Human Detection System

Hassan Mahmood Khan
School of Electrical Engineering and Computer Sciences (SEECS)
National University of Sciences and Technology (NUST)
Islamabad, Pakistan
hkhan.bee17seecs@seecs.edu.pk

***Abstract*— People counting has significant importance in computer vision technology. With the recent pandemic, the value of people counting solutions has increased substantially, to implement guidelines pertaining to social distancing. This python implemented project primarily makes use of OpenCV & Imutils libraries for object recognition with the aid of machine learning.**

***Keywords*— HOG Descriptor, Support Vector Machine, Non-maximum Suppression, Image Pyramids**

I. Introduction

Practical scenarios generally contain a plethora of complex scenes, including interacting and overlapping targets. These are difficult imaging conditions as they make the task of object recognition and tracking a challenging one.

This project aims to tackle the task of human recognition in still images. The human body has similar body structure regardless of gender, build, race etc. At the basic level, we all have a head, arms, legs, hip etc. We can employ computer vision to extract these features and feed them to machine learning algorithms to detect and track humans in images.

In this paper, I propose a human detection and tracking scheme based on the pre-trained Histogram of Oriented Gradients (HOG) Descriptor and Linear Support Vector Machine (SVM) included in the OpenCV library. This model is based on Dalal-Triggs algorithm, which automatically detects pedestrians in images, and can be used for detection in both images and video stream; however, the scope of my project is limited to images only.

II. Literature Review

Human detection based on Scale Invariant Feature Transform (SIFT) and Speeded Up Robust Feature (SURF) are decent methods which produce high quality features. However, there is a catch. These are computationally intensive and require high processing power for use in real-time applications of varying difficulty. [1]

You Only Look Once (YOLO) is another novel approach that has gained popularity when it comes to object recognition. It models the task as a regression problem, whereby it predicts bounding boxes and their probabilities using a neural network over the entirety of the image in a single evaluation. [1]

For multiple target tracking, particle filtering has shown some promise as it able to integrate object recognition, color tracking and occlusion handling. [2]

III. Methodology

My paper proposes a simple, computationally efficient human tracking model which can be made to work with existing solutions.

A. *Histogram of Oriented Gradients (HOG)*

The HOG feature descriptor is generally used to extract features from an image to facilitate object detection. It focuses on the structure of an object by computing the gradient and orientation of the edges, for localized regions, unlike standard edge features which determine if a cell is an edge or not. Finally, a histogram is computed for each of these localized regions. [3]

B. *Calculating the Histogram of Gradients (HOG)*

- Preprocessing the image by resizing width to height ratio to 1:2 i.e., (64 x 128 standard). This allows for simpler calculations.
- Calculating gradients for each pixel in both, the x and y direction. This results in two new matrices, with gradients in the x-direction being stored in one, and in the y direction in the other.
- The magnitude and direction are calculated for the gradients found in the last step for each pixel value. This can be found using the Pythagoras Theorem.
- HOG is computed for the entire image by dividing it into 8 x 8 cells. Thus, we obtain features for these smaller cells which basically represent the entire image.
- The HOG features are still sensitive to overall lighting. To remove this effect, we normalize the gradients by taking 16x16 blocks.

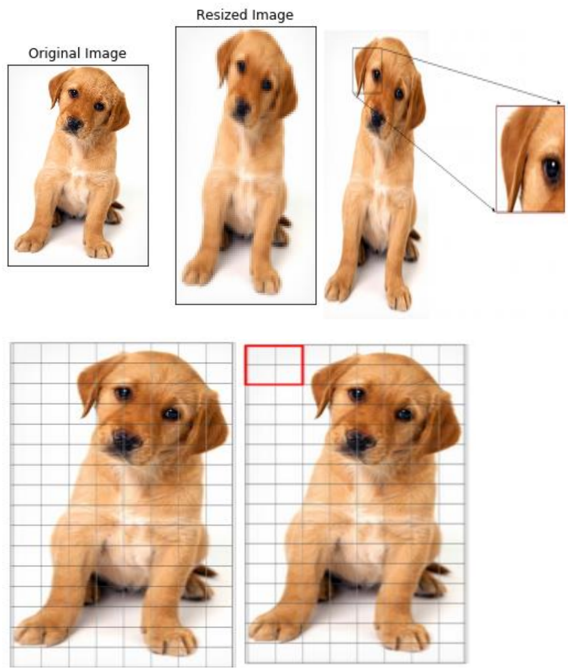

Fig. 1. These images depict the step by step procedure of calculating HOG feature descriptors.

C. Support Vector Machine (SVM)

SVM is a machine learning algorithm that is essentially used for classification or regression problems whereby it creates a hyper plane to separate data classes. It takes data as input and draws a line to separate them if possible, in our case positive samples of data that contain objects i.e., human beings, from negative samples that do not contain humans. This is an example of linear classification, since there are only two classes to predict.

The algorithm locates the hyperplane i.e., separating line, by maximizing the distance between support vectors i.e., margin. The scenario where this margin is maximum is our optimum case.

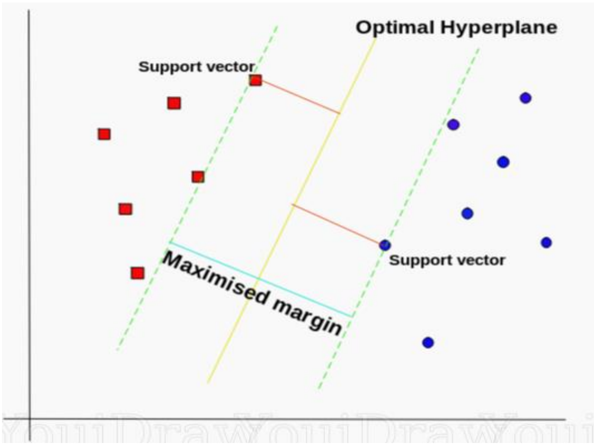

Fig. 2. Linear Support Vector Machine (SVM)

D. Combination of HOG and SVM

The HOG descriptors combine with the linear SVM to ensure human detection is possible.

- Initially, for each image multiple scales are calculated using an image pyramid, which is another name for multiple image-scale representation.
- For each image and each possible scale, we apply a sliding window across the image i.e., a bounding rectangle over the image of a pre-defined size.
- HOG descriptors are then computed for these samples and fed to the linear SVM. The classifier then works to categorize these windows as an object detected.
- If the classifier correctly classifies a window as an object detected or not detected – it constitutes as a positive and negative, respectively.
- However, if it incorrectly classifies a window for an object, this is known as a false-positive. The probability of false-positives is recorded to further train the classifier for better accuracy.

Let us expand the last bullet point referring to false-positives. False-positives can significantly affect the accuracy of the model, therefore we sort these based on their probability of occurrence and re-train our SVM for greater accuracy.

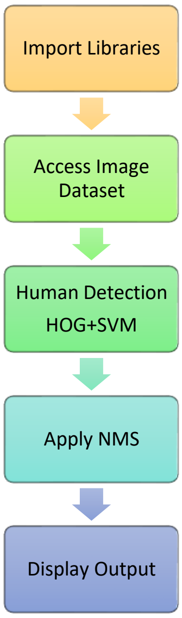

Fig. 3. General flowchart of the entire model displayed in five basic steps.

E. Non-Maxima/Maximum Suppression (NMS)

One of the key problems with object detection in images is multiple overlapping bounding boxes around the object of interest. Though these boxes might be correct, however, we do not want multiple instances of a single object.

This technique works to remove redundant and overlapping boxes by suppressing the smallest bounding boxes and retaining the largest one. This gives us the true detection.

Fig. 4. Example of before and after applying non-maximum suppression

IV. Results and Discussion

Human detection is achieved via the proposed approach mentioned above. Results obtained demonstrate the model's ability to work well with images having a higher degree of complexity i.e., humans in the foreground and background.

However, it is also observed that the proposed solution requires a great degree of detail when it comes to assigning the right parameters for the bounding frame, and their tuning.

Another parameter that is absolutely critical to the computational speed and correct output of the algorithm is *scale*; an input parameter for the image pyramid. It controls how the image is resized at each layer, with a smaller value resulting in greater number of layers and a larger doing just the opposite. A very small value results in wastage of processing power and increases the likelihood of false-positives. Whereas, a large one would produce lesser number of layers yielding faster results but at the cost of losing accurateness. Therefore, the value of *scale* I have chosen lies well between the two thresholds i.e., 1.05. [4]

Another key issue with the HOG descriptor is the fact that it is not rotation invariant, which means the model fails to correctly detect from rotated images. This is a significant limitation, especially in scenarios where we acquire images from various sources without proper inspection.

Overall, results produced are satisfactory and computationally efficient without compromising accuracy.

Fig. 3. Example of human detected correctly and no human detected.

VI. FUTURE WORK

The model proposed in this paper is human detection in still images only. However, this can be extended to work with video streams, real-time captured images via webcam etc. Work can be done to improve the robustness of the model when catering to blurred or tampered images. The proposed model may also be combined with various existing solutions to tackle challenges of the modern day.

V. REFERENCES

[1] Medium. 2021. *Object Detection Techniques In Computer Vision*. [online] Available at: <https://medium.com/swlh/object-detection-techniques-in-computer-vision-7c169771fb15>

[2] S. Maskell, M. Rollason, N. Gordon and D. Salmond, “Efficient Multitarget Tracking using Particle Filters,” Journal Image and Vision Computing, 21(10), pp. 931- 939, September 2003.

[3] Singh, A., 2021. *Feature Descriptor | Hog Descriptor Tutorial*. [online] Analytics Vidhya. Available at: <https://www.analyticsvidhya.com/blog/2019/09/feature-engineering-images-introduction-hog-feature-descriptor/>

[4] Rosebrock, A., 2021. Image Pyramids With Python And Opencv - Pyimagesearch. [online] PyImageSearch. Available at: <https://www.pyimagesearch.com/2015/03/16/image-pyramids-with-python-and-opencv/>