Human Detection and Counting

S. Madhuri[1#] P. Anusha[2#] S. Chendu Priya[3#]

[#]*UG Student, Department of Computer Science and Engineering*
Prathyusha Engineering College, Chennai

Abstract— A Population growth is rising in modern times. Due to the population's linear increase, many individuals now frequent public spaces. Thus, this technique will give the number of people in a specific area's malls, supermarkets, etc. Some businesses exclusively depend on the timing and schedule of their customers. Therefore, by creating a system for counting people, our work satisfies the problem and offers a solution. Therefore, a mobile single shot detector (SSD) network and centroid tracker are presented. For better feature extraction, this model swaps out the VGG16 base network for a Mobile Network, and for classification, it links with six convolutional layers following the base network. In order to calculate the centre of a bounding box, the centroid tracking algorithm uses bounding box coordinates from an object detector SSD. Each person will receive an ID when the centroid is calculated, and this operates the dataset with training and testing the data.

Keywords— Human Detection, Signal Shot detector, Object detector, Mobile network

I. INTRODUCTION

There are numerous methods for finding and counting persons in a situation. Manual counting is not a recommended method due to the crowded nature of most public spaces, such as malls and shopping centres. Currently, heat sensors and tally counters are the main technologies utilised in public settings for people counting. These systems are limited to counting people who are travelling through a specific area and cannot be expanded. These systems are currently getting less reliable because to blockage. Two people passing through the tally counter simultaneously will result in erroneous results [1]. As a result, research has focused on vision-based counting because it is relatively cheap, provides more information, and doesn't directly interact with people who move. The basic concepts behind people counting via video are tracking and detection. People in the scene must be located during detection, and then their locations must be tracked throughout the video sequence [2].

There are currently three ways for counting people using videos: body component and body form detection, and backdrop modelling. As a result, research has focused on vision-based counting because it is relatively cheap, provides more information, and doesn't directly interact with people who move [1].

The basic concepts behind people counting via video are tracking and detection. People in the scene must be located during detection, and then their locations must be tracked throughout the video sequence. There are currently three ways for counting people using videos: body component and body form detection, and backdrop modelling [6].

II. LITERATURE REVIEW

To detect objects in static photos by components, Anuj Mohan, Constantine Papageorgiou, and Tomaso Poggio suggested a broad example-based system. The method is

demonstrated by creating a system that finds individuals in congested environments [1]. In particular, the system uses four different example-based detectors to identify only the parts of a person's anatomy in a photograph, such as the head, left and right arms, and consequently the legs, as opposed to the entire body. The system then confirms that the components have been detected and are in the correct geometric configuration [5].

A. Existing System

Machine learning algorithms have a difficult time accurately counting the instances of objects in a given image or video frame. To count people, cars, and other items, a number of solutions have been created, but none of them is ideal [3]. Naturally, since we are discussing image processing, a neural network appears to be an appropriate tool. You can read a description of many strategies, frequent issues, difficulties, and cutting-edge solutions in the subject of Neural Networks object counting below [2]. With video examples provided at the conclusion of the post, the existing Faster R-CNN network model will be utilized to count items on the street as a proof of concept. But it moves slowly and is less precise [5].

B. Proposed System

The technology under consideration uses SSD and deep neural networks. Numerous people have flocked to popular locations as a result of direct population growth [6]. These frameworks will subsequently provide the inclusion of the unique region shopping malls, general stores, and so forth. Some businesses solely depend on the client strategy and its planning. Our work now resolves the problems and provides a solution by designing a framework for group counting. Thus, a centroid tracker and the single short identifier portable net are suggested [7]. In order to more effectively remove highlights, this model substitutes the VGG16 basis organisation with a Mobile Net organisation. Following the basic organisation, it associates with six convolutional layers for order [8]. The centroid uses the item identifier SSD and the c bouncing box facilities to compute the jumping box's focal point. Every person will receive an ID after the centroid is established, and this model uses the dataset by preparing and evaluating the data [10].

C. Modules Description

Here we used many modules for making this project. These are the modules which are used for implementation.

(a) Python 3

Python 3.0 was released in 2008. Although this version is supposed to be backward incompatible, later on many of its important features have been backported to be compatible with version 2.7.

(b) PIL

The python Imaging Library or PIL is an amazing Python library used for image processing. This library provides so many features for working on image using python

(c) NumPy

NumPy is a library for the Python programming language, adding support for large, multi-dimensional array and matrices, along with a large collection of high-level mathematical functions to operate on these arrays.

(d) Cv2

OpenCV-Python is a library of Python bindings designed to solve computer vision problems.

(e) Argparse

Python argparse() is one of the recommended python module that take care of multiple scripting need in an automated fashion by enabling the developer to create reproducible scripts right away from the jupyter

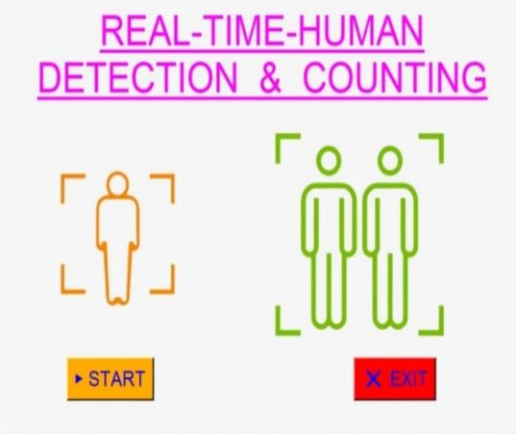
notebook code. argparse() enables the user to provide values for the variables during the runtime process

(f) HOG (Histogram of oriented Gradients)

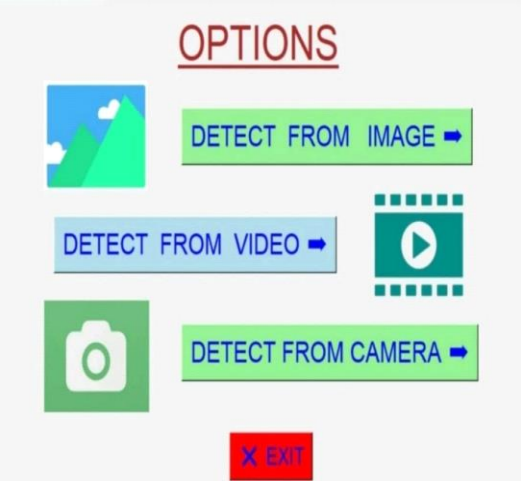
Histogram of Oriented Gradients (HOG) is a features descriptor used in image processing, mainly for object detection. A feature descriptor is a representation of an image or an image patch that simplifies the image by extracting useful information from it

III. Detection Process

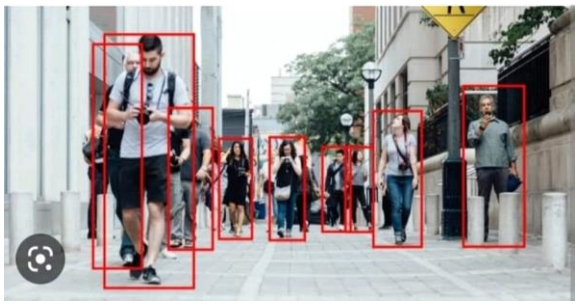
(a) Screen Shot



(b) Options



(c) Detection



IV. Conclusion

Therefore, using the input whether it be video or images, we are able to identify and count the human movements. And by employing HOG descriptor instead of the Haar cascade classifier, we will achieve greater accuracy.

Since we are taking any image, video, or picture with a camera, we are able to distinguish persons and accurately count them.

References

[1] Programming computer vision with python, 1st Edition,jan Eric Solem,2012, O'Reily

[2] Learning OpenCv, Adrian Kachler and Gray Rost Bradski, 2008, O' Reily.

[3] Deep Learning with Tensorflow,Giancarlo Zaccone ,Md.Rezaul Karim ,Ahmed Menshawy,2017

[4] Python GUI Programming with Tkniter,Alan D.Moore,2018.

[5] Python Standard,Fredrik Lundh,2001, O'Reil.

[6] AVSS 2005. Real time event detection solutions for enhanced security and safety in public transportation. http://www-dsp.elet.polimi.it/avss2005/.

[7] H. Chui and A. Rangarajan. A new point matching algorithm for non-rigid registration. Computer Vision and Image Understanding, 89(3):114–141, March 2003.

[8] D. Gavrila. Pedestrian detection from a moving vehicle. In Proc. 6th European Conf. Computer Vision, Dublin, Ireland, volume 2, pages 37–49, 2000.

[9] J. Giebel, D.M. Gavrila, and C. Schnorr. A bayesian frame- ¨ work for multi-cue 3d object tracking. In Proc. 8th European Conf. Computer Vision, Prague, Czech Republic, 2004.

[10] I. Haritaoglu, D. Harwood, and L. S. Davis. HYDRA: Multiple people detection and tracking using silhouettes. In IEEE International Workshop on Visual Surveillance, 1999.