

**TITLE OF THE RESEARCH WORK**

**Video Analytics in Surveillance Videos**

(*Font: Times New Roman, Size 18, Bold*)

Submitted by

Name : **Kushal Jagadeesha** Seat No. :

**Research Centre**

M.S.Ramaiah Institute of Technology, Bengaluru

|  |  |
| --- | --- |
| **Under the Guidance of** | |
| Guide | **Co-Guide** |
| Name: Dr.K.Indira | Name: Dr.Dinesh Ramegowda |
| Designation: Professor | Designation: |
| Affiliation:VTU | Affiliation: |

**05-June-2018**

**Department of Electronics and Communication**

**M.S. Ramaiah Institute of Technology- 560054**

Table of Contents

[INTRODUCTION: 3](#_Toc518286707)

[LITERATURE REVIEW: 4](#_Toc518286708)

[OBJECTIVE OF THE PROPOSED RESEARCH WORK 6](#_Toc518286709)

[STUDY AREA AND METHODOLOGY 6](#_Toc518286710)

[EXPECTED OUTCOME 7](#_Toc518286711)

[REFERENCES 7](#_Toc518286712)

# INTRODUCTION:

A lot of automated surveillance systems, like CCTV cameras, have been installed at various public places. These public places include walking pathways, traffic junctions, college campuses, school campuses, railway stations, airports etc. These cameras have been installed to prevent any undue incidents in these public places. These incidents can be: un-necessary congregation of people, fight between two/three people, chain snatching, accidents, vehicle/ vehicles moving in wrong direction, traffic violations smoking in non-smoking zone, terrorism prevention and mitigation etc.

In-order to capture these unwarranted incidents, practically, at-least one human operator should be monitoring one CCTV camera 24/7. Effectively, if there are one thousand cameras installed, at least one thousand human resources are required to monitor the cameras. Additionally, such incidents occur very rarely and are un-common. So, in-order to practically short-list all such incidents, the Governing agency required lot of human resources, as well as the human operators need to spend their whole time in viewing boring and un-interesting footages. Moreover, humans are prone to commit errors because of interruptions and exhaustion. It would be highly impossible for the human operator to analyze these traffic videos and come up with qualitative and quantitative insights that might help in actual cause.

Given this situation, it is impractical to find so many human operators, as many as video surveillance cameras. Also, it would be very stressful and laborious for the officer to continuously view, analyze and conclude meaningful insights from the surveillance cameras that mostly consists of similar repetitive and monotonous footages.

These needs and challenges motivate to bring up an automatic video analytics solution to provide meaningful insights from surveillance cameras. A significant part of video surveillance program is video analytics, computer algorithms used to automatically alert an officer/operator of a probably un-warranted incident. Video analytics is primarily related to the development of models and algorithms for crowded urbane scene analysis. Due to maturing analytic engines and the exponential increase in camera and server processing power, analytics can be used by many different kinds of users and in a variety of environments. Video analytics development is proceeding in three directions [6]:

1. visual event modeling and algorithmic studies to detect track and classify visual events, especially anomalies.
2. development of data fusion models and cooperative algorithms to leverage multiple observations of the same visual phenomenon and improve system performance
3. combining analytic and hardware designs to facilitate real-time video analytics, called edge processing, all the algorithms are implemented in the camera to reduce network bandwidth requirements.

Automatic abnormal event detection is a challenging task because of the constraints like camera resolution, occlusion in the scene, data availability, wide range of abnormalities, ambiguity in appearance, perspective distortion, etc.

An abnormal event is identified as an event which does not occur normally. These events can be classified as traffic jams, breaking of traffic rules, accidents, non-vehicular (animals, cattle) objects in the middle of the road, vehicle break-down, chain snatching, theft, etc. These events are un-common events in any traffic and road condition and do not occur frequently. The proposed work presents an approach to identify, classify and localize these abnormal events from the real-time CCTV footage of traffic on Indian roads.

# LITERATURE REVIEW:

This chapter is sub-divided into three sections: Section 1 consists of researches carried out for event identification and classification based on visual modeling. Section 2 consists of research activities conducted on fusion of visual information from different camera sources, called as multi-camera video analytics. Section 3 consists of literature work performed to combine signal level and semantic level processing to enhance overall system performance.

**Section 1: Literature survey on algorithms of event detection based on visual modeling**

Based on the density of objects in the video frames, different approaches have been developed to detect anomalous events in the video footage. Tracking based approaches are used when the density of number of objects are less in the video frames, whereas object motion based approaches are used in the case of high object density video footages.

**Tracking based approaches:**

In [5], [Shifu Zhou](https://www.semanticscholar.org/author/Shifu-Zhou/2527043) and et.al., uses KLT feature tracker to obtain trajectories, and these trajectories are grouped to form the motion patter of the crowd. Later, the trajectories are modeled using Multi-Observer Hidden Markov Model (MOHMM) to determine of the frame is normal or abnormal. In the journal “Abnormal Behavior Detection via Sparse Reconstruction Analysis of Trajectory”[6], B-spline curves are used for motion trajectories of objects via fixed length vectors. Then the vectors are classified as normal or abnormal using sparse reconstruction analysis.in which the classifier is constructed with sparse linear coefficients by computing L1-norm minimization.

**Object motion based approaches:**

In [1], spatio-temporal features, based on optical flow information that are used to model general concepts, such as orientation, velocity, and entropy, are used to detect anomalous events in video footages with crowd. Normal patterns in videos are trained. During test, events that significantly differ from the normal patterns are classified as anomalous. Md. Haidar Sharif in [2] has proposed an approach using eigen values to detect the flows and events in a crowd. Spatio-temporal features of two consecutive frames are extracted to retain appearance and velocity information. Later polynomial function of different degrees are fitted by analyzing Zero flow, single flow and multiple flows and interesting (anomalous) events are detected using frame basis. In the paper, Abnormal Event Detection in Video Using Motion and Appearance Information [3], the video sequence is divided into spatio-temporal non-overlapping local patches. Later for each patch motion features like velocity and acceleration of optical flow and appearance features like texture and gradients are extracted. These features are trained using minimum distance classifier. Classification based on local features is performed to avoid perspective distortion. Yong Shean Chong et.al. in [4], makes use of deep nets to efficiently detect anomalies in crowded scenes. The input video is subjected to spatial encoding using convolution filters and the the temporal features are detected using temporal encoders namely convLSTM (Long Short Term Memory) nets. Later each frame is classified as normal or anomalous using thresholding method.

**Section 2: Literature survey on multi-camera video analytics**

Multi-camera video analytics relates to the field of fusion of signal data (video frames) from different sources and developing co-operative algorithms on fused data. Zhong Zhang et.al., in their work in [7] describes a system that fuses tracking information from multiple cameras. They present a method to calibrate the multi-camera system by using the visual information in the site-map. Different steps like data sharing, system design and cross calibration techniques are proposed for multi-view camera surveillance. (Zhu, Shao and Lin, 2013) in their paper [8] present a local segment similarity voting scheme to solve the multi-view action recognition problem. A random forest classifier is used to map the local segments to their corresponding prediction histograms. The results are base-lined using multi-view IXMAS dataset. A fusion based multiple hypothesis tracking algorithm is presented by (Ying, Zhang and Xu, 2014) in the paper [9]. Here appearance features like local motion pattern and repulsion—inertial model for multiple object tracking is combined to describe objects. A likelihood calculation framework is proposed to incorporate the similarities of appearance, dynamic process and local motion pattern.

**Section 3: Literature survey on articles used to combine signal level and sematic level processing**

Articles and papers in this area of research aim to reduce the “sematic gap”. In any imaging sensor, the content of interest is defined in very abstract terms related to how humans interpret video imagery, but the data is defined in very physical terms related to the imaging device. This difference in representation of data is termed as sematic gaps. The algorithms proposed in this area attempts to narrow down the sematic gap by moving the algorithms to physical data domain present in imaging sensory devices.

# OBJECTIVE OF THE PROPOSED RESEARCH WORK

The aim of this research work is to identify and alarm abnormal incidents (or events), like chain snatching, accidents, vehicle breakdown, and non-vehicular objects that occur on Indian roads. Later, these events are reported to the operator in real-time so that necessary action can be discharged on time.

The objectives of the research work are listed below:

1. To survey existing literature for various state-of-the-art anomalous detection algorithms and compare them with their understanding of their advantages & limitations.
2. To collect suitable video data for the development of algorithm.
3. To design and implement suitable computer vision and machine learning algorithm detecting useful insights from the video frames
4. To test the algorithm on the benchmarking datasets
5. To compare the result and performance of the algorithm with other existing methods.
6. To develop a software prototype for integration of the developed algorithm with actual surveillance systems.

# STUDY AREA AND METHODOLOGY

This section describes the methods and methodology used for the research work to achieve the proposed objectives.

Literature review of the existing video analytic algorithms will be carried out by referring reviewed journals, books, manuals and related documents. Based on the literature survey, video dataset for development of the system will be procured. The video dataset will be obtained by using

* real-time traffic videos(from Bangalore Traffic Police) OR
* available datasets on the internet or specific dataset from a colleges/university

A suitable video processing workflow (proof of concept algorithm) will be designed for the selected video dataset. The processing workflow will be based on the inputs from the literature and the dataset used for development and benchmarking. Proof of concept algorithms will be developed using existing software libraries like OpenCV, Python, CNN, Scikit. Processing algorithms will be developed in C/C++ and/or Python.

The algorithm will be constantly tested against the benchmarking datasets for result tabulation and comparison against the existing methods. Once the algorithm is fixed, the parameters of the algorithm will be fine-tuned to improve the results. After final modifications, the developed solution will be deployed on one of the production servers for analyzing the actual behavior.

# EXPECTED OUTCOME

The current research work is being pursued in order to reduce the manual load and human error on the Government body to maintain the traffic sense. The research work after realization is aimed to assist Government in the maintenance of the road safety principles with minimal cost. This will assist the Government body to reduce crime, prevent traffic congestion that will eventually result in better living for the citizens.

# REFERENCES

1. Colque, R., Caetano, C., de Andrade, M. and Schwartz, W. (2017). Histograms of Optical Flow Orientation and Magnitude and Entropy to Detect Anomalous Events in Videos. *IEEE Transactions on Circuits and Systems for Video Technology*, 27(3), pp.673-682.
2. Menejes Palomino N., Cámara Chávez G. (2018) Abnormal Event Detection in Video Using Motion and Appearance Information. In: Mendoza M., Velastín S. (eds) Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications. CIARP 2017. Lecture Notes in Computer Science, vol 10657. Springer, Cham
3. Chong Y.S., Tay Y.H. (2017) Abnormal Event Detection in Videos Using Spatiotemporal Autoencoder. In: Cong F., Leung A., Wei Q. (eds) Advances in Neural Networks - ISNN 2017. ISNN 2017. Lecture Notes in Computer Science, vol 10262. Springer, Cham
4. Zhou, S., Shen, W., Zeng, D., Zhang, Z.: Unusual event detection in  
   crowded scenes by trajectory analysis. In: ICASSP, IEEE International  
   Conference on Acoustics, Speech and Signal Processing - Proceedings. vol.  
   2015-Augus, pp. 1300–1304. Institute of Electrical and Electronics Engineers Inc. (Aug 2015)
5. Li, C., Han, Z., Ye, Q., Jiao, J.: Abnormal behavior detection via sparse  
   reconstruction analysis of trajectory. In: Proceedings - 6th International  
   Conference on Image and Graphics, ICIG 2011. pp. 807–810 (2011)
6. Regazzoni, C., Cavallaro, A., Wu, Y., Konrad, J. and Hampapur, A. (2010). Video Analytics for Surveillance: Theory and Practice [From the Guest Editors. *IEEE Signal Processing Magazine*, 27(5), pp.16-17.
7. Zhong Zhang, Andrew Scanlon, Weihong Yin, Li Yu, Péter L. Venetianer. Video Surveillance using a Multi-Camera Tracking and Fusion System. Workshop on Multi-camera and Multi-modal Sensor Fusion Algorithms and Applications - M2SFA2 2008, Oct 2008, Marseille, France. 2008. <inria00326754
8. Zhu, F., Shao, L. and Lin, M. (2013). Multi-view action recognition using local similarity random forests and sensor fusion. Pattern Recognition Letters, 34(1), pp.20-24.
9. Ying, L., Zhang, T. and Xu, C. (2014). Multi-object tracking via MHT with multiple information fusion in surveillance video. *Multimedia Systems*, 21(3), pp.313-326.

|  |  |  |
| --- | --- | --- |
| Candidate’s Signature | Co-Guide’s Signature | Guide’s Signature |