IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 8, June 2022
Impact Factor: 6.252

Vision Based Daily Activity Monitoring in an Indoor Environment

Prof. Poonam Nagale[1], Kshitij Sabale[2], Omkar Patole[3], Prasad Patil[4], Aman Bharuka[5]
Professor, Department of Computer Engineering, NBN Sinhgad School of Engineering, Pune[1]
Students, Department of Computer Engineering, NBN Sinhgad School of Engineering, Pune[2,3,4,5]

Abstract: *Video surveillance is an active area of research. Video surveillance systems have always been spread widely and are common indifferent environments. Video surveillance is a primary component in providing security atbanks, casinos, airports, and other institutions. More recently, government agencies, businesses, and even schools are turning toward video surveillance as a means to increase public security. This system is a review of many existing video surveillance systems. With the growing size of security footage, it becomes important that surveillance systems are able to support security in tracking and monitoring activities. The aim of the surveillance systems is to classify, detect and track targets. In this project, we have described object recognition, detection, and ways to use such a system.*

Keywords: Vision, Artificial Intelligence, Machine Learning, Convolution Neural Network, YOLO, Vehicle Crashdetection, Fall Detection, Social Distance Monitoring, etc.

I. INTRODUCTION

Due to exponential increase in crime rate, surveillance systems are being put up in malls, stations, schools, airport sect. With the videos being captured all the time from these cameras, it is hard to manually monitor them to detect suspicious activities. So, there is a great demand for intelligent surveillance system. The proposed work automaticallydetects multiple anomalous activities in videos and reports them to the respective authorities. The proposed frameworkincludes three main features: Crash detection, fall detection and social distance detection. Using these features with a surveillance system would be a smart approach than having to manually monitor surveillance cameras.

Surveillance system using videos is a very important in the field of security. The task of surveillance is to detect andtrack moving objects in the video sequence. Now-a-days, surveillance system using video is an important security assetfor commercial, law enforcement and military applications. In video surveillance, detection of moving objects from a video is necessary for object classification, target tracking, activity recognition as well as behavior understanding. Simple systems detect motion in a camera's field of view. More advanced systems offer features such as object identification, tracking and analysis, perimeter intrusion detection, and abandoned/removed object recognition. Surveillance system using video is used to monitoring of the behavior, activity or other information generally of people in a specific area. The application of video surveillance is now not only limited to provide security of an area. Such systems are now implemented in different sectors like hospital for monitoring patient, in industry and process plant to monitor the activity of the production line etc. Generally, a video surveillance system consists of a video camerafor capturing video footage and a monitor to see the capturing footage. Surveillance system has been a popular securitytool for years.

II. LITERATURE SURVEY

[1] Fall Detection for Elderly People using Machine Learning. The system detects falls by classifying different activities into falls and non-falls and notifies an elderly relative or caregiver in an emergency. The SisFall database with a variety of multi-participant activities is used to calculate important features. Machine learning algorithms likeSVM and decision tree are used to detect falls on the basis of calculated features. The system achieves an astonishingaccuracy of up to 96% by using decision tree algorithm.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 8, June 2022
Impact Factor: 6.252

[2] Fall Detection System for the elderly. The system defines two major components: a wearable device and a cell phone. The wearable has the capability of communicating with a cell phone can be located in a 100ft radius. Once, thewearable device detects a fall, it sends an alert to the cell phone; then the cell phone alerts to the emergency contacts defined by the user. The main idea is to avoid the need to carry a cell phone regularly. Additionally, our system has apanic button that can be used to alert emergency responders in the event that a user feels that a fall may occur.

[3] Vehicle Accident Detection System using Internet of Things. The idea of accomplishing this was to install a sensor and a microcontroller in the car so that to calculate vehicle movement. The system is built using a vibrating sensor to determine the collision impact of an accident and a gyro sensor to determine the x-y displacement of the vehicle. When an accident occurs, the instantaneous coordinates of the vehicle will be captured using a GPS module and transmitted to the emergency response department via a GSM module.

[4] Vehicle Accident Detection & Alert System using IoT and Artificial Intelligence. This paper analyzes and suggests how IoT can be used in this regard in a way that can save thousands of lives. Along with IoT, we have included machine learning and image processing techniques to accurately identify road hazards. Sensors like accelerometer, gyroscope, camera, etc. provide data to the same microprocessor the sensor data with the machine learning model and determines if there is an accident or not and if it is, the device sends the related metrics to the server through the internet.

[5] Social Distancing Detection with Deep Learning Model. The paper presents a methodology for detecting social distance using in-depth reading to explore distance between people to reduce the impact of this coronavirus virus. A vision tool was developed to warn people to keep a safe distance from each other by checking the video feed. The videoframe from the camera is used as input, as well as a pre-trained open-source discovery model based on the YOLOv3 algorithm used for pedestrian detection.

III. CHALLENGES

In cases where the wearable of any one person is not being detected, it could hinder the whole agenda of the project making the project vulnerable.

1. **Cost Effective Methodology.**
   In order to do any of the detection, using traditional approach would require development of a hardware. This hardware should be wearable or could be easy to carry. This hardware would have various sensors, wires, processorsto feed the data to the neural network to generate an output. This would increase the overall cost of the project. Hence a major challenge was to make this project as cost effective as we could.
2. **Challenges for Fall Detection**
   There are several challenges in detecting human fall. The traditional way to detect falls or any kind of human activity was to generate time series data by using a gyroscope or accelerometer sense. Using the time series data, the machine learning model is trained and then changes are comparedand if matched it would tell if the person has fallen or not. The Major challenge is the system requires the use of wearables. The subject has to wear the device at all times and has to keep it on. Moreover, the device needs some power and thedata must be constantly fed to the neural network. The aim of our project does not require us to use any wearable and should detect any human activity and classify it as the person falling or tripping.
3. **Challenges for Vehicle Crash Detection**
   There are several proposed methodologies. Most of them focus on using safety features deployed in cars as a key part of the system. Such systems would have to be installed in the car in order to detect their crash. These systems would be mostly electronic. Hence, they would need timely maintenance. Since if any of these parts break or don't function, the whole system would fail if unfortunately, a crash occurs. Another major challenge in detecting car crashes is that such systems would be present in newer cars but won't

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 8, June 2022
Impact Factor: 6.252

be present in older cars. Since older cars run predominantly on roads and the road presence of transport vehicles like trucksand lorries is a lot more, these trucks and lorries are pretty old and outdated. So, these won't have the System.

4. **Challenges for Social Distance Detection**
   The main objective for Social Distance Detection is detecting humans and people first. The primary objective itselfrequires Detection of human beings. The same can be done using wearables and to detect if any wearable is close to each other. The problem with this is that every person in the population has to be wearing the wearable.

IV. DATASET

Common Objects in Context (COCO) Common Objects in Context (COCO) is a database that aims to enable future research for object detection, instance segmentation, image captioning, and person key points localization. COCO is a large-scale object detection, segmentation, and captioning dataset. Features of the COCO dataset includes- Object segmentation with detailed instance annotations, Recognition in context. Super pixel stuff segmentation, Over 200'000 images of the total 330'000 images are labelled, 1.5 Mio object instances, 80 object categories, the "COCO classes", which include "things" for which individual instances may be easily labelled (person, car, chair, etc.), 91 stuff categories, where "COCO stuff" includes materials and objects with no clear boundaries (sky, street, grass, etc.)that provide significant contextual information., 5 captions per image, 250'000 people with 17 different key points, popularly used for Pose Estimation.

YOLOv4

YOLO is an abbreviation for the term 'You Only Look Once'. This is an algorithm that detects and recognizes various objects in a picture (in real-time). Object detection in YOLO is performed as retreat problem and provides class opportunities for captured images.

YOLO, the algorithm uses convolutional neural networks (CNN) to detect real-time objects. As the name suggests, the algorithm requires only one forward stream through the neural network to find objects.

This means a prediction throughout the image is created by running a single algorithm. The CNN is used to predict various class probabilities and bounding boxes simultaneously. The YOLO algorithm consists of various variants. Some of the common ones include tiny YOLO and YOLOv3. YOLO forms a key role in our project.

Proposed Approach

YOLO algorithm runs using the following three steps:

1. Residual blocks
2. Bounding box regression
3. Intersection Over Union (IOU)

1. **Residual Blocks**
   Initially, the input image is divided into different grids. Each one of the grids has a dimension of S x S. The following example image shows how an image is divided into different grids.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 8, June 2022
Impact Factor: 6.252

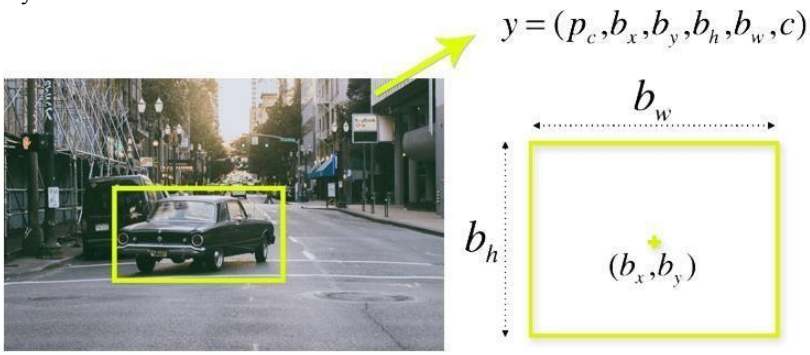
If we observe in the image above, there are many different grid cells of equal dimension. Every grid cell has the responsibility to detect objects that appear inside of them. For an example, if an object center appears within a certain grid cell, then this cell will be responsible for detecting that particular object.

2. **Bounding Box Regression**

A bounding box can be defined as an outline that highlights any particular object in an image. Every bounding boxin the image consists of the following things:

- Width (bw)
- Height (bh)
- Class (for example, car, person, flower pots, traffic lights, etc.)-This is represented by the letter
- Bounding box center (bx,by)

The following image shows an example of a bounding box. The bounding box has been represented by a yellowoutline.

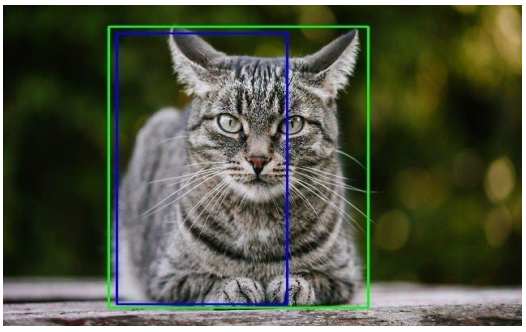


YOLO uses a single bounding box regression to predict the height, width, center, and class of objects. In the imagepresent above, it shows the probability of the object that appears in the bounding box.

3. **Intersection Over Union (IOU)**

Intersection over union (IOU) is a technique in object recognition and detection that tells us how boxes overlap. YOLO uses IOU to provide an output box that surrounds the objects perfectly.

Each grid cell is responsible for predicting the bounding boxes and their confidence scores. The IOU is equal to 1if the predicted bounding box is the same as the real box. This mechanism eliminates bounding boxes that are not equalto the real box. The following image provides a simple example of how IOU works.



In the image above, there are two bounding boxes, one in green and the other one in blue. The blue box is thepredicted box while the green box is the real box. YOLO ensures that the two bounding boxes are equal.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 8, June 2022
Impact Factor: 6.252

V. THE SYSTEM STEPS

For the system to execute the following steps are done.

1. Installation of necessary libraries like opencv, darknet, matplotlib.
2. Download and install YOLOv4 in the target system.
3. Once the installation is complete, Test the trained model on a test video.
4. Upon successful installation, the model is capable of detecting each and every object in the video with aconfidence score and a bounding box.

1) For Fall Detection The steps for fall detection

1. Once the model is loaded, we need to instruct the model to only detect classes related to human beings.
2. Upon successful execution of Step1, we get a clear detection of human beings in a video input.
3. For every detection, a bounding box appears around the person.
4. The dimensions of the bounding box are then checked and an output is reached.
5. The Major check is if the length of the bounding box ever becomes greater than the width of the bounding box.A fall is detected.
6. Due to inconsistencies of the model, sometimes false positives can be detected.
7. To tackle the above problem, we have made sure that if the fall is detected continuously for a threshold value,then the next step is taken place.
8. A screenshot is taken of the output and is then sent to the email of the admin issuing an alert.

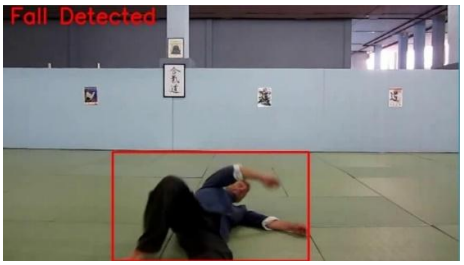
2) For Vehicle Crash Detection

1. Once the model is loaded, we need to instruct the model to only detect classes related to vehicles.
2. Upon successful execution of Step1, we get a clear detection of vehicles in a video input.
3. For every detection, a bounding box appears around the vehicle.
4. The dimensions of the bounding box are then checked and an output is reached.
5. The Major check is if the bounding boxes of two or more vehicles intersect, a crash is detected,
6. A screenshot is taken of the output and is then sent to the email of the admin issuing an alert.

VI. RESULT

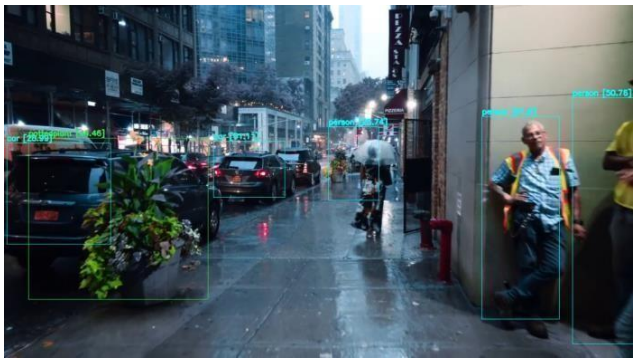
With the use of the above methodologies, we got the following output.

The above image gives us an overview of how the crash detection mechanism works.



The above screenshot gives us an overview of how the fall detection mechanism works.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 8, June 2022
Impact Factor: 6.252



The above screenshot gives us an overview of how YOLOv4 detects objects.

The above screenshot gives us an overview of how the social distance detection system works.

VII. CONCLUSION

Video surveillance is an active area of research. Video surveillance systems are widespread and common in many environments. This system is review of many existing video surveillance systems. With the growing quantity of security video, it becomes vital that video surveillance system be able to support security personnel in monitoring and tracking activities. The aim of the surveillance applications is to detect, track and classify targets. On classifying it as suspicious it alertsthe system and therefore, we can detect such activity and also prevent them.

REFERENCES

[1] Ľuboš OVSENIˊK, Anna KAŽIMÍROVÁ KOLESÁROVÁ Ján TURAˊ N, "Video Surveil- lance System" Acta Electrotechnical et Informatica, Vol. 10, No. 4, 2010, pp. 46–53
[2] A. Lipton, et al, Moving Target Classification and Tracking from Real-time Video - IEEEWACV, 1998, pp.8-14.
[3] Mehrsan Javan Roshtkhari, et al, "An on-line, real-time learning method for detecting anomalies in videos usingspatio-temporal compositions."-2013 Elsevier Inc.
[4] Sarita Chaudharya, et al, "Multiple Anomalous Activity Detection in Videos" -2017, Procedia Computer Science125 (2018) 336345
[5] Chen Change Loy, et al, Detecting and discriminating behavioural anomalies - 2010Elsevier Ltd.
[6] A. Abrantes, J. Marques, J. Lemos, "Long Term Tracking Using Bayesian Networks" in IEEE ICIP, 609-612,2002.
[7] A. Cavallaro, et al, "Objective Evaluation of Segmented Quality Using Spatio-Temporal Context"-in Proc. ofIEEEICIP, 22-25 September 2002.
[8] D. Comaniciu, et al, "Kernel-Based Object Tracking" IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 25, No. 4, April 2003.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 8, June 2022
Impact Factor: 6.252

[9] F. Oberti, et al," ROC curves for performance evaluation of video sequences processing systems for surveillanceapplications"- DIBE-University of Genoa.
[10] G. Foresti, C. Regazzoni , "New Trends in Video Communications, Processing and under-standing in Surveillance Applications"-in ICIP 2001, October 7,2001, Thessaloniki.
[11] I. Cox, S. Hingorani," An Ecient Implementation of ReidsMultiple Hypothesis Tracking Algorithm and its Evaluation for the Purpose of Visual Tracking"-IEEEPAMI, 138-150,1996.
[12] I. Hariataoglu, et al," W4: A Real-Time Surveillance of People and Their Activities"-IEEE PAMI, 22, No. 8,August 2000.
[13] Intel Corporation, "Integrated Performance Primitives for Intel Architecture" -Reference Manual, U.S.A., 2002.(http://developer.intel.com)
[14] Olivier Barnich and Marc Van Droogenbroeck, "ViBe: A Universal Background Subtraction Algorithm for Video Sequence", Member, IEEE, VOL. 20, NO. 6, JUNE 2011