

# Survey On Application Of Event Detection In The Videos And Learning The Human Behaviour In The Videos

\*ShikhaSinghal

# AbhishekSinghal

Amity School of Engineering and Technology, Noida, India

**Abstract:** -Video Event Detection is one of the most essential components for different kinds of domain functions of video information system. Now a day, security is the main issue of the human so in this survey we learn about the event detection of the human behaviour in the crowd like fighting of three persons, walking together, running etc. or in the airport, bank, play ground etc.

**Keyword:-** event detection, human recognition, detect human group.

## I. Introduction

At the beginning of the 21th century, data compression and physical storage of the videos are the main problem in the huge video data set. Videos are very important to get any types of information by using the internet videos like lecture on any topic, etc. For the large amount of videos we are studied about the data set of YouTube, KTH etc. In this Survey paper we are going through the recognize the event in the collection of human activity and we can recognize the activity of the human activity automatically analyzed the online activities and continuous recognize the human activity by detecting in the beginning and the ending of the activity in the videos. For the security purpose surveillance event detection is also very important. It can detect the unusual human behaviour in the complex areas like airports; banks etc and we also understand and detect the human behaviours in the videos.

## II. Related Work

For more than past, many researcher were working on detecting the event and recognize the events and human behaviour in the videos. In the past decade, a lot of work has been done in action recognition. JialieShen *et al.* [1] used a subspace technique to achieve fast and accurate video event detection. This approach is capable of discriminating different classes and preserving the intramodel geometry of samples with identical classes. Michel Merleret *et al.* [2] proposed semantic model vectors, an intermediate level semantic representation and also presented an end-to-end video event detection system and studied early and late feature fusion across the various approaches. J. C. SanMiguel *et al* [7] presented a feedback-based approach for event detection in video surveillance that improves the detection accuracy and dynamically adapts the computational effort depending on the complexity of the analysed and a structure is also proposed based on defining different levels of details for the analysis performed and estimating the complexity of data being analysed. A ruled based system is also designed to manage the interaction between feedback strategies. Alexia Briassouli *et al* [8] proposed that the system first find the pixels where the activities occurs by changing the statistics of luminance between the frames and also find the ratio based change detection technique where the activities occurs. J. K. Aggarwal *et al.* [9] discussed the statistical approaches, syntactic approaches, and description-based approaches for hierarchical recognition and also discussed the recognition of human-object interactions and group activities. H. Li A. Achimet *et al.*[10]

studied the problem of automatic anomaly detection for surveillance applications and a general framework for anomalous event detection in uncrowded scenes has been developed. Weilun Lao *et al.* [11] studied a flexible framework for semantic analysis of human behaviour from a monocular surveillance video, captured by a consumer camera and also proposed the framework was calculated and explore the combination of using trajectory and posture recognition in order to improve the semantic analysis of the human behavior.

Svetlana Lazebnik *et al* [20] proposed a method for recognizing scene categories based on global geometric. Weiyao Lin *et al* [24] presented a novel approach for automatic recognition of group activities for video applications and use an asynchronous hidden Markov model (AHMM) to model the relationship between people and also proposed a group activity algorithm which can handle both symmetric and asymmetric group activities and also make the detection of hierarchical interactions between two people.

### III. Event Recognition

Event Recognition purposes is to identify the action and goals of one or more from a collection of human activity and the environmental conditions. Example, group events include people fighting, people being followed, crowd analysis, people walking together, etc. Event Recognition methods can be categorized into Model based approaches and Appearance based techniques. Event detection is one of the most important components for the many types of domain application of video information system. The main functionality of event detection is to extract event inside the large scale collection. The event detection method is divided into two steps 1) generating videos content representation and 2) decision making process for detection [1]-[5]. Video information has many unique features and they are as: - i) Rich segmentation, ii) Large volume, iii) High dimensionality and iv) Complex internal structure. The main work is to get a fast and effective video recognition framework which is based on a novel subspace. It has two main components- video pre-processing and event detection component. After this a new method is used to reduce a dimension of the segment is called a Modularity Mixture Framework (MMF). MMF is a linear discriminate algorithm based on geometry preserving projection [5,6].

Video data produced in the form of image and video has increased in few years. With the growth of electronic storage capacity, necessity of large digital video libraries has also increased. Automatic video surveillance holds a great impact for the security and to prevent criminal behaviours or traffic accidents has become more and more popular in many places, such as airports, train stations, critical intersections, etc. Fields like surveillance, monitoring and security have received considerable attention lately, and many methods are developed for the detection of motion, activities, objects and events in video and other multimodal content. Video event detection is a key stage in semantic video content analysis. Feedback-based approach for event detection in video surveillance is to improve the overall performance of the system and its efficiency. An important structure of feedback based approach is based on two key ideas: i) availability of levels of detail for analysis (LoD) and ii) the complexity estimation of the data analysed [7]. In this we find the pixels where activities are occur by processing the higher order statistics of luminance changes between accumulate video frames. Once the pixels are extracted then the ratio based change detection techniques are applied to find at which frames activities occur. A novel approach for the detection of pixels where activity occurs in video and extraction of the frames when changes occur in the video that is where activities begin and/or end. Sequential change detection techniques are

then applied to the luminance variations between video frames, to find at which frames changes occur. The activity masks are then used to use only the active pixels to detect times of changes, providing lower computational cost and higher system reliability. The statistics of each active pixel's inter frame illumination variations are processed via sequential likelihood ratio testing to detect changes in it. The changes correspond to the beginning or ending of events [8].

#### **IV. Recognize the human activities**

Human activity recognition is an important area of computer vision research today. The aim of human activities recognition is to automatically analyze ongoing activities from an unknown video. The continuous recognition of human activities must be performed by detecting starting and ending times of all occurring activities in the videos [9]. Human activities are divided into four types:- gestures, actions, interactions and group activities. Gestures is defined as the motion of the person e.g. "Raising a leg". Actions are the combination of the many gestures e.g. "walking". Interaction involves the two or more person involvement e.g. "three persons walking". Group activities are those activities in which there are multiple persons are involved e.g. "a group of person walking"[10].

#### **V. Understand the human behaviour in video surveillance**

To understand the human behaviour analysis based on video surveillance, we need to examine not only the motion of the people, but also the posture of the people, as the posture gives the important clues to understand the human behaviour. To understand and recognize the various human motion first understand the scene. For the better understanding, we explore the combination of using trajectory and posture recognition to improve the semantic analysis of human behaviour [11]. Accurate detection and efficient recognition of various human postures plays a role in human recognition in the scene. Posture representation technique can be classified into two parts: - appearance based and shape based methods. Appearance based approaches utilize the intensity or color configuration within the whole body inferspecific body parts [12]. Shape based posture can be divided into two parts: contour based and silhouette based methods. Contour based approach [13], [14], different body parts are located, employing the external points detected along the contour, or internal points estimated from the shape analysis. Silhouette based approach utilize silhouettes in human movement analysis [15]. Human activity can be regarded as a temporal process in which human silhouettes continuously change with time. A 3-D reconstruction scheme is used for scene understanding so the actions of the persons can be analysed from different views. The semantic analysis have four processing levels:- 1) a preprocessing level including background modelling and multiple detection, 2) an object based level performing trajectory estimation and posture classification, 3) an event based level for semantic analysis, and 4) a visualization level including camera calibration and 3-D scene reconstruction. Human behaviour recognition method can include single person, multiple person interactions, person vehicle interactions, and person-facility/location interaction. All the pre-processing steps before human recognition are as:- 1) Motion Detection, 2) Object Classification, and 3) Object Tracking [15]. Classification and terminology for the human behaviour is similar to Video and Image Retrieval Analysis Tool (VIRAT). VIRAT divides the human behaviour in two parts as events and activities. Events define as a single low level spatiotemporal

entity that cannot be decomposed e.g. Person standing and person walking. An activity refers to a composition of multiple events e.g. A person loitering. Behaviour include the both events and activities. Human behaviour recognition has focused on several workshops such as Visual Surveillance, Event Mining [16]-[17], and Event Detection and Recognition [18]-[19].

For scene recognition categories it based on the global geometric and its work by partitioning the image into increasingly sub-regions and find the histograms of local features of each sub region. For the best result spatial pyramid is a simple and very efficient to enhance an order less bag of features image representation and it improved the performance on scene categorization tasks. Bag of feature represent an image as an orderless collection of local features. Bof disregard all information about the spatial layout of features and also it has some limitations also. They are incapable of capturing shape or segmenting an object from its background. A more efficient approach is to augment a basic bag of features representation with pair wise relations between neighbouring local features. Other strategy for increasing robustness to geometric deformations is to increase the level of invariance of local features and also computing histograms of local features at increasingly fine resolutions [20].

#### VI. Detecting human group behaviour

Detecting human group behaviour or human interactions has attracted increasing research interest [21]-[22]. Group events including people walking together, people fighting, terrorists launching attacks in groups etc. For the security purpose it is important to automatically detect human activities. So it have some problems for group event detection and given as follows:- i) Group Event Detection with a varying number of group members, ii) Group Event Detection with a hierarchical activity structure, iii) Clustering with an Asymmetric Distance Metric. To solve these problems [23] he used a symmetric-asymmetric activity structure (SAAS), group representative (GR), seed representative centered (SRC), and group representative based activity detection (GRAD) algorithms.

#### VII. Understand the human behaviour in the crowded area

For the complex environment or crowded area we need to detect a person on this so we use many cameras to detect an individual. To detect an individual from the crowded area we use the tracking by detection approach with nonnegative discretization. Nonnegative discretization is partitions the detected data points onto non-overlapping group. Nonnegative discretization is guarantee a person detection output to only belong to exactly one individual. This framework take the advantage of all important cues like color, person detection, face recognition and non background information to perform tracking but the cues are not always reliable so we used the localization and tracking algorithm which uses these features to perform robust tracking. Tracking algorithm is used to handle the complex indoor scenes consisting of different rooms, many walls and corridors and in this first to utilize face recognition for tracking. Trackers not used color information [24]-[25] will have difficulty avoiding identity switches when multiple people come very close together and split up. This algorithm followed the tracking by detection paradigm which handle the re-initializations naturally and avoids excessive model drift. To utilize semi-supervised learning techniques to perform multi camera multi object tracking. If two person detection

which have the spatial temporal neighbours and have same appearances then it is very common that two points have the same person and it is good for the manifold assumption so to uncover the manifold structure of detected data points by leveraging local learning technique. Sometime the manifold assumption is not good for tracking [25].

To find the abnormal behaviour in the crowded scenes firstly we find the low level visual features extraction and event classification by one class SVM. For abnormal event detection low level motion features could also be used like walking, running and climbing. In this HOFs (Histogram of the orientation of optical flow) are calculated in dense grids of the gradients of the image at a single scale without dominant orientation alignment.

### VIII. Conclusion

In this survey we learnt about the behaviour of human in crowded area and detect the events in the videos. In this survey we used many methods to detect and recognize the human behaviour in the videos and scenes. How we can understand the event and activities and also detect the many features of the image like a single person, group event, background understanding of the scene, ect.

### IX. References

- [1] L. Xie, S.-F. Chang, A. Divakaran, and H. Sun, *Unsupervised Mining of Statistical Temporal Structures in Video, Video Mining*, A. Rosenfeld, D. Doermann, and D. Dementhon, Eds. Norwell, MA: Norwell, 2003.
- [2] D. Xu and S. Chang, "Visual event recognition in news video using kernel methods with multi-level temporal alignment," presented at the IEEE Int. Conf. Computer Vision and Pattern Recognition (CVPR), 2007.
- [3] M. Shyu, X. Xie, M. Chen and S. Chen, "Video Semantic Event/concept detection using a subspace-based multimedia data mining framework," *IEEE Trans. Multimedia*, vol. 10, no. 5, p. 569, Oct. 2008.
- [4] D. Xu, S. Lin, S. Yan, and X. Tang, "Rank-one projections with adaptive margins for face recognition," presented at the IEEE Int. Conf. Computer Vision and Pattern Recognition (CVPR), 2006.
- [5] JialieShen, Dacheng Tao, and Xuelong Li, "Modality Mixture Projections for Semantic Video Event Detection," presented at the IEEE Trans. On Circuits And Systems for Video Technology, Vol. 18, No. 11, p. 1587-1596 November 2008.
- [6] MicheleMerler, Bert Huang, LexingXie, Gang Hua, and ApostolNatsev, "Semantic Model Vectors for Complex Video Event Recognition," presented at the IEEE Trans. On Multimedia, Vol. 14, No.1, p.88-101 February 2012.
- [7] J.C. SanMiguel and J.M. Martinez, "Use of feedback strategies in the detection of events for video surveillance," published at IET Computer Vision on 12<sup>th</sup> April 2010.
- [8] Alexia Briassouli and IoannisKompatsiaris, "Statistical Processing of Video for Detection of events in space and time," p. 1293-1296, ICME 2008.
- [9] J. K. AGGARWAL and M. S. RYOO, "Human Activity Analysis: A Review," *ACM Computing Surveys*, Vol. 43, No. 3, Article 16, April 2011.
- [10] H. Li, A. Achim and D. Bull, "Unsupervised Video anomaly detection using feature clustering," *IET Signal Process*, Vol. 6, Iss. 5, p. 521-533, 2012.

- [11] Weilun Lao, Jungong Han, and Peter H.N. de With, "Automatic Video-Based Human Motion Analyzer for customer Surveillance System," *IEEE Trans. Automatic Video-Based Human Motion Analyzer for Consumer Surveillance System*, p. 591-598, April 15, 2009.
- [12] S. Park and K. Aggarwal, "Simultaneous tracking of multiple body parts of interacting persons," *Computer Vision and Image Understanding*, vol. 102, pp. 1-21, 2006.
- [13] H. Fujiyoshi, A. Lipton and T. Kanade, "Real-time human motion analysis by image skeletonization," *IEICE Trans. Information and System*, vol. 87, pp. 113-120, 2004.
- [14] P. Peursum, H. Bui, S. Venkatesh and G. West, "Robust recognition and segmentation of human actions using HMMs with missing observations," *EURASIP Journal on Applied Signal Processing*, vol. 13, pp. 2110-2126, 2005.
- [15] P. Peursum, H. Bui, S. Venkatesh and G. West, "Robust recognition and segmentation of human actions using HMMs with missing observations," *EURASIP Journal on Applied Signal Processing*, vol. 13, pp. 2110-2126, 2005.
- [16] R. Nevatia, T. Zhao, and S. Hongeng, "Hierarchical language based representation of events in video streams," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog. Workshop*, 2003, vol. 4, pp. 39-46
- [17] R. Hamid, Y. Huang, and I. Essa, "ARGMode—Activity recognition using graphical models," in *Proc. IEEE Comput. Vis. Pattern Recog. Workshop*, 2003, vol. 4, pp. 38-44.
- [18] R. Nevatia, J. Hobbs, and B. Bolles, "An ontology for video event representation," in *Proc. IEEE Workshop Event Detection Recog.*, 2004, p. 119.
- [19] C. Rao and M. Shah, "View-invariant representation and learning of human action," in *Proc. IEEE Workshop Detection Recog. Events Video*, 2001, pp. 55-63.
- [20] SvetlanaLazebnik, CordeliaSchmid, Jean Ponce, "Beyond Bags of Features: Spatial Pyramid Matching for recognizing Natural Scene Categories,"
- [21] D. Zhang, D. Gatica-Perez, S. Bengio, and I. McCowan, "Modeling individual and group actions in meetings with layered HMMs," *IEEE Trans. Multimedia*, vol. 8, no. 3, pp. 509-520, Jun. 2006.
- [22] D. Wyatt, T. Choudhury, and J. Bilmes, "Conversation detection and speaker segmentation in privacy-sensitive situated speech data," in *Proc. Interspeech*, 2007, pp. 586-589.
- [23] Weiyao Lin, Ming-Ting Sun, RadhaPoovendran, and Zhengyou Zhang, "Group Event Detection with a varying number of group members for video surveillance" *IEEE Trac. On Circuit and Systems for video technology*, Vol. 20, no. 8, p. 1057-1076, August 2010.
- [24] A. Andriyenko, K. Schindler, and S. Roth, "Discrete continuous optimization for multi-target tracking," *CVPR*, 2012.
- [25] J. Berclaz, F. Fleuret, E. Turetken, and P. Fua., "Multiple object tracking performance: the clear mot metrics," *IEEE TPAMI*, 2011
- [26] Shoou-I Yu, Yi Yang, Alexander Hauptmann, "Harry Potter's Marauder's Map: Localizing and Tracking Multiple Persons of interest by Nonnegative Discretization,"