AALBORG UNIVERSITY
STUDENT REPORT

Suspicious Behavior Detection in Video Surveillance

Project Report

Group 843

Aalborg University
MSc in Vision, Graphics and Interactive Systems

AALBORG UNIVERSITY
STUDENT REPORT

School of Information and Communication Technology
Aalborg University

Title:
Suspicious Behavior Detection in Video Surveillance

Theme:
Computer Vision

Project Period:
Spring Semester 2018

Project Group:
843

Group members:
Ana Rita Viana Nunes
Atanas Atanasov Nikolov

Supervisors:
Kamal Nasrollahi
Mohammad Naser Sabet Jahromi

Pages Number: 9

Contents

Chapter 1

Introduction

One of the research topics in Computer Vision that has become of interest in recent years is the detection of abnormal activities in surveillance videos.

Now, more than ever, public safety has become a great concern. Therefore, public spaces need to be monitored for any individual(s) that might be acting out of the ordinary with the intent of doing harm to someone else or even with the intent of stealing something.

Thus, the need for automated surveillance systems has increased, with the objective of assisting security officers in performing their job more efficiently.

Computer Vision algorithms can be used to detect and track moving targets in a video feed and extract features that allow a system to analyze and classify such targets as performing abnormal or suspicious behaviors and send an alert to security officers.

The classification of such behaviors is not linear and so research needs to be done in order to develop a system that can accurately detect different types of suspicious behaviors.

1.1 How are Anomalous and Suspicious behavior different?

Anomalous behavior or anomalies are patterns in data that are different from predefined normal behaviors (normal model). Figure 1.1 shows an example of anomalies in a two-dimensional data set. There are two normal regions N_1 and N_2, all points that are lying outside those regions are considered anomalies. Those patterns might be a result of malicious activity, system failure or noise in the data. Therefore, the relevance of anomalies to real life is a key feature in anomaly detection.

Novelty detection is also closely related to anomaly detection [16, 17]. Unlike anomaly detection, novelty detection is concerned with finding new or previously unknown patterns that a system was not aware of during training. The main

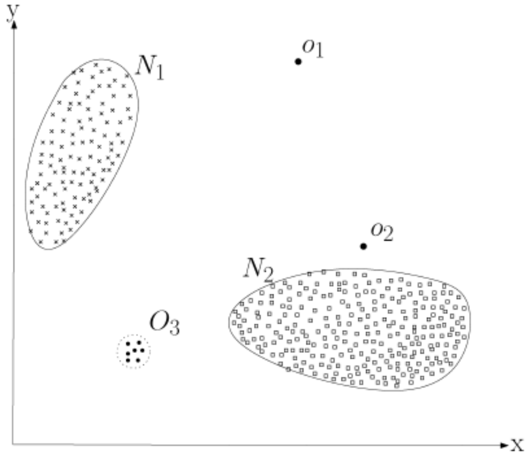

Figure 1.1: An example of anomalies in two-dimensional data set.

difference between the two is that novel patterns are typically later incorporated in the normal model.

While anomalous behavior can be described as behavior which differs from the expected, suspicious behavior is not that simple to model. Suspicious behavior is detected through subjective interpretation and presents a challenge even for human observers. Often the ability to predict how a situation will develop is helping in detecting suspicious behavior. Human observers rely on their experience often described as 'sixth sense' or 'gut feeling', to correctly detect suspicious behavior [18]. Context is another very important part of suspicious behavior detection. Behavior that is considered normal can become suspicious in a different context. Because of that, the normal behavior model needs to be updated over time as the context changes. This makes labeling data for training nearly impossible and it is also impossible to generate a data set capturing all possible human behaviors.

1.2 Challenges

Since anomalies are considered deviations from the normal model, detecting them should be straight forward. One solution would be to define a normal region and all observations outside that region are abnormal. However, several issues arise when trying to implement that approach.

- It is very difficult to define a normal region, that correctly represents every possible normal behavior. In addition, the separation between normal and abnormal is not always clearly defined.

- When the anomalies are caused by malicious actions, those actions are usually made to appear normal.

- The normal behavior is prone to change over time. A model which is currently a good representation of normal behavior might not be sufficient in the future.

- It is hard to find labeled data for training models.

- Detecting suspicious behavior is very subjective and depends on how actions are interpreted.

- Given behavior might be considered suspicious in one context but normal in another and because the context changes the system needs to account for that.

Because of those challenges it is not easy to create a general solution for detecting suspicious behavior. Instead most solutions are focused on very specific formulation of the problem.

1.3 Problem Statement

To Do

1.4 Related Work

The detection of suspicious behavior by automatic systems has raised a lot of interest in recent years. To be able to accurately detect suspicious behavior a system needs to have an understanding of general human behavior but also needs to understand the context in which this behavior is presenting itself.

Wiliem et al. [1] propose a context-based system for detecting suspicious behavior. This work considers that a system like this needs to have three main components. Firstly, it needs to continuously extract and learn contextual and human behavioral information from the video stream. Hence, a context space model is introduced in which the systems designers select important information that can describe a context but also allows the system to distinguish between two different instances of contexts. Secondly, it exploits contextual information in making decisions by introducing the use of a data stream clustering algorithm that enables the system to continuously update its knowledge from the data. This algorithm is capable of retrieving knowledge learned from a specific context. Lastly, the system incorporates an inference algorithm that combines contextual information and the system's knowledge to make decisions. The system also incorporates human observers' input in the decision making.

To test the system, experiments were done using the CAVIAR data set [19] and a private data set. The experiments showed that the system made accurate detections due to using previously acquired knowledge relevant to the context. The system was even able to detect unexpected events.

A lot of researchers resort to machine learning for detecting suspicious behavior which mostly relies on having reliable standard data sets for training and testing which, sometimes, might be hard to acquire. To escape this issue Elhamod and Levine [2] propose a semantics-based solution which is based on human reasoning and logic. This solution elaborates a mathematical framework based on abstract descriptions shown by Fuentes and Velastin [3] for detecting suspicious behavior and also builds up on previous work by the same authors [4].

Their proposed framework tracks people and luggage in a scene. Behaviors and events happening in a scene are semantically recognized by extracting object and inter-object motion features. The context of the investigation was the detection of suspicious behavior in public transport areas.

While earlier articles focus on detecting only one type of behavior [5, 6] this work analyzes different types of behavior relevant to the context such us: abandoned and stolen objects, fighting, fainting and loitering.

Detected objects in a scene are classified as being animate (e.g. people) or inanimate (e.g. luggage) which are the semantic entities associated with the events described. The extracted features are divided into single-object and inter-object. The single-object features include position, speed, direction and merged, while distance, alignment and speed difference are the inter-object features. The method follows the concept presented by Bird et al. [6] which uses motion features to classify objects into four categories: *unknown U*, *abandoned object O*, *person P*, and *still person SP*. The motion features are calculated and recorded in historical records and behaviors are semantically defined and detected by checking the records against predefined rules and conditions.

Public data sets such as BEHAVE [20], CAVIAR (PETS 2004) [21] and PETS 2006 [22] were used to test the framework. The results showed that the framework successfully detects all of the events discussed.

Methods for detecting suspicious behavior are usually developed for specific types of behaviors. To improve the accuracy of the system many methods require complex feature extraction algorithms that do not allow for the systems to be used in real-time. To solve this, Mu et al. [7] present a fast method for detecting suspicious behavior such as wandering, trailing, chasing and falling down. The proposed method is based on the extraction of motion vectors from a video stream to obtain the necessary features to classify behaviors in a video.

7-D features $\{\theta, V, \sigma_\theta, \sigma_V, E_\theta, E_V, Inter_{Dj}\}$ are extracted from a frame to describe each target detected. θ and V represent the average direction and velocity of the target; σ_θ and σ_V represent the direction variance and the velocity variance; E_θ and E_V represent entropy of direction and velocity; $Inter_{Dj}$ is the interesting degree of inter-frame $Inter_D$. The direction, velocity and inter-frame difference proved to be the most important features. A Support Vector Machine (SVM) is used to learn and classify the input videos.

The results from the experiments are compared with those in [8–11] and this method, in the majority of aspects, shows significant improvements when compared with the other methods.

In order to increase the efficacy of a video-surveillance control center for a shopping mall, in comparison with traditional methods, Arroyo et al. [12] investigated the detection of suspicious behavior in shopping malls. The analyzed risk situations in this context were a shop entry or exit of people, loitering events that can lead to theft and unattended cash desk situations.

The proposed approach employs an innovative tracking method that manages occlusions based on SVM kernels to compute distances between appearance features such as GCH (Global Color Histogram), LBP (Local Binary Pattern) and HOG (Histogram of Oriented Gradients). With these features, color, texture and gradient information are combined to obtain a robust visual description of people in videos.

The analysis of people's entrance or exit from the shops is important in the detection of crowded situations when a lot of people enter a shop at the same time or when people run away when exiting. For this, the line of the entrance to the shop is manually placed and the directions of people passing by are analyzed.

For loitering detection, risk zones are specified and loitering is detected whenever someone is in an area for a period of time longer than a defined threshold.

Unattended cash desks are a risky situation because someone might try to steal money from the cash register if no one is looking. So an alarm is given if a person is detected loitering around an unattended cash desk.

To evaluate the performance of the system, the publicly available CAVIAR data set was used for testing occlusion situations and the system showed to be slightly superior to multiple others state-of-the-art methods [13–15]. The system was also tested on a private data set and it was concluded that it could detect suspicious behaviors in the intended context.

State-of-the-art approaches need to take into consideration the context in which they want to detect suspicious behavior in order to accurately do so. There is still a lot of research that can be done to unveil new ways and approaches on how to implement even more efficient systems.

1.5 Organization

To Do

Bibliography

[1] A. Wiliem, V. Madasu, W. Boles and P. Yarlagadda. *A suspicious behaviour detection using a context space model for smart surveillance systems*. Computer Vision and Image Understanding, Elsevier, October 2011.

[2] M. Elhamod and M. D. Levine. *Automated Real-Time Detection of Potentially Suspicious Behavior in Public Transport Areas*. IEEE Transactions on Intelligent Transportation Systems, vol. 14, no. 2, June 2013.

[3] L. M. Fuentes and S. A. Velastin. *Tracking-based event detection for CCTV systems*. Pattern Anal. Appl., vol. 7, no. 4, pp. 356–364, December 2004.

[4] M. Elhamod and M. D. Levine. *A real time semantics-based detection of suspicious activities in public scenes*. Proc. 9th Conf. CRV, Toronto, ON, Canada, pp. 268–275, 2012.

[5] N. D. Bird, O.Masoud, N. P. Papanikolopoulos, and A. Isaacs. *Detection of loitering individuals in public transportation areas*. IEEE Trans. Intell. Transp. Syst., vol. 6, no. 2, pp. 167–177, June 2005.

[6] N. Bird, S. Atev, N. Caramelli, R. Martin, O. Masoud, and N. Papanikolopoulos. *Real time, online detection of abandoned objects in public areas*. Proc. IEEE ICRA, pp. 3775–3780, 2006.

[7] C. Mu, J. Xie, W. Yan, T. Liu and P. Li. *A fast recognition algorithm for suspicious behavior in high definition videos. Multimedia Systems*. Vol. 22, Iss. 3, 2016-6, p. 275–285, June 2016.

[8] K. Schindler and L. v Gool. *Action snippets: How many frames does human action recognition require?* IEEE Conference on Computer Vision and Pattern Recognition, 2008. CVPR 2008.

[9] A. Gilbert, J. Illingworth and R. Bowden. *Fast realistic multi-action recognition using mined dense spatio-temporal features*. Conference on Computer Vision, 2009 IEEE 12th International, 2009.

[10] A. Yao, J. Gall and L. V Gool. *A Hough Transform-Based Voting Framework for Action Recognition*. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2010.

[11] S. Sadanand and J. J. Corso. *Action Bank: A High-Level Representation of Activity in Video*. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 2012.

[12] R. Arroyo, J. J. Yebes, L. M. Bergasa, I. G. Daza and J. Almazán. *Expert video-surveillance system for real-time detection of suspicious behaviors in shopping malls*. Expert Systems with Applications, Elsevier, June 2015.

[13] T. Zhao and R. Nevatia. *Tracking multiple humans in crowded environment*. Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004., Vol. 2, 2004, p. II–406-II-413 Vol.2, July 2004.

[14] B. Wu and R. Nevatia. *Tracking of Multiple, Partially Occluded Humans based on Static Body Part Detection*. IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Volume 1 (CVPR'06), Vol. 1, 2006, p. 951–958, July, 2006.

[15] L. Li, W. Huang, I. Y. H. Gu, R. Luo and Q. Tian. *An efficient sequential approach to tracking multiple objects through crowds for real-time intelligent CCTV systems*. IEEE Transactions on Systems, Man and Cybernetics Part B, 38, 1254–1269, July 2008.

[16] M. Markou and S. Singh. *Novelty detection: A review-part 1: Statistical approaches*. Sig. Proc. 83, 12, 2481–2497, 2003.

[17] M. Markou and S. Singh. *Novelty detection: A review-part 2: Neural network based approaches*. Sig. Proc. 83, 12, 2499–2521, 2003.

[18] Wells, Helene. A., Allard, Troy, Wilson, Paul. *Crime and CCTV in Australia: Understanding the Relationship*. Centre for Applied Psychology and Criminology: Bond University, Australia, 2006.

[19] EC Funded CAVIAR project/IST 2001 37540.
`http://homepages.inf.ed.ac.uk/rbf/CAVIAR/`

[20] S. J. Blunsden, R. B. Fisher. *The BEHAVE video dataset: ground truthed video for multi-person behavior classification*. Annals of the BMVA, Vol 2010(4), pp 1-12, 2010.

[21] PETS-ECCV. 2004.
`http://www-prima.imag.fr/PETS04/caviar_data.html`

[22] PETS 2006 Benchmark Data. 2006.
`http://www.cvg.rdg.ac.uk/PETS2006/data.html`