Advances in Parallel Computing Algorithms, Tools and Paradigms
D.J. Hemanth et al. (Eds.)


doi:10.3233/APC220063

# Abnormality Detection in Human Action Using Thermal Videos

Kanimozhi Soundararajan[a,1], Anbarasi Soundararajan[b] and Sai Ramesh L[c]
[a,c]*Department of IST, College of Engineering, Anna University, Chennai*
[b]*Department of Computer Science, Meenakshi College for Women, chennai, India*

**Abstract.**Anomaly detection is a challenging task in the surveillance system due to the factors like extracting appropriate features, inappropriate differentiation among the normal vs abnormal behaviours, the sparse occurrence of abnormal activities and environmental variations. In the dark environment, detection of human actions is still difficult as more features for recognizing the key point are not visible. Hence the proposed work is focused on overcoming the environmental variations task that too in a less bright environment by using thermal videos. Variations in the actions can be easily identified as it works on the property of infrared radiations. For recognizing actions, the skeleton-based approach is used as it helps with the joint-wise segregation of human parts, resulting in more accuracy. The motion pattern of humans in the thermal video is tracked to classify the level of abnormality.

**Keywords:** Abnormality detection, Thermal Video, Action identification, (3DCNN)

## 1. Introduction:

Artificial intelligence, cognitive modelling, and neural networks are information processing paradigms inspired by the way biological neural systems process data. Artificial intelligence and cognitive modelling try to simulate some properties of biological neural networks. In the artificial intelligence field, it has been applied successfully to action and speech recognition, image analysis and adaptive control to construct software agents (in computer and video games) or autonomous robots. Among them human abnormal action recognition is a widely studied research area.Over the past years, a huge deal of work has been done in this as it attracted the attention of numerous researchers from different fields. It creates increased demand for public safety and security at crowded places such as shopping malls, restaurants, hospitals, and stations. Humans are often at the centre of such interactions and detecting human actions is an important practical and scientific problem. In order to develop more dynamic virtual worlds and smarter robots [1], it is necessary to teach machines to capture, understand and replicate these interactions. The information that is

[1]Kanimozhi Soundararajan, Department of IST, College of Engineering, Anna University, Chennai. E-mail: kanimozhiist21@gmail.com.

need to be learnt is widely available in the form of large video collections. The growing number of anomalies happening in indoor and outdoor environments calls for accurate and robust action recognition systems. These anomalies could vary from theft, destruction of public property or even fighting innocents. Identification of human features that are partially visible in the captured image is in examination stage only. Hence the goal of this work is

- To create a system that can identify human's actions
- The ability to automated understanding of action behaviour.
- Annotating actions as normal or abnormal using thermal videos.

## 2. Related Works:

### *2.1. Data Pre-processing*

Novel threshold based approach [2] using a new colour space model is proposed and embed it into background subtraction method for the retrieval of expected image regions. It uses colour distortion and brightness distortion of each pixel to detect the changes. The colour distortion considers the vector's position in the space so that it can assemble its features effectively. Moreover, this method removes the moving shadow also. By applying it, possibility of extracting entire foreground object is comparatively higher.

### *2.2 Action Recognition*

An end-to-end alert system [3] for real-time crime detection in low-light environments is developed unlike Closed-Circuit Television (CCT). The system uses feeds processed in real-time by an optical-flow network, spatial-temporal networks and a Support Vector Machine (SVM) to identify shootings, assaults and thefts. A method was proposed [4] to analyze global terrorist, threats, illegal migration an intensified concern for the security of citizens and protect people as well as their property. Due to the weather conditions and at night times RGB cameras [5] do not perform well, thermal cameras have become an important component of sophisticated video surveillance systems. Hence they use thermal images with convolutional neural network models originally intended for detection in RGB images. Another method [6] utilized near-infrared (NIR) and thermal cameras to solve the problem of visibility in dark environment. Compared to NIR cameras, thermal cameras enable long- and short-distance objects to be visible without an additional illuminator.

### *2.3 Abnormal Action Identification*

To identity abnormal behaviours [2] it is necessary to detect and recognize abnormal running movement along with its spatiotemporal parameters. Meanwhile, to obtain more accurate spatiotemporal parameters and improve the real-time performance, a multi-target tracking algorithm based on the intersection area among the minimum enclosing rectangle of the moving objects, is presented. With the increasing number of surveillance cameras [7] in both indoor and outdoor locations, there is a grown demand for an intelligent system that detects abnormal events. Although human action recognition is a highly researched topic in computer vision. Abnormal behaviour detection is lately attracting research attention a lot. Indeed, several systems are proposed in order to ensure human safety. Diverse methods [8] that are abound for building intelligent vision systems aimed at scene understanding and making correct semantic inference from the observed dynamics of moving targets. The anomalies [9] could vary from theft, destruction of public property or even fighting innocents. A new algorithm based on machine learning paradigm is developed to detect human actions and to label them as normal or abnormal. Tele-health care applications [10] have gained much attention in the field of ubiquitous computing. With availability of affordable wearable sensors, it is possible to recognize human activities. Abnormal activities are unexpected events that occur in random manner. Multi-class SVM based approach is most widely used method to recognize human activities.

## 3. System Design

Human actions that appear in videos vary due to their speed of motion, camera view, posture and pose variations make it as a challenging task. A good method of representing actions [11] should be accurate for estimating and characterizing actions effectively.

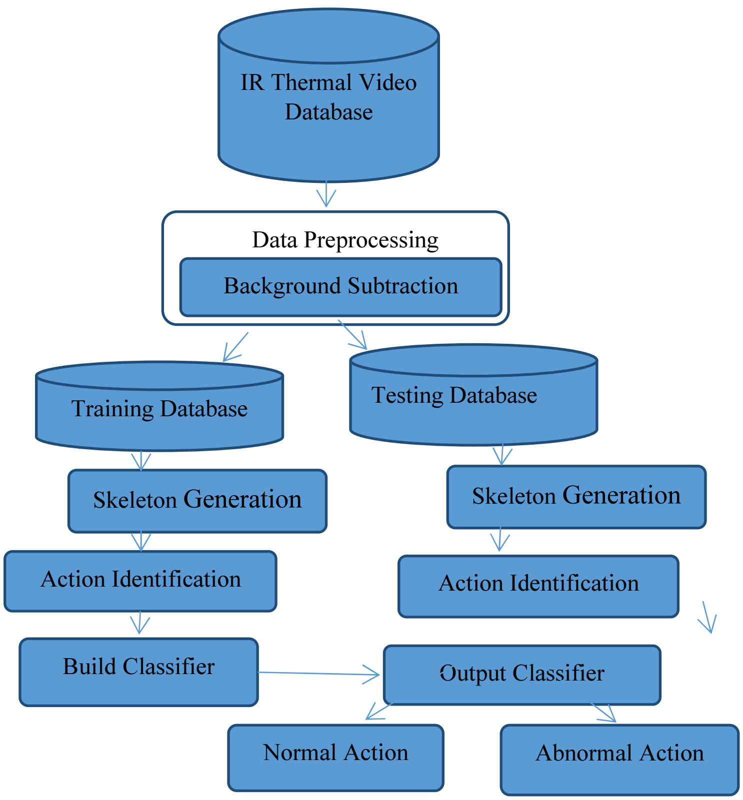


**Figure** 1: Proposed Abnormality Detection System Architecture

So an action video is to be turned into a feature vector that extracts representative and biased knowledge about human actions along with minimizes variations. Thus it increases the recognition efficiency. Detected representation is classified into global features and local features based on spatial and temporal variations. With help of data set, as represented in **Figure 1** data pre-processing is done where background subtraction is performed. Later by splitting database into training and testing skeleton generation for human action identification is taken place. With its resulted outcome actions are identified using 3DCNN and LSTM. Visual-temporal features help in classifying the final outcome either as normal or abnormal actions [12].

### *3.1 Action Identification*

Videos are taken as a sequence of frames. It is fed into CNN to extract the object information from the images. Open pose method [13] is used to fix the keypoint pointing to the human body joints. Skeleton structure with its keypoint feature values is used to decide the posture labels. In order to decide its corresponding actions, 3DCNN [14] method is used. It decides the action class based on the motion detection, motion tracking and post-processing information taken from the video.

### *3.2 Abnormality Recognition*

Abnormal activities are detected [15] by ruling out all possible activities that can be performed from the current activity and report the level of abnormalities in it. The proposed system recognizes 12 different activities such as fight, jump, punch, hug, walk, etc. As shown in **Figure 2**, CSV file is generated with identified action features. Followed by calculating the probability of data by the class they belong to, the so-called base rate [16]. Summarize data by class and calculate the Gaussian Probability Density Function.

$$f(x) = \frac{(1)}{\sqrt{2\pi\sigma_x{}^2}} e^{\frac{-(x-m_x)^2}{2\sigma_x^2}} \qquad \text{Eq-1}$$

Where $m_x$ is the mean value, $\sigma_x^2$ is the derived varience value. The predicted class helps in deciding whether action is normal or abnormal.

## 4 Experimental Setup

### *4.1 Dataset*

InfAR dataset consists of 600 video sequences captured by infrared thermal imaging cameras. Fight, Handclapping, handshake, hug, jog, jump, punch, push, skip, walk, wave 1 (one-hand wave), and wave 2 (two-hand wave) are included in the dataset. All clips have a fixed to a frame rate of about 30 FPS and resolution of 293 x 256. Totally it consists of 12 actions 600 clips at the duration of 4s.

### *4.2 Training Setup*

Training was carried out in colab platform with GPU processor. All methods used images with size of 224 × 224 pixels for training and testing. The initial learning rate is set to 3x104 and decayed over multiple epochs with patience set to 3. The learning rate was determined based on the epoch number. Initially videos are separated as clips which then undergo frame separation process. This is done for better understanding of human postures. Segmentation taken place at the region [17] where there is occurrence of humans in the frame. 3DCNN is employed to detect the action [18] happens in the clips. So a sequence of frames at the intervals of 5 each is chosen and fed as input to finalize the action label in the 3DCNN. This method recognizes the label semantically with the sequence of frames itself. At the end, the actions are identified with the normalized input values and fed into CSV file.

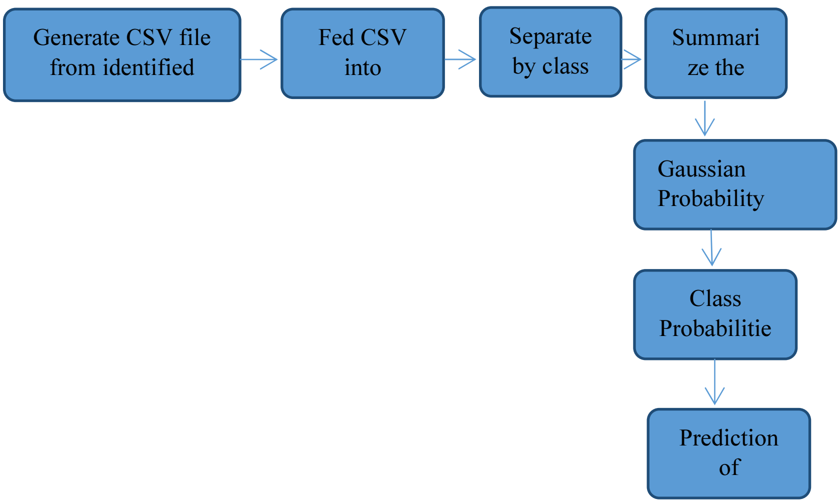


**Figure 2:** Abnormality Recognition Flow Chart

**Algorithm 1 Abnormal action identification**

**Input:** Thermal Video
**Output:** Video frames with action probabilities and abnormality message.

1. Read input video
2. Create destination frame folder
3. **while** not end of file **do**
4. Read frame using cam.read()
5. Write the extracted frames in destination frame folder
6. counter+=1
7. Create a dataframe with appended values for action labels
8. Open the directory in which frames are stored
9. Split the data for training and testing by slicing it.
10. Reshape the dataset-validationDatasetNew.reshape()
11. Feed the data set into CNN and LSTM - network.Model(inputs = input, outputs = output)
12. Load the model - 12. loadModel('/content/drive/MyDrive/cnnlstmmodelne w3.h5')
13. Predict the action with probabilities - tf.nn.softmax (prediction model)
14. Classify the action as abnormal or normal using classifier- predict (summary,row)
15. End While

### 4.3 Abnormal Action Identification

An end-to-end pipeline is designed by integrating 3DCNN with LSTM [8], followed by a time series pooling layer [19] and a softmax layer to predict the activities in video. The proposed system is then trained for abnormality classification using the algorithm-1 in which the possibility of actions along with its probability values is displayed. A CSV file containing 7 columns with action1, probability1, action2, probability2, action3, probability3, and behaviour either as abnormal-0 or normal-1. In **Table 1**, Actions like the fight, push and punch are interpreted as 0 while other actions as 1 as a resultant element.

**Table 1**: Action values normalizing algorithm

| |
|---|
| If(probability > =16) |
| Then probability=3 |
| else if (probability > =11 and < =15) |
| Then probability =2 |
| else |
| Then probability =1 |

### 4.4 Results

The abnormality detection results are assessed by predicting the right action within the given video which is tested for 9 classes. This system is developed in the aim that anomaly during the dark environment. So the work is carried out using thermal video with classified class highlighted over it.


**Figure**. **3**(a): Fight and Hug

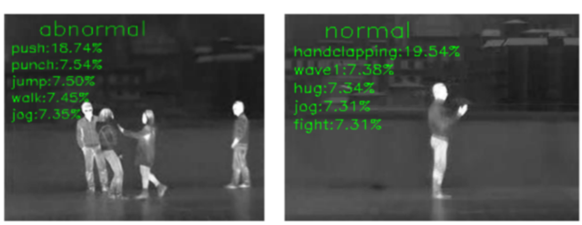


**Figure**. **3**(b): Push and Handclap

Based on the higher probability values that are retrieved for each possible action related to that clip the final decision is taken. Figure. 3(a) shows the screenshot result of two actions like fight and hugging with a probability of 17.63 % and 13.35 %. It helps in deciding the final targeted annotation is abnormal or normal. Similarly, push and hand-clapping score the greater values in Figure. 3(b). Hence the targeted class is displayed over it.

## 5 Performance Evaluation

**Table-2** shows the metrics of each class in comparison with other similar methods. Recall as mentioned in Eq-2 is the ratio of the total number of correctly classified positive examples to that of total positive examples.

**Table 2.** Performance comparison with state-of-art methods

| Methods | Precision | | Recall | | F1-score | |
|---|---|---|---|---|---|---|
| | Normal | Abnormal | Normal | Abnormal | Normal | Abnormal |
| Residual RNN | 0.78 | 0.43 | 0.87 | 0.79 | 0.84 | 0.7 |
| CNN-LSTM single frame | 0.89 | 0.5 | 0.72 | 0.8 | 0.6 | 0.62 |
| Thermal AAI | 0.82 | 0.53 | 0.50 | 0.83 | 0.62 | 0.65 |

Precision refers to the total number of correctly classified positive examples by the total number of predicted positive examples which is measured based on Eq-3. F-measure helps in having a measurement that represents both precision and recall values based on Eq-4.

$$Recall = \frac{TP}{(TP+FN)} \qquad \text{Eq-2}$$

$$Precision = \frac{TP}{(TP+FP)} \qquad \text{Eq-3}$$

$$F\ measure = \frac{(2*precision*recall)}{(Recall+Precision)} \qquad \text{Eq-4}$$

Here TP denotes True Positive value, TN denotes True Negative value, FP denotes False Positive value, FN denotes False Negative value.

$$ROC\ Curve = 1 - specificity(x - axis)\ and$$

$$sensitivity\ \ (y - axis) \qquad \text{Eq-5}$$

To measure the performance of a binary classifier the receiver operating characteristic (ROC) curve is used and its result is shown in **Figure-4**. It is created by plotting the true positive rate (TPR) (or recall) against the false positive rate (FPR) Eq-5.

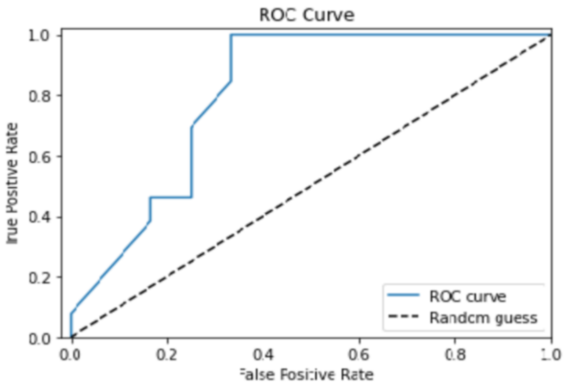


**Figure 4.** ROC curve for the proposed method

## 6 Conclusion and Future work

Abnormality in human action identification using thermal videos focused on increasing the awareness of the public administration about the events taking place in it, promptly reacting to problem situations, and as a result increasing the safety of residents. The proposed system achieved this using 3DCNN and LSTM along with algorithm-1. Different type of actions like fight, hand clapping, handshake, hug, jog, jump, punch, push, skip, walk, wave is taken into consideration from the dataset. Thermal video is given as input in order to detect the abnormality in a dark environment. Future enhancement of this work can be deployed in an IoT infrastructure that parses the results to the cloud. With the help of user interface like a mobile application, notification can be sent to the nearly related persons with video of the crime scene to corresponding authorities. Citizens will then have a public app that enables law enforcement to push crime alerts based on user proximity.

## References

[1] B.-J. Lee, J. Choi, C.-Y. Lee, K.-W. Park, S. Choi, C. Han, D.-S. Han, C. Baek, P. Emaase, and B.-T. Zhang. Perception-action-learning system for mobile social-service robots using deep learning. In Proceedings of the AAAI Conference on Artificial Intelligence, volume 32, 2018

[2] L. Jia and Y. Liu. A novel thresholding approach to background subtraction. In 2008 IEEE Workshop on Applications of Computer Vision, pages 1–6, 2008

[3] M. Potter, H. Gridley, N. Lichtenstein, K. Hines, J. Nguyen, and J. Walsh. Low-light environment neural surveillance. In 2020 IEEE 30th International Workshop on Machine Learning for Signal Processing (MLSP), pages 1–6, 2020

[4] M. Kristo, M. Ivasic-Kos, and M. Pobar. Thermal object detection in difficult weather conditions using yolo. IEEE Access, 8:125459–125476, 2020

[5] G. Batchuluun, D. T. Nguyen, T. D. Pham, C. Park, and K. R. Park. Action recognition from thermal videos. IEEE Access, 7:103893–103917, 2019

[6] M. Ramanathan, W. Yau, and E. K. Teoh. Human action recognition with video data: Research and evaluation challenges. IEEE Transactions on Human-Machine Systems, 44(5):650–663, 2014

[7] A. B. Mabrouk. Abnormal behavior recognition for intelligent video surveillance systems: A review. Expert Systems with Applications, 91:480–491, 2018.

[8] O. P. Popoola and K. Wang. Video-based abnormal human behaviour recognition—a review. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 42(6):865–878, 2012.

[9] O. A. Elsayed, N. A. Mohamed Marzouky, E. Atef, and M. A.. Salem. Abnormal action detection in video surveillance. In 2019Ninth International Conference on Intelligent Computing and Information Systems (ICICIS), pages 118–123, 2019

[10] A. Palaniappan, R. Bhargavi, and V. Vaidehi. Abnormal human activity recognition using svm based approach. pages 97–102, 2012.

[11] M. Asadi-Aghbolaghi, A. Clap´es, M. Bellantonio, H. J. Escalante,V. Ponce-L´opez, X. Bar´o, I. Guyon, S. Kasaei, and S. Escalera. Deep learning for action and gesture recognition in image sequences: A survey.In Gesture Recognition, pages 539–578. Springer, 2017

[12] Ahir, P. K. Pateriya, D. Kaur, and M. K. Rai. A review on abnormal activity detection methods. In 2018 International Conference on Computational Techniques, Electronics and Mechanical Systems (CTEMS), pages 521–526. IEEE, 2018.

[13] B. Fernando, E. Gavves, J. M. Oramas, A. Ghodrati, and T. Tuytelaars. Modeling video evolution for action recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 5378–5387, 2015.

[14] D. Singh and C. K. Mohan. Graph formulation of video activities for abnormal activity recognition. Pattern Recognition, 65:265–272, 2017.

[15] Y. Fu, T. Liu, and O. Ye. Abnormal activity recognition based on deep learning in crowd. In 2019 11th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC), volume 1, pages 301–304. IEEE, 2019.

[16] G. Morales, I. Salazar-Reque, J. Telles, and D. D´ıaz. Detecting violent robberies in cctv videos using deep learning. In AIAI, 2019.

[17] W. Liu, C. Zhang, H. Ma, and S. Li. Learning efficient spatial-temporal gait features with deep learning for human identification. Neuroinformatics, 16(3):457–471, 2018

[18] H.-B. Zhang, Y.-X. Zhang, B. Zhong, Q. Lei, L. Yang, J.-X. Du, and D.-S. Chen. A comprehensive survey of vision-based human action recognition methods. Sensors, 19(5):1005, 2019

[19] P. Wang, H. Liu, L. Wang, and R. X. Gao. Deep learning-based human motion recognition for predictive context-aware human-robot collaboration. CIRP annals, 67(1):17–20, 2018.