# 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. For this, 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 dataset and a private dataset. The experiments showed 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 to detect suspicious behavior which most times relies on having reliable standard data sets for training and testing. 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 on previous work by the same authors [4].

Their proposed framework tracks people and luggage in a scene and, based on temporal properties, behavior and events are semantically recognized by employing 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 analyses 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 features such as position, speed, direction and merged, and inter-object features such as distance, alignment and speed difference. The method follows the concept presented by Bird et al. [6] which uses motion features to classify objects into four categories: unknownU, abandoned objectO, personP, and still personSP. 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, CAVIAR (PETS 2004) and PETS 2006 were used to test the framework. The results showed 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 {θ, V, σθ, σV, Eθ, EV, InterDj} are extracted from a frame to describe each target detected. θ and V represent the average direction and velocity of the target; σθ and σV represent the direction variance and the velocity variance; Eθ and EV represent entropy of direction and velocity; InterDj is the interesting degree of inter-frame InterD. The direction, velocity and interframe 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 dataset 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 dataset 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.

# 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.