Detection of Suspicious Human Activity based on CNN-DBNN Algorithm for Video Surveillance Applications

Alavudeen Basha A
Research Scholar
School of Electrical Engineering, VIT Univeristy
Vellore, India
alavudeenbasha16phd0492@gmail.com

Parthasarathy P
Research Scholar
School of Electrical Engineering, VIT Univeristy
Vellore, India
parthasarathy.p@vit.ac.in

Vivekanandan S
Associate Professor
School of Electrical Engineering, VIT Univeristy
Vellore, India
svivekanandan@vit.ac.in

Abstract— **Detection of suspicious human actions in automated video surveillance applications, is of great practical importance. Those kind of unusual activities in human is very difficult to acquire and classify to predict. In our proposed work, automatic tracking and detecting unusual movement's problems in closed circuit videos was resolved. Firstly, the videos are converted into frames. Then from the obtained frames, humans are detected from the video using a background subtraction method. Then the features are extracted using a convolutional neural network (CNN). The features thus extracted are fed to a Discriminative Deep Belief Network (DDBN). Labeled videos of some suspicious activities are also fed to the DDBN and their features are also extracted. Then the features extracted using Convolutional Neural Network (CNN) are compared against these features extracted from the labeled sample video of classified suspicious actions using a Discriminative Deep Belief Network (DDBN) and various suspicious activities are detected from the given video and results shows increase accuracy of 90% for the proposed framework for classification.**

Keywords— ***Closed Circuit TV, Convolutional Neural Network, Discriminative Deep Belief Neural Network.***

I. INTRODUCTION

The monitoring of conduct, exercises, or other evolving data, for the most part of individuals or spots to influence, overseeing, coordinating, or ensuring them is named as surveillance. The observation methodologies can fuse recognition from a detachment by strategies for electronic apparatus, for instance, closed circuit TV (CCTV) cameras, or catch of electronically transmitted information, for instance, Internet movement or phone calls, and it can consolidate essential, for the most part low-advancement procedures, for instance, human knowledge administrators and postal square endeavor [1-4]. Various affiliations and people are sending video perception structures at their zones with Closed Circuit TV (CCTV) cameras for better security. The acquired video data is significant to keep the risks beforehand the bad behavior truly happens. These chronicles furthermore transform into a not too bad legitimate verification to recognize offenders after the occasion of bad behavior. For the most part, the video feed from CCTV cameras is seen by human overseers. These overseers screen various screens without a moment's delay chasing down unpredictable activities [5].

This is a costly and wasteful method for observing. The procedure is costly in light of the fact that the administrators are on a finance of the association and wasteful on the grounds that people are inclined to mistakes. A human administrator can't productively monitored numerous screens at the same time. Likewise, grouping of an administrator will decrease definitely over the long haul. One of the techniques to adapt to this issue is to utilize robotized video observation frameworks (video investigation) rather than human administrators. Such a framework can screen various screens all the while without the inconvenience of dropping fixation [6, 7]. The capacity of a robotized reconnaissance framework is to draw the consideration of observing work force to the event of a client characterized suspicious conduct or episode when it occurs. Perceiving human activities in reality condition discovers applications in an assortment of spaces including smart video observation, client properties, and shopping conduct investigation. In any case, precise acknowledgment of activities is an exceptionally difficult assignment because of numerous variables, for example, jumbled foundations, impediments, and perspective varieties, and so forth. The vast majority of the present methodologies make certain presumptions (e.g., little scale and perspective changes) about the conditions under which the video was taken. In any case, such suspicions only from time to time

978-1-5386-8190-9/19/$31.00 © 2019 IEEE

hold in reality condition. Also, the vast majority of the strategies pursue a two-advance methodology in which the initial step figures highlights from crude video outlines and the second step learns classifiers in light of the got highlights. In certifiable situations, it is once in a while recognized what highlights are critical for grouping the undertaking or action within reach since the selection of highlights is exceedingly issue subordinate. Particularly for human activity acknowledgment, diverse activity classes may show up drastically extraordinary as far as their appearances and movement designs as various activity classes may appear to be unique when done by various identities [8-10].

Deep learning models are a class of machines that can take in a progression of highlights by building abnormal state highlights from low-level highlights. Such learning machines can be prepared utilizing either administered or unsupervised methodologies, and the subsequent frameworks have been appeared to yield aggressive execution in different territories like visual question acknowledgment, human activity acknowledgment, characteristic dialect handling, sound arrangement, mind PC association, human following, picture rebuilding, de-noising, and division undertakings. The convolutional neural systems (CNNs) are a kind of deep learning models in which trainable channels and nearby neighborhood pooling activities are connected alternatingly on the info pictures, bringing about a chain of command of progressively complex highlights [11, 12]. It has been demonstrated that, when prepared with fitting regularization, CNNs can accomplish prevalent execution on visual question acknowledgment assignments. What's more, CNNs have been appeared to be invariant to specific varieties, for example, posture, lighting, and encompassing mess [13].

In this paper we plan a framework to deal with the bizarre conduct in video. We are utilizing CNN with DBNN for feature extraction and classification. Movement design are dissected and utilized as highlights. Not at all like past work, we pursued idea of overwhelming set in which predominant conduct is dealt with as typical and less prevailing conduct is considered as bizarre. Limited number of past work thought about various exercises in single video for identification, yet the proposed framework can used for classification.

II. RELATED WORK

In this segment, the ongoing work in the field of programmed suspicious or abnormal movement acknowledgment is talked about. There are assortment of structure accessible which can be utilized for identification of bizarre conduct in observation recordings without human intercession. Different scientists have utilized distinctive systems as per application or occasion. Each acknowledgment framework is for the most part having following advances: pre-preparing, highlight extraction, question following and conduct understanding. In pre-handling step, clamor evacuation and frontal area is extricated. In highlight extraction, a few highlights of the question is removed from the edges of video. After component extraction protest direction is set up based on separated element. Behavioral understanding is an urgent and critical advance on acknowledgment framework as in this progression conduct of question is identified and based on that conduct arrangement of occasions are finished. Last advance give result as a few occasions has a place with typical and different has a place with strange in peculiar occasion recognition in recordings yet sadly, less work is done in various odd occasion location. The proposed system identify and perceive different occasions and give occasion mark to them as ordinary or strange. In prior research on bizarre occasion recognition [14].

Some researchers have proposed a novel way to deal with recognize peculiar conduct and overwhelming conduct and also prevailing set hypothesis has been given. Later SVM algorithm is utilized for characterization of typical and odd occasions [15]. In this work he has given an audit paper that portray technique to tackle issue of strange occasion recordings portrayal. Idea of Conditional Restricted Boltzmann Machine, Independent Component examination for better element extraction as ordinary technique isn't anything but difficult to learn finish highlight portrayals and profoundly adapted moderate element investigation idea are talked about. In action acknowledgment undertaking most imperative and basic assignment is of conduct understanding was exhibited an overview paper, in that creator depicted profiling of conduct in recordings for abnormality identification in itemized. Diverse body parts are utilized for signal and feeling acknowledgment of human [16-19].

Behavior recognition is an expansive term that covers various classifications of exercises, which require diverse methods for location. For instance, swarm conduct, for example, swarm development, requires systems that catch the general qualities of the group instead of the people in it. This methodology centers on consequently hailing suspicious conduct out in the open transportation frameworks. These sorts of conduct may happen over a huge timeframe [20]. They regularly include in excess of one protest; hence, such issues as discovering directions, character following, and question arrangement must be tended to. Red-green-blue (RGB) shading foundation demonstrating to extricate moving districts are proposed here. This technique is appropriate for the continuous reconnaissance framework in view of the quick calculation and is vigorous against the ecological impacts. To identify the moving items, RGB BM with another affectability parameter was utilized to remove moving areas, morphology plans to take out commotions, and blob-marking to aggregate the moving articles. To track the gatherings of the moving items, a following calculation was proposed comprising of the expectation of the situation of each gathering, the acknowledgment of a similar gathering, and the ID of recently showing up and vanishing gatherings [21].

The method of recognizing moving items from the info picture comprises of the extraction organize in light of RGB BM and morphology, and the gathering stage in light of blob-marking. As a rule, the extraction of moving districts from successive pictures is done by utilizing BM. This sort of BM includes the loss of picture. In directed methodology of irregularity location input tests are marked as both typical and bizarre examples. This technique is intended for the conduct whose properties are pre-characterized and appearance, speed, movement or direction go about as signs to order them as ordinary and bizarre signals. Second methodology is Semi-directed that just require ordinary information for preparing the framework. This strategy is additionally partitioned into sub-classifications as manage based class and model based classification [22, 23]. Third methodology is Unsupervised, in which neither typical nor atypical examples are required as predecessor precedents for preparing. In this, arrangement is done in light of suppositions that oddities are less recurrence of event when contrasted with ordinary conduct event. To solve these kind of issues, a deep learning algorithm like convolutional NN and DBNN was proposed in this work.

III. PROPOSED WORK

The proposed architecture is based on convolutional and recurrent neural networks for feature extraction and DBNN for action classification.

A. Architecture description

The primary neural system is a convolutional neural system with the motivation behind removing abnormal state highlights of the pictures and lessening the multifaceted nature of the information.

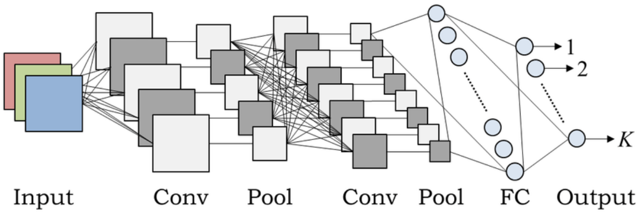


Fig. 1. General architecture of Convolutional NN [4]

We utilized the above showing model in fig. 1 to apply the procedure of exchange learning. Present day protest acknowledgment models have a huge number of parameters and can take a long time to completely prepare. For an arrangement of classifications like Image-Net and retrains from the existing weights for new classes a transfer learning method is more suitable.

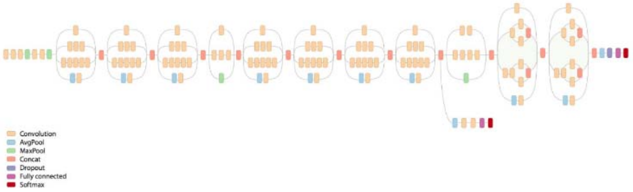


Fig. 2. Inception model [5]

In fig. 2 it clearly shows that the intermittent neural system is utilized in the second neural system to understand the arrangement of the activities and this system has a LSTM cell in the primary layer, trailed by two concealed layers and the yield layer is a three-neuron layer with delicate max initiation, which gives us the last grouping.

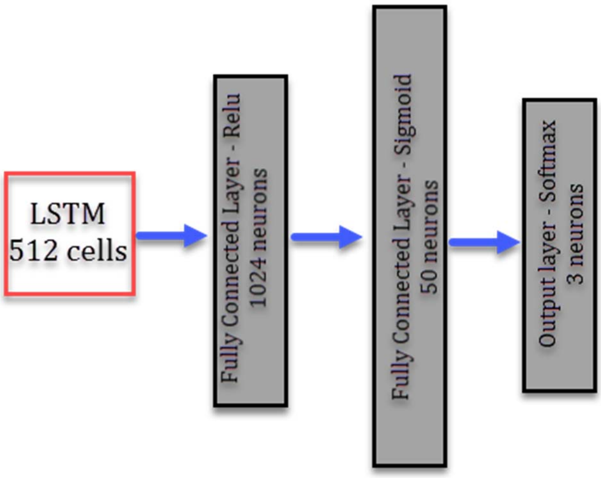


Fig. 3. Recurrent neural network [3]

In the initial step we remove the edge intervals and by making this we can do the initial demonstration. By utilizing transfer learning system we extricating the consequence of the last pooling layer, which is a vector of 2,048 qualities (abnormal state highlight outline). To do as such, we are not thinking about single casings to make our last expectation. We take a gathering of casings to characterize not the edge but rather a portion of the video. We consider that breaking down three seconds of video at once is sufficient to make a decent forecast of the action that is going on right then and there. Here we stored fifteen element maps produced by the origin demonstrate forecast, the likeness three seconds of video. At that point, we link this gathering of highlight maps into one single example, which will be the contribution of the intermittent neural layer shown in fig. 3, to get the last characterization of our framework. The video classification architecture is shown in below fig. 4 for extracting the data's.

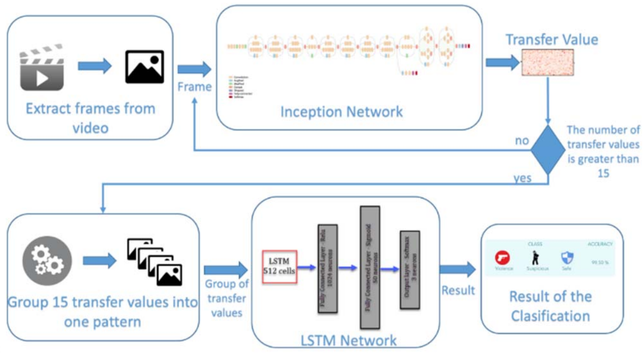

Extract frames from video
Frame
Inception Network
Transfer Value
The number of transfer values is greater than 15
no
yes
Group 15 transfer values into one pattern
Group of transfer values
LSTM 512 cells
Fully Connected Layer - Relu 1024 neurons
Fully Connected Layer - Sigmoid 50 neurons
Output layer - Softmax 3 neurons
LSTM Network
Result
CLASS
ACCURACY
Violence
Suspicious
Safe
Result of the Clasification

Fig. 4. Video classification architecture [5]

B. Discriminative Deep Belief Network

The highlights separated utilizing the convolutional neural system are looked at against the highlights extricated from named test video of ordered activities. That is utilizing a Discriminative Deep Belief Network shown in below fig. 5, the framework will be prepared utilizing either administered or unsupervised calculations to arrange the perceived activities into client indicated suspicious exercises. For the preparation reason test marked video informational collections are utilized [5].

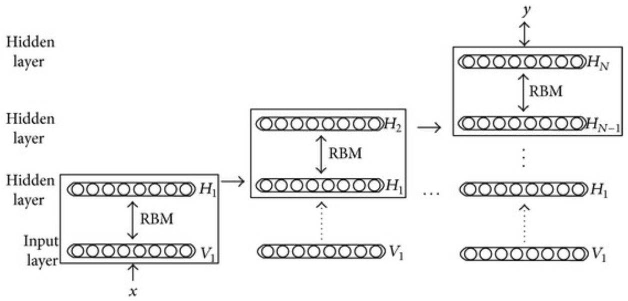


Fig. 5. General Architecture of DBNN [7]

Arrangement under deficient marked information is an outstanding difficult issue. Tragically, this is probably going to happen since getting the named information is regularly troublesome, costly or tedious. For instance, in substance based picture recovery, a client for the most part represents a precedent picture as a question and requests that the framework return comparable pictures. For this situation, there are numerous unlabeled pictures existing in a database, yet there is just a single marked precedent, i.e. the inquiry picture. To address this issue, semi directed realizing, which utilizes expansive measure of unlabeled information together with named information to assemble better students, has pulled in more consideration. Commonplace semi-directed techniques include: self-preparing, Expectation-Maximization (EM) calculation with generative blend models, transductive help vector machine, diagram based strategies, and co-preparing. Right now, the greater part of semi-directed procedures utilize shallow engineering to demonstrate the issue, for example, kernelized straight model. As contended by a few specialists, profound design, made out of numerous levels of non-straight activities, is relied upon to perform well in semi-managed learning due to its capacity to demonstrate hard man-made brainpower undertakings. Weston basically utilized shallow semi-directed calculations to profound design by connecting them to any layer of the engineering as regularizers. Also, the exact approval for genuine characterization assignments yielded focused execution. Enlivened by the investigation of semi-managed learning and profound design, this paper proposes a novel semi-regulated classifier named discriminative profound conviction systems (DDBN), in view of a delegate profound calculation profound conviction systems (DBN).

IV. RESULTS AND DISCUSSION

This segment incorporates the tests performed on the proposed approach, the dataset utilized, the natural conditions, setup, imperatives forced. Besides, the aftereffects of various analyses are figured and are additionally contrasted and past works in peculiar action recognition. The corresponding datasets was described below.

A. Datasets

i. *Numerous datasets are accessible among which the datasets utilized for the task is an ongoing dataset. The constant datasets are gathered from PETS 2007, CAVIAR, and different recordings from YouTube and so forth. The ongoing dataset comprises of three suspicious activity classes, for example, unattended baggage, battling, and lingering.*

ii. *The framework portrayed here is fit for ordering a video into three classes:*

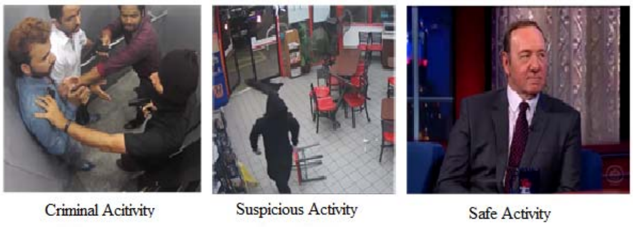
iii. *Criminal or brutal action*

iv. *Potentially suspicious*

v. *Safe*

B. Training Dataset

The dataset utilized for preparing the system is involved 150 minutes of screening partitioned into 38 recordings. A large portion of these recordings are recorded on surveillance cameras of stores and distribution centers. The general example for the data classes and various group activities are shown in the fig. 6 and fig. 7. The aftereffect of taking a casing, each 0.2 seconds in length, is having a dataset of 45,000 edges for preparing — the likeness 3000 portions of video, thinking about that as a fragment of a video speaks to three seconds of it (or 15 outlines). The entire dataset was named by us and separated into gatherings: 80% for preparing and 20% for testing. As should be obvious, the last dataset is quite little. Be that as it may, because of the exchange learning procedure, we can get great outcomes with less information. Obviously, for the framework to be more precise, it's smarter to have more information; that is the reason we continue taking a shot at getting an ever increasing number of information to enhance our framework.


Criminal Acitivity

Suspicious Activity

Safe Activity

Fig. 6. Data classes' example [5]

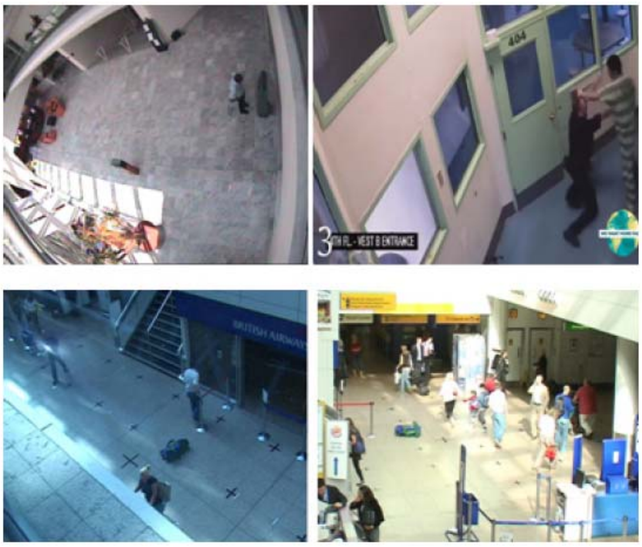


Fig. 7. Group of various activities [5]

C. Implementation

The entire framework was executed with Python 3.5. We utilize Open-CV for Python to section the video in casings and resize them to 200x200px. When we have every one of the casings, we make an expectation on the beginning model utilizing every one of them. The consequence of every expectation is an "exchange esteem" speaking to the abnormal state include delineate from that particular casing. We spare that in the exchange esteems variable and its individual marks in the name prepare variable. After video classification it is associated with a surveillance camera and continue breaking down the video progressively, and the minute the framework recognizes criminal or suspicious action, it could actuate a caution or alarm the police. Moreover, framework prepared with the fitting information is utilized to identify various types of exercises; for instance, with a camera situated in a school where your target could be to recognize harassing. By considering these factors. Separate them into gatherings of 20 outlines.

Coding 1

```
fixture_num=20
total = 0
joint_shift=[ ]
for i in range(int(len(shift_values)/fixture_num)):
   inc = total+fixture_num
   joint_shift.append([shift_values[count:inc],labels_train[count]])
   total =inc
```

Coding 2

Transfer values and their labels are available now, by utilizing these values and labels we can train the recurrent neural network and the implementation was done using Keras as follows,

```
from keras.models import consecutive
from keras.layers import Dense, Activate
from keras.layers import LSTM
chunk_size = 2048
n_chunks = 15
rnn_size = 512
model = consecutive ()
model.add(LSTM(rnn_size, input_shape=(n_chunks, chunk_size)))
model.add(Condensed (1024))
model.add(Activate('relu'))
model.add(Condensed(50))
model.add(Activate(sigmoid))
model.add(Condensed (3))
model.add(Activate('softmax'))
model.compile(loss='mean_squared_error', optimizer='adam',metrics=['accuracy'])
```

Coding 3

```
data =[]
target=[]
epoch = 1500
batchS = 100
for i in joint_transfer:
   data.append(i[0])
   target.append(np.array(i[1]))
model.fit(data, target, epochs=epoch, batch_size=batchS, ve
rbose=1)
```

The program shows about the construction of the model and after training it should save as,

```
model.save("rnn.h5", overwrite=True)
```

Since the model is completely prepared, we can begin to order recordings. We composed the accompanying front-end where you can transfer a video and start grouping it continuously. You can perceive how the classes are continually changing, and additionally the particular precision for that class. These qualities always refresh like clockwork until the point when the video is finished. The below fig.8 shows the error rate calculation for DB neural network.

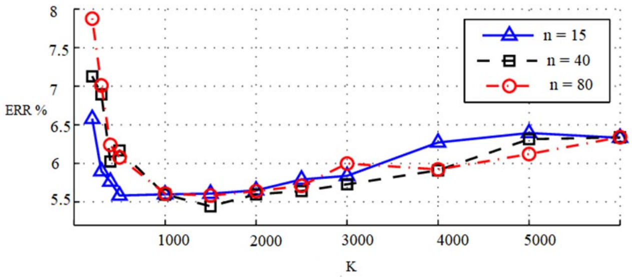


Fig. 8. Error rate calculation of DBNN

V. CONCLUSION

The work proposed a suspicious movement discovery from the observation video utilizing convolutional neural system for highlight extraction and a discriminative profound conviction arrange for activity characterization. Contrasted and the past works, the proposed approach accomplishes better characterization by profound learning-based model. After people are identified utilizing a foundation subtraction technique, seven casings are chosen in which the region of the bouncing boxes ascertained for the people distinguished are bigger among all. From these seven chose outlines, 33 include maps are extricated which are in five unique channels characterized by, dark channel, slope x, inclination y, optflow-x and optflow-y channels. These 33 include maps are given as contribution to the CNN which restores a 128D highlights in a solitary vector. At that point this yield is nourished to a DDBN which is utilized to order the perceived activities into typical and suspicious activities via preparing the framework utilizing semi regulated learning technique. The profound learning model guarantees more precision and lesser false positives.

REFERENCES

1. In Su K, Hong Seok Choi, Yi Kwang Moo, Choi Jin Young, and Kong Seong G. Intelligent visual surveillance — a survey. International Journal of Control, Automation, and Systems, 8:926–939, 2010.

2. Valera, M. and Velastin, S.A. Intelligent distributed surveillance systems: a review. IEE Proc. . Vis. Image Signal Process., Vol. 152, No. 2, , pp. 192.204, April 2005

3. Hannah M. Dee and Sergio A. Velastin. How close are we to solving the problem of automated visual surveillance? : A review of real-world surveillance, scientific progress and evaluative mechanisms. Machine Vision and Applications, 19:329–343, September 2008.

4. Vallejo, D., et al. A cognitive surveillance system for detecting incorrect traffic behaviors. Elsevier. Expert Systems with Applications. 2009

5. SAGEM et al. Integrated surveillance of crowded areas for public security. Website, 2007.

6. Gouaillier V and Fleurant A. Intelligent video surveillance: Promises and challenges technological and commercial intelligence report. Technical report, CRIM and Technopole Defence and Security, 2009.

7. Weiming H, Tieniu T, Liang W, and S. Maybank. A survey on visual surveillance of object motion and behaviors. Systems, Man and Cybernetics, Part C, IEEE Transactions on, 34(3):334–352, 2004.

8. Hampapur, L. Brown, J. Connell, S. Pankanti, A. Senior and Y. Tian, “Smart surveillance: applications, technologies and implications”, IBM T.J. Watson Research centre, Mar 2008.

9. Duque D.Santos H. , and Cortez P. . Prediction of abnormal behaviors for intelligent video surveillance systems. In Computational Intelligence and Data Mining, 2007. CIDM 2007. IEEE Symposium on, pages 362–367, 2007.

10. Mohannad Elhamod, Member, Martin D. Levine. Automated Real-Time Detection of Potentially Suspicious Behavior in Public Transport Areas. IEEE Trans. Intelligent Transportation Systems. Vol. 14, 2013.

11. Jong Sun Kim, Dong Hae Yeom, Young Hoon Joo. Fast and Robust Algorithm of Tracking Multiple Moving Objects for Intelligent Video Surveillance Systems. IEEE Trans. Consumer Electronics, 2011.

12. Qian Zhang and King Ngi Ngan. “Segmentation and Tracking Multiple Objects Under Occlusion From Multiview Video,” IEEE Trans. Processing, Vol. 20, No. 11, 2011.

13. Shuiwang Ji, Wei Xu, Ming Yang. 3D Convolutional Neural Network for Human Action Recognition. pattern analysis and machine intelligence, IEEE Trans. Vol. 35, 2013.

14. P Shusen Zhou, Qingcai Chen and Xiaolong Wang, “Discriminative Deep Belief Networks for Image Classification” in Proc. IEEE. Image Processing, 2010.

15. Ke, S.R., Thuc, H.L.U., Lee, Y.J., Hwang, J.N., Yoo, J.H., Choi, K.H., 2013. A review on video-based human activity recognition. Computers 2, 88–131.

16. Lu, H., Li, H.S., Chai, L., Fei, S.M., Liu, G.Y., 2012. Multi-feature fusion based object detecting and tracking, in: Applied Mechanics and Materials, Trans Tech Publ. pp. 1824–1828.

17. Mahadevan, V., Li, W., Bhalodia, V., Vasconcelos, N., 2010. Anomaly detection in crowded scenes., in: CVPR, p. 250.

18. Ning, J., Zhang, L., Zhang, D., Yu,W. Joint registration and active contour segmentation for object tracking. IEEE Transactions on Circuits and Systems for Video Technology 23, 1589–1597, 2013

19. Nurhadiyatna, A., Jatmiko, W., Hardjono, B., Wibisono, A., Sina, I., Mursanto, P., 2013. Background subtraction using gaussian mixture model enhanced by hole filling algorithm (gmmhf), in: IEEE International Conference on Systems, Man, and Cybernetics, IEEE. pp. 4006–4011, 2013

20. Wang, T., Chen, J., Snoussi, H. Online detection of abnormal events in video streams. Journal of Electrical and Computer Engineering 2013, 20, 2013

21. Xiang, T., Gong, S., Video behavior profiling for anomaly detection. IEEE transactions on pattern analysis and machine intelligence 30, 893–908, 2008

22. Xu, Z., Tsang, I.W., Yang, Y., Ma, Z., Hauptmann, A.G. Event detection using multi-level relevance labels and multiple features, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 97–104, 2014

23. Zhang, Y., Lu, H., Zhang, L., Ruan, X. Combining motion and appearance cues for anomaly detection. Pattern Recognition 51, 443–452, 2016