International Journal of Future Generation Communication and Networking
Vol. 13, No. 1s, (2020), pp. 76- 83

Real-time Activity Detection & Recognition in Video

Ketki Salunkhe[1], Priyanka Rajaram[2], Samidha Raut[3] and Samidha Kurle[4]
[1, 2, 3, 4]*Department of Computer Engineering, University of Mumbai,*
Atharva College of Engineering, Malad, Mumbai, India

Abstract

Real-time video recognition and identification is the ability to analyze video automatically to identify and evaluate temporal events that are not dependent on a single image. It is the process by which a video is processed, data gathered and data analyzed to obtain domain-specific information. Object Detection is the method of identifying instances of real-world objects. It takes into consideration to recognize, localize, and detect multiple objects within an image, in video real-time. This paper is about a security monitoring system that uses the identification activity of the human body to build improved surveillance and also detect suspicious objects. Nevertheless, no unusual circumstances are detected by people, rather computers that analyze the recorded pictures and identify suspicious behavior or incidents. In this paper, we have described a system to distinguish and perceive suspicious activity. We are using the OpenPose Python API for our conceptual development.

Keywords: *Real-time video recognition and detection, object detection, security monitoring system, OpenPose*

1. Introduction

The real-time activity recognition and detection through video do the job of analyzing video sequences to detect and recognize the activities and the objects [7]. This paper is about a security monitoring system that uses the identification of activity of the human body and the detection of items to establish surveillance security. Humans do not identify unusual circumstances, though, but computers that analyze the photos collected and recognize suspicious behavior or incidents and artifacts [5]. We've used OpenPose Python. The system records recordings and screens a human's actions. The human face and history of human behavior play a significant role in the recognition of individuals.

Video surveillance has become a major concern in everyday life nowadays with the increasingly growing needs of citizen's protection and personal assets [9]. A consequence of these requests has resulted in cameras being mounted almost everywhere. Most current video surveillance systems have one function, they need a human operator to track continually, displaying the images collected by the cameras. We can use a method called OpenPose that records videos and identifies suspicious activities. Much of this research is devoted to the visual analysis of human actions and the identification of unusual human movements and suspect objects in exam centers [5, 7].

2. Related work

Cao Zhe et al. [1] introduced a constant way to deal with recognition of 2D posture of numerous individuals in a picture. The proposed approach utilizes a non-parametric model, called Part Affinity Fields (PAFs), to figure out how to associate parts of the body with entities in the picture. This bottom up approach accomplishes high accuracy and precision progressively, paying little heed to the number of individuals on the image.

Tomas Simon et al. [2] introduced a methodology that utilizes a multi-camera framework to train fine-grained indicators for key impediment inclined focuses, for example, hand joints. This offers an method

ISSN: 2233-7857 IJFGCN

to improve the efficiency of a key-point detector utilizing a multi-camera set-up. Making the technique sufficiently strong to work with less cameras can be considered as future work.

Shih-En Wei et al.[3] have demonstrated a comprehensive method to show how convolutional systems can be executed into the posture framework outline structure for learning picture highlights and setting subordinate spatial models for present estimation undertakings. This leads to the implicit simulation of long-range dependencies in organized prediction activities such as articulated pose estimation between variables. This demonstrated that a sequential architecture made up of convolutional networks is capable of implicitly learning a spatial model for pose by expressing even more sophisticated uncertainty-preserving beliefs between levels. Nevertheless, the management of several individuals in a single end-to-end system was a demanding solution issue.

Wei Niu et al.[4] developed a system for the monitoring and identification of human behavior for applications including outdoor video surveillance. This work has enhanced robustness by offering frameworks for smart control and fail-over based on low-level movement tracking algorithms such as frame separation and movement identification and correlation of tracking features. On the other hand, it dealt with the problem of real-time monitoring and action recording in a rigorous way.

Rajat Singh et al.[5] suggested the use of an advanced Gaussian mixture model for the segmentation of artifacts from the background. To trace it out, it distinguishes a person carrying or leaving an item and removes the object from the entity. The tracking algorithm sees the person as a whole from frame to frame, it doesn't track the individual parts like limbs.

Seyed Yahya Nikoueia et al.[6] used the CNN edge model to incorporate abstraction and fuzzy decision-making. The calculations for separating highlights from approaching video streams are sent on an edge framework which successfully decreases overhead correspondence and allows for outsourcing of the decision-making cycle to the level of fog.

Gowsikhaa D et al. [7] explained a methodology for suspicious human activity detection through face recognition. In this paper, initially Gabor filter is used to pre-process the video. Background estimation and foreground extraction are done and Artificial Neural Networks are used to detect human faces. The outcome is then used to recognize head movement. Further a blend of Motion Detection, Edge Detection and Skin Color Detection is performed to distinguish the hands of students. The marked faces and hands are boxed out by the system in a green color. When faces or hands exceed a certain threshold, the machine switches the color of the box from green to red, signaling that unusual behavior is taking place. Also under varying lighting and illumination conditions the system can be further improved to achieve acceptable results.

Karishma Pawar and Vahida Attar[8] intended to test and evaluate profound learning strategies for the identification of anomalous activity based on video. Deep learning approaches are contrasted from viewpoints of both precision-based anomaly detection and real-time anomaly detection driven computation. As a part of the analysis, the graphical taxonomy was established based on types of anomalies, extent of detection of anomalies and anomaly estimation for identification of anomalous behavior.

Ramachandran Sumalatha et al.[9] used optical flow to obtain the pattern of the motion vector in the flow and the neural deep convolution network is used to train the images and to obtain the correct expected value. The input videos are converted into frame sequences, and after collecting images that were analyzed to train and check the captured data set in the datastore, ROI is extracted from the sequential

ISSN: 2233-7857 IJFGCN

structure.

T. Senthil and G. Narmatha [10] focused on automated detection of the face from a pre-processed surveillance video. First, the foreground object is extracted and the Haar cascades sense the face region. Activities are then categorized to determine whether they are normal or irregular, depending on the direction of the face, the identification of hand contact using the background subtraction, and the Gaussian Mixture Model (GMM). This technique detects unusual behaviors that are frequently observed, such as the swapping of objects, peeping into other people's answer sheets, and individuals sharing information during the test.

The approaches described above observe some instances of failure. Some fail when there are multiple people close by. In extremely crowded photos where people overlap, one of the solutions appears to combine various people's annotations, while ignoring others that make the greedy multi-person parsing fail. Another tracking algorithm often considers the person as a whole from frame to frame, it doesn't track individual parts like limbs, hands. In some cases, the approach fails to detect them if the person is far away from the camera. While in other cases, the time taken to calculate the result is also more.

Our use of OpenPose provides greater precision than previous approaches, particularly for face and hand key-point detection, generalizing faces and hands with better to occluded, blurred, and low resolution. The technique suggested may be highly effective provided that the arrangement will be an enclosed area during the test.

3. Proposed system

The system captures the real-time video as input and further processes to detect and recognize the objects and human activities. The overview of the proposed approach to detect human activities and objects is shown in Figure 1.

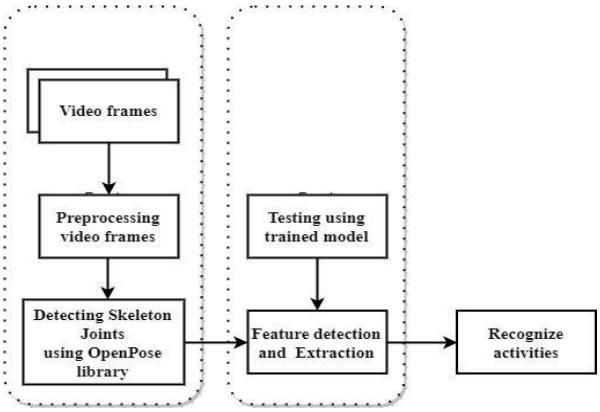

Figure 1. System Architecture

1. Pre-processing

●Background extraction: For human segmentation from the background, or elimination of either noise, might be required.

●Creation of a bounding box: It generates boundary boxes for each human present in the image. For detecting Human Pose, each bounding box is then evaluated separately.

2. Feature Extraction

Feature extraction used for generating derived values from raw data, which can be used as input to a learning algorithm. Explicit features contain conventional Computer Vision based features such as

ISSN: 2233-7857 IJFGCN

Oriented Gradient Histogram. Implicit features correspond to deep learning based feature maps such as outputs from Deep Convolutional Neural Networks.

3. Inference

● Confidence Maps: The certain way of predicting joints location is through the creation of confidence maps for each joint. Trust maps are a probability distribution over the image, representing the confidence of the joints located in each pixel.

● Bottom-Up Approach: Bottom-up approaches involve first detecting the parts or joints for one or more humans in the image, and then assembling the parts and associating them with a particular human. In short term this algorithm initially predicts all the body parts / joints present in the image.

● Top-Down Approach: Top-down pose estimation can be split into approaches based on the generative body-model and deep learning. A body model based generative approach involves trying to fit a body model onto the picture, enabling the final prediction to be human-like.

4. Post-processing

There are many algorithms, including both bottom-up and top-down techniques, that have no relation constraints on the final performance. The output pose from any Pose Estimation pipeline is passed through an algorithm of learning, which scores each pose based on its possibility.

4. Methodology

OpenPose

OpenPose is a bottom-up approach for identifying multi-person human pose estimation. It first detects parts that belong to each person in the image, then assigns parts to distinct individuals. The characteristics are then fed into two parallel divisions of convolutional layers. The first branch forecasts a series of 18 confidence charts, each of which depicts a particular part of the human pose skeleton. Successive stages are used to refine each branch's predictions. Bipartite graphs are built between pairs of parts using the component trust maps. Weaker ties in the bipartite graphs are pruned using the PAF values.

Kalman Filter

Kalman Filter helps us to model tracking based on an object's location and distance, and to forecast where it is likely to be. This uses Gaussians to predict future positions and velocities. When a new reading is obtained it may use the likelihood to add the measurement to its forecast and update itself. It's memory-light and running fast. And as it uses both position and motion velocity it has better results than the centroid-based tracking.

Deep SORT

It records not only the size, the velocity of what that person looks like. Deep sort enables us to attach this functionality to the tracking logic by computing deep features for each bounding box and using the similarities between deep features as well.

- **Dataset:**

We have trained our very own dataset for object detection as well as on activity detection and tested it on videos.

Object detection dataset(1000+images of various suspicious object such as Cellphone, Paper-cheat, Smartwatch).

Activity detection dataset(x and y coordinates of hands, limbs, and face are considered to segregate it with different actions such as stand, walk, bend, wave).

ISSN: 2233-7857 IJFGCN

5. Results

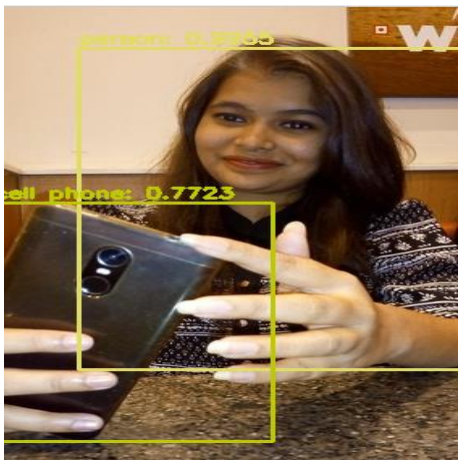

Figure 2. Real-time Object Detection

Figure 2 shows the real-time object detection. At the point when objects such as cellphones, books, smartwatch, and paper cheats are detected, a bounding box alongside a warning message is generated.

ISSN: 2233-7857 IJFGCN

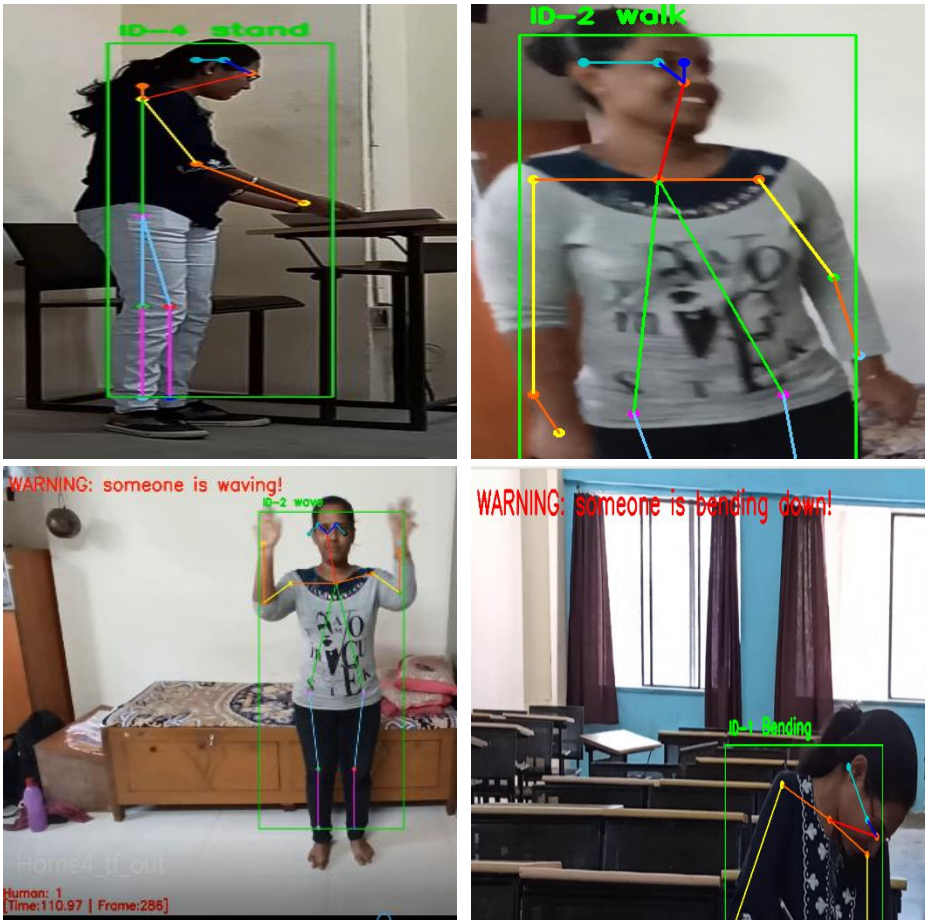

Figure 3. Key-points Detection Using OpenPose

As shown in figure 3, the system detects and recognizes continuous human exercises, for example, standing, walking, waving, bending. The activities such as waving and bending down are considered as suspicious activities in examination centers, so warning message is generated when these activities are detected.

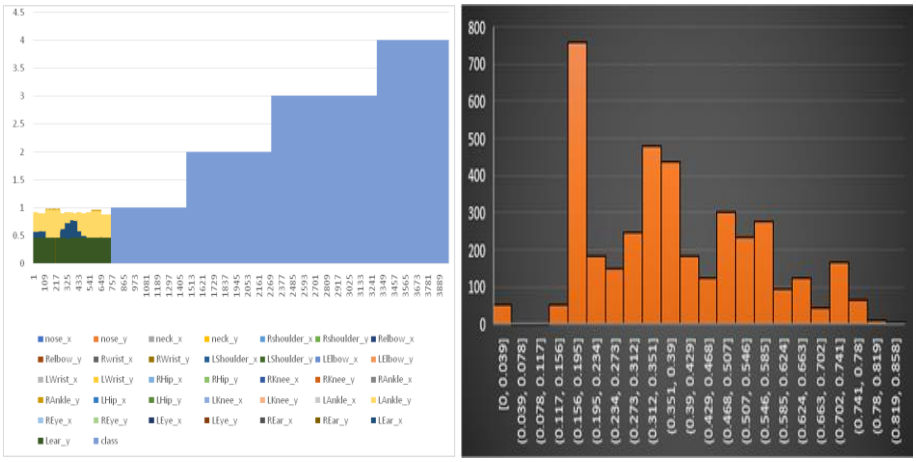

Figure 4. Graphs of Dataset

ISSN: 2233-7857 IJFGCN

6. Applications

- The proposed system can be used for bank robbery detection.
- It can also be used for developing a patient monitoring system.
- Detecting and reporting suspicious activities at the railway station is also one of the important applications of the proposed system.
- It can also be used for developing Security related applications such as Defense, Military, checking at airports.

7. Future scope

Potential avenues for future research incorporate utilizing multiple cameras' perspectives and exploring strategies for keeping up object identities in the tracker better. Typically real-world situations are more complex in terms of the number of people involved in the events, than the circumstances we referenced here. In this way, to think about taking care of these complexities, sophisticated algorithms are required. The methodologies that we presented for video observation show promising outcomes and can be utilized as a basis for advanced video surveillance research.

8. Conclusion

This paper proposes an automated video monitoring system that can analyze people's actions and recognize suspicious practices. It provides a high-quality analysis of body poses and preserves efficiency regardless of the number of people. We have described in this paper the key-point association which encodes both the position and orientation of human appendages. Second, we have checked images on our own dataset which we generated for detection of suspicious objects. Thirdly, we used a deep sort algorithm to detect human activities in video or real-time. The information extracted through the video stream will be valuable in performing suspicious incident analysis. We also developed a working prototype to exhibit the capacities of our proposed system when applied to a pragmatic system.

Acknowledgement

We would like to express our sincere thanks to Prof. Samidha Kurle, for her co-operation and guidance. We would also like to thank our Computer Department HOD, Prof. Suvarna Pansambal, our Project Coordinators Dr. Mamta Meena, Prof. Shweta Sharma, and all the Computer Department staff who have directed us to grow this project concept throughout.

References

[1] Zhe Cao, Gines Hidalgo, Tomas Simon, Shih-En Wei, and Yaser Sheikh, "OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields", IEEE Transactions on Pattern Analysis and Machine Intelligence- 2019.

[2] Tomas Simon, Hanbyul Joo, Iain Matthews, Yaser Sheikh, "Hand Keypoint Detection in Single Images using Multiview Bootstrapping", IEEE Conference on Computer Vision and Pattern Recognition (CVPR) – 2017.

[3] Shih-En Wei, Varun Ramakrishna, Takeo Kanade, Yaser Sheikh, “Convolutional Pose Machines", IEEE Conference on Computer Vision and Pattern Recognition (CVPR) -2016.

[4] Wei Niu, Jiao Long, Dan Han, and Yuan-Fang Wang, “Human Activity Detection and Recognition for Video Surveillance”, IEEE International Conference on Multimedia and Expo (ICME) (IEEE Cat. No.04TH8763), 2004.

[5] Rajat Singh, Sarvesh Vishwakarma, Anupam Agrawal, M.D Tiwari, “Unusual activity detection for video surveillance ”, Proceedings of the First International Conference on Intelligent Interactive

ISSN: 2233-7857 IJFGCN

Technologies and Multimedia - IITM '10, 2010.

[6] Seyed Yahya Nikoueia, Yu Chena, Alexander Avedb, Erik Blaschb, Timothy R. Faughnanc, "I-SAFE: Instant Suspicious Activity identiFication at the Edge using Fuzzy Decision Making", presented at the Fourth ACM/IEEE Symposium on Edge Computing, Washington DC,2019.

[7] Gowsikhaa D, Manjunath, Abirami S, "Suspicious Human Activity Detection from Surveillance Videos", (IJIDCS) International Journal on Internet and Distributed Computing Systems. Vol: 2 No: 2, 2012.

[8] Karishma Pawar, Vahida Attar, "Deep learning approaches for video-based anomalous activity detection".

[9] Sumalatha Ramachandran, Lakshmi Harika Palivela and C. Giridharan, "An intelligent system to detect human suspicious activity using deep neural networks".

[10] T. Senthil Kumar and G. Narmatha, "Video Analysis for Malpractice Detection in Classroom Examination", Conference on Soft Computing Systems, Advances in Intelligent Systems and Computing 397, 2016.

ISSN: 2233-7857 IJFGCN