International Journal of Advanced Science and Technology
Vol. 29, No. 03, (2020), pp. 5386- 5392

Detection of Violence in Videos using Hybrid Machine Learning Techniques

Sumitra Kisan [1*], Baishnabi Sahu [2], Abhijit Jena [3],
[4]Sachi Nandan Mohanty

[1] *Department of Computer science engineering and Information technology, Veer Surendra Sai University of Technology, Burla, Sambalpur, India*
[2] *Department of Computer science engineering and Information technology, Veer Surendra Sai University of Technology, Burla, Sambalpur, India*
[3]*Department of Computer science engineering and Information technology, Veer Surendra Sai University of Technology, Burla, Sambalpur, India*
[4]*Department of Computer Science & engineering, ICFAI Foundation for Higher Education, Hyderabad*
[1]*skisan_cse@vssut.ac.in* [2]*baishnabisahu.98@gmail.com*
[3]*abhijitjena12@gmail.com* [4]*sachinandan09@gmail.com*

Abstract

Surveillance videos are able to capture a variety of real-world anomalies. All kinds of activities like walking, talking, riding a vehicle are considered normal behaviour. But any sort of activity that does not adhere by the definition of normal pattern is considered as abnormal or anomalous. It can be any irregular behaviour like yelling, theft in public, breaking into a house, accident and many more. In this work, the intention is to develop a model that will take videos captured from CCTV cameras on streets, ATMs, police station, roads, hospitals, railway stations, etc., as input and classify them as normal or abnormal (contains anomaly at some point of time). A video which contains anomaly is labelled as positive and normal videos are labelled negative. Change in motion vector, is taken as the key for classification. This research, aims to find several spatial and temporal feature descriptors like Histogram of Oriented Gradients, Histogram of Optical Flow and then use the extracted features in classifiers such as CNNs for HOG and HOF. The aim the proposed method is to return positive hits for portions of video sequences that contain violent content.

Keywords: *Anomalous activity detection, convolutional neural network, optical flow, Histogram of Oriented Gradients, Histogram of Optical Flow*

1. Introduction

Computer vision and image processing are one of the prime domains of artificial intelligence. Over the years, many advancements have been made in these fields to help computers “see” and “recognize” images as humans do. Image processing is basically a method of performing certain operations on an image so as to obtain an enhanced image or extract useful information from it. Simply, it is a type of signal processing where input is an image and output may be an image or characteristics associated with the image.

Before beginning with image and video classification and diving deep into its techniques, the first thing to understand is what an image is exactly. For computers, an image is a large array of numbers. These are composed of a finite number of elements called pixels, each of which has a particular value between 0 and 255. Technically, an image is a two-dimensional function f(x, y) where x and y are the plane coordinates and amplitude for any pair of (x, y) gives the intensity of the image at that point. In image classification, an image is assigned a label from a set of categories. However, this task

ISSN: 2005-4238 IJAST

faces numerous challenges like variation in illumination, viewpoint, background clutter, deformation, etc.

A major but yet unexplored application of image and video processing is the automated understanding of events in public places through surveillance cameras [14,15,16]. Surveillance in public areas using CCTV is common in many areas around the world. Closed-circuit television (CCTV), video cameras are used to transmit a visual information to a particular place, on a given set of monitors through secure transmission methods. Recent studies have found that CCTV installations helped in reducing crime by 24-28% in public streets and urban railway stations. It also inferred that CCTV could decrease unruly crowd behaviour in football stadiums and theft in supermarkets/ merchant stores. However, there was no evidence of CCTV having desirable effects in facilities where constant monitoring wasn't done or wasn't practical for a human point of view.

With all these advancements happening around, automated visual surveillance has been one of the most sought after research areas in computer vision. And is also one of the most challenging tasks. Hence a computer vision algorithm has to be developed that is capable of detecting violence automatically, so that alert systems can be triggered in case of unmonitored systems, and quick action or response can be taken. Surveillance cameras are installed almost everywhere but there are not really useful unless monitored. In this research work, an approach to detect violence in real-time videos is proposed.

This paper thus aims at creating a method for the real-time detection and alarming the personnel about the start or possibility of occurrence of such an event. The work then aims at extending this concept to other forms of violence like violent crowd behaviour, mob lynching, riots, etc. Objective is to create an algorithm that automatically detects presence of firearms or violent crowd behaviour in a real-time surveillance footage and to alert the administrator of the detection of such an event.

The whole paper has been organized in the following manner. In section 2, some basic concept of video processing that has been used in the work are discussed. In the next section, i.e. section 3, some of the existing methodologies that have been implemented and analysed are reviewed. Section 4 & 5 present the actual problem statement and our proposed approach, respectively.

2. Background Study

Video is a visual media product featuring moving images, with or without audio, that is recorded and digitally stored.

Frames per second: Casing rate, the quantity of still pictures per unit of time of the video, ranges from six or eight edges for every second (outline/s) for old mechanical cameras to at least 120 edges for every second for new expert cameras. Internationally acknowledged spilling/broadcasting models determine 25 and 29.97 edge/s.

Aspect ratio: Aspect ratio portrays the corresponding connection between the width and stature of video screens and video picture components. All well-known video designs are rectangular, thus can be portrayed by a proportion height and width.

Pre-processing: Pre-processing involves extracting candidate frames from the video sequence to be used for training and testing. Image enhancement is done to increase contrast and bring out details in the frames. It acts as a pre phase to feature extraction.

Feature Extraction: Feature extraction is the stage where necessary feature vectors from the frames are extracted to give spatial information in the frames, like edge detection, background subtraction, etc. Features are also extracted from consecutive frames to provide temporal information regarding movement of actors within the frames.

Classification techniques: There are numerous techniques to classify images and videos. Each one has its own mathematical way of computing results. Some of the well-known techniques are neural networks, support vector machine, k-nearest neighbours, logistic regression, radial basis function network, etc. At times, the training data has to be

ISSN: 2005-4238 IJAST

modified for better training and results. Most commonly used is dimensionality reduction technique Principal Component Analysis (PCA) and LDA.

3. Literature Review

Research on complex event detection and recognition has been an active topic in the field of computer vision in the recent years. Studies on this topic can be grouped in terms of the methods used, modelling techniques and domain taken into consideration and event types among others.

The lowest level jobs in the detection of complex events are pixel-level operations. Some studies (eg. Tian et al, 2010) use pixel-based operations to solve event detection problems. In these studies, user understandable event representation is not given much importance, so they lack high-level inference capability.

Some other studies generate reliable input data for high-level inference, including background subtraction, spatial and temporal domain object-tracking. Some studies use rule based methods, which are based on specific set of rules laid down by a domain expert with regard to how a violent situation can unfold in a sequence of frames. The disadvantage of these methods is that they cannot handle uncertainty due to their non-probabilistic approach to deal such events.

Surveillance Video Analysis System (Kardas, 2018) [7] gives an Interval-Based Spatio-Temporal Method. A set of feature models named Threshold Model, which reflects the spatio-temporal motion analysis of an event, is kept as the first model. As the second model, Bag of Actions (BoA) model is used in order to reduce the search space in the detection phase. Markov Logic Network (MLN) model, which provides understandable and manageable logic predicates for users, is kept as the third model.

Waqas Sultani et al [1] described a novel framework for detection of anomalies in videos using deep multiple instance ranking. Multiple instance ranking is a type of supervised learning method. Instead of feeding a set of instances which are individually labelled, the model is fed a set of labelled bags, each containing many instances. For example, in a simple case of multiple-instance binary classification, the bag is labelled negative if all the instances in it are negative. On the other hand, bag is labelled positive if all the instances in it are positive. Here, anomaly detection is considered as a regression problem rather than classification.

Vijay Mahadevan et al [2] developed a novel framework to detect anomalous events in crowded scenes. They used the concept that anomalies are events of low-probability with respect to a model acquainted with normal crowd behaviour. There are two types of normalcy spatial and temporal. Temporal normalcy reflects that normal events are recurring in nature. They keep happening over and over again in time, unlike abnormal events that are rare. A particular event being labelled as normal or abnormal depends on the situation. For example, an ambulance moving at a speed of 60 kmph in a stretch of highway is pretty normal. But the same ambulance moving at 60 kmph in a highly congested highway is dangerous. Hence, abnormal.

4. Problem Statement

Surveillance cameras are set up almost everywhere like ATMs, shops, hospitals, restaurants, etc. for security purposes. They are meant to monitor and keep track of all sort of activities happening around, can be both desirable and undesirable. Any activity which does not conform to the “human definition of normal” can be marked as abnormal or anomalous or undesirable. These include abusing animals, harassing people, fighting, robbery, explosion, accident and assault. However, monitoring capability of human in-charge of it has not kept pace. As a result of which there is a strong deficiency in the utilization of these cameras. Their real potential goes wasted. Moreover it is not practically possible for a person to keep an eye on the camera footage all the time.

ISSN: 2005-4238 IJAST

Algorithms currently being used are specific to domain and events. Many methods require human intervention. Generally, event detection is a two-step process consisting of low and high level jobs. [8] At the low level, the spatio-temporal features are extracted, which provide positions of actors in space and time. At high level, inference and prediction of events is done. The major problems with the existing approaches is the ability to deal with uncertainty as the data coming from low levels may not always be reliable due to occlusions or movement of camera. This must be compensated at the high level to avoid false event detection.

Another major problem is the problem of performance. With the increase in the volume of videos, the learning process might not be subject to strict performance restrictions, but the inference process needs to be fast enough to generate real time results.

The specifics of domain is also a major problem in the existing applications where event models are defined by domain experts, which is not feasible in all situations.

Therefore, developing an intelligent computer vision algorithm to automatically detect anomaly in videos is a pressing need. The goal of this system will be to timely signal or alert about the occurrence of any activity that deviates from normal patterns. If such an algorithm is installed into cameras and run in real time, all sort of anomalous or violent events can be detected right away and remedies can be taken as soon as possible. This can be advanced by adding some sort of alarming system that can blare alarms to signal the occurrence of an anomalous event.

5. Proposed Method

Anomaly detection has been modelled using various different algorithm and all of them have exceled in their own ways. Humans have the ability to sense occurrence of events around them as soon as they acknowledge some sudden change in position or motion of objects. Take for example, a person on bicycle is crossing a road and a speeding car hits him. The person on bicycle will definitely be thrown away in some other direction and deviated from the path he was supposed to take. This brings about sudden change in motion of the bicycle in x and y directions. The crux of our proposed approach lies in tracking this change in motion vectors of all entities in the video[4]. This algorithm is based on the assumption that spatial information does not change in 1 second. And there has to be some significant change in motion after some sort of crime or any other unusual activity takes place. Having understood the need of the problem, we proceed to an algorithm that can automatically detect occurrence of unusual or anomalous activities in videos. The algorithm involves below mentioned steps.

1. Computing change in motion vectors in videos using optical flow.
2. Recognizing activity and number of people involved.
3. Train a two-stream convolutional neural network.
4. Classify the videos as being normal or abnormal.

The algorithm aims to address certain problems in the existing methodologies like differences created due to camera motion by subtracting mean vector from displacement fields obtained from optical flow.

Computing change in motion vector using optical flow

At first, motion of all the entities in the video has to be tracked. This is done using optical flow[3]. Optical flow is the pattern of apparent motion of image objects between two consecutive frames caused by movement of the object. This is a 2D vector field where each one is a displacement vector showing the movement of points from first frame to second. It is based on the following assumptions-

1. The pixel intensities of an abject (in the image) do not change between two consecutive frames.
2. Neighboring pixels have similar motion.

ISSN: 2005-4238 IJAST

Let us consider a pixel (x, y). At time t, let the intensity of that pixel be I(x, y, t). In the immediate next frame taken after dt time, the pixel moves by distance (dx, dy). Since our assumption says pixel intensities do not change in between frames, we can say

$$I(x, y, t) = I(x + \Delta x, y + \Delta y, t + \Delta t)$$

Taking Taylor series expansion of the right-hand side of the above equation, removing common terms and dividing both sides of the equation by dt we get,

$$I_x u + I_y v + I_t = 0 \qquad (1)$$

Where,

$$I_x = \frac{\partial I}{\partial x} \; ; \; I_y = \frac{\partial I}{\partial y} \quad \text{and} \quad u = \frac{\partial x}{\partial t}; v = \frac{\partial y}{\partial t}$$

The above equation (numbered as 1) is the Optical flow equation, where I_x , I_y and I_t are partial derivatives of the intensity function with respect to x, y and t. And u and v refer to velocity components along x and y directions.

Optical flow can be computed in two ways – sparse optical flow and dense optical flow.

- Sparse optical flow is a method where optical flow is determined only for some specific points (or pixels) in the frame and not for all the pixels. Those specific points are generally corners present in the image. With this algorithm we get the coordinates of those specific entities in consecutive frames. This implementation is assumed to be faster but less efficient. Otherwise known as Lucas-Kanade method.
- Dense optical flow is a method where optical flow is computed for each and every pixel in the frame. Change in motion is tracked for all pixels. With this algorithm, we get the magnitude of change in motion along with its direction. Magnitude corresponds to the value plane. Direction corresponds to the hue value. This method is also known as Farneback's algorithm.

A video is actually a sequence of frames at certain frames per second. Let's say frame rate is set to 30 fps i.e. 30 frames per second. Temporal information is taken using optical flow for every consecutive frame. This is cumulative. Change in motion computed for every frame is added up. For 30 fps frame rate, the final optical flow result will be something like,

OF(1,2)+ OF(2,3)+ OF(3,4)+ OF(4,5)+ …..+ OF (28, 29) + OF(29,30)

Where OF(x, y) means optical flow between frame number x and y.

Recognizing activity and number of people involved

The first layer acts as a checkpoint to detect normal walking/running of a certain n number of people in order to save unnecessary computation in the second stage, which is largely compute intensive. This stage acts as a measure to reduce the number of false alarms and to wake up the second layer only when necessary.

Activities in videos can be recognized by deep learning techniques. If any activity involving a quantity of people/actors >n, then the second stage is triggered to check if it leads to an anomalous behaviour.

This stage is expected to reduce the amount of computation required by the entire system and hence save time in real life implementations.

ISSN: 2005-4238 IJAST

Training a two Stream Convolutional Network

Videos can be broken down to spatial and temporal components. Spatial components consist of the information in a single frame, like the nature of a scene or the actors/objects present in the frame. The temporal components give information about the motion across several frames, conveying the movement of camera and the objects in the video.

In order to capture both the components of a video, a two stream Convolutional Network is used [9,10]. Spatial stream Convolutional Network is used on singular video frames, efficiently performing activity recognition from static images. Optical flow Convolutional Network uses the stacked optical flow displacement fields between several consecutive frames as the input.

Using the processes mentioned above, we have extracted the necessary motion descriptors for video sequences and trained the CNN using these inputs.

With this information, we'll come up with a binary classification indicating occurrence/non-occurrence of violent/abnormal activities.

The proposed method primarily focusses on reducing the complexity of computation that is generally seen in other pre-existing methods. Since the algorithm has to be used in systems that monitor CCTVs, where the computing environment may not be capable of running very complex algorithms, the focus is given on reducing the weight of heavy computation along with the number of possible false alarms as well.

5. Conclusion

Having understood the need to develop an automated system, we have proposed a deep learning approach to automatically detect unusual activities in public areas, using convolutional neural networks. Given the existing methodologies in this field, we believe our method will definitely prove to be efficient. Although there are a number of methods under development and in use they rely on heavy processing in highly advanced environments, we aim to develop an algorithm that can produce efficient results in a normal computing environment.

Acknowledgments

We would like to thank our university, Veer Surendra Sai University of Technology for helping us and providing us with necessary materials to complete our research work.

References

1. Waqas Sultani, Chen Chen, Mubarak Shah. Real-world anomaly detection in surveillance videos. Conference: IEEE/CVF conference on computer vision and pattern recognition. DOI: 10.1109/CVPR.2018.00678, 2018.
2. Cem Direkoglu, Melike Sah, Noel E. O'Connor, abnormal crowd behavior detection using novel optical flow based features. 14th IEEE International conference on advanced video and signal based surveillance, 2017.
3. V. Mahalakshmi, K.P. Anumol, M.Kokiladeepa, Mrs. S. Archana. Abnormal visual events detection using optical flow orientation histogram, ICEETS, 2016.
4. R. Parvathy, Soumya Thilakan, Meenu Joy, K.M. Sharma. Anomaly detection using motion patterns computed form optical flow, IEEE, DOI: 10.1109/ICACC.2013.18. 2013.
5. C. Bergeron, J. Zaretzki, C. Breneman, K.P. Benett. Multiple instance ranking, ICML, Proceedings of the 25th International conference on machine Learning, pages 48-55, 2008.
6. A. Karpathy, G. Toderici, S. Shetty, Li fei-fei, Thomas Leung, Rahul Sukthankar. Large scale video classification with convolutional neural networks, IEEE conference on computer vision and pattern recognition (CVPR). DOI: 10.1109/CVPR.2014.223, 2014.
7. Karani Kardas, Nihan Kesim Cicekli. (2018) Surveillance Video Analysis System.
8. P. Zhou, Qinghai Ding, Haibo Luo, Xinglin Hou. (2018) Violence detection in surveillance video using low-level features.

ISSN: 2005-4238 IJAST

9. Fath U Min Ullah, Amin Ullah, Khan Muhammad, Ijaz Ul Haq, Sung Wook Baik. (2019) Violence Detection Using Spatiotemporal features with 3D Convolutional Networks.
10. Karen Simonyan, Andrew Zisserman. (2017) Two-Stream Convolutional Networks for Action Recognition in Videos.
11. L.Kratz and K. Nishino. Anomaly detection in extremely crowded scenes using spatiotemporal motion pattern models, IEEE conference on computer vision and pattern recognition (CVPR). Pages: 1446-1453, 2009.
12. Gunnar Farneback, two-frame motion estimation based on polynomial expansion, 2005.
13. D. Chen, P. Huang. Motion based unusual event detection in human crowds, in journal of visual communication and image representation, 22(2), pages: 178-186, 2011.
14. Y. Cong, J. Yuan and J. liu, abnormal event detection in crowded scenes using sparse representation. Pattern recognition, 46(7): 1851-1864, 2013.
15. W. Li, V. Mahadevan, and N. Vasconcelos. Anomaly detection and localization in crowded scenes, TPAMI, 2014.
16. R. Mehran, A. Oyama, M. Shah. Abnormal crowd behaviour detection using special force model, IEEE conference on computer vision and pattern recognition (CVPR). Pages: 1446-1453, 2009.

Bibliography

Dr. Sumitra Kisan is working as an Assistant Professor in the department of Computer Science & Engineering, Veer Surendra Sai University of Technology (VSSUT), Burla. She has completed her B.Tech in Computer Science & Engineering from Veer Surendra Sai University of Technology, Burla(formerly University College of Engineering, Burla) in 2007 and M.Tech from Indian School of Mines, Dhanbad in 2011.She has done her Ph.D from Utkal University, Bhubanewsar. She has eight years of experience in teaching and research. Her areas of research are Image Processing (Fractal Analysis, Image Segmentation, and steganogrphy), Network Security and cryptography. She has published her research in many International and national journals as well as conferences.

Prof. Dr.Sachi Nandan Mohanty, received his Postdoc from IIT Kanpur in the year 2019 and Ph.D. from IIT Kharagpur in the year 2015, with MHRD scholarship from Govt of India. He has recently joined as Associate Professor in the Department of Computer Science & Engineering at ICFAI Foundation for Higher Education Hyderabad. His research areas include Data mining, Big Data Analysis, Cognitive Science, Fuzzy Decision Making, Brain-Computer Interface, and Computational Intelligence. He has received 3 Best Paper Awards during his Ph.D at IIT Kharagpur from International Conference at Benjing, China, and the other at International Conference on Soft Computing Applications organized by IIT Rookee in the year 2013. He has published 20 SCI Journals. As a Fellow on Indian Society Technical Education (ISTE), The Institute of Engineering and Technology (IET), Computer Society of India (CSI), Member of Institute of Engineers and IEEE Computer Society, he is actively involved in the activities of the Professional Bodies/Societies.

ISSN: 2005-4238 IJAST