

LEARNING ALGORITHMS FOR EVENTS DETECTION FROM TRAFFIC SURVEILLANCE VIDEOS

*Thesis submitted in the partial fulfillment
of the requirements for the Degree of
Doctor of Philosophy (Ph.D.)*

by

Anima Pramanik

(17IM91P01)

Under the supervision of

Prof. J. Maiti

Department of Industrial and Systems Engineering, IIT Kharagpur

and

Prof. Sankar K. Pal

Center for Soft Computing Research, ISI Kolkata



Department of Industrial & Systems Engineering

Indian Institute of Technology Kharagpur

June, 2024

©Anima Pramanik. All Rights Reserved

Abstract

Name of the student: **Anima Pramanik**

Roll No: **17IM91P01**

Degree for which submitted: **Doctor of Philosophy (Ph.D.)**

Department: **Department of Industrial & Systems Engineering**

Thesis title: **LEARNING ALGORITHMS FOR EVENTS DETECTION
FROM TRAFFIC SURVEILLANCE VIDEOS**

Thesis supervisor(s): **Prof. J. Maiti, and Prof. Sankar K. Pal**

Month and year of thesis submission: **June, 2024**

The usefulness of computer vision (CV) in detecting events has been established in diverse domains of application (i.e., military, security, life, and medical). However, its exploration in detecting real-time traffic events in road transportation domain is limited. This necessitates the investigation of CV algorithms in traffic safety domain. This research aims at the development of various learning algorithms for detecting the real-time traffic anomalous events (i.e., pre-events, current-events, and post-events) and taking the on-time corrective actions in order to enhance the traffic safety. The thesis addresses four objectives as: to develop various learning algorithms for (i) both foregrounds and multi-class objects detection and tracking, (ii) vehicle-related traffic pre-events detection, (iii) traffic current-events detection, and (iv) person-linked traffic anomalies detection.

The first objective addresses the use of granular computing in foregrounds and multi-class objects detection and tracking. Based on the object of interest used for traffic anomaly detection, both unsupervised and supervised algorithms are developed for foregrounds and multi-class objects detection and tracking, respectively. Here, foregrounds detection and tracking (based on unsupervised learning) are done using the spatio-neighborhood temporal granulation, and distance- and area-based data association, respectively. On the other hand, two supervised algorithms, namely granulated RCNN (G-RCNN) and multi-class

deep SORT (MCD-SORT) are developed for multi-class objects detection and tracking, respectively. Efficacy of both of the unsupervised and supervised algorithms for foregrounds and multi-class detection and tracking are demonstrated over various benchmark data (PASCAL VOC, Urban Tracker, YouTube8M, and MOT15) and one real-life traffic data, namely Video Analytics Road 2019 (VAR19). Moreover, superiority of the developed algorithms over some recent state-of-the-art is also proved.

The second objective comprises of two sub-objectives: (i) development of a conceptual framework for video surveillance system in enhancing the traffic safety management and (ii) development of various unsupervised algorithms for the detection of vehicle-related traffic pre-events, such as speed violation, one-way/oncoming traffic, overtaking, illegal parking, roundabout violation, and wrong drop off location of passengers. The developed unsupervised algorithms for the detection of aforesaid traffic pre-events are modeled using various rules based on some defined spatial and temporal features. The efficacy of aforesaid algorithms is demonstrated over 80 real-time traffic videos containing both normal and abnormal (i.e., traffic pre-events) scenarios. A comparative study is also conducted to prove the superiority of developed algorithms for aforesaid traffic pre-events detection.

The third objective addresses the use of spatio-temporal features (STFs), rough fuzzy granulation, and Z-numbers for handling the uncertainty issue arising in between two traffic current-events, namely collision and near-miss. Various salient STFs are defined corresponding to each detected foreground region and used for obtaining the crisp granules over the video frames. These granules represent the approximate anomaly-prone regions. As there is no definite class (collision, near-miss, and/or normal scenario) assigned to these regions, fuzziness or uncertainty may inherently present in these granules, thereby resulting in fuzzy information granules. The said uncertainty has been modeled by defining two rough fuzzy granules (RFGs) (one is for collision event and another is for near-miss event) along with their roughness scores. For each RFG, Z-number is computed based on the membership value of its roughness score to ensure a higher degree of reliability in the detection of anomaly/event-class. All these characteristics constitute an unsupervised algorithm, namely Z-numbers-based spatio-temporal rough fuzzy granulation (Z-STRFG). The efficacy of Z-STRFG has been demonstrated over 130 real-time traffic videos containing collisions, near-misses, and normal traffics. The superiority of Z-STRFG over some

state-of-the-art is also proved through extensive experiments.

In the fourth objective, two semi-supervised algorithms, namely Z-numbers-based spatio-temporal granulation (Z-STG) and deep transfer learning (Z-DTL) are developed for the detection of four types of person-linked anomalies (PLAs) which are categorized into three classes: (i) pre-events (i.e., causes: (pedestrian crossing violation (PCV) and pedestrian proximity to moving vehicle (PPMV)) of current-events), (ii) post-event (i.e., consequence (pedestrian/driver fall (PF)) of current-events), and (iii) personal protective equipment violation (PPE-V) (i.e., violation of using the hard-hat which can be responsible for mitigating the impact of consequence). Z-STG addresses the issue of detecting the occluded or blurred pedestrian crossing as well as its violation. Z-STG consists of two steps. In the first step, various spatio-color and temporal features are defined and used for extracting the pedestrian crossings in terms of crisp granules. Whereas, in the second step, object-level features are fused with the detected pedestrian crossing features for generating various rules in order to detect the PCV and PPMV. On the other hand, Z-DTL for fall and PPE-V detection deals with the issues of low illumination, occlusion, and varying behavior of the traffic events from video to video. The Z-DTL incorporates a new deep network, namely G-RCNN3 with a two-dimensional bidirectional long short-term memory (Bi-LSTM) using transfer learning. G-RCNN3 is an advanced version of G-RCNN. The output layer (FC3) of G-RCNN3 contains three dimensional feature maps. Each dimension represents the feature map corresponding to either person, fall-event, or hard-hat (PPE). The 1st- and 3rd-dimensional feature maps are used for re-training of G-RCNN3. Whereas, the 1st- and 2nd-dimensional feature maps are used for training and finetuning of Bi-LSTM. In both Z-STG and Z-DTL, Z-number is computed to ensure the reliability of the detected PLA-class. The superiority of Z-STG and Z-DTL over some state-of-the-art is also established through extensive experiments.

The thesis has made contributions from both theoretical development and practical implication points of view. From the theoretical development point of view, the thesis contributes as follows: (i) an unsupervised algorithm is developed for foregrounds detection and tracking using the commonality between spatio-color contour-based neighborhood and temporal granules, and distance- and area-based data association, respectively, (ii) a

supervised detector, namely G-RCNN is developed for multi-class objects detection by incorporating the concept of spatio-temporal granulation within deep CNN, (iii) a supervised tracker, namely MCD-SORT is developed for multi-class tracking by restricting the data association between detected object(s) and track-lets of same categories, (iv) various unsupervised algorithms are developed for detecting vehicle-related real-time traffic pre-events, (v) an unsupervised algorithm, namely Z-STRFG is developed for real-time traffic current-events detection by handling the uncertainty issue arising between collisions, near-misses, and normal scenarios, (vi) a semi-supervised algorithm, namely Z-STG is developed for detecting person-linked traffic pre-events (i.e., PCV and PPMV) using spatio-temporal granulation, rule base system, and Z-numbers, (vii) another semi-supervised algorithm, namely Z-DTL is developed for detecting two person-linked anomalies which are consequence (i.e., fall) of current-events and violation of using the mitigating factor (i.e., PPE-V), (viii) the research demonstrates a way of incorporating the recently emerged granular computing in deep learning network for on-line object tracking and classification directly from the input videos, and (ix) the effectiveness of granulation in detecting various traffic pre-events and current-events, is demonstrated in this research. On the other hand, from practical implication point of view, the thesis contributes as follows: (i) developed algorithms can automatically detect anomalous events from 24×7 live stream traffic videos, (ii) after detection, a video summary of the anomalous event can be generated automatically and used for future training, (iii) reduction of human effort in order to search the traffic anomalous events from huge amount of videos, and (iv) after detection, an alarm signal can be sent automatically to the traffic control room for taking on-time corrective actions in order to either stop re-occurring the current-events (i.e., collision and near-miss) or mitigate the impact of consequence for same current-events, thereby enhancing the traffic safety.

From the analyses presented in the thesis, some key findings are obtained. They are: (i) commonality between spatio-color contour-based neighborhood and temporal granules represents the foreground regions in videos more effectively as compared to spatio-color and temporal granules, (ii) region of interest (RoI)-pixels detected by the developed G-RCNN are more representative of object classes as compared to features generated by RCNN and its variants, (iii) comparative results signify the relevance of spatio-temporal information

over spatial information for the task of video object localization, (iv) G-RCNN is found to be the best among state-of-the-art video object detectors with mean average precision (mAP) of 80.6% on the generated data, namely VAR19, (v) among all the combinations of supervised detectors and trackers considered in the experiment, G-RCNN + MCD-SORT is found to be the best, (vi) algorithms for the detection of vehicle-related traffic pre-events and scene-based traffic current-events are modeled using foreground-level features; therefore, foregrounds detection and tracking is the prior task in such cases, (vii) person-linked traffic anomalies are related to a specific object class (i.e., person); therefore, in such cases, multi-class objects detection and tracking is the prior task, (viii) the developed conceptual framework for traffic safety management is generic in nature and can be used in any road context to detect vehicle-related traffic pre-events, (ix) for traffic current-events detection, the use of RFGs and Z-numbers in handling the uncertainty issue is more effective than using only salient cost functions, (x) the use of Z-numbers for each RFG in ensuring the reliability of the detected current-event (collision and near-miss) is unique, (xi) Z-STRFG is found to be the best among state-of-the-art video summarization and deep CNN-based traffic anomaly detection techniques with accuracy of 96.8% over a new traffic data, namely Anomaly 2020 (An20), (xii) in Z-STG, the incorporation of Z-numbers with spatio-temporal granulation is found to be very effective in detecting blurred/occluded pedestrian crossing and its violation (i.e., PCV and PPMV), (xiii) in Z-DTL, the fusion of G-RCNN3 with Bi-LSTM using transfer learning and Z-numbers shows an adequate performance in PF and PPE-V detection, and finally, (xiv) experimental results prove the effectiveness of Z-STG and Z-DTL over some state-of-the-art for person-linked traffic anomaly detection.

Keywords: Traffic safety management, Computer vision, Multi-object detection and tracking, Vehicle-related traffic pre-events detection, Traffic current-events detection, Person-linked traffic anomaly detection, Granulation, Rough fuzzy inference system, Uncertainty handling, and Rule-base system.