MAT JOURNALS

Journal of Computer Science Engineering and Software Testing
e-ISSN: 2581-6969
Volume 5 Issue 2

A Survey on Unusual Event Detection in Videos

Seemanthini. K[1], Akhila. S. Jain[2], Ashwin. M. Bharadwaj[2], Hema. B. M[2], Meghana. P. V[2]
[1]Assistant Professor, [2]UG Students
[1,2]Department of Information Science and Engineering, Dayananda Sagar Academy of Technology and Management, Bangalore, Karnataka, India
Email: *akhila.jain67@gmail.com*
DOI: *http://doi.org/10.5281/zenodo.3333430*

Abstract

As the usage of CCTV cameras in outdoor and indoor locations has increased significantly, one needs to design a system to detect the unusual events, at the time of its occurrence. Computer vision is used for Human Action recognition, which has been widely implemented in the systems, but unusual event detection is lately entering into the limelight. In order to detect the unusual events, supervised techniques, semi-supervised techniques and unsupervised techniques have been adopted. Social force model (SFM) and Force field are used to model the interaction among crowds. Only normal events training samples is not sufficient for detection of unusual events. Double sparse representation has been used as a solution to this, which includes normal and abnormal training data. To develop an intelligent video surveillance system, behavioural representation and behavioural modelling techniques are used. Various machine learning techniques to identify unusual events include: Graph modelling and matching, object trajectory based, object silhouettes based and pixel based approaches. Kullback–Leibler (KL) divergence, Quaternion Discrete Cosine Transformation (QDCT) analysis, hidden Markov model (HMM) and histogram of oriented contextual gradient (HOCG) descriptor are some of the models used are used for detecting unusual events. This paper briefly discusses the above mentioned strategies and pay attention on their pros and cons.

Keywords: *Video surveillance, Machine Learning, Unusual event detection*

INTRODUCTION

As the concern regarding public safety and security has been constantly increasing day by day, constant research is going on for detection of Unusual Events in the field of Computer Vision. CCTV cameras have now become a part of life, with their installations seen at places like airports, malls, railway stations, traffic signals and so on. Unusual events were detected using traditional systems too where a human operator would observe the videos captured by the CCTV camera, who would then find out the unusual events. But this task becomes very tedious for the worker, especially if he has to inspect the videos captured by many cameras at once. Hence, efforts have been made to detect the unusual events captured through a surveillance camera. This has seen great progress over the recent years, thus freeing the operators from their tedious work, and thus reducing the labour cost to great extent. Unusual Event Detection (UED) in a crowded scene is quiet difficult and challenging too because of many factors. These factors take the shape of blocking of CCTV cameras, noise associated with the videos, diversity and complexity of events that might occur in real world and finally unpredictability of events. Regardless of these difficulties, UED aims to detect the unusual events in a video sequence through pattern recognition. Pattern recognition works by identifying the various patterns that have some deviation from an already defined normal pattern of a model. The classification of an event as

an unusual event depends on 2 factors. The first is, how a normal event is defined or modelled and the second is, which event description is applied to it. The main factor that influences UED is event description. Event description can be defined as the arrangement of the raw input data into many aspects that represents the properties of a video. In order to capture the motion and appearance of an event, HOG descriptor is commonly used. HOG descriptor is a traditional pixel-wise descriptor. But the HOG descriptor has a disadvantage of not being able to capture contextual data which is required to discover certain important features in the video scenes.

Hence, a novel idea has been proposed, which processes the surveillance videos, to detect unusual events in crowded scene. Unusual events can be simply defined as those that differ from the current dominated behaviour. The proposed approach detects unusual events at both pixel and block-level. In the pixel level approach, Gaussian mixture models have been used to detect unusual events whereas in the Block Level approach, entire crowd is divided into a number of groups or blocks, based on detection of pedestrians. Once these blocks have been created, one can easily detect the events using a Social Force Model [2].

Yet another idea has been proposed to detect the unusual events. In this method, adjacency-matrix based clustering (AMC) is used to cluster the human crowd into groups. First, Optical flow is estimated, followed by AMC clustering. Once this is done, the behaviour of the crowed with respect to the attributes, position, orientation and crowd size, are characterized through the force field model. Finally, the anomaly in the crowd is predicted. [3].

Another approach is based on AMDN. AMDN is nothing but Appearance andMotion Deep Net. AMDN is based on deep neural network, which automatically learns the feature representations of an image. A fusion framework is used to combine early fusion and late fusion, which represents the motion patterns. Appearance and motion features are learnt using denoising auto-encoders. Based on these features, anomaly score of each input is calculated using a SVM model. Later, the late fusion strategy combines the predicted score and hence computes the unusual events. [4].

Another approach is to use sparse representation method. This is widely used to detect abnormalities in a crowd. One advantage of this method is that, it represents the crowd motions with high dimension features. A double sparse representation method can be used for higher accuracy. Two sparse representation classifiers are used in this method, and the results for each video sample are different.

In [6], behaviour representation and the behaviour modelling techniques are presented. These 2 methods are important for any video surveillance. Certain techniques with respect to feature extraction and crowd behaviour are also reviewed in this paper. Classification methods for behaviour modelling have also been proposed. [5].

MACHINE LEARNING

Machine learning is a branch of Artificial Intelligence, wherein the system will automatically learn and improver from the previous experience without the necessity to program explicitly. It mainly promotes one to write programs which access the data and learn things themselves. In simple terms, the final aim is to enable the computers to learn automatically without the need for human intervention and assistance. Machine learning includes to data mining and Bayesian predictive

modelling concepts. The system takes data as input and then uses an algorithm to analyse and then predictions are made. Machine learning is used in variety of tasks like predictive maintenance, fraud detection, portfolio optimization, and automatize tasks. Types of machine learning algorithms are:

1) Supervised Learning: supervised learning is a concept of approximation of function, where basically system is trained with an algorithm and at end selection of the function with best input data. Here the human experts will provide computer with training data that can be input or predictors and from the given data the system will be able to learn the patterns. Common Algorithms are Linear Regression, Naive Bayes, Support Vector Machines (SVM), Nearest Neighbour, Neural Networks and Decision Trees.

2) Unsupervised Learning: The system will be trained with unlabelled data. The system might be capable to teach new things after learning patterns from input data. The algorithms are useful in cases where the human experts don't have enough knowledge to what to look for in the input data. Common Algorithms are k-means clustering, Association Rules.

VIDEO PROCESSING

Digital video processing is used to improve the quality of the images captured. Video processing applications include astronomy, medicine, image compression, sports, rehabilitation, motion pictures, surveillance, production industries, robot control, TV productions, biometrics, photo editing, and so on. Categories of Video Processing:

1) Image Processing :
   Image processing is a field of signal processing, which includes segmentation. Segmentation is a method of enhancing the object while suppressing the image.
2) Machine Vision :
   Machine Vision refers to the use of video processing in the industries.
3) Computer Vision:
   Vision possessed by humans is called human vision, whereas vision possessed by a computer is called computer vision. Computer vision uses advance algorithms which are similar to what a human can perform.

UNUSUAL EVENT DETECTION

Anomaly Detection is a technique used for identification of unusual events. These are event which do not match the expected behavior. Those are identified is often said to be anomalies or outliers. Detect abnormal behavior of equipment in a manufacturing plant can be done using sensor data as temperature, pressure and humidity.

Detection and prevention of fraudulent spending by understanding normal customer spending amounts, time and locations between transactions.

Most Common Anomaly Detection Algorithms:

1) K-Means:
   Clustering techniques is an approach for detection of anomaly. Clusters of "normal" characteristics is identified and if the distance between a new points of cluster and all other clusters is too high, it is considered as an anomaly.
2) One-Class Support Vector Machines:
   If we need to plot the data in an n-dimensional space, One-Class Support Vector Machines (SVM) can be used which attempts to identify the region where most cases lie, then those are considered "normal".
   When there is a new data then it is labeled as "normal" or an "anomaly" depending on how close it is to the "normal" boundary.
3) Auto encoders:

An Auto encoder is a technique used for reduction of dimensionality. It is a form of neural networking where the first part of the network is called as the encoder and

reduces the input to a lower dimension.
The second part of the network is called the decoded part and aims for reconstruction of the original input.
The aim is to create a model which will have the same input and output. A new data point can be passed to the model and if the error between the input and the output is too high then it can be considered as an anomaly.
Video surveillance can be used in wide areas such as:
1. Live Cams
2. Extreme Environmental Conditions
3. Time Lapse
4. Wildlife and Zoo Assistance
5. Crowd Control
6. Traffic Flow

DEEP LEARNING

Deep learning allows training an AI system to make predictions from the given set of inputs. Both supervised and unsupervised learning are used for training the systems.
Deep learning is a branch of machine learning, which is based on learning data representations, contradicting the task-specific algorithms.

APPLICATIONS

1) Image recognition :
   Image recognition is a subset machine vision, is the ability of an application to recognize objects, people, places and actions in images.
   Systems use machine vision technologies with a camera and artificial intelligence application to achieve recognition of the image.
2) Natural language processing :
   Natural language processing (NLP), a branch of Artificial Intelligence (AI), can simply be defined as the ability of the system to understand human language.
   The emerging natural language processing approaches are on Deep learning techniques and the concept of artificial intelligence is also used to understand the program's capacity.
3) Image restoration:
   Image Restoration gives the estimating the original image from the corrupted image. Corruption can be motion blur, camera miss-focus and noise.

EXISTING WORK

Prajwal B.K et al.[1] conclude, in present day applications, tracking object in low resolution video is a tedious task due to loss of distinctive aspect in the visual outward look of moving target objects.

Quing Ge et al.[2] proposed to study the dynamics of the pedestrians in a crowd, a new technique called Social Force Model(SFM) was introduced. Block-based SFM is a technique to detect anomalies in a crowd, in single pixels, which makes use of interactive and personal desire force. GSFM (General Social Force Model) studies the behaviour of the pedestrians in a crowd, which is based on the principle that, "Due to presence of Actual Force, the velocity of a person changes because each person has a defined mass". The term Actual Force specified previously has 3 components: (i) Interactive Force with respect to the pedestrians (ii) Individual driving force and (iii) Interactive Force with respect to the surroundings.

The BSFM will compute the social force associated with each block that has been segmented and therefore is better than the GSFM. The proposed method is based on "Tracking by Detection". This methoduses template-matching along with SIFT algorithm (detecting features) for tracking each block.

Chen et al. [3] conclude, a ''crowd'' can be described as a few people in a considerably small area. The motion and dynamics of the crowd and their interactions provides a lot of information about human activities.

In order to characterize the interactions among the crowds, optical flow is estimated to group the crowd using unsupervised clustering method. Lucas–Kanade (LK) algorithm is proposed to estimate the optical flow. In order to detect an unusual event happening in a crowd, a method known as force field model is used. Main idea here is that, as the size of the cluster increases or as it becomes dense, they will be able to represent the features more precisely. In order to accomplish this goal, adjacency matrix-based clustering is projected. The dynamics and the trends in motion of the crowd can be predicted using a model called as Force Field, which models the human crowd.

Xu et al. [4] proposed Adjacency Matrix Basedframework in order to detect unusual events. This method is composed of 2 important processes. In the first step, Stack de-noising auto-encoders (SDAEs) are used to capture the correlation between the motion and appearance features. In the second step, three separate one-class SVMs (support vector machines) are used to represent various types of features. After theSVM model is trained, three anomaly scores will be computed for a given video sample. The SVM score is obtained with an innovative late fusion scheme. In this paper, given a test frame, a sliding window technique is used for classifying each path as a usual or unusual event.

Liu et al. [5] presented, detecting unusual events using training dataset only for the usual eventsis not adequate. Double sparse representation method was proposed in order to overcome this issue. This method uses both normal and abnormal samples. Given below is the description of two sparse representation models. Each model is a classifier.Fuzzy measures are used to describe the significance amongtheclassifiers. There are three important steps in the proposed system: preparation, training, andtesting. In the first step, two dictionaries are considered. One is used for normal dataset and the other for abnormal dataset. After this, the abnormality of the test samples is evaluated by using the double sparse representation. Both the dictionaries are updated with the samples which have high confidence, in the Training process. After training this model, the test samples are automatically predicted as normal or abnormal by using the proposed double sparse representation, which becomes the testing process. Thus, double sparse representation can be used to detect unusual events by adopting supervised learning techniques.

Zagrouba et al. [6] concluded, with the increase in the number of surveillance cameras in public places, there is a need to develop an intelligent video surveillance system that automatically detects unusual events from large number of videos from multiple cameras and raises an alarm when something dangerous happens. This system makes use of two stages of video processing: behavioural representation and behavioural modelling.

The first stage is Behavioural representation, whichcaptures the suitable features of the target object. This further consists of two steps. Initially, based on the low-level features, the area of interest is found in the scene. Then, each target object is provided with a description. Different aspects such as shape, texture and so on are used in representing the target object. Multiple features such as global, local and optical flow features of the moving object are used to detect and describe the motion of the object with respect to time. Local features are first detectedin the frame. Motions in the frame are described using Global features. Details about the global motion are extractedusing the Optical flow. Specific information regarding the humanaction is

provided using Behavioural modelling.Later it determines whether the behaviour is usual or unusual.

Chen et al. [7] concluded, human crowds modeling is one of the major issue and concern in video surveillance because of their unpredictable behavior. The position of an isolated region that comprises of an individual person or a set of occluded persons is detected using background subtraction method.

Paul et al. [8] proposed, a visual surveillance system for accurate detection of human beings is critical for various areas of application such as unusual event detection, human gait categorization and person recognition.

Yang et al.[10] proposed a model which supports the spatio-temporal data analysis and also exhibits its practical use. A large-scale spatio-temporal data sets are used which provides a prospect to study and analyze the behavioral patterns to comprehend the interactions between physical and cyber events. Spatial statistics and machine learning techniques are combined together in this framework. One of the vital approaches in data modeling for spatial data in statistics, the Morisita index is used in this work. High computational complexity is prevented by employing meta-Morisita index.

Masaki et al. [11] concluded that, one of the methods that can detect precise human behaviors in crowded video surveillance scenes is described in this paper.This system can recognize specific activities and behaviors based on the trajectories that are created by tracking and detecting people in a particular video. This system detects people using an SVM classifier and HOG descriptor. It also tracks the regions by calculating two-dimensional color histograms. Verification techniques such as calculating optical flows and backward tracking are contributed to robust recognition of unusual behaviors.KTH motion sequences are commonly used for motion recognition. In this system, a method is developed to detect specific human behaviors in crowded video surveillance sequences.

Yan et al.[12]proposed the real world events will occur in crowded situations. This becomes a challenging task for the existing systems to detect unusual events. This is due to thediverting motion of the other objects and difficulty in segmenting the actors from the background.
The above systemhas the following main ideas:

1. Effectivelymatchingevent representation against the spatio-temporal video volume.
2. Augmentation of features based on shape using optical flow.
3. Instead of treating an event as an entity, they are individually matched in this approach.

The experiments on human behavior and activities, such as waving in a crowd or picking up a dropped object show reliable detection with few false positives.

Yigithan et al. [13] concluded that in the approach, an instance based machine learning algorithm and a system for classification of real-time objects and human activity recognition is considered.In the proposed method, object silhouettes are used forthe classification of human objects present in a scene which is supervised by a stationary camera. For the segmentation of objects, an adaptive background subtraction model is used. In order to classify objects into previously defined classes like boxing, walking, running and kicking, a supervised learning technique based on template matching is used.Moving regions in the video are detected which will correspond to different real world objects like pedestrians, automobiles, clutter, etc.

Collins et al.[14] proposed,VSAM Integrated Feasibility Demonstration (IFD) contract is used. They also have developed technology which automatically understands video which enables a human operator to monitor all activities using a network of video sensors.The main goal is to collect the real time information automatically in order to improve the situational awareness. Automated video surveillance is one of the significant research areas in the commercial sector. A 24-hour monitoring and analysis of video surveillance is required to alert security officers.

Cassol et al.[15] proposed, crowds occur in a variety of circumstances, such as sporting matches, public concerts and so on. The crowd movements occur in an orderly manner under certain conditions, but few problematic situations can lead to disastrous effects. A computer vision technique is proposed for identifying the motion trend changes in human crowds. By using 2D motion histograms, the proposed system can identify the global and local effects. This method is evaluated on public data sets and also on datawhich is generated by using crowd simulation algorithm.

Guo et al.[16] concluded that in this work, a technique to detect abnormal events based on saliency attention mechanism of human visual system is proposed. Temporal and spatialanomaly saliencies are considered by representing the pixel value in each and every frame as a quaternion, which contains weighted components that are composed of the following features - intensity, contour, motion-speed and motion-direction. For each such quaternion frame, Quaternion Discrete Cosine Transformation (QDCT) and signature operationsare applied.Inverse QDCT and Gaussian smoothing are used to develop a spatio-temporal anomaly saliency map. By multi-scale analysis, it can be observed that the abnormal events appear in the areas with high saliency scores.

Xu et al.[17] proposed, Video surveillance cameras have been extensively installed in public areas for security purposes. To detect rare events, a weakly-supervised approach using Kullback–Leibler (KL) divergence is formulated. This technique utilizes the sparse nature of the target events to its advantage.It guarantees the existence of a decision boundary to separate samples that contain the target event from the events that do not. This classifier requires only a decision threshold, and hence its use compared to other weakly supervised techniques is much simplified.

Elise et al. [18] proposed a model which contains the hidden markov models and Bayesian frameworks. This paper derives the variation based learning for Dirichlet HMM. Reducing the constraints over the data and splitting the problem into independent sub problems has resulted in the enhancement of the algorithm's performance. The system is robust to alter detectionswhich are the anomalies usually seen ona larger scale.

Cosar et al. [19] proposes a novel approach for group behavioural analysis and abnormal behaviour detection in video using trajectory pixel based methods.Contrasting these approaches, an integrated pipeline that integrates the output of object trajectory analysis and pixel-based analysis is proposed for abnormal behaviour interpretation. This enables the detection of abnormal behaviours related to speed and direction of object trajectories. It also detects thecomplex behaviours related to finer motion of each object. A fully unsupervised method uses basic trajectory features such as speed and duration, and also fine motion features to represent body

movements is used. Therefore, this method outperforms the trajectory-based and pixel-based approaches, by detecting different types of abnormalities, from basic to complex events.

Hu et al. [20] proposed, initially one has to find the gradient among any two local regions. Then, a histogram of oriented contextual gradient (HOCG) descriptor is constructed for detecting unusual events. This HOCG descriptor can be described as a distribution of contextual gradients of different sub-regions in various directions. This can be used to characterize the compositional perspective of events effectively. The HOCG descriptor has the following significant features –It is simple, flexible, distinguishable, and is capable of capturing an event's contextual information. Based on qualitative and quantitative analyses of the results it can be observed that the performance of HOCG descriptor is better than the pixel based HOG descriptors. This method is proportional to the state of art methodswithout the use of any complicated modelling techniques.

CONCLUSION

Different techniques are approached for event detection. They include behaviour representation and behaviour modelling, block-based social force model, double sparse representation and so on. These techniques are mostly based on the enhancement of low resolution video by super resolution techniques. The above examined techniques have shown high computation costs, but the techniques used are efficient in detecting unusual events from a video surveillance camera. Hence, it becomes more useful and computationally efficient if low resolution video camera is used to detect unusual events.

REFERENCES

1. Prajwal, B. K., Sreeharsha, M. S., Raghulan, R., Babu, S. S., & Kiran, M,"Detection of unusual events in low resolution video for enhancing ATM security and prevention of thefts". In Smart Technologies for Smart Nation (SmartTechCon), International Conference on (pp. 1139-1143). IEEE, 2107.
2. Ji, Qing-Ge, Rui Chi, and Zhe-Mingdetection and localisation in the crowd scenes using a block-based social force model." IET Image Processing 12.1: 133-137, 2017.
3. Chen, Duan-Yu, and Po-Chung Huang. "Motion-based unusual event detection in human crowds" Journal of Visual Communication and Image Representation 22.2: 178-186, 2011.
4. Xu, D., Yan, Y., Ricci, E., & Sebe, N.,"Detecting anomalous events in videos by learning deep representations of appearance and motion",Computer Vision and Image Understanding, 156, 117-127, 2017.
5. Liu, P., Tao, Y., Zhao, W., & Tang, X,"Abnormal crowd motion detection using double sparse representation". Neuro-computing, 269, 3-12,2017.
6. Mabrouk, A. B., & Zagrouba, E.,"Abnormal behaviour recognition for intelligent video surveillance systems: A review". Expert Systems with Applications, 91, 480-491, 2018.
7. Chen, Duan-Yu, and Po-Chung Huang, "Visual-based human crowds behaviour analysis based on graph modelling and matching." IEEE Sensors Journal, 13.6 2129-2138, 2015.
8. Paul, Manoranjan, Shah ME Haque, and Subrata Chakraborty, "Human detection in surveillance videos and its applications-a review." EURASIP Journal on Advances in Signal Processing, 176, 2016.
9. Li, Weixin, Vijay Mahadevan, and Nuno Vasconcelos. "Anomaly detection and localization in crowded scenes." IEEE transactions on pattern analysis and machine intelligence 36.1: 18-32, 2015.

10. Yang, Zhao, and Nathalie Japkowicz. "Anomaly behaviour detection based on the meta-Morisita index for large scale spatio-temporal data set." Journal of Big Data 5.1: 23, 2018.
11. Takahashi, Masaki,"Robust recognition of specific human behaviours in crowded surveillance video sequences." EURASIP Journal.
12. Ke, Yan, Rahul Sukthankar, and MartialHebert,"Event detection in crowded videos." Computer Vision,IEEE 11th International Conference on. IEEE, 2017.
13. Dedeoglu and Yigithan,"Silhouette-based method for object classification and human action recognition in video." European Conference on Computer Vision. Springer, Berlin, Heidelberg, 2015.
14. Collins, Robert T., Alan J. Lipton, Takeo Kanade, Hironobu Fuji Yoshi, David Duggins, Yanghai Tsin, David Tolliver., "A system for video surveillance and monitoring", VSAM final report: 1-68., 2015.
15. de Almeida, I. R., Cassol, V. J., Badler, N. I., Musse, S. R., & Jung, C. R. "Detection of global and local motion changes in human crowds". IEEETransactions on Circuits and Systems for Video Technology, 27(3), 603-612., 2017.
16. Guo, H., Wu, X., Cai, S., Li, N., Cheng, J., & Chen, Y. L "Quaternion discrete cosine transformation signature analysis in crowd scenes for abnormal event detection" Neurocomputing, 204, 106-115, 2016.
17. Xu, J., Denman, S., Fookes, C., & Sridharan, "Detecting rare events using Kullback–Leibler divergence: A weakly supervised approach" Expert Systems with Applications, 54, 13-28.,2016.
18. Epaillard, Elise, and Nizar Bouguila"Variational Bayesian Learning of Generalized Dirichlet-Based Hidden Markov Models Applied to Unusual Events Detection." IEEE transactions on neural networks and learning systems 99: 1-14., 2018.
19. Coşar, S., Donatiello, G., Bogorny, V., Garate, C., Alvares, L. O., & Brémond, F. "Toward abnormal trajectory and event detection in video surveillance." IEEE Transactions on Circuits and Systems for Video Technology, 27(3), 683-695, 2017.
20. 20. Hu, X., Huang, Y., Duan, Q., Ci, W., Dai, J., & Yang, H. (2018). "Abnormal event detection in crowded scenes using histogram of oriented contextual gradient descriptor". EURASIP Journal on Advances in Signal Processing, 54, 2018.

Cite this article as: *Seemanthini. K, Akhila. S. Jain, Ashwin. M. Bharadwaj, Hema. B. M, & Meghana. P. V. (2019). A Survey on Unusual Event Detection in Videos. Journal of Computer Science Engineering and Software Testing, 5(2), 29–37. http://doi.org/10.5281/zenodo.3333430*