Journal of Informatics
Electrical and Electronics Engineering, 2022,
Vol. 03, Iss. 01, S. No. 006, pp. 1-8
ISSN (Online): 2582-7006

Object Detection Using Various Camera System

Pushpendra Tripathi[1], Pawan Singh[2]

Amity School of Engineering and Technology, Amity University, Lucknow, Uttar Pradesh, India[1,2]
pushpendratripathi303@gmail.com[1], pawansingh51279@gmail.com[2]

How to cite this paper: P. Tripathi, P. Singh, "Object Detection Using Various Camera System," *Journal of Informatics Electrical and Electronics Engineering* (*JIEEE*), Vol. 03, Iss. 01, S No. 006, pp. 1–8, 2022.

http://doi.org/10.54060/JIEEE/003.01.006

Received: 05/04/2022
Accepted: 24/04/2022
Published: 25/04/2022

Abstract

Multiple cameras use to simultaneously view an object from multiple angles and at high resolutions detect using real time tracking for surveillance and security management. The component key of tracking for surveillance system are extracting the feature, background subtraction and identification of extracted object. Video surveillance, object detection and tracking have drawn a successful increased interest in recent years. An object tracking can be understood as the problem of finding the path (i.e. trajectory) and it can be defined as a procedure to identify the different positions of the object in each frame of a video.

Keywords

Real time object tracking; Open CV; Real time object detection; object identification; surveillance system and background subtraction.

1. Introduction

A computer vision method for detecting things in images or movies is called object detection. Deep learning and machine learning algorithms have made significant progress in object detection. We can easily detect individuals, objects, settings, and visual features when we look at a photograph or watch a video. The purpose of computer vision is to detect an image and its contents by using picture processing algorithms to solve some of its duties, similar to the human brain. Object detection is a computer vision and image processing technology that compares numerous consecutive frames from a movie with various ways to see whether any objects are detected. Detecting and detecting an object in a digital image has become one of the most popular commercial applications to help users save time and effort. This technique was developed a year ago, but it still has to be improved in order to reach the desired goal in a more efficient and accurate manner. The goal of this project is to use a few procedures like colour processing and outline detection to detect and allocate the item. Image recognition is the ability of software to recognize objects, places, people, writing, and actions in images in the context of machine

vision. The concept of highest-level application of Image Processing and Computer Vision is real-time human body recognition and tracking in the physical environment. At many locations, such as airports, train stations, and offices, we can track objects. Extraction of a feature, background image subtraction, and identification of extracted feature are the fundamental components of real-time tracking's identify to path. The region-based technique and the boundary-based approach are both used for moving object detection. Optical flow and background removal are the most popular region-based techniques. By eliminating predicted backdrop models from photos, the background subtraction method finds real-time moving objects. However, estimating the background models via the background subtraction method takes a lengthy time [1].

2. Related Works

Computer vision is a branch of AI and computer science that seeks to calculate a visual comprehension of the world, and it lies at the heart of Hayo's strong algorithms. It is a field that entails the processing, analysis, and comprehension of high-dimensional data from the actual world in order to provide numerical and symbolical information. It is a science and machine technology that allows it to extract information from photos. It was created in 1999 by Intel Corporation. Open cv stands for open-source computer vision library. It allows us to alter photos and videos, as well as store and retrieve them. It can read and write images, change images, recognizes human faces and their features, and create augmented reality (augmented reality). Its most notable characteristic is its versatility, as it is compatible with practically all programming languages. Computer vision is a branch of artificial intelligence (AI) that allows computers and systems to extract useful information from digital photos, videos, and other visual inputs and act or make recommendations based on that data. If artificial intelligence allows computers to think, computer vision allows them to perceive, observe, and understand the world around them. Human vision is quite comparable to computer vision; yet humans are one step ahead. The benefit of human vision is that it has a contextual lifecycle that allows it to learn how to identify objects, how far away they are, if they are moving, and if there is a problem with the image. Computer vision teaches machines to execute these tasks, but it must do it faster than the retina and optic nerve [1-5].

3. Proposed System

A model that has been trained on a previous problem and may be used to tackle additional problems in the same domain is known as a pre-trained model. These models' architecture can be tweaked slightly, allowing you to fine-tune them to your application's requirements. The use of pre-trained or delegated learning models in object recognition-based applications has made them popular. In no particular order, these are some of the most popular pre-trained object identification models.

3.1 MobileNet SSD

SSD (Single Shot MultiBox Detector) is a popular object detection technique. Faster RCNN is generally faster. This article quickly explains object identification, the TensorFlow API, neural network concepts, and how the SSD architecture works. The following is a step-by-step implementation of a Mobilenet V2 SSD using the TensorFlow API and trained on a COCO dataset. In this tutorial, you'll learn how to identify each class from the COCO dataset's classes. With a little perseverance, you should be able to implement your own SSD after that. This article follows the structure of the original paper. Object recognition is a computer technique that deals with the recognition of instances of semantic objects in a computer vision and image processing context [5-9].

3.2 SSD MobileNet Architecture

The SSD architecture is a single convolutional network that learns to anticipate and classify bounding box locations in a single run. SSD can thus be trained from beginning to end. The foundation design of SSD networks is followed by numerous convolution layers, With SSDs, many objects in a picture can be detected with just one photo, but with a Regional Proposal Network (RPN)-based technique like the R-CNN series, it takes two. One provides geographical suggestions, while the other determines the proposal's theme. As a result, SSDs outperform the two-tier RPN-based strategy [9,10].

4. Methodology

CNNs are frequently used in image identification, image classification, object detection, face recognition, and other applications. CNN image classification takes an image as input, processes it, and categorises it. Another sort of neural network is the CNN, which may be used to help machines see things and perform tasks like picture classification, recognition, and object detection. A convolutional neural network (ConvNet / CNN) is a deep learning system that takes an input image and assigns importance (learnable weights and biases) to different characteristics / objects in the image, allowing them to be distinguished from one another. In comparison to other classification techniques, ConvNet requires far less preprocessing. ConvNet can learn these filters / characteristics, whereas primitive approaches necessitate well-trained and manual filter creation. ConvNet was inspired by the arrangement of the visual cortex and has a comparable architecture to the connecting network of neurons in the human brain. Individual neurons can only respond to stimuli in a specific portion of the visual field called a receptive field. The visual cortex is covered entirely by a group of such fields. Convolution is used to extract high-level characteristics from a picture, such as edges. ConvNet does not need to have only one convolutional layer. The first ConvLayer is usually in charge of capturing low-level details like edges, colors, and gradient alignment. The architecture adapts to increasing levels of functionality by adding layers, resulting in a network that is well-understood. The dataset has the same number of photos that ours does. Operation outcomes can be divided into two categories. On one hand, the convolution feature's dimensions are smaller than they were previously. The dimensions of the input, on the other hand, are expanded or remain the same. In the first example, valid padding is used, and in the second situation, the same padding is used. The pooling layer, like the convolution layer, reduces the spatial size of the convolution features. Through dimension reduction, the processing power required to process the data is reduced. It also aids in the extraction of essential characteristics that are rotation and position invariant, as well as the model's training process.

4.1 Convolutional Layer

In CNN's architecture, the convolution layer is crucial. To utilize a 3x3 or 5x5 filter, one must first enter the image. The green filter is dragged over the input image, which is shown in blue pixel by pixel, starting at the top left. As you walk across the image, the filter multiplies its values with the image overlay values, then adds them all together to provide a single value for each overlay. When the input images include several channels, the kernel has the same depth as the input image (red, green, blue). The stacks Kn and In ([K1, I1], [K2, I2], [K3, I3]) are multiplied by a matrix, and the results are displayed are prestressed and joined to form a narrow channel at depth Each neuron in the output array has many overlapping receptive fields. Typically, the first ConvLayer reaches a low-level state. Gradient alignment, border, color, and so on are all examples of properties. By adding layers, the theme enables high-level functionality and provides us with a network that understands all the photos in the data set. During transmission, the core travels the length and breadth of the image. Make a drawing of the problem's reception area. Activation map is a two-dimensional representation of a picture's response to any spatial position in the image.

$$W_{out} = \frac{W - F + 2P}{S} + 1$$

4.2 Nonlinearity Layer

In CNN layers, the activation function is critical. Another mathematical function called the activation function provides the filter output. The most widely utilized activation function in CNN functions Extraction is Unit Linear Rectified, which stands for Unit Linear Rectified. The trigger function is mostly used to end output neural networks, such as yes or no. The activation function converts initial values from è1 to 1 or 0 to 1 and so on (depends on activation function). Activation Functions are divided into two categories:

i. Linear Activation Function Uses the function F(x) = CY. Take the ticket and It multiplies it by the constant c (weight of each neuron) and produces the output signal proportional to the input. The linear function can be better than the step function because there is only the answer yes or no and not the multiple answers.

ii. Nonlinear activation functions in modern neural networks, nonlinear activation functions are used. They allow the model to create complex mappings between them. network inputs and outputs, which are crucial for learning and modelling complex data, including images, video, audio, and non-linear or high-dimensional data recordings.

4.3 Fully Connected Layer

A forward neural network is all a fully linked layer is. The network's lowest tiers have fully connected layers. A fully connected layer gets its input from the last pooling or convolution layer's output layer, which is flattened before being used as input. Flattening the output after the last pool or convolution layer entails unwrapping all values from the output into a vector (3D array). Adding an FC layer to the convolutional layer output is a simple way to learn nonlinear combinations of high-level features. The FC layer learns a non-linear function in this space.

We cover using multiple cameras to record video and create a video data set. There will be brief information on research questions involving multiple cameras.

- **Frame generation**- Here we have presented an algorithm for generating frames from a video sequence. The steps are given as follows.

  1.Play the video file using aviread () for an interlaced AV file format or mmread () for another supported file format.

- **Object detection**- Object detection is the process of determining the area of interest based on user requirements. Here we have proposed the object detection algorithm using the frame difference method (one of the background subtraction algorithms). The steps are given as follows:

  i. Play all frames generated from video stored on a variable or storage medium.

  ii. Convert them from a color image to a grayscale image using rgb2gray ().

  iii. Calculate the difference as | frame i - frame i-1| >Th.

  iv. If the difference is greater than a threshold Th, then the value is considered to be part of the foreground, otherwise part of the background.

  v. Update the value of i by increasing it by one.

  vi. Repeat steps 3-5 until the last image.

  vii. Finish.

- **Post Processing-** The object found in the preceding phase may cause a connection difficulty and/or have holes that prevent the object from being returned. As a result, post-processing is required to address the issue of dealing with holes and pixel connectivity in object space. One of the post-processing methods is mathematical morphological analysis, which involves enhancing the segmented image to achieve the desired output. We applied erosion and dilation iteratively in the proposed method so that an object emerges clearly in the front while the rest of the superfluous areas are deleted. Morphological techniques are helpful for maintaining important image components. These components can include the object's boundaries, region, shape, and skeleton, among others.
- **Object Representation-** To represent the item, we are utilizing a centroid and a rectangular shape to cover the object's border. Find the object's width Wi and height Hi by extracting the pixel coordinates Px(max) and Px(min) that have the maximum and least X coordinate values relative to the object after calculating the center of gravity. In the same way, compute Py(max) and Py(min) for Y coordinates. Calculate the width and height of the object in the i-th frame that is supplied as the i-th segment.
- **Use of Multiple Cameras-** The purpose is to record video and evaluate the trajectory of the object using numerous cameras. Using several cameras is beneficial for two reasons. The first purpose is to calculate object depth information in order to track resolution and occlusion, while the second goal is to enhance the field of view necessary for a single camera due to the sensor's limited field of view. Research). In order to capture or follow a moving object across a vast region, we employ multiple fixed cameras to enhance the field of view.

5. Implementation

Object detection is the process of identifying various items in photographs. R-CNN has recently become well-known for object detection. Ren suggested a quick R-CNN for item detection in 2015 as an improvement over R-CNN (a completely linked convolutional neural community) for feature extraction, which can comprehend the border and score of items at several points at the same time. Similarly, Dai proposed using absolutely linked CNNs to find things based on their location in 2016. Gidaris and colleagues Describe an object identification technique based on Deep CNN, a multi-area technique that aids in the learning of semantic functions. Gidaris' technique recognizes elements with reasonable accuracy in the PASCAL VOC 2007 and 2012 datasets.

5.1 Image Classification

CNN is extensively used for photo classification tasks. Medical pictures are one among CNN's critical uses, particularly for cancer diagnosis and using histological images. CNN to diagnose breast cancer images and compared the outcomes to a pre-educated community with a dataset the use of homemade descriptors. To deal with the hassle of sophistication skewness, statistics augmentation is used within the 2nd section. There are several famous image class pre-educated networks to be had. Image category can be clean if a labelled dataset may be produced for the target photograph.

5.2 Video Processing

The temporal and spatial statistics of moving images are used by video processing systems. Many researchers have used CNN to solve problems related to video processing. For example, a genuine border detection system based on CNN has gained popularity. In Tong's approach, TAGs are formed using CNN. TAGs are blended in competition with a single shot to annotate that precise movie during the examination. Wang employed 3-d CNN and LSTM to differentiate activity inside a movie in 2016. Frizzi recruited CNN in an exclusive deal in 2016 to find out about a few emergencies, such as fire or smoke, in the video. According to Frizzi, CNN structures can extract prominent attributes as well as complete the classification task. In move-

ment popularity, however, accumulating geographical and chronological details is a time-consuming process. Shi Y proposed a three-stream based structure in 2017 to address the inadequacies of existing function descriptors. This shape may extract spatial–temporal characteristics from both short and long-term motion in a movie. CNN divides their technique into components and employs bi-directional LSTM to recognize interest within the video, as mentioned in their study. The sixth section of the film is first used to extract abilities from the first portion. In the second segment, the bi-directional LSTM framework is employed to apply sequential statistics among frame abilities.

5.3 Pictures with low Frequency

ML researchers have used CNN-based image enhancement algorithms to improve picture resolution. A deep CNN-based approach for detecting objects in low-resolution images. Chevalier et al. were the ones who introduced LR-CNN. For low-resolution photo classification. Kawashima and colleagues Describe any additional deep learning-based method that uses convolutional layers and an LSTM layer to determine action from low-resolution thermal pictures.

5.4 CNN for several Dimensional Data

Three-dimensional form models have grown more widely available and easier to get, making 3-D data crucial for the improvement of item classification. CNN is used by cutting-edge trending procedures to overcome this difficulty. CNN based on volumetric representations and CNN based entirely on multi-view representations are two new types of CNN. As proven by actual results from these two types of CNN, existing volumetric CNN architectures and algorithms are unable to fully leverage the power of 3-d representations. This work aims to improve both volumetric CNN and multi-view CNN based on a detailed analysis of current approaches. To that end, amazing volumetric CNN community structures are presented. In addition, we look at multi-view CNN, which employs multi-decision filtering in 3D. Overall, every volumetric CNN and multi-view CNN outperforms current methods. Extensive experiments are provided to examine the underlying design choices, allowing us to have a better idea of the distance between object categorization algorithms and 3-D facts.

5.5 Object Computation

Counting devices in images is one of the most essential challenges in computer vision. What's wrong with a wide range of programmed connected to microbiology (e.g., bacterial colony counting), monitoring (e.g., people counting), agriculture (e.g. net or vegetable counting), pharmaceuticals (e.g. tumor cell counts on histopathology images), and nature conservation? (For example, counting animals). Counting items is a simple operation for humans, but it may be challenging for computers. Mobilenet SSDs that have been pre-qualified can count cell counts based on usage. The lower prediction level has been deleted, and the functionality has been passed to a class layer ahead.

Real-time object detection methods will be compared next. It's worth noting that algorithm selection is influenced by the use case and application; different algorithms excel at different tasks (e.g., Beta R-CNN shows best results for Pedestrian Detection).

- SSD- SSD is a popular one-stage detector that can distinguish between numerous classes. Using a single deep neural network, the approach finds objects in images by discretizing the output space of bounding boxes into a series of default boxes with varied aspect ratios and scales per feature map position. the object detector evaluates each default box for the existence of each object type and modifies the box to better fit the shape of the object. To handle objects of various sizes, the network also incorporates predictions from numerous feature maps with varied resolutions. the SSD detector is simple to programmed and integrate into software systems that need to detect objects. SSD provides substantially better accuracy than other single-stage algorithms, especially with smaller input images.

- YOLOR- In 2021, YOLOR was introduced as a new object detector. The algorithm trains the model using both implicit and explicit information. As a result, YOLOR can learn a general representation and use it to fulfil many jobs. Kernel space alignment, prediction refinement, and multi-task learning are used to incorporate implicit knowledge into explicit knowledge. YOLOR obtains significantly better object detection results with this strategy. The MAP of YOLOR is 3.8 percent higher than the PP-YOLOv2 at the same inference speed when compared to other object detection algorithms on the COCO dataset benchmark. The inference speed has been enhanced by 88 percent when compared to the Scaled-YOLOv4, making it the quickest real-time object detector currently available [11-12].

6. Result

The model's test photos taken using MobileNet SSD. The rectangular box depicts the most accurate approach for object recognition, which is real-time object recognition. All functional systems are compiled or run using the Python programming language and the OpenCV library. Python libraries are an open-source platform for creating pre-trained data models and identifying objects. Feature extraction at various scales can increase the accuracy of large object detection, but it does not provide good speed accuracy for small object detection. To build a network, we employ a variety of packages as well as the TensorFlow GPU. The goal of the data model prior to training is to ensure that the model is accurate. The Tensorflow directory, the MobileNet SSD function extractor, the Tensorflow object recognition API, and the Anaconda virtual environment are all used in the experiment. All of this setting enables us to recognise objects more accurately in real time. We increase the number of standard boxes with lower confidence and focus on boxes with high reliability to attain high precision. The trust value is used by these deep, multi-layered neural networks to improve the process of detecting accurate packets. The Anaconda Virtual Environment and the Tensorflow Object Discovery API are used. To minimise channels and feature maps, the technique adds a width multiplier and a resolution multiplier. The proposed method uses aspect ratio to create real-time object recognition. SSD algorithm consists of large amounts of data, an easy-to-train model, and faster GPUs that can detect and classify multiple objects in an image with high precision. We put together a test video of the indoor data set. There are many complex elements in the video dataset, including a complex environment with system.

- **Threshold Selection-** Choosing a constant threshold may be useful in some situations, but it may provide nonsensical outcomes in others. The use of the threshold has been demonstrated. Part a represents the real image, part b represents the absolute frame difference with no threshold, resulting in a thin outline of the person (object) on a black background, part c represents the lower part of the frame difference threshold frame, resulting in a loss of the main information, which is an object, disappearing from the background, and finally the d-part represents the difference of frames with a higher threshold, resulting in addition, and finally the e-part represents the difference of frames.
- **Experimental Result Prediction about the Object-** To obtain maximal coverage of machine vision applications, we tested our suggested approach with a variety of video datasets created in the environment. It is vital to keep an eye on items when numerous cameras are utilized to enhance the camera's field of view and monitor a vast region. In this application, there is a high risk of losing the moving item, which means the object will not enter the next camera view or another object will be disturbed, causing predictions on the object to be lost. For each paired camera combination, such as the item's area, the height-to-width ratio in following frames, and the object in the last frame of the first camera, in the first frame of the second camera.

7. Conclusion and Future work

Object recognition technique that recognizes things from a photo using deep learning neural networks. Most robotic and computer vision systems have an object recognition feature. Although there have been significant advancements in recent

years, and some current methodologies have been incorporated into driver aid technologies, we are still a long way from obtaining human-level performance, particularly when it comes to open world learning. It's worth noting that object recognition isn't widely used in many areas where it may be quite useful. Object identification systems are becoming increasingly vital as mobile robots and other autonomous machines become more common. Finally, we must remember that object identification systems are required for nanobots or robots that explore locations that humans have never seen, such as deep portions of the ocean or other planets, and that the recognition systems must learn new classes of items as they are discovered. The ability to learn in real time and in the open environment is critical in such situations. To achieve high real-time precision for object detection, the researchers used an upgraded SSD technique as well as a deep sublayer neural network layer. Our system performs well on both still images and videos. The proposed model has an accuracy of more than 79.8%. These deep neural networks take feature information from images and videos, then use feature mapping to identify the class label. Our program's main purpose is to improve the SSD technique for object detection by selecting standard boxes with the best aspect ratio values. We offer a new method for detecting and tracking multi-camera objects in videos in this paper. We investigated four distinct object detection methods and developed a modified frame differencing methodology to reduce mistake detection rates. The detection of non-rigid objects and their surveillance using many cameras were the focus of this research. The method is put to the test on several video datasets. The identified object's center of gravity and rectangle shape surrounding the object border are used to depict it. This could come in handy in surveillance systems. In the future it is planned to accelerate the processing and analysis speed of the detected object. Other video elements including edges, color, and texture will be included in future studies. In addition, we will attempt to develop a strong classifier tracking system for classifying object state and attributes.

References

[1]. "TensorFlow White Papers", *TensorFlow*, 2020, [online] Available: https://www.tensorflow.org/about/bib

[2]. Y. Xiao et al., “A review of object detection based on deep learning,” Multimed. Tools Appl., vol. 79, no. 33–34, pp. 23729–23791, 2020.

[3]. OpenCV Library, “Home,” OpenCV, 09-Feb-2021. [Online]. Available: https://opencv.org/. [Accessed: 28-Jan-2022].

[4]. “Keras: The Python deep learning API,” Keras.io. [Online]. Available: https://keras.io/. [Accessed: 28-Jan-2022].

[5]. D. Li, J. Zhao, and Z. Liu, “A novel method of multitype hybrid rock lithology classification based on convolutional neural networks,” Sensors (Basel), vol. 22, no. 4, p. 1574, 2022.

[6]. M. Garifulla et al., “A case study of quantizing convolutional neural networks for fast disease diagnosis on portable medical devices,” Sensors (Basel), vol. 22, no. 1, p. 219, 2021. *Iop.org*. [Online]. Available: https://iopscience.iop.org/article/10.1149/10701.15533ecst/meta. [Accessed: 23-Jul-2022].

[7]. S. Squarepants, “Bitcoin: A peer-to-peer electronic cash system,” *SSRN Electron. J.*, 2008.

[8]. E. Feyen, J. Frost, L. Gambacorta, H. Natarajan, and M. Saal, “Fintech and the digital transformation of financial services: implications for market structure and public policy,” *Bis.org*, 2021. [Online]. Available: https://www.bis.org/publ/bppdf/bispap117.pdf. [Accessed: 27-Jan-2022].

[9]. N. Mangla and P. Rathod, “Unstructured data analysis and processing using Big Data tool -Hive and machine learning algorithm -linear regression,” *Iaeme.com*. [Online]. Available: https://iaeme.com/MasterAdmin/Journal_uploads/IJCET/VOLUME_9_ISSUE_2/IJCET_09_02_006.pdf. [Accessed: 27-Jan-2022].

[10]. W. Shi, “Recommendation systems: A review,” Towards Data Science, 23-Feb-2020. [Online]. Available: https://towardsdatascience.com/recommendation-systems-a-review-d4592b6caf4b. [Accessed: 28-Jan-2022].

[11]. N. Srivastava, U. Kumar and P. Singh (2021) Software and Performance Testing Tools. Journal of Informatics Electrical and Elec-tronics Engineering, Vol. 02, Iss. 01, S. No. 001, pp. 1-12, 2021.

[12]. A. Singh, P. Singh, “Image Classification: A Survey. Journal of Informatics Electrical and Electronics Engineering”, vol. 01, Iss. 02, S. no. 2, pp. 1-9, 2020.