ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 3, March 2017

Detection of Unusual Activity from a Video Set

Chinmay Pednekar [1], Aman Kumar[2], Aradhana Dhumal[3], Dr. Saylee Gharge[4], Dr. Nadir N. Charniya[5], Ajinkya Valanjoo[6]

B.E Students, Dept. of E.X.T.C., Vivekanand Education Society's Institute of Technology, Mumbai, India[1,2,3]

Associate Professor, Dept. of E.X.T.C. ,Vivekanand Education Society's Institute of Technology, Mumbai, India[4]

Professor, Dept. of E.X.T.C. ,Vivekanand Education Society's Institute of Technology, Mumbai, India[5]

Assistant Professor, Dept. of E.X.T.C., Vivekanand Education Society's Institute of Technology, Mumbai, India[6]

ABSTRACT: Unusual event detection is a growing demand to process a plethora of a video set. It comes under the area of artificial intelligence. Suppose we are given a long duration of footage then using optimal algorithms "unusual" activities can be traced out. Unusual events are rare, difficult to describe, hard to predict and can be subtle. However, given a large number of observations it is relatively easy to verify if they are indeed unusual. Our project helps detect the unusuality in videos using machine learning algorithms, thus saving time for organizations and individuals who would have to go through the entire footage instead. The algorithm exploits inherent redundancy of videos and constructs a sparse combination learning matrix to identify possible abnormal events.

KEYWORDS: Unusual; Sparse Learning; Video set; Features

I. INTRODUCTION

This research project presents a technique for detecting unusual activity in a video set using machine learning algorithms. .Recently, there has been growing interests in developing systems to automatically analyze video data. Of the many possible tasks, detecting unusual events from video set is of considerable practical importance. In some event detection applications, events of interest occur over a relatively small proportion of the total time. Machine learning and probabilistic techniques are widely applied in this area.

Most of the activity recognition systems predefine and model the anomalous activities so that the system can recognize whether the activities detected are anomalous or not. Others learn the unusual activity patterns either in supervised or unsupervised manner and then recognize unusual activities based on their dissimilarity from the usual ones. Traditional image processing algorithms to detect abnormalities are heavy on computations and are thus slow and require very powerful systems to be run on.

Sparsity based techniques convert these heavy computations into smaller costless least square optimizations which speeds up the process and provides faster detection rates as sparse matrix data structure is very generous in terms of memory usage.

II. RELATED WORK

In [1] author has discussed about two approaches to detect an unusual activity in a video. The first method is a model based approach that detects unusual activity in two steps. In the first step, image features like speed and shape of the object are extracted and computed. In the second step, models are developed for normal activities based on the extracted features. The unusual activities are detected by matching a video set against a set of normal models. The frames or

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 3, March 2017

segments which do not match with the normal set of events is considered as an abnormal or unusual event. The second approach is an unsupervised approach which does not make use of models. It detects abnormal events based on the fact that the occurrence of normal events in a sample data set would be greater than abnormal events. So if there is no similar event then the event is declared as abnormal event. Thus this approach works on the basis of ability to compare two events and measure their similarity.

There are several methods proposed for detection of abnormality as in [2] the author uses unsupervised approach and detects unusual activity entirely on the basis of spatio-temporal volumes. The foundation of this approach is that an anomalous behavior has low likelihood of occurrence. Densely constructed spatio-temporal video volumes help to learn the video events. The dominant frequently occurring events are characterized by organizing the video volumes into large contextual graphs. By decomposing spatio-temporal contextual information into unique spatial and temporal contexts the framework learns the models of the dominant spatial and temporal events.

A new technique for detection of abnormal activities based on the concept of Gaussian mixture model (GMM) discussed in [3] creates a statistical model based on the blob detector which analyzes foreground function of cells. Different motion characteristics are extracted in this method by using GMM as a position vector. In addition to proposed method ,the author has also mentioned about the use of the differential method of Lucas and Kanade to estimate abnormal events observed in a surveillance video. The presented algorithm in this method helps to accelerate the process of abnormal motion detection based on a local adjustment of the velocity field by calculating the light intensity between two images to detect the abnormal movement.

The powerful semantic description of video imagery used by John Y. A. Wang in [4] provided by Image segmentation results in better image understanding. In this paper an iterative framework for segmentation of video data is used. Thus with spatio temporal segmentation they have produced a layered image representation of video for video processing application where in video data can be described as a set of moving layers.

A.Niranjil Kumar and Dr. C. Sureshkumar in [5] have discussed about a new approach towards analysis of crowd activities in surveillance videos. The authors have divided the computer vision based analysis algorithm into three categories; people counting, people tracking and crowd behavior analysis. Their recent contribution for real time detection of abnormal event uses the motion of computational attenuation which quantifies motion saliency.

Spatio Temporal video attention detection technique [6] is used for detecting the regions corresponding to both interesting objects and actions in video sequences. Both spatial and temporal saliency maps are constructed and fused in a dynamic fashion to produce the overall spatiotemporal attention model. In the temporal attention model, motion contrast is computed based on the planar motions between images, which is estimated by applying RANSAC on point correspondences in the scene.

The technique of sparse coding is used to extract features for a supervised task such as classification in [7]. In this paper they have proposed a sparse coded net and a feed forward model that integrates sparse coding and task driven output layers. Their results showcased significant improvement over semisupervised learning as well as the superior classification performance.

III. IMPLEMENTED METHOD OF SPARSE BASED LEARNING

For an input video sequence, the implemented approach(of sparse combination learning) extracts set of consecutive frames which are analyzed to define an event. The frames which are extracted are resized into a different scale as [8] and uniformly partitioned each layer to form a set of non-overlapping patches. All patches are equal in area. Corresponding regions in 5 continuous frames are stacked together to form a spatial-temporal cube. The image pyramid so formed contains the local information in layer having fine resolution so that the layer has wider dimension along with global information in low resolution i.e the layer being resized to a smaller dimension.

A video is represented as a set of cuboids, these cuboids residing in a video set define an event or a sequence. As the algorithm scans through the entire video sequence, the video is broken into a set of events, each represented by a group

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 3, March 2017

of cuboids. Specifically, the video is represented as $X = \{X_1,..., X_M\}$, with each event X_i composed of a group of cuboids, i.e.,$X_i = \{X_{1_i},..., X_{n_i}\}$, where n_i is the total number of cuboids within the sliding window.

Having obtained the spatio-temporal cubes, we compute 3D gradient features on each CUBE . These features in a video sequence are processed separately according to their spatial coordinates. Only features at the same spatial location in the video frames are used together for training and testing. These features are sobel derivative of each image frame considering 3 parameters- x-position, y-position and t time.

3.1 Learning Combinations on Training Data

For each cube location, 3D gradient features in all frames are denoted as $X = \{x_1,...,x_n\} \in R^{p\times n}$, gathered temporally for training. Our goal is to find a sparse basis combination set $S = \{S_1, . . . ,S_k\}$ with each $S_i \in R^{p\times s}$containing s dictionary basis vectors, forming a unique combination. Each S_i belongs to a closed and bounded set, which ensures column-wise unit norm to prevent over-fitting.

The sparse combination learning has two goals. The first goal is to find K basis combinations, which has a small reconstruction error t as [9]. The second goal is to make the total number K of combinations is small enough based on redundant video information. It is obvious that if we use large $K$the it will make reconstruction error close to zero letting the unusual events go unnoticed.

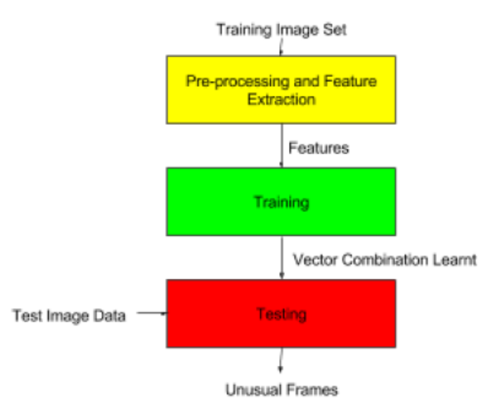
3.2 Optimization for Training

The two goals discussed above contradict each other in a sense. Reducing K could increase reconstruction errors. It is not optimal to fix K(as stationary) as well, as content may vary among videos. This problem is addressed with maximum representation strategy. It automatically finds K while not wildly increasing the reconstruction error t. A set of combinations with a small K is obtained by setting a reconstruction error upper bound λ uniformly for all elements in S. If the reconstruction error for each feature is smaller than λ, the coding result is with good quality.

3.3 Testing Phase

With the learned sparse combinations $S = \{S_1, . . . ,S_k\}$ in the testing phase with new data x, we check if there exists a combination in S fitting the reconstruction error upper bound. It can be quickly achieved by checking the least square error for each S_i

IV. DATA FLOW DIAGRAM



ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 3, March 2017

V. RESULTS

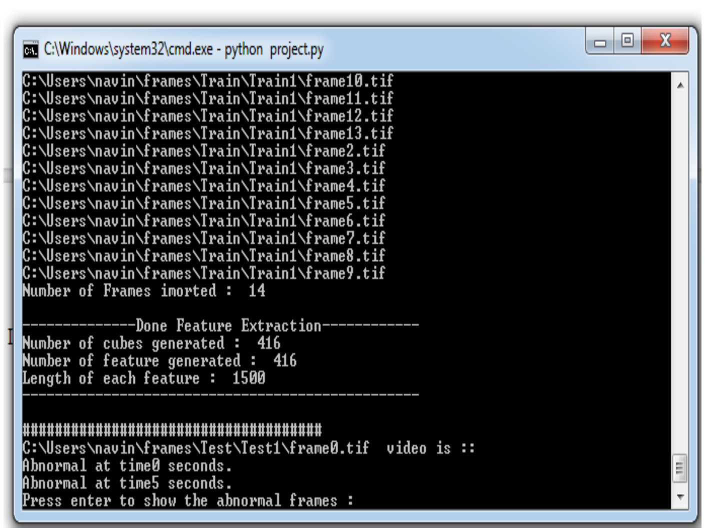
A video input is accepted from the user and then divided in separate image frames. Features of each image is extracted. Though the algorithm gives the best result when all the image frames are considered for comparison with the frames just preceding and just exceeding it, there is a considerable wastage of time in such implementation. More number of images need to be stored and corresponding number of features for each frame has to be calculated. This increases the computational load and also demands for more memory storage. Hence the feature extraction processes is a tradeoff between the number of features extracted for comparison and the computational load on the processing system.

In this project features are extracted from each and every frame considering the fact that Time is not a constraint for implementation as it will lead to more accurate results.

5.1 Snapshots

5.1.1 Input Training Dataset

Video is converted into frames and the initial frames are given as input to Training Algorithm assuming the frames exhibit "Usual" events.



5.1.2 Output on console window

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 3, March 2017

After the testing algorithm is run the program shows "Unusual" frames in the video along with their time of occurrence

VI. CONCLUSION AND FUTURE WORK

An abnormal event detection method is presented and implemented via sparse combination learning. This approach directly learns sparse combinations, which increase the testing speed hundreds of times without compromising effectiveness.

In a time where surveillance cameras are being used everywhere, effectively checking it for any abnormal event would be a bottleneck. Thus a fast and intelligent method to check theses surveillance cameras is at most required. It would help in cutting down a lot of work to be done by people struggling to monitor it and would help it taking faster actions during those situation by integrating these with alarms and other important actions like informing the police or calling an ambulance.

The project can be further improvised using a totally Unsupervised approach where there is no need of creating a training model of usual events instead the Unusuality is detected by comparison with previous frames. Even though Unsupervised approach will quite heavy on computations it can be proved very beneficial in terms of the flexibility it can offer. It can be even taken forward to networking domain where the anomaly in video sets can be uploaded on the cloud so that it can be observed directly by the officials present in the control room. The output frames that are abnormal(as detected by the algorithm) will be published on cloud via WiFi module which will be interfaced with Arduino/Raspberry-pi controller.

REFERENCES

[1] Zhong, Hua, Jianbo Shi, and Mirkó Visontai. "Detecting unusual activity in video." Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on. Vol. 2. IEEE, 2004.

[2] Javan Roshtkhari, Mehrsan, and Martin D. Levine. "Online dominant and anomalous behavior detection in videos." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2013.

[3]Mariem, G., E. Ridha, and Z. Mourad. "Detection of Abnormal Movements of a Crowd in a Video Scene." International Journal of Computer Theory and Engineering 8.5 (2016): 398.

[4]Wang, John, and Edward H. Adelson. "Spatio-temporal segmentation of video data." IS&T/SPIE 1994 International Symposium on Electronic Imaging: Science and Technology. International Society for Optics and Photonics, 1994.

[5]Kumar, A. Niranjil, and C. Sureshkumar. "Abnormal Crowd Detection and Tracking in Surveillance Video Sequences." International Journal of Advanced Research in Computer and Communication Engineering 3.9 (2014).

[6]Zhai, Yun, and Mubarak Shah. "Visual attention detection in video sequences using spatiotemporal cues." Proceedings of the 14th ACM international conference on Multimedia. ACM, 2006.

[7]Gwon, Youngjune, et al. "Sparse-coded net model and applications." Machine Learning for Signal Processing (MLSP), 2016 IEEE 26th International Workshop on. IEEE, 2016.

[8]Cong, Yang, Junsong Yuan, and Ji Liu. "Sparse reconstruction cost for abnormal event detection." Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on. IEEE, 2011.

[9]Lu, Cewu, Jianping Shi, and Jiaya Jia. "Abnormal event detection at 150 fps in matlab." Proceedings of the IEEE International Conference on Computer Vision. 2013.

[10]Sak, Hasim, Andrew W. Senior, and Françoise Beaufays. "Long short-term memory recurrent neural network architectures for large scale acoustic modeling." Interspeech. 2014.

[11] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." Advances in neural information processing systems. 2012.

[12] Pham, Duc Truong, Stefan S. Dimov, and Chi D. Nguyen. "Selection of K in K-means clustering." Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science 219.1 (2005): 103-119.