Grenze International Journal of Computer Theory and Engineering , Special issue

Human Group Tracking and Anomalous Event Detection

Seemanthini.K [1] and Dr.Manjunath.S.S [2]
[1-2]Dayananda Sagar Academy of Technology and Management, Bangalore, India
Email: be.outstanding@gmail.com, mnj_ss2002@yahoo.co.in

***Abstract*— This paper presents a framework that recognizes the small human group and to detect the event in the video. This framework is utilized for robotized little human gathering occasion discovery inside of social or open spot environment furthermore serves to recognize a fording wrong doings, for example, Railway station, Traffic, collages, office etc. The proposed framework aims to automatically extract foreground human group without any user interaction or the use of any training data and identifies the event in the group. In the proposed method, the coarse foreground extraction is obtained by using the motion and the edge information of an object. Then, the human group is extracted by using the horizontal/ vertical filling scheme based on the coarse foreground extraction. If group is formed, features are extracted using features extraction algorithms from particular frame. Finally, event in particular frame is classified using unsupervised classifiers. The proposed method can be applied to video object segmentation and further video editing and retrieval applications.**

***Index Terms*— Foreground Extraction, Human group Extraction (HOE), Visual Saliency, Unsupervised Classification and Event detection.**

I. INTRODUCTION

Over the recent years, detecting human beings and recognizing events in a video scene of a surveillance system is attracting more attention due to its wide range of applications in abnormal event detection, human gait characterization, person counting in dense crowd, person identification, gender classification, fall detection for elderly people, etc. The scenes obtained from a surveillance video are usually with low resolution. Most of the scenes captured by a static camera are with minimal change of background. Objects in the outdoor surveillance are often detected in far field. Most existing digital video surveillance systems rely on human observers for detecting specific activities in a real-time video scene. However, there are limitations in the human capability to monitor simultaneous events in surveillance displays. Hence, human event analysis in automated video surveillance has become one of the most active and attractive research topics in the area of computer vision and pattern recognition.

An intelligent system detects and captures motion information of moving targets for accurate object classification. The classified object is being tracked for high-level analysis. In this study, it focuses on detecting humans and to consider recognition of their complex activities. Human detection is a difficult task from machine vision perspective as it is influenced by a wide range of possible appearance due to changing articulated pose, clothing, lighting and background, but prior knowledge on these limitations can improve the detection performance.

Grenze ID: 01.GIJCTE.3.4.43

New video cameras are installed daily all around the world, as webcams, for surveillance, or for a multitude of other purposes. As this happens, it becomes increasingly important to develop methods that process such data streams automatically and in real-time, reducing the manual effort that is still required for video analysis. Of particular interest for many applications is the behavior of persons, e.g., for traffic safety, surveillance, or sports analysis. As most tasks at semantically higher levels are based on trajectory information, it is crucial to robustly detect and track people in dynamic and complex real-world scenes. However, most existing multiperson tracking methods are still limited to special application scenarios. They require multi-camera input, scene specific knowledge, a static background, or depth information, or are not suitable for online processing.

The main challenge when using an object detector for tracking is that the detector output is unreliable and sparse, i.e., detectors only deliver a discrete set of responses and usually yield false positives and missing detections. Thus, the resulting association problem between detections and targets is difficult.

Several recent algorithms address this problem by optimizing detection assignments over a large temporal window in an offline step [1], [2], [3], [4]. They use information from future frames and locate the targets in the current frame with a temporal delay or after the entire sequence have been observed. In contrast, Sequential Monte Carlo methods offer a framework for representing the tracking uncertainty in a causal manner. By only considering information from past frames, such approaches are more suitable for time-critical, online applications.

Moreover, there may be both individual and gathering activity in the same scene, it is much harder to speak to and perceive such situations. Punctuation models have been generally utilized as a part of the complex visual occasion acknowledgment lately. To apply sentence structure models for occasion acknowledgment, normally low-level components are firstly removed from features and after that characterized to an arrangement of terminal images, i.e. visual occasion primitives.

In a video like group of humans present, recognizing the human activities is not an easy task, the human activities in the videos consist of number of frames and each frame may have the different events or action such as hug, kick, punch and normal activities and so on. Unsupervised classifier are used for learning method and labels are known, Hence the instances such as kick, hugging, punch, other activities like features are extracted and events are detected. Before detecting the events the people in the video are detected and tracked to find group formation. The proposed work is used to represent both individual and group events, hence it overcomes the drawback of existing system, and it is based on cognitive linguistic method which uses unsupervised learning method.

II. LITERATURE SURVEY

[J. Berclaz et al, 2006] [1] has taken three or four videos which are synchronized at eye level and from different angles. Here, the author has showed the effective usage of dynamic programming to accurately follow up to six individuals across thousands of frames in spite of significant occlusions. In addition, this paper also derived metrically accurate trajectories for each one of them. The main contribution is to show that multi-person tracking can be reliably achieved by processing individual trajectories separately over long sequences, provided that a reasonable heuristic is used to rank these individuals and avoid confusing them with one another. In this way, author achieved robustness by finding optimal trajectories over many frames while avoiding the combinatorial explosion that would result from simultaneously dealing with all the individuals.

[M. Andriluka et al, 2008] [2] has combined the advantages of both detection and tracking into a single framework. The approximate articulation of each person is detected in every frame based on local features that model the appearance of individual body parts. Prior knowledge on possible articulations and temporal coherency within a walking cycle are modeled using a hierarchical Gaussian process latent variable model (hGPLVM). The author showed that how the combination of these results improves hypotheses for position and articulation of each person in several subsequent frames. The experimental results that demonstrate how this allows detecting and tracking multiple people in cluttered scenes with reoccurring occlusions.

[B. Leibe, K. Schindler et al,2008][3] The author presented a novel approach for multi-object tracking which considers object detection and space-time trajectory estimation as a coupled optimization problem. The approach is formulated in a Minimum Description Length hypothesis selection framework, which allows the

system to recover from mismatches and temporarily lost tracks. Tracking is achieved by performing model selection after every frame. At each time instant, the approach searches for the globally optimal set of space-time trajectories which provides the best explanation for the current image and for all evidence collected so far, while satisfying the constraints that no two objects may occupy the same physical space, nor explain the same image pixels at any point in time. Successful trajectories hypotheses are then fed back to guide object detection in future frames. The optimization procedure is kept efficient through incremental computation and conservative hypothesis pruning. The authors has evaluated this approach on several challenging video sequences and demonstrate its performance on both a surveillance-type scenario and a scenario where the input videos are taken from inside a moving vehicle passing through crowded city areas.

[Nor SurayahaniSuriani et al, 2008][4] Has proposed a survey which aims to detect and characterize a sudden event, which is a subset of an abnormal event in several video surveillance applications. This paper discusses the following in detail: (1) the importance of a sudden event over a general anomalous event; (2) frameworks used in sudden event recognition; (3) the requirements and comparative studies of a sudden event recognition system and (4) various decision-making approaches for sudden event recognition. The advantages and drawbacks of using 3D images from multiple cameras for real-time application are also discussed. The paper concludes with suggestions for future research directions in sudden event recognition.

[Y. Li, C. Huang et al, 2009] [5] has proposed a learning-based hierarchical approach of multi-target tracking from a single camera. This approach automatically select various features and corresponding non-parametric models, and combine them to maximize the discriminative power on training data using Hybrid Boost algorithm. A hybrid loss function is used in this algorithm because the association of tracklet is formulated as a joint problem of ranking and classification: the ranking part aims to rank correct tracklet associations higher than other alternatives; the classification part is responsible to reject wrong associations when no further association should be done. Experiments are carried out by tracking pedestrians in challenging datasets. The comparison of approach is done with state-of-the-art algorithms to show its improvement in terms of tracking accuracy.

[Weiyao Lin et al, 2009][6] Presents a novel approach for automatic recognition of group activities for video surveillance applications. This paper proposed to use a group representative to handle the recognition with flexible or varying number of group members, and use an asynchronous hidden Markov model (AHMM) to model the relationship between two people. Furthermore, a group activity detection algorithm which can handle symmetric and asymmetric group activities, and demonstrate that this approach enables the detection of hierarchical interactions between people is done. Experimental results showed the effectiveness of the approach.

[Joshua Candamo et al][2010] has described the current state-of-the-art image-processing methods for automatic-behavior-recognition techniques, with focus on the surveillance of human activities in the context of transit applications. The main goal of this paper to provide researchers in the field with a summary of progress achieved to date and to help identify areas where further research is needed. The author provided a thorough description of the research on relevant human behavior-recognition methods for transit surveillance. Recognition methods include single person (e.g., loitering), multiple person interactions (e.g., fighting and personal attacks), person– vehicle interactions (e.g., vehicle vandalism), and person–facility/ location interactions (e.g., object left behind and trespassing). A list of relevant behavior-recognition papers is presented, including behaviors, data sets, implementation details, and results. In addition, algorithm's weaknesses, potential research directions, and contrast with commercial capabilities as advertised by manufacturers are discussed.

[Sofia Zaidenberget al][2011] has presented a complete method for detecting and tracking groups of people in video-surveillance videos and to recognize behaviors of groups of people in the subway in order to alert subway security. The system is composed of 3 main layers: the detection of people in the video, the detection and tracking of groups among the detected individuals and the detection of events and scenarios of interest based on tracked actors (groups). The proposed method detects people in images using a LBP (Local Binary Pattern)-based people detector. The detected targets are linked by a Frame-to-Frame tracker before the groups are detected and tracked themselves. Events of interest are encoded in scenarios using a descriptive language and recognized based on properties of groups. The segmentation method (S) uses background subtraction and blob detection. Persons are detected by merging blobs of pixels. The HD method is used for human detection, which only detects persons entirely visible in the view. This paper majorly focused on group tracking and event detection layers.

[HarisUddin Sharif et al, 2011][9] Event detection in video surveillance is an important task for the places of both private and public. This paper mainly addressed a detection based method for video event detection, which is heavily based on optical flow techniques. Optical flow techniques track low level information like points of interest. The tracked interest points are grouped into several clusters using k-means algorithm. To find the principle components of each cluster, geometric means of location, direction, and displacement of the interest points of each cluster are estimated. And then they are used as the principle detecting components of each cluster rather than the individual feature points. Based on these components each cluster is defined either high activity cluster or low activity cluster. High activity cluster has high probability to contain a video event. Final decision of the video event is then performed by the thresholding of displacement and direction.

[Yimeng Zhang et al, 2012][10] has addressed the problem of group-level event recognition from videos. The events of interest are defined based on the motion and interaction of members in a group over time. Example events include group formation, dispersion, following, chasing, flanking, and fighting. To recognize these complex group events, the author propose a novel approach that learns the group-level scenario context from automatically extracted individual trajectories. First step performed a group structure analysis to produce a weighted graph that represents the probabilistic group membership of the individuals. Then extracted features from this graph to capture the motion and action contexts among the groups. The features are represented using the "bag-of-words" scheme. Finally, the method uses the learned Support Vector Machine (SVM) to classify a video segment into the six event categories. The implementation builds upon a mature multi-camera multi-target tracking system that recognizes the group-level events involving up to 20 individuals in real-time.

[Dongping Zhang et al, 2012][11] describe a fully automatic approach for detecting human groups and analysing their behavior in videos. The author has proposed clustering algorithm which involves human tracking based on particle trajectories of pedestrian which are clustered if they are similar in position, velocity, and direction by sociological studies and appears to be robust for group detection. This method describes that extracting accurate and continuous trajectories is the key point for successful detection of human groups. It was found that particle trajectories in the algorithm of particle motion estimation in foreground segmentation are much reliable. Finally, the author analysis the character of normal behavior in human crowd to detect anomalies of human crowd(s) presented in the scenes. Experimental result demonstrates that this method is effective in detecting anomalous events for public surveillance videos.

[Manoranjan Paul et al, 2013][12] has proposed a detection of human beings accurately in a visual surveillance system which is one of the major topics off vision research due to its wide range of applications. The first step of the detection process is to detect an object which is in motion. Object detection could be performed using background subtraction, optical flow and spatio-temporal filtering techniques. Once detected, a moving object could be classified as a human being using shape-based, texture-based or motion-based features. A comprehensive review with comparisons on available techniques for detecting human beings in surveillance videos is presented in this paper. The characteristics of few benchmark datasets as well as the future research directions on human detection have also been discussed.

[M.Sivarathinabala et al, 2013][13] introduces an intelligent analysis of single person activity to enrich the current video surveillance systems through an automatic identification of abnormal behavior of the person. The first step in this technique is to pre-process the captured video which contains the features of geometry information of the target. Second step is to analyze the images, which is required to classify the target and extract relevant information to analyze the motion of targets. Activity/ Behavior of the human are analyzed in the intelligent analysis module. If the abnormal behavior is found in the scene, the system automatically compresses the video using H.264 coding and sent to the mobile phone of the user. The mobile phone can decode them and play the surveillance video.

[MyoThida et al, 2013][14]This chapter presents a review and systematic comparison of the state of the art on crowd video analysis. The aim of this chapter is to propose a critical review of existing literature pertaining to the automatic analysis of complex and crowded scenes. The literature is divided into two broad categories: the macroscopic and the microscopic modelling approach. The effort is meant to provide a reference point for all computer vision practitioners currently working on crowd analysis. Later the author discusses the merits and weaknesses of various approaches for each topic and provides a recommendation on how existing methods can be improved.

[ShivaniArbat et al, 2014][15] has proposed a reflexive and efficacious replay detection method to detect event and classify them in order to generate highlights in soccer game. This method is based on logo detection and replay recognition. Firstly, this method detects some logo-transitions and extracts some logo-

samples from them during the video. Afterwards, the logo-template threshold value from these samples has been collected. Then, it uses this template threshold to detect the other logos in the same video. After all logos are detected, the video can be divided into segments by taking logos as boundaries. Thus these detected replay segments can be expended for further analysis of the video.

[Du Tran et al, 2014][16] has proposed a novel approach for detecting complex and dynamic events. The author has developed a Max-Path algorithm which results in global optimal solution. This global optimal solution improves the smoothness of the event and thus eliminates the false positives and alleviates missed due to occlusions and the low image quality. In addition, Max-Path's results in low complexity which makes it efficient to search for spatiotemporal paths in a large 5D space of spatiotemporal, scale, and shape. Due to its low complexity, Max-Path search has been used as an efficient method for fast inference and complex structured learning problem or applied to improve the speed performance. In conclusion, this paper strongly benefits to the computer vision literature a novel approach for video event detection and localization, especially for the complex and dynamic events.

[A.Niranjil Kumar et al, 2014][17] has proposed a method for abnormal crowd detection and tracking. This method consists of three major modules: Background Modeling, Blob Analysis and Crowd Detection and Tracking. Computer vision based crowd analysis algorithm is used for people tracking and crowd behavior analysis. First, foreground/background segmentation is performed and then, objects are tracked with a combination of blob matching and particle filtering techniques. The experimental results show that the proposed method achieves good accuracy.

[Pooja N S et al, 2015][18] has proposed a paper to detect suspicious group events in video streams. To address this problem, the author has proposed a method based on the image descriptor and classification approach. The method starts by optical flow computation followed by HOFO descriptor generation. This method can be evaluated using image descriptor with publically available dataset. HOFO descriptor is used as features to detect the suspicious moving queues .Support vector machine (SVM) is used for the classification approach. The method applied on several Benchmark Dataset to detect suspicious event. The proposed method gives the promising results with the good accuracy.

[TianWanget al, 2015][19] has proposed a novel algorithm to detect abnormal events in video streams. The algorithm is based on the histogram of the optical flow orientation descriptor and the classification method. The details of the histogram of the optical flow orientation descriptor are illustrated for describing movement information of the global video frame or foreground frame. By combining one-class support vector machine and kernel principal component analysis methods, the abnormal events in the current frame can be detected after a learning period characterizing normal behaviors. The difference abnormal detection results are analyzed and explained. The proposed detection method is tested on benchmark datasets, and then the experimental results show the effectiveness of the algorithm.

III. PROBLEM STATEMENT

Detecting human beings and recognizing event in a video scene of a surveillance system is attracting more attention due to its wide range of applications in abnormal event detection, human gait characterization, person counting in dense crowd, person identification, gender classification, fall detection for elderly people, etc. The proposed work is used to represent both individual and group events, hence it overcomes the drawback of existing system, and it is based on cognitive linguistic method which uses unsupervised learning method.

Traditional Video Surveillance systems based on human interaction are ineffective as the number of cameras exceeds the capability of human operators to monitor them. The task is time expensive for the human operator demanding full attention in real time to detect events and leading to low accuracy in detections due to fatigue, lack of observation or lapse in concentration. Another limitation entails the case when there are not human operators monitoring cameras in real time and the video is recorded to be analyzed afterward if any abnormality is reported (i.e. theft, murder, etc) to be used as an evidence. This work investigates systems capable to detect events automatically in video surveillance applications reducing or suppressing human interaction with the system and reporting alerts based on the events detected

IV. OBJECTIVES

The main objective of this research work is to build up a framework that recognizes the small human group

and to detect the event in the video. This framework is utilized for robotized little human gathering occasion discovery inside of social or open spot environment furthermore serves to recognize a fording wrongdoings, for example, Railway station, Traffic, collages, office etc.
The major challenges in small human group action recognition system are,

1. Human object tracking and detection in the video.
2. Small human group detection in the video.
3. Analyzing small human group actions and event detection.
4. In the background subtraction stage, to overcome the partial occlusion in the tracking and feature extraction.
5. The performance of the event detection depends on these previous stages and also on the choice offeature for action representation.

Finally, this research work would build up a complete framework for small human group tracking and event detection in the video.

V. PROPOSED METHODOLOGY

The proposed method contains various processing steps like raw video collection, preprocessing, people detection and tracking, group detection, feature extraction, unsupervised classification and event detection. The main problem in the current work is to propose a system for detecting and tracking a group of people to identify the event in the group.
The following figure shows the block diagram of the proposed work.

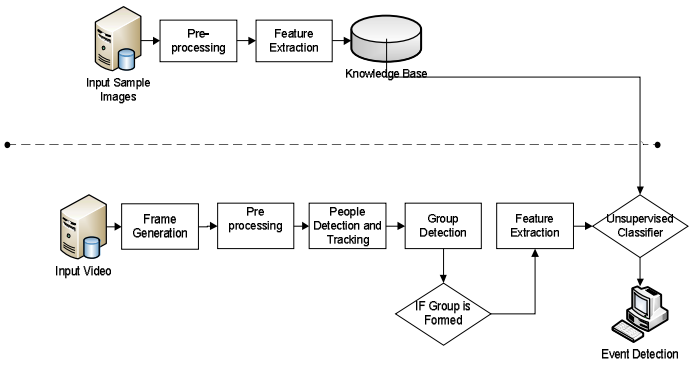

Figure 1: Block Diagram Of Proposed System

The proposed methodology consists of two phases i.e., training phase and testing phase. Initially, in the training phase main features are extracted from the event samples and Machine Learning Algorithm is trained using an extracted feature which creates the knowledge of different events.
In the testing phase, first the raw video is taken as an input; frames are generated and preprocessed to the input video. In the preprocessing stage, gray conversion and video denoising is performed to input video.Next, small group of people is detected and tracked using standard algorithm. If group is formed, features are extracted using features extraction algorithms from particular frame. Finally, event in particular frame is classified using unsupervised classifiers.

VI. EXPECTED OUTCOMES

Since video surveillance has received many attention over the last years and is a major research topic in computer vision. Detecting the event of human beings accurately in a visual surveillance system is crucial for diverse application areas including abnormal event detection, human gait characterization, congestion analysis, person identification, gender classification and fall detection for elderly people etc. The research aim is to build fully or semi-automation system for detecting and tracking the small group of people and also aims to detect the event in the group to assess new video surveillance techniques and apply them to the various applications.
The main expected outcome is showed below:

1. To automatically detect the salient event in the input video.
2. To obtain the starting and end time of the event and to mark the event in the video.

The future work includes extending the existing algorithms to develop more robust algorithms so that it should work on a dynamic background changes and overcomes static area human activity. By using the unsupervised classifier and feature extraction method, human objects are tracked and events are detected more accurately than the existing system and also it should work on large human crowd instead of low environment.

REFERENCES

[1] M. Andriluka, S. Roth, and B. Schiele. People-tracking-by-detection and people-detection-by-tracking. In IEEE Comp. Vision and Pattern Rec., 2008.

[2] J. Berclaz, F. Fleuret, and P. Fua. Robust people tracking with global trajectory optimization. In IEEE Comp. Vision and Pattern Rec., 2006.

[3] B. Leibe, K. Schindler, N. Cornelis, and L. V. Gool. Coupled object detection and tracking from static cameras and moving vehicles. IEEE T. Pattern Anal. and Machine Intell., 30(10):1683–1698, 2008.

[4] Y. Li, C. Huang, and R. Nevatia. Learning to associate: Hybridboosted multi-target tracker for crowded scene. In IEEE Comp. Vision and Pattern Rec., 2009.

[5] Nor SurayahaniSuriani, AiniHussain and MohdAsyrafZulkifley, "Sudden Event Recognition: A Survey", Sensors, Issue 3, PP 9966-9998, 2008.

[6] Weiyao Lin, Ming-Ting Sun, RadhaPoovendran and Zhengyou Zhang, "Group Event Detection for Video Surveillance", IEEE, PP 2830 – 2833, 2009.

[7] Joshua Candamo, Matthew Shreve and Dmitry B. Goldgof, "Understanding Transit Scenes: A Survey on Human Behavior-Recognition Algorithms", IEEE Transactions On Intelligent Transportation Systems, VOL. 11, NO. 1, MARCH 2010

[8] Sofia Zaidenberg , Bernard Boulay , Carolina Garate and Duc-PhuChau, "Group interaction and group tracking for video-surveillance in underground railway stations", "International Workshop on Behaviour Analysis and Video Understanding, 2011.

[9] HarisUddin Sharif, Aloke Kumar Saha, KaziShamsulArefin and Haidar Sharif, "Event Detection from Video Streams", ISSN, VOLUME 01, ISSUE 02, 2011.

[10] Yimeng Zhang, WeinaGe Ming-Ching Chang and Xiaoming Liu, "Group Context Learning for Event Recognition", Applications of Computer Vision (WACV), IEEE, PP 249 – 255, 2012

[11] Dongping Zhang, Feiyu Chen and Chao Tong, "Particle Motion-based Abnormal Events Detection in Group-Level Crowd", Journal of Convergence Information Technology, Volume 7, Number 14, 2012.

[12] Manoranjan Paul, Shah M E Haque and SubrataChakraborty, "Human detection in surveillance videos and its applications - a review", Journal on Advances in Signal Processing 2013

[13] M.Sivarathinabala and S.Abirami, "An Intelligent Video Surveillance Framework for Remote Monitoring", International Journal of Engineering Science and Innovative Technology (IJESIT), Volume 2, Issue 2, 2013.

[14] MyoThida, Yoke Leng Yong, Pau Climent-Pérez, How-lung Eng and Paolo Remagnino," A Literature Review on Video Analytics of Crowded Scenes", Intelligent Multimedia Surveillance, 2013

[15] ShivaniArbat, ShashiKumariSinha, Shikha, BeenaKhade, "Event Detection In Broadcast Soccer Video By Detecting Replays ", International Journal of Scientific & Technology Research, Volume 3, ISSUE 5, 2014.

[16] Du Tran, Junsong Yuan and David Forsyth, "Video Event Detection: From Subvolume Localization to Spatiotemporal Path Search", IEEE Transactions on Pattern Analysis and Machine Intelligence, VOL. 36, NO. 2, 2014.

[17] A.Niranjil Kumar and C.Sureshkumar, "Abnormal Crowd Detection and Tracking in Surveillance Video Sequences", International Journal of Advanced Research in Computer and Communication Engineering, Vol. 3, Issue 9, September 2014

[18] Pooja N S, "Suspicious Group Event Detection for Outdoor Environment", International Journal of Modern Trends in Engineering and Research, Volume 0X, Issue 0Y, 2015.

[19]TianWang and HichemSnoussi, "Detection of Abnormal Events via Optical Flow Feature Analysis", Sensors,Issue 5, PP 7156-7171, 2015.