

Anomalous Human Activity Recognition in Surveillance Videos

Raksha S, B G Prasad

Abstract: *This paper is a survey on different approaches for Human Activity recognition which has utmost significance in pervasive computing due to its many applications in real-life. Human-oriented problems such as security can be easily taken care of by detecting abnormal behavior. Accurate human activity recognition in real-time is challenging because human activities are complicated and extremely diverse in nature. The traditional Closed-circuit Television (CCTV) system requires to be monitored all the time by a human being, which is inefficient and costly. Therefore, there is a need for a system which can recognize human activity effectively in real-time. It is time-consuming to determine the activity from a surveillance video, due to its size, hence there is a need to compress the video using adaptive compression approaches. Adaptive video compression is a technique that compresses only those parts of the video in which there is least focus, and the rest is not compressed. The objective of the discussion is to be able to implement an automated anomalous human activity recognition system which uses surveillance video to capture the occurrence of an unusual event and caution the user in real-time. So, the paper has two parts that include adaptive video compression approaches of the surveillance videos and providing that compressed video as the input to detect anomalous human activity*

Index Terms: *Human Activity Recognition, Adaptive Video compression, Vision-based Human Activity Recognition, Anomaly detection.*

I. INTRODUCTION

The idea of activity recognition in pervasive computing is of utmost importance because of its varied applications in real-life, particularly to handle human-centric problems such as security. The purpose of an activity recognition system is to recognize the basic daily life activities of human beings. Due to the diversity and complexity in human activities, the accuracy of human action recognition becomes challenging. Construction of activity models that identify and classify various human activities follow several approaches. This research discipline attracts video processing and machine learning communities as it finds applications in several fields of studies like medicine and healthcare, human-computer interaction, crime investigation and security systems. Applications of human activity recognition are not limited to

Revised Manuscript Received on July 5, 2019.

Raksha S, Department of Computer Science and Engineering, B.M.S College of Engineering, Bangalore, India, raksha95128@gmail.com.

Dr. B G Prasad, Department of Computer Science and Engineering, B.M.S College of Engineering, Bangalore, India.

health care and security. It as an ongoing and open research topic in computer vision that includes instances of Behavioural Biometrics, video analysis, animation, and synthesis. Human activity recognition system uses sensors for interpretation of gestures or motion of the human being, thus identifying the action that the human body makes. Understanding human activities hold within itself a recognition of the activity and pattern discovery of that activity. The initial part makes use of a predefined activity model to correctly detect human activity. Thus, there is a need to construct a high-level conceptual model for implementing the pervasive system for human activity recognition. Moreover, activity pattern discovery does not require predefined models as it uses only a few low-level sensor data that is captured in order to find the unknown patterns. Though there is a major difference in the two techniques they have a common target of improving the performance of human activity recognition systems. These techniques are also supportive of each other. They combine to enhance the performance by using the discovery of the activity pattern to define recognized activity.

Applications like Behavioural Biometrics, video analysis for surveillance, human-computer interaction, animation and synthesis are all based on the research of vision-based Human Activity Recognition. This review highlights the advances in activity recognition approaches, especially for the activity representation and classification methods. The recognition of visual patterns, which are spatiotemporal and semantic, in vision-based activity recognition is a major problem in computer vision for video surveillance. Determination of the spatial and temporal events in a video and then the automatic analysis is possible through video analytics. The output of the video content analysis machine is used to construct other functionalities that include behaviour analysis for situation awareness in terms of security, they may include identification of anomalous activities like theft, fight, assault, abuse, burglary, road accidents, and vandalism.

When we have a surveillance video, it is possible to perform both compression and activity detection and they can be combined to improve the system that recognizes activities to perform in real-time. The generic flow of the anomalous human activity recognition system is depicted in Fig. 1. Video compression has several approaches that have been followed so far, a few approaches concentrate on the compression of the whole video with the only intention



of reducing the size. In such an approach there will be loss of some useful information.

Thus, a more reasonable compression technique would be the Adaptive video compression [7]. Adaptive video compression compresses only those parts of videos in which there is least focus, eliminating the compression of the parts that have the objects of interest. The video compression standard can be controlled with the intention of giving higher fidelity encoding to those elements of the video that are significant and semantically meaningful. This is achieved by combining low-level features of the video with low computational costs. This transmitted representation will thereby allocate lower bits to the background elements which are not significant in this case. The basis of this method is to selectively smoothen individual frames so that it preserves those features of the image that contain the semantically interesting objects which could be motion of a human being in the case of Human activity recognition. Each discrete frame in a video is adaptively compressed using the low-level features eliminating the insignificant parts and this result can be used as the pre-processing state by inserting it into the video coding pipeline. Adaptive video compression for a surveillance video will be most efficient as the camera will be still, thus keeping the background constant. Hence when a motion is detected it is considered for activity recognition and then classified into normal or anomalous activity.

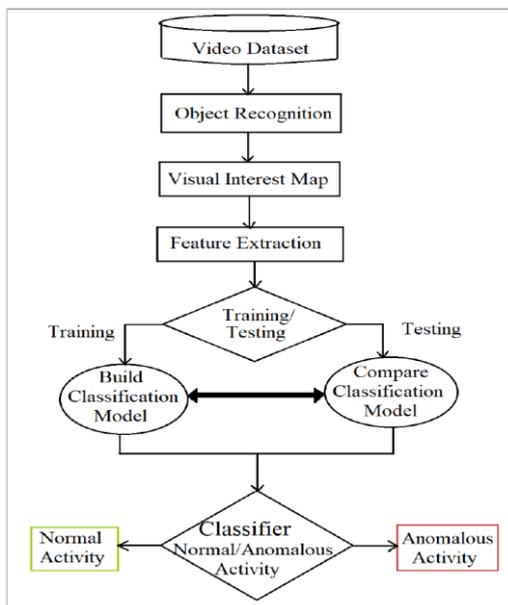


Fig. 1. Anomalous Human Activity Recognition

Human activities can be categorized as normal and anomalous activity. Anomalous activity is when a human deviate from a normal behaviour and tends to cause harm to other human beings, himself or to the surrounding environment. Unusual activities are a result of mental or physical discomfort. The extreme research that is being carried out in terms of recognition of human activities has led to the rise in the interest of anomalous human activity

recognition. There are many existing approaches for detecting anomalous behaviour of human beings based on the type and speed of motion, interaction between objects of interest and their sizes. The intention of this review is to point the different available approaches for anomaly detection, which would help to overcome existing challenges and result in a real-time automated anomalous activity recognition system using some innovation. The usage of libraries like OpenCV, TensorFlow, Keras, Theano are required to implement such a real-time system. Training the system requires large datasets and one such dataset is the UCF crime dataset. This dataset covers around 13 different kinds of anomaly including abuse, burglary, theft, assault along with normal event.

The remaining of this paper is structured to give a description of the related literature in Section II, and the next section presents the applications and need of Anomalous Human Activity Recognition. The following section depicts the challenges involved in implementing an anomalous activity recognition system. The last section holds the conclusion of the review.

II. LITERATURE REVIEW

Applications of Human activity recognition are vast and varied. Recognizing an action performed by a human being can be useful in hospitals for observing patients and for providing safety and security in public places. Growth in different kinds of crime rates have made it compulsory for the installation of the CCTV (Close Circuit Television) cameras. But then again there is need for manpower to monitor the surveillance cameras. Thus, a lot of research has been done to automate the process of surveillance and to detect anomalous human activity. As discussed in the research paper [1], application of activity recognition in an examination hall, to reduce workload of invigilation, makes use of the Viola Jones algorithm to detect and identify objects, which is the human face in this case. Features are represented using integral image which is thereby computed using pixel-based operation. The computation of the integral image is the result of summation of the left, top and affected pixel. Extraction of the interesting point uses the SURF (Speed Up Robust Features) algorithm. This method also matches and finds corresponding features. Classification of the extracted features makes use of AdaBoost. Cascade is another classifier that is used for classification that is much more complex. Thus, this system detects any suspicious human activity in an examination hall.

A. Recurrent Neural Networks

A Recurrent neural network is a deep learning concept that belongs to the class of deep neural network. It sees application in visual imagery analysis. A variation of multilayer perceptron is used, designed with the intention of minimizing the preprocessing. The convolutional neural networks are supervised, and they use labels as learning signals. Anomalies occur very rarely as compared to usual



events. Hence it is a meticulous job to manually monitor the surveillance and identify anomalous behavior. In addition to it, anomaly is very ambiguous and vaguely defined because an action can or cannot be an anomaly depending on the surrounding environment.

It is totally normal to run in a park and abnormal to run around in a restaurant. In one of the research papers done for anomaly detection, the task is traced as a binary classification problem, usual and unusual [18] which has been proven effective in terms of accuracy. This methodology requires very little or no supervision at all and hence unlabeled videos are efficient to train the system. Introducing autoencoders [2] with the intension of replacing low-level features by learned hierarchical features to minimize labor required for feature engineering resulting in data representation to support effective machine learning methodology for sparse coding. Furthermore, it has been replaced with autoencoders, thereby removing the separation between feature extraction and learning a model for those features. Hierarchical feature learning is achieved by allowing autoencoder to have several layers of hidden units. This autoencoder is used for dimensionality reduction by marking the total sum of input units to be higher than the total sum of output units. Training of an autoencoder happens in an unsupervised manner using back propagation. Recurrent Neural Networks (RNN) have a working like that of a feedforward network but there is temporal dependency between inputs as it is a sequence unlike

The traditional feedforward neural network. Here the output not only depends on previous input but also on the whole history of inputs. But the problem is that the gradients vanish because of which it is not practical to look back for long sequences. This results in the introduction of LSTM (Long Short-Term Memory) that incorporates the forget gate which happens to be the recurrent gate. This prevents vanishing and explosion of back propagated errors. Thus, it is made practical to work on longer sequences that are stacked together for higher level information capture. Therefore, this solution for anomaly detection uses the idea of combining the temporal sequencer with the spatial feature extractor, resulting in development of end to end model that is trainable for anomaly detection which requires only semi-supervised videos.

B. Probabilistic Neural Networks

Human activity recognition finds an important application in security system as it helps to detect abnormal events. There has been enough research on different techniques suitable for anomaly detection and one of them is using PNN (Probabilistic Neural Networks) [3] that classifies the actions into normal and abnormal categories. Tracking of the object through Blob analysis makes use of LBP (Local Binary Pattern). There has been diverse contribution in the field of machine learning, activity recognition, particularly anomaly detection for smart surveillance systems. The main intention of the study is to broadly distinguish the existing procedures and bring attention to their key challenges. HMM (Hidden

Markov Model) is one of the techniques used for suspicious behavior detection which is robust against brightness changes, variations caused due to climatic or weather changes in outdoors and background noise. This mainly focuses on movement detection and object tracking.

C. Hidden Markov Model

The technique that is built on the probabilistic neural network strategy uses the background subtraction method to adjust variations in the bitrate of the video streams. Since this follows multiple layers PNN (Probabilistic Neural Network), the probabilities for classification of targeted and non-targeted actions is picked by the previous layer output participating in the transfer function. This paper suggested that PNN method for anomaly detection gives superior results as compared to other techniques like [1] HMM and Dynamic Bayesian network, whose performance needs to be tested in unstructured situation. [19] Bayesian and optical flow method is used for movement detection. Classification and tracking are done according to the object's aspect ratio. Another technique which helps in activity recognition is the latent Dirichlet allocation [20] model, which is constructed by coupling the background subtraction procedure with the hidden Markov model. However, this method did not prove to be suitable to be used in a crowded scenario. It includes extensive research on Saliency detection. Even the KNN classifies [21] did not achieve efficient classification of the anomalous activities and that led to the suggestion of using detectors that are capable of handling higher number of attributes.

D. Generative Adversarial Networks

Several review papers suggest different methodologies for detection of abnormal activity. Extending the survey done in the past [8] focuses on contextual abnormal human behavior system particularly for video surveillance. It is desirable to have an automated surveillance system that provides security through an intelligent real time detection of suspicious activities of a human being. There are different authors who have done an impressive survey on the variety of techniques available for anomaly detection. Highlights of this review includes the recent trends in Suspicious Behavior recognition. It is challenging to detect abnormality in a video sequence as there is always some amount of ambiguity in defining anomaly. There could be false alarm generation even when activity is normal or an activity that is abnormal can be purposely intended to seem normal by the criminal. Though there has been significant advancement in this field there are few benchmarks against which the performance of different methods is evaluated. There are very few datasets available for training the anomaly detection system, therefore using the footage of CCTV cameras throughout the world would be a standard dataset. The deployable system should be computationally fast as it must work in real-time and a system with reduced computational requirement is best suitable.

Advancement in deep generative models results in development of methods like Generative Adversarial



Networks (GANs) [29]. On the basis of information that was previously determined the representation and categorization of anomalies is made and different unsupervised learning models are suggested accordingly, which include the generative models and the spatial temporal predictive models which are reconstruction based. Training samples obtained from normal distribution is ensured to have the least reconstruction error in the Reconstruction model [9]. Spatio-temporal predictive model views the videos as the spatial-temporal time-series. This is meant to minimize the prediction error.

These simple studies are used to analyze and distinguish the various approaches to evaluate the spatial-temporal anomaly detection for appropriate criteria. Also, the state of art techniques for methods based on deep learning are reviewed, because although the information available for surveillance applications are huge in number it does not contain annotation that maybe required for supervised learning. Thus, it can be considered that there are not many datasets available to train the anomaly detection system. Hence there is a need for techniques that use deep learning for unsupervised and semi-supervised learning of anomaly detection systems.

Automated Human behavior analysis system is a core part of building an intelligent system they can have a perspective just like human beings and it must be able to interact with humans. Most of the previous algorithms for activity recognition is based on supervised learning, whose performance depends on the standard of annotation and inconsistency in human activity occurrences. Hence a dynamic clustering algorithm that is fast and robust is proposed to process a nonexclusive multi-dimensional time arrangement [10]. Clustering of a data driven style occurs, and it dynamically evolves with time. The algorithm follows two steps, the first one being an initialization step, where the robust spectral clustering method is used [30,31] to learn the co-variances and centers of initial pattern in an unsupervised fashion. The step the follows this is the online evolution step, where in every arriving feature is either considered as one of the already existing features or it is considered as a new recognition. The work prior to this was the Aligned Cluster Analysis (ACA) for alignment of the time series [32]. Specification of the cluster count along with action segment's minimum and maximum length leads to an alternate advanced version of ACA. A dynamic program is solved over the whole stream making the approach unsuitable for fast and online video segmentation.

E. Convolutional Neural Network

The previously existing two-stream CNNs [22] for action recognition achieves the state of arts performance but it is inappropriate for real-time deployment as it is computationally expensive. It consists of spatial network and temporal network which considers the RGB image and discretionary stream individually as the input. Here the most expensive step is the calculation of optical flow, even if implemented efficiently. Thus, the major drawback that is stopping real-time detection in the classical two-stream methodology is the optical flow calculation. As a solution to

this work, optional flow is replaced with motion vector for description of the local block movement with respect to the corresponding reference frame [23]. It explores motion vector in the deep convolutional neural network framework. However, the low resolution and imprecise description of fine movements in motion vector causes the depletion of recognition execution performance when directly trained using those motion vectors. In the process of relieving this problem, several training methods for enhancement of the motion vector [4] have been introduced. The facts and features extracted in optical flow CNN must be transferred and used to guide the motion vector CNN as they are both inherent correlated. This can boost the performance of the optical flow CNN. The strategies used to achieve this include teacher's initialization, supervision transfer and a combination of the two. Network initialization largely impacts the final performance. But there is a possibility of losing the knowledge that is initially transferred through the period of the fine-tuning process and thus leads to introduction of supervision transfer which keeps count of the additional supervision from teachers' net when the Motion vector CNN is being trained. The last strategy combines both initialization that includes teacher's parameter and supervision transfer for the performance enhancement.

Quite a lot of investigation has been made in the area of activity recognition from video sequences. Animated Pore Template (API) has been described in one of the researches [24] for detection of short term, long term and contextual actions. This whole work revolves around pore template [25, 26]. Shape and motion templates are the result of classification of the pose template. It is emphasized on SVM (Support Vector Machine) algorithm which is built on the semi-supervised structure. This model suits well for most applications despite the 2D view dependency. Studying the various methodologies for human activity recognition and anomaly detection in particular it is clear that CNN's show an excellent performance for visual recognition. CNN's first perform local feature extraction from the sequence of frames generated using the video and then combine these features in order to obtain more complex higher order features. Extracting features using the 3D convolution [5] is one in which multiple frames that are continuous will be stacked to form a cube to which a 3D kernel is convolved. Capturing of the motion information is done by connecting feature maps in convolutional layer to multiple frames on the previous layer that is contiguous. Extraction of a single type of feature is possible by the 3D convolutional kernel from frame cube since there is replication of kernel weights across the entire cube. This model is trained using back propagation and supervised algorithm of 80-90%. Thus, the 3D-CNN model gives a promising performance in recognizing abnormal activity of human beings.

Another latest and most effective technique for activity detection is the contextual multi-scale region convolutional 3D Network [6] which involves a temporal feature pyramid representing activities of corresponding temporal scales. Each of the levels in the pyramid detect activities of the temporal scale using the activity classifier and activity proposal detector. Thus, the architecture of (CMS-RC3D) the proposed model comprise



fine components and they are the network input, feature extraction network, temporal feature pyramid, activity proposal and activity classifier network. The GPU memory limits the arbitrary length of the untrimmed video which happens to be the network input. The network for feature extraction can follow any applicable architecture including two stream networks [22, 21] C3D [5, 28]. However, the temporal and spatial features are important for video representation. Without any constrained settings the activities in the videos have varied temporal scales, this led to the creation of temporal feature pyramid, with different levels based on different temporal scales. This allows the detector to efficiently detect activities of different temporal scales in a single forward pass in contrast to the previous model which needed several forward passes for different scales that resulted in the inevitable increase in interface time.

As the model holds detectors that can detect activities within all temporal range in a single shot making it computationally efficient.

III. APPLICATIONS OF ANOMALOUS HUMAN ACTIVITY RECOGNITION

Smart video surveillance is essential for smart security to ensure public safety. An anomalous human activity recognition system therefore identifies and alerts people when there is a potential threat. Detection of abnormal behaviour can be applied in various scenarios depending on the surrounding environment.

A. College or University Surveillance

In a college environment anomalous activity may include students involving in a fight, abuse, ragging or dealing drugs. These activities are to be brought to the notice of concerned authorities and this can be achieved using a smart surveillance system that recognises abnormal activities.

B. Crowded Public Places Surveillance

An anomalous human activity recognition system can be used in banks, airports, railway stations, bus-stops, shopping malls and so on where there is a need for continuous monitoring of surveillance that is practically not possible.

C. Road -Accident Detection

The death rate caused by road accidents is gradually increasing. One possible way to help avoid death on road is to get timely help to the victims. The smart surveillance system can detect road accidents in real-time and call for an ambulance.

D. Patients' Surveillance in Hospitals

Patients who are unconscious can be monitored using the Activity recognition system to detect any kind of motion

rather than having an attendant next to the patient for constant observation.

E. Security Breach Detection at Homes

Smart security systems can be installed at homes for theft detection. Timely and automatic detection of an intruder using the surveillance videos helps provide safety in homes rather than using the surveillance videos for crime investigation after the occurrence of theft or burglary.

IV. CHALLENGES IN ANOMALOUS HUMAN ACTIVITY RECOGNITION

- The accuracy of activity recognition keeps decreasing due to the variety in human motion patterns with respect to difference in subjects and time
- The orientation and position obtained from sensors and wearable devices must be dependable
- Human beings are capable of multi-tasking and during the transition between tasks it is difficult for any algorithm to recognize the activity
- The availability of datasets for training and testing the anomaly detection model is minimal. However, one can use the UCF Crime dataset [14] which covers almost 13 anomalous events.
- Camera jitter can lead to degradation in the quality of the video captured resulting in a low-resolution video frame
- The pixels of the objects of interest may be identical to the pixels of other elements resulting in camouflage and hence it becomes difficult to capture the object of interest
- Changes in weather and climate may result in dynamic variations in brightness, contrast and other important factors of the video.
- Chances of disturbances like birds sitting in front of the camera or leaves of trees blocking the cameras' view when it is too windy will affect the activity recognition system as it cannot capture the objects of interest.

V. CONCLUSION

After researching the several methods that exists for human activity recognition it is noticed that the best suitable methods will be the ones that follow the unsupervised learning model, taking into consideration the kind of datasets available. Datasets available in large amounts are never labeled and hence Unsupervised or Semi supervised Deep Learning models must be selected. However, one cannot depend on a single methodology for activity recognition because the same algorithm works differently on different datasets. One must also consider the recognition performance and accuracy. The result of activity recognition is time consuming due to the size of the video hence it is better to adaptively compress the video before passing it for feature extraction. However, it is not advisable to suggest one algorithm as the best for



implementation as each of them have their own advantages and disadvantages. Moreover, this is a field seeing rapid growth and there is a need to learn, research and analyze the best suitable method for the implementation.

ACKNOWLEDGMENT

The work reported in this paper is supported by the college through the TECHNICAL EDUCATION QUALITY IMPROVEMENT PROGRAMME [TEQIP-II] of the MHRD, Government of India.

REFERENCES

- G. O. Young, "Synthetic structure of industrial plastics (Book style Wassima Aitfares, Abdellatif kobbane, Abdelaziz Kriouile, "Suspicious Behaviour Detection of People by Monitoring Camera" *IEEE 2016*
- Yong Shean Chong, Yong Haur Tay, "Abnormal Event Detection in Videos using Spatiotemporal Auto encoder" arXiv:1801.03149v2 [cs.CV] Jan 2017
- Karuna B. Ovhal, Sonal S. Patange, Vaishnavi K. Tarange, Vijay A. Kotkar, "Analysis of anomaly detection techniques in video surveillance," *IEEE conference June 2018*
- B. Zhang, L. Wang, Z. Wang, Y. Qiao, and H. Wang, "Real-time Action Recognition with Enhanced Motion Vector CNNs," Apr. 2016
- Sathyashrisha, Pushparaj, Sathivel Arumugam, "Using 3D Convolutional Neural Network in Surveillance Videos for Recognizing Human Actions" *The International Arab Journal of Information Technology, Vol. 15, No. 4, July 2018*
- Yancheng Bai, Huijuan Xu, Kate Saenko, Bernard Ghanem, "Contextual Multi-Scale Region Convolutional 3D Network for Activity Detection" arXiv:1801.09184v1 [cs.CV] 28 Jan 2018
- Andrew D. Bagdanov, Marco Bertini, Alberto Del Bimbo, Lorenzo Seidenari, "Adaptive Video Compression For Video Surveillance Applications" *IEEE conference June 2011*
- Oluwatoyin P. Popoola, Member, IEEE, and Kejun Wang "Video-Based Abnormal Human Behaviour Recognition—A Review" *IEEE conference April 2016*
- B Ravi Kiran, Dilip Mathew Thomas, Ranjith Parakkal "An overview of deep learning-based methods for unsupervised and semi-supervised anomaly detection in videos" arXiv:1801.03149v2 [cs.CV] 30 Jan 2018
- Yan Zhang, He Sun, Siyu Tang, Heiko Neumann "Temporal Human Action Segmentation via Dynamic Clustering" arXiv:1803.05790v2 [cs.CV] 18 Mar 2018
- Rensso Mora Colque, Carlos Caetano, Victor C. de Melo, Guillermo Camara Chavez and William Robson Schwartz "Novel Anomalous Event Detection based on Human-object Interactions" In Proceedings of the 13th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2018) - Volume 5: VISAPP, pages 293-300 ISBN: 978-989-758-290-5
- N. Kumaran, Dr. U. Srinivasulu Reddy "Article Location and Following in Group Condition - A Survey" Procedures of the Global Gathering on Innovative Registering and Informatics (ICICI 2017) IEEE Xplore Agreeable - Part Number: CFP17L34-Craftsmanship, ISBN: 978-1-5386-4031-9 Number: CFP17L34-ART, ISBN: 978-1-5386-4031-9
- Ong Chin Ann, Lau Bee Theng "Human Activity Recognition: A Review" *IEEE International Conference on Control System, Computing and Engineering, 28 - 30 November 2014, Penang, Malaysia*
- Waqas Sultani, Chen Chen, Mubarak Shah "Real-world Anomaly Detection in Surveillance Videos" arXiv:1801.04264v2 [cs.CV] 31 Mar 2018
- N. Patil, Prabir Kumar Biswas "A Survey of Video Datasets for Anomaly Detection in Automated Surveillance" *Sixth International Symposium on Embedded Computing and System Design (ISED) 2016*
- Angela A. Sodemann, Matthew P. Ross, and Brett J. Borghetti "A Review of Anomaly Detection in Automated Surveillance" *IEEE transactions on systems, man, and cybernetics—part c: applications and reviews, VOL. 42, NO. 6, November 2012*
- J. Yin and Y. Meng, "Abnormal behavior recognition using self-adaptive hidden Markov models," in *Pro. 6th Int. Conf. Image Anal. Recognit.*, Jul. 6-8, 2009, pp. 337-346. Zhou, S., Shen, W., Zeng, D., Fang, M., Wei, Y., Zhang, Z.: Spatio-temporal convolutional neural networks for anomaly detection and localization in crowded scenes. *Signal Processing: Image Communication* 47,358-368 Sep 2016
- Neha Gaba, Neelam Barak and Shipra Aggarwal "Motion Detection, Tracking and Classification for Automated Video Surveillance" *IEEE conference April 2016*
- Vijay Mahadevan, Weixin Li, Viral Bhalodia and Nuno Vasconcelos "Anomaly Detection in Crowded Scenes" *IEEE conference March 2010*
- Ramin Mehran and Alexis Oyama "Anomaly in road side scene" *IEEE conference March 2017*
- K. Simonyan and A. Zisserman. Two-stream convolutional networks for action recognition in videos. In *NIPS'14*, pages 568-576, 2014.
- V. Kantorov and I. Laptev. Efficient feature extraction, encoding, and classification for action recognition. In *CVPR'14*, pages 2593-2600, 2014.
- Yao B., Liu Z., Nie B., and Zhu S., "Animated Pose Templates for Modelling and Detecting Human Actions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 3, pp. 436-452, 2013.
- Huo F. furthermore, Hendriks E., "Numerous Individuals Following and Posture Estimation with Impediment Estimation," *PC Vision and Picture Understanding*, vol. 116, no. 5, pp. 634-647, 2012.
- Huo F. and Hendriks E., "Real Time Multiple People Tracking and Pose Estimation," in Proceedings of the 1st ACM International Workshop on Multimodal Pervasive Video Analysis, Firenze, pp. 5-10, 2010.
- K. Simonyan and A. Zisserman. Two-stream convolutional networks for action recognition in videos. In *Advances in neural information processing systems*, pages 568-576, 2014.
- D. Tran, L. Bourdev, R. Fergus, L. Torresani, and M. Paluri. "Learning spatiotemporal features with 3d convolutional networks" In Proceedings of the IEEE international conference on computer vision, pages 4489-4497, 2015.
- I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Advances in neural information processing systems*, 2014, pp.2672-2680.

AUTHORS PROFILE



Raksha S is working Computer Science & Engineering at B.M.S. College of Engineering is an autonomous engineering college in Basavangudi, Bangalore, India, she has completed her from Master of Technology from B.M.S.C.E.



Dr. B G Prasad is working as Professor at Department of Computer Science and Engineering, B.M.S College of Engineering, Bangalore, India. He did his Ph.D. from IIT, New Delhi. His Research Areas are, Integrated Visual Feature-Based Indexing for Content Based Image Retrieval.