Abnormal Pedestrian Trajectory analysis based on arbitrary-length clustering


Published in:

Document Version:
Peer reviewed version
Abnormal pedestrian trajectory analysis based on arbitrary-length clustering

Diane Murdock and Jesus Martinez del Rincon

The Centre for Secure Information Technologies (CSIT), Queen’s University Belfast, UK

Abstract

This paper examines the use of trajectory distance measures and clustering techniques to define normal and abnormal trajectories in the context of pedestrian tracking in public spaces. In order to detect abnormal trajectories, what is meant by a normal trajectory in a given scene is firstly defined. Then every trajectory that deviates from this normality is classified as abnormal. By combining Dynamic Time Warping and a modified K-Means algorithm for arbitrary-length data series, we have developed an algorithm for trajectory clustering and abnormality detection. The final system performs with an overall accuracy of 83% and 75% when tested in two different standard datasets.

Keywords: Trajectory Analysis, Dynamic Time Warping, Clustering, Abnormal detection

1 Introduction

Video surveillance of public spaces in order to study activities and behaviours is becoming increasingly popular in today’s society. The benefit of processing and analysing surveillance data automatically is becoming increasingly obvious since allows activity recognition and anomaly identification. This is due to the need for heightened public safety and crime prevention by law enforcement agencies in vast areas and camera networks. In particular, the study of trajectory patterns corresponding to erratic or obscure movements and activities, such as wandering or loitering, may suggest suspicious and abnormal behaviour in public spaces.

A new generation of pedestrian detectors and trackers, able to provide an unprecedented accuracy even in dense and moderately crowded sequences, allows the development of human behaviour analysis systems that can cater from their outputs. Thus, abnormal behaviour detection based on the analysis of accurate trajectories provided by automated systems has emerged in the literature [Morris and Trivedi, 2009, Zhang et al., 2006]. While incorporating other sources of information in addition to the simple raw trajectory data, such as local motion descriptors [Datta et al., 2002] or spatio-temporal data [Robertson and Reid, 2006] may provide more precise information about the types of actions and give better results in the detection of abnormal behaviour, there are also disadvantages and they make the system more dependent on the environment and camera setup.

In this context, the present paper focuses on trajectory analysis of pedestrians in public spaces. To distinguish between normal and abnormal trajectories, two main processes are applied. Our proposed framework first defines what is meant by a normal trajectory in a given scene, and then classifies as abnormal every trajectory that deviates from this normality. Given a set of ‘normal’ trajectories, these are grouped into clusters according to information encapsulated into the trajectories elements. A distance able to take into consideration the specific characteristic of human motion patterns, such as different lengths, speeds and jittering, is crucial in such process. Trajectories that are close to the cluster mean are considered normal, while all trajectories that lie further away are considered abnormal.
1.1 State of the art

In trajectory analysis, unsupervised learning provides a versatile and effective approach given the lack of prior information regarding the different types of possible trajectories for each given scenario. In this context, clustering has been proved as a standard approach to group and classify trajectories. Among the different clustering options, diverse approaches such as k-means, fuzzy k-means, graph mining, Expectation-Maximization, Self-organised maps and Hidden Markov models have been evaluated with no significant differences between them [Morris and Trivedi, 2009]. On the contrary, the choice of the distance used to compare trajectories shows a bigger relevance and it relates directly with the resulting clustering [Zhang et al., 2006]. For instance, Euclidean distance allows measuring the dissimilarity between two trajectories but it is limited to trajectories with exactly the same number of elements, which is unusual in pedestrian analysis.

As a consequence, a preprocessing step is usually required prior to clustering. Since human trajectories are diverse in their execution, a normalization process is required to allow clustering algorithms to properly group trajectories based on their motion patterns and shape rather than by speed variations, global direction, fragmentation or punctual differences.

In order to address this limitation, the simplest solution consists in applying a geometric transformation to raw trajectory data prior to its classification. In [Sillito and Fisher, 2008, Majekka, 2009], raw trajectories are approximated by cubic spline curves with a fixed number of points. In [Johnson and Hogg, 1996], a slightly different approach was proposed, where each trajectory is converted into a sequence of 4-dimensional flow vectors composed of 2D position coordinates and velocity of the tracked object, followed by a clustering algorithm based on neural networks to obtain a finite set of prototype vectors. These approaches allow trajectories to be represented by the same number of attributes in order to facilitate the posterior analysis. However, they entail a set of assumptions, such as the required number of points or the order of the spline approximation, which is not constant for different motion patterns and largely unknown a priori for a give scene.

Other approaches [Bashir et al., 2007] made use of dimensionality reduction techniques such as principal component analysis (PCA) to select automatically the most relevant elements before clustering. This permits to initially overestimate the number of elements in a trajectory that is then projected in lower dimensional space. However, a fixed number of points should be still decided initially and an interpolation/subsampling may be required for every trajectory. In general, they suffer to distinguish speed variations [Morris and Trivedi, 2009].

As alternative, the use of more complex distance metrics that allow comparing two set of data of arbitrary number of elements can be applied. Hausdorff distance has been proposed [Junejo et al., 2004], but since only captures the minimum distance between two shapes, the subtle information encapsulated in human trajectories and required for its classification may be lost [Zhang et al., 2006] leading to poor classification rates. Better results have been achieved with the use of Dynamic Time Warping (DTW) [Ratanamahatana and Keogh, 2004] and Longest Common Subsequence ( LCSS) [Buzan et al., 2004]. The extra complexity of these distance metrics allow simpler clustering algorithms to be employed in the later stage. However, while the aforementioned distances allow comparing trajectories of different number of elements they may bring new problems such as the removal of contextual information, such as global direction, entrance and exit in the scene, etc., helpful for the classification of trajectories.

Moreover, only a minority of approaches in the literature have addressed the detection of abnormal behaviour, but limited to classify trajectories into known classes.

2 System Architecture

The aim of our proposed system is to detect abnormal trajectories. In order to achieve this goal, the meaning of what a normal trajectory in a given scene is should be first defined. Figure 1 depicts our proposed framework. Our system is composed of a k-means-based clustering algorithm, which has been modified in order to allow arbitrary length trajectories to be compared and grouped using DTW as a distance metric. The proposed modifications also allowed trajectories of different element numbers to be averaged and their variability to be measured, which is a novel contribution regarding other implementations. The resulting cluster means will
represent the normal trajectory prototypes and their learned variation, in the form of standard deviation, will be finally used in testing to classify and separate normal and abnormal trajectories.

Figure 1: Framework overview

The input data to our system are the raw trajectories provided by any automated human pedestrian detector and tracking system. We define a trajectory as a \( n \times d \)-dimensional time series composed of a sequence of \( n \) elements of \( d \)-dimensions associated with a pedestrian or an object moving through a scene. In order to fully exploit the information provided by the sourcing automated system, each element is composed by the 2D spatial coordinates and their corresponding instant velocities as attributes. Thus a trajectory \( A \) is defined as follows:

\[
A = (a_1, a_2, ..., a_n) = \left( (x_1, y_1, v_{x1}, v_{y1}), (x_2, y_2, v_{x2}, v_{y2}), ..., (x_n, y_n, v_{xn}, v_{yn}) \right)
\]  

(1)

where \( n \) is the number of elements, \((x, y)\) is each of the 2D points composing the trajectory and \((v_x, v_y)\) their corresponding instant velocities.

2.1 Dynamic Time Warping

DTW is a time series alignment algorithm which aims to align two time series of coordinates by warping the time axis iteratively until an optimal match between the two time-series has been found [Ratanamahatana and Keogh, 2004]. In this manner, two time series that are similar but locally out of sync can be aligned in a nonlinear manner. The \( d_{DTW} \) distance between two trajectories \( A_1 \) and \( A_2 \) of lengths \( n_1 \) and \( n_2 \) respectively is described as

\[
d_{DTW}(A_1, A_2) = \min \left\{ \sqrt{\sum_{k=1}^{\min(n_1,n_2)} w_k} \right\}
\]

(2)

where \( w_k \in W_{A_1,A_2} \) is the matrix element \( C(i, j) \) that belongs to the \( k^{th} \) element of the warping path \( W_{A_1,A_2} \), a continuous set of matrix elements that represent the mapping -or minimum path- between \( A_1 \) and \( A_2 \). Each matrix element is calculated recursively following:

\[
C(i, j) = d(a_i, a_j) + \min(C(i-1, j-1), C(i-1, j), C(i, j-1))
\]

(3)

being \( d(a_i, a_j) \) the Euclidean distance between two trajectory elements.

In our framework, DTW was chosen above other alternatives such as LCSS due to DTW sensitiveness to outliers. Since our aims is to detect anomalies, this sensitivity would draw attention to abnormalities in the data.
2.1.1 Directional normalisation

Since we aim to define normality versus abnormality rather than thoroughly classify normal trajectories, trajectories from a point X to a point Y will be considered identical to those from Y to X for the sake of compactness and better resource management. However, given that in DTW compares sequentially each element in the trajectory without having an overall look at the trajectory, DTW will return a large distance even when two trajectories are identical element by element. This means that both trajectories could be assigned to different clusters. To avoid this undesired effect for our application, the global direction of the trajectories is aligned before applying DTW by flipping each trajectory if \( d(a_1^1, a_2^n) < d(a_1^1, a_2^L) \) so their closest extreme points are located at the initial position of the data series.

2.2 Trajectory Clustering

We base our system on a modification of the k-means algorithm that allows times series of different lengths to be clustered. Given a number \( K \) of expected clusters, the algorithm first assigns each trajectory to a cluster using DTW such that the distance between the trajectory and the cluster mean \( \hat{A}_k \) is the minimum distance.

\[
    k_i = \arg\min_k \{d_{DTW}(A_i, \hat{A}_k)\} \tag{4}
\]

Then, each cluster average is recalculated using all the correspondingly assigned trajectories. This algorithm iterates until convergence, i.e. cluster means do not change, or a maximum number of iterations is reached.

However, while the cluster means \( \hat{A}_k \) are easily initialised by randomly selecting \( K \) trajectories in the training set, the computation of a trajectory average given sets of elements of different lengths is not trivial. In order to recalculate the new cluster means, our method uses the mappings \( W \) provided by DTW between the longest trajectory \( A_L \) and every other trajectory \( A_i \) in its cluster. Once the correspondence between the longest reference trajectory elements to every other element within the trajectories in the cluster is known, the cluster mean can be computed as the concatenation of element averages with different number of elements using eq. 6. Figure 2 illustrates this procedure with a graphical representation.

\[
    \hat{A}_k = (\hat{a}_k^1, \hat{a}_k^2, ..., \hat{a}_k^n_L) \tag{5}
\]

\[
    \hat{a}_k^n \propto \sum_{i \in K} \left( \sum_{w \in W_{A_L,A_i}} a_i^w \frac{w}{n_i} \right) \tag{6}
\]

Figure 2: Mapping correspondence between three trajectories and averaging procedure

Two alternatives were tested for this paper to define the longest trajectory. First, the longest trajectory in a cluster was defined as the trajectory that possess the largest number of elements \( n_i \) so

\[
    L = \arg\max_i \{n_i\} \tag{7}
\]

Second, the longest trajectory was defined as the trajectory with the largest physical distance between its start and end point

\[
    L = \arg\max_i \{d(a_1^1, a_i^n)\} \tag{8}
\]

The comparison between these two approaches will be analysed in the experimental section.
2.3 Defining normality and detecting abnormality

Once the trajectory clustering has converged to a solution, each cluster mean $\hat{A}_k$ can be used as a prototype for each normal trajectory type. By selecting the minimum distance, a new trajectory $T$ can be assigned and classified to a specific type or normal trajectory, so

$$ pred = \arg\min_k \{ d_{DTW}(T, \hat{A}_k) \} $$

(9)

While the previous equation predicts the most likely classification assuming normality, abnormal trajectories will still be considered as part of a normal cluster. To avoid it, a trajectory will be classified as abnormal if it lies outside a given distance or threshold from its assigned cluster. We propose a threshold which is different for each of the normal clusters and proportional to the expected variance learned during training. This variance is calculated in a similar manner to the cluster averages, where the previously generated mappings $W$ are used to calculate the standard deviation around the means of each of the clusters, using the equation:

$$ STD_{A_k} = \sqrt{\frac{1}{n_L} \sum_{n=1}^{n_L} std_{a_k}^2} $$

(10)

$$ std_{a_k}^2 \propto \sum_{i \in k} (\sum_{w \in W_{A_k}^i} (a^n_i - \hat{a}_k^n)^2) $$

(11)

Finally, abnormal trajectories are detected by applying thresholding:

$$ Abnormal = \begin{cases} 1 & \text{if } d_{DTW}(T, \hat{A}_{pred}) > \alpha \cdot STD_{A_{pred}} \\ 0 & \text{otherwise} \end{cases} $$

(12)

where $\alpha$ is a constant chosen empirically.

3 Experimental Results

3.1 Datasets and experimental setup

Two different datasets have been used to validate the experiments and ensure that the conclusions are not depending on scenario or camera setup. The first dataset is the Edinburgh Informatics Forum Pedestrian Database [Edinburgh, 2016] which contains trajectories of detected targets of people walking observed from a zenithal view. Since manual annotation is needed to validate our experiments, a subset with the first 150 trajectories in the file tracks.24Aug were selected and classified as 'normal' or 'abnormal', obtaining a split of 66% normal and 34% abnormal trajectories. As annotation criterion, for a trajectory to be considered normal, it must start and end at an entry and exit in the scene with little diversion. Any trajectory that does not meet these criteria is considered abnormal.

The second dataset is the Oxford Real-Time Surveillance Town Centre Dataset [Oxford, 2016], which consists of a video surveillance sequence of in a town centre on a busy pedestrianised shopping street with several points of interest and shop entrances. The first 150 trajectories were selected and annotated for our evaluation, obtaining a 76.7% - 23.3% normal versus abnormal split.

3.2 Results

Different variations of our methodology were tested: using only 2D coordinates as attributes for each element (version 1), using 4D attributes for each element of the trajectory (2D coordinated + instant velocity) (version 2), using the directional alignment explained in section 2.1.1 (version 3 and 4). Version 3 and 4 differ in their
Results are shown in Figure 4 for the Edinburgh dataset, using an 80%-20% split between training and testing (similar to [Majecka, 2009]) but respecting the original normal-abnormal ration between them. Parameter $\alpha$ was set to 180 while the number of normal cluster $K$ was equal to 8. It can be seen how each version improves the previous one, demonstrating the individual value of each contribution. Thus, adding velocity to the comparison of each individual element improved the overall $d_T W$ and the mapping $W$ between trajectories. This is due to the fact that DTW algorithm fully removes the global structure and position of the trajectories which results on a loss of the trajectory structure. On the contrary, the explicit addition of velocity information preserves better the relations between trajectory elements. Direction normalisation also improves the result by reducing the number of clusters needed for defining normality and avoiding a multiplicity of cluster with similar trajectories. Finally, the best results are obtained when using as a reference for each cluster the longest physical trajectory rather than the trajectory with the largest number of elements. This is because humans tend to wander around points of interest in the scene for a long time which may result on extremely long outliers which should not be taken as reference to avoid corruption. Qualitative effect of each of this modifications are depicted in Fig. 5.a, b and c.

Previous results were obtained with manually-tuned parameter values $\alpha$ and $K$, so certain degree of overfitting to teh scenario is expected. To evaluate the sensitivity of our method to those parameters, parameter values were varied over a range for our best framework (version 4). Results are reported in Fig. 6. While the final performance varied according to the particular values, it can be observed how good performance is still achieved over a large range of values, which makes the algorithm easy to tune.

In order to evaluate the generality of our framework, our system (version 4) was evaluated on the Oxford dataset. A 50%-50% split between training and testing was used and parameters were set to $\alpha = 600$ and $K = 4$. Lower performance is achieved in this second dataset, which may be explained by the more limited definition of longest trajectory used as reference for each cluster: version 3 defines it as the trajectory with most elements while version 4 uses the trajectory with the largest physical distance (see eq. 7 and 8).
of this scenario, where most normal trajectories are linear motions along the street. This also justify the use of a larger $\alpha$ value, to ensure that normal trajectories are still included as such.

Finally, our system was compared against [Majecka, 2009], one of the scarce methods in the literature that targets abnormal trajectory detection rather than trajectory classification. The same testing set (20% of the Edinburgh dataset) was used in the comparison. Results could not be generated with this system in the Oxford dataset due to the very different scenario and the lack of a training model. These results are summarised in Table 1. Our method achieved the same performance than [Majecka, 2009] but with a much simpler and reduced training set (only 120 training trajectories versus a few thousand). Opposite to [Majecka, 2009], our methodology does not require supervised learning and abnormal trajectories must be removed. Furthermore, since our system does not apply any geometric transformation to the trajectories, it can also be easily applied to other scenarios with different camera perspectives.

4 Conclusion

In this paper, an unsupervised framework for abnormal trajectory analysis has been proposed. The system proposed an extension of k-means clustering that allows dealing with human trajectories of varied and arbitrary sizes. Our system has been evaluated in two different scenarios obtaining good performance in both, being the results at state-of-art level in the standard Edinburgh dataset, and showing advantages regarding previous methodologies in terms of simplest and fully unsupervised training as well as easier extension to different environments and camera perspectives.

As future work, an extension to automatically determine the number of cluster using Expectation maximization will be proposed.

References

Table 1: Result comparison over the 2 datasets.

<table>
<thead>
<tr>
<th>Method</th>
<th>Edinburgh</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th>Oxford</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Acc</td>
<td>Err</td>
<td>Sens</td>
<td>Spec</td>
<td>Acc</td>
<td>Err</td>
<td>Sens</td>
<td>Spec</td>
<td></td>
</tr>
<tr>
<td>Ours (V4)</td>
<td>0.83</td>
<td>0.17</td>
<td>0.87</td>
<td>0.71</td>
<td>0.75</td>
<td>0.25</td>
<td>0.78</td>
<td>0.63</td>
<td></td>
</tr>
<tr>
<td>[Majecka, 2009]</td>
<td>0.83</td>
<td>0.17</td>
<td>0.87</td>
<td>0.71</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
</tbody>
</table>


