. 1097–1105.

[10] M. Halkidi, Y. Batistakis, and M. Vazirgiannis, "On clustering validation techniques," *Journal of Intelligent Information Systems*, vol. 17, pp. 107–145, 2001.

[11] H. Wang, J. Ondej, and C. OSullivan, "Trending paths: A new semantic-level metric for comparing simulated and real crowd data," *IEEE Transactions on Visualization and Computer Graphics*, vol. 23, no. 5, pp. 1454–1464, May 2017.