# 

# REAL-TIME HUMAN DETECTION AND COUNTING SYSTEM

## A Project Work Synopsis

*Submitted in the partial fulfilment for the award of the degree of*

## BACHELOR OF ENGINEERING

**IN**

### COMPUTER SCIENCE (HONS.) WITH SPECIALISATION IN ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING

**Submitted by:**

**JASLEEN MINHAS:** 20BCS6897

**SAKSHAM RAJ:** 20BCS6903

**PRABHJOT SINGH:** 20BCS6895

**LAKSHAY CHAWLA:** 20BCS6900

**Under the Supervision of:**

**Dr. Parul Goyal**



**CHANDIGARH UNIVERSITY, GHARUAN, MOHALI – 140413,**

**PUNJAB**

**November, 2022**

**Table of Contents**

Abstract i

1. **INTRODUCTION\* 4**
   1. Problem Definition 4
   2. Project Overview/Specifications\* 4
   3. Hardware Specification 4
   4. Software Specification 5
2. **LITERATURE SURVEY 6**
3. **PROBLEM FORMULATION 8**
4. **RESEARCH OBJECTIVES 9**
5. **METHODOLOGY 9**
6. **TENTATIVE CHAPTER PLAN FOR THE PROPOSED WORK 10**
7. **REFERENCES 14**

### *ABSTRACT*

Human face counting is generally done in our daily lives for keeping a record of number of people in a particular area, or people passing by a particular area. Precisely this task is done by a human (generally with a hard counter), but nowadays with the advancements in computational intelligence and computer vision, this project is based on an automatic people counting system based on face detection through artificial intelligence and machine learning techniques.

This project investigates and reports benchmarks for detecting and counting humans through real time images, videos and camera. This is very useful in various image processing and performing computer vision tasks. This schemes have been implemented in Python programming language, and using various tech-stacks like OpenCv, imultis, numpy, argparse, hogsvm algorithm, tkinter, etc.

#### INTRODUCTION

* 1. **Problem Definition**

Human detection/counting systems are very necessary in the modern world. The need to develop intelligent security systems that can monitor footage and ‘understand’ what they see is highly important. Rather than have human personnel sit all day observing CCTV cameras an intelligent system would be able to detect an object (in this case human) of maybe a specific number and raise an alarm under any of the above conditions according to some timed program. Retail shops might also need to have information on the number of people going in or out of it’s stores in order to monitor effectiveness and impact of its market and make surveys and evaluations. Organisations/hobbyists might want to use human traffic information to: (i) limit crowds (crowd control) (ii)monitor building or seating capacity (iii) Management of queues.

The main problem around us we can see is the ‘population problem’. And directly or indirectly this is causing chaos or traffic in any one or the other place. This so called chaos may sometimes even lead to quarrel or even suffocation sometimes.

* 1. **Project Overview**

A human counter is a device used to measure the number and direction of people traversing a certain passage or entrance per unit time. Many technologies are used in counting persons; infra-red beams, computer vision, thermal imaging and pressure sensitive mats. [1]Each technology has its merits and demerits although no system has a one hundred per-cent accuracy factor. Of all the afore-mentioned counter technologies the computer vision system is employed in this work. Computer vision is a field that includes methods for acquiring, processing analysing and understanding image sand, in general high dimensional data from the real world in order to produce numerical or symbolic information e.g.in the form of decisions. A theme in the development of this field has been to duplicate the abilities of human vision by electronically perceiving and understanding an image. This image understanding can be seen as disentangling of symbolic information from image data using models constructed with the aid of geometry, physics, statistic sand learning theory. [2]Advantages of Computer vision technologies Relatively high accuracy(95%) Flexible to customize. Highly scalable Easy to integrate with other systems Disadvantages High Cost High power consumption Accuracy can be affected by shadow, floor background, poor image quality, differing light levels.

In solution to our problem, we will build an AI and Machine Learning based model. This model will help us to easily get the population (no. of persons) or traffic criteria in any real time video of any place. Also we will get the idea of which area is more or less populated or trafficked as compared to normal traffic. In this model will allow user to upload any real time video and model will show us the traffic condition of the area in video. This model will also give suggestion for future about what should be the normal traffic in that area. Thus like-wise using this model one can take on control of traffic and which will prevent various chaos and accidents. And will also help to regulate traffic flow in that area.

* 1. **Hardware Specification**

**Operating systems supported**

* Mac
* Windows
* Linux

**Application Program supported:**

* Google Chrome
* Apple Safari
* Mozilla Firefox

**RAM required**

* 2GB minimum, 3 GB recommended
* 4 GB recommended for parallel processing with 8 or more cores
  1. **Software Specification**

**Software**

* VS Code
* Python interpreter

**Languages**

* Python

**Libraries**

* Opencv
* Imutils
* Numpy
* Argparse
* Hogsvm algo
* Tkinter

#### 2 LITERATURE REVIEW

*Schofield et all* [1], Paper is based on counting people in video images using neural networks, in this paper the proposed model is based on background identification concept and subtracts the background image in order to identify number of objects or people, and facilitated to locate and track the objects/people and resulting in counting the number of objects in a frame, but this generally had a limitation if there were uncommon types of backgrounds. Zhao et all [2] is based on recording the number of people passing by a particular area, in takes a video as an input and works on the assumption that the objects are people, a scale-invariant Kalman filter is proposed and employed which is based on the strategy to count people by their face trajectories to quantize the number of people passing by. *Tsong et all [3],* also takes video as the input and satisfies the limitiations caused by Zhao et all, by using the concept of moving object segmentation. *Edeoghon et all [4]*, uses MATLAB, and is based on Viola-Jones Algorithm, and MATLAB provides proper tool and support for this algorithm. It is more similar to the functionality of Convolutional Neural Networks. The Viola-Jones Algorithm converts the facial data into M by 4 Matrix which represents features of facial data, and with this concept it finds out number of faces in a scene.

**Literature Review Summary**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Year and citation** | **Article Title** | **Purpose of the study** | **Tools/ Software used** | **Comparison of technique done** | **Source (Journal/ Conference)** | **Evaluation parameters** |
| *Schofield et all* | *A system for counting people in video images using neural networks to identify the background scene* | *Counting people in video images* | Artificial Neural Networks | Background Identification | 1996 Elsevier Science Ltd | Neural Network Logical Model |
| Zhao et all | *A People Counting System based on Face Detection and Tracking in a Video* | *Counting people in video based inputs* | Scale-invariant Kalman filter | Scale-invariant Kalman filter | LIRIS, UMR5205 | Object Detection |
| *Tsong-Yi Chen et all* | *A People Counting System Based on Face-Detection* | *Counting people in video based images* | - | Moving object segmentation | 2010 Fourth International Conference on Genetic and Evolutionary Computing | Moving Object Segmentation |
| *Edeoghon et all* | *Design and Implementation of a Human Detector and Counting System using MATLAB. 5* | *Counting and detecting faces in images* | MATLAB | Viola-Jones Algorithm | International Journal of Electronics Communication and Computer Engineering | Convolutions |

#### 3 PROBLEM FORMULATION

Computer vision is a field that includes methods for acquiring, processing analysing and understanding image sand, in general high dimensional data from the real world in order to produce numerical or symbolic information e.g.in the form of decisions. A theme in the development of this field has been to duplicate the abilities of human vision by electronically perceiving and understanding an image. This image understanding can be seen as disentangling of symbolic information from image data using models constructed with the aid of geometry, physics, statistic sand learning theory. [2]Advantages of Computer vision technologies Relatively high accuracy(95%) Flexible to customize. Highly scalable Easy to integrate with other systems Disadvantages High Cost High power consumption Accuracy can be affected by shadow, floor background, poor image quality, differing light levels.

A human counter is a device used to measure the number and direction of people traversing a certain passage or entrance per unit time. Many technologies are used in counting persons; infra-red beams, computer vision, thermal imaging and pressure sensitive mats. [1]Each technology has its merits and demerits although no system has a one hundred per-cent accuracy factor. Of all the afore-mentioned counter technologies the computer vision system is employed in this work.

In solution to our problem, we will build an AI and Machine Learning based model. This model will help us to easily get the population (no. of persons) or traffic criteria in any real time video of any place. Also we will get the idea of which area is more or less populated or trafficked as compared to normal traffic. In this model will allow user to upload any real time video and model will show us the traffic condition of the area in video. This model will also give suggestion for future about what should be the normal traffic in that area. Thus like-wise using this model one can take on control of traffic and which will prevent various chaos and accidents. And will also help to regulate traffic flow in that area.

#### 4 OBJECTIVES

The proposed work aims to create an Human detection and counting ML Model. The proposed aim will be achieved by dividing the work into the following objectives:

1. Create a GUI to select images, video or open camera
2. Detect humans and count them from the image, video, and camera

#### 5 METHODOLOGY

The following methodology will be followed to achieve the objectives defined for the project work:

1. Detailed study of about Computer vision
2. Detailed study about the algorithms used for the Human detection
3. Creating GUI for an python file using Tkinter
4. Selecting files from the computer and perform detection on that

**6 TENTATIVE CHAPTER PLAN FOR THE PROPOSED WORK**

**CHAPTER 1: INTRODUCTION**

* 1. **Problem Definition**

Human detection/counting systems are very necessary in the modern world. The need to develop intelligent security systems that can monitor footage and ‘understand’ what they see is highly important. Rather than have human personnel sit all day observing CCTV cameras an intelligent system would be able to detect an object (in this case human) of maybe a specific number and raise an alarm under any of the above conditions according to some timed program. Retail shops might also need to have information on the number of people going in or out of it’s stores in order to monitor effectiveness and impact of its market and make surveys and evaluations. Organisations/hobbyists might want to use human traffic information to: (i) limit crowds (crowd control) (ii)monitor building or seating capacity (iii) Management of queues. The main problem around us we can see is the ‘population problem’. And directly or indirectly this is causing chaos or traffic in any one or the other place. This so called chaos may sometimes even lead to quarrel or even suffocation sometimes.

* 1. **Project Overview**

A human counter is a device used to measure the number and direction of people traversing a certain passage or entrance per unit time. Many technologies are used in counting persons; infra-red beams, computer vision, thermal imaging and pressure sensitive mats. [1]Each technology has its merits and demerits although no system has a one hundred per-cent accuracy factor. Of all the afore-mentioned counter technologies the computer vision system is employed in this work. Computer vision is a field that includes methods for acquiring, processing analysing and understanding image sand, in general high dimensional data from the real world in order to produce numerical or symbolic information e.g.in the form of decisions. A theme in the development of this field has been to duplicate the abilities of human vision by electronically perceiving and understanding an image. This image understanding can be seen as disentangling of symbolic information from image data using models constructed with the aid of geometry, physics, statistic sand learning theory. [2]Advantages of Computer vision technologies Relatively high accuracy(95%) Flexible to customize. Highly scalable Easy to integrate with other systems Disadvantages High Cost High power consumption Accuracy can be affected by shadow, floor background, poor image quality, differing light levels.

In solution to our problem, we will build an AI and Machine Learning based model. This model will help us to easily get the population (no. of persons) or traffic criteria in any real time video of any place. Also we will get the idea of which area is more or less populated or trafficked as compared to normal traffic. In this model will allow user to upload any real time video and model will show us the traffic condition of the area in video. This model will also give suggestion for future about what should be the normal traffic in that area. Thus like-wise using this model one can take on control of traffic and which will prevent various chaos and accidents. And will also help to regulate traffic flow in that area.

**CHAPTER 2: LITERATURE REVIEW**

*Schofield et all* [1], Paper is based on counting people in video images using neural networks, in this paper the proposed model is based on background identification concept and subtracts the background image in order to identify number of objects or people, and facilitated to locate and track the objects/people and resulting in counting the number of objects in a frame, but this generally had a limitation if there were uncommon types of backgrounds. Zhao et all [2] is based on recording the number of people passing by a particular area, in takes a video as an input and works on the assumption that the objects are people, a scale-invariant Kalman filter is proposed and employed which is based on the strategy to count people by their face trajectories to quantize the number of people passing by. *Tsong et all [3],* also takes video as the input and satisfies the limitiations caused by Zhao et all, by using the concept of moving object segmentation. *Edeoghon et all [4]*, uses MATLAB, and is based on Viola-Jones Algorithm, and MATLAB provides proper tool and support for this algorithm. It is more similar to the functionality of Convolutional Neural Networks. The Viola-Jones Algorithm converts the facial data into M by 4 Matrix which represents features of facial data, and with this concept it finds out number of faces in a scene.

**CHAPTER 3: BACKGROUND OF PROPOSED METHOD**

Human detection[2] is the task of locating all instances of human beings present in an image, and it has been most widely accomplished by searching all locations in the image, at all possible scales, and comparing a small area at each location with known templates or patterns of people.

In this we can use various predefined methods and can detect the human in any image, video and can even get various factors like accuracy, each detections counting, etc.

Some common methods are :

* **Using Haar Cascade Classifier:**

Here we make use of .xml file for human detection, and using that we detect the humans in real time videos and images

* **Using HOG(Histogram of Oriented Gradients):**

Here we make used of predefined functions and with that we detect, and this case gives some how better accuracy as compared to Harr Cascade Classifier. In our model we have used this method.

* **Using Tensorflow:**

TensorFlow is an open-source API from Google, which is widely used for solving machine learning tasks that involve Deep Neural Networks. And again this method gives even better accuracy than above two methods.

page12image4849088

**CHAPTER 4: METHODOLOGY**

The following methodology will be followed to achieve the objectives defined for the project work:

1. Detailed study of about Computer vision
2. Detailed study about the algorithms used for the Human detection
3. Creating GUI for an python file using Tkinter
4. Selecting files from the computer and perform detection on that

We have implemented the project for three different cases: Images, Camera and Video

1. **Detection & Counting through Image**

This section works with real time images. Here will allow user to select any real time image from the local system and then user can detect the humans in it. And along with that it also gives the count of humans detected.

**2) Detection & Counting through Video**

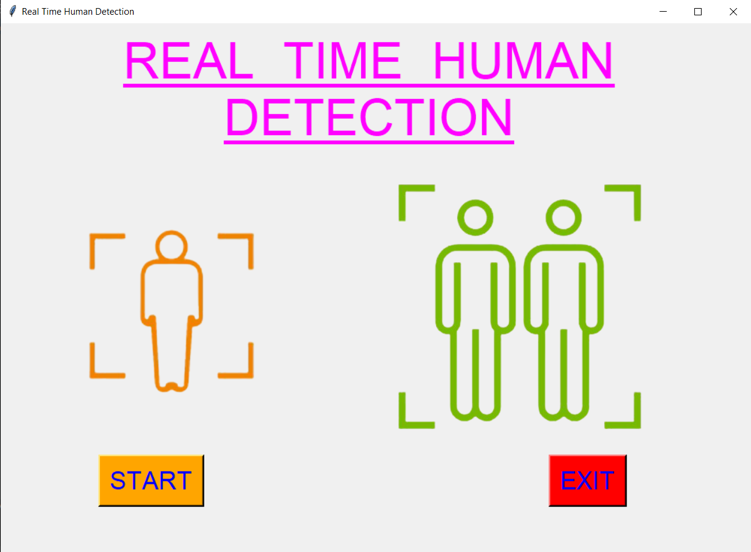
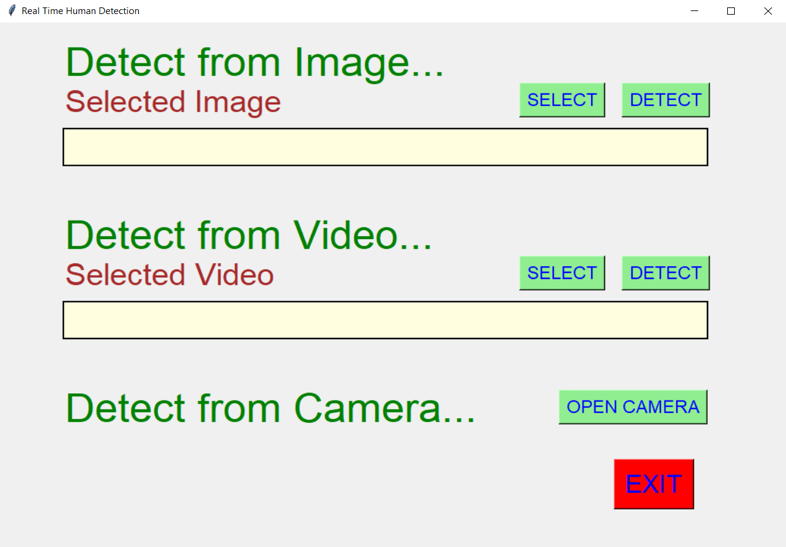
This section works with real time videos. Here will allow user to select any real time video from the local system and then user can detect the humans in it. Now in case of video, since it is running, while the detection process is going on user will be able to see the detected peoples and their count for each frames per second of the video.

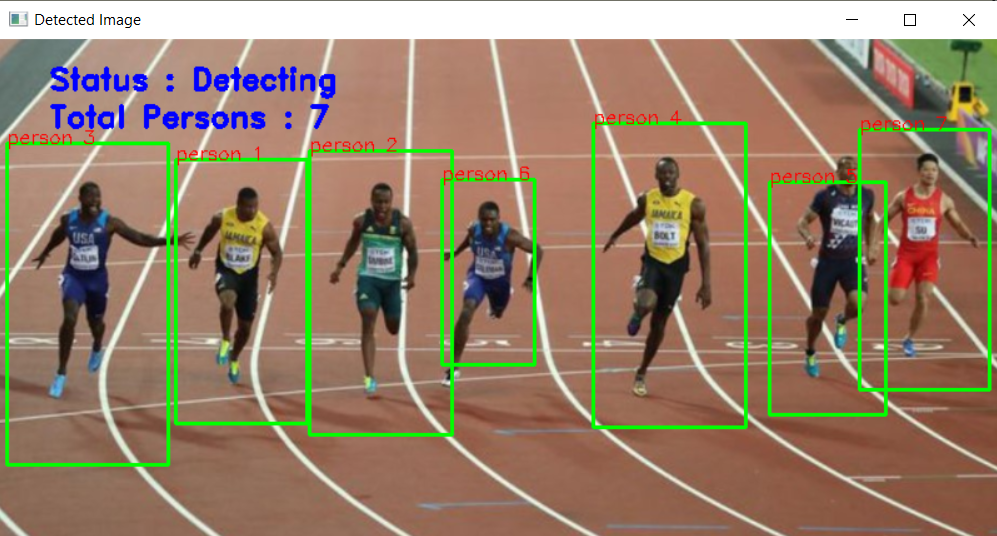
**3) Detection & Counting through Camera**

This section works somehow similar to case of video. Here user will be asked to first open the webcam, and it will detect humans that will comes in that webcam during the detection process.

**CHAPTER 5: RESULTS AND DISCUSSION**

* Using Histograms of Oriented Gradients (HOG)**:** a feature descriptor used in computer vision and image processing for the purpose of object detection. We used this method, but got no measurable difference in accuracy as in Haar Cascade Classifier method.
* Tkinter – for frontend GUI window
* PIL – for adding images to the tkinter GUI window
* Messagebox from tkinter – to show any message using dialog box
* Filedialog – to select images and video using select buttons.
* Imutils - a series of convenience functions to make basic image processing functions such as translation, rotation, resizing, skeletonization, and displaying Matplotlib images easier with OpenCV and Python.
* Argparse – this module makes it easy to write user-friendly command-line interfaces.

** **

****

**CHAPTER 6: CONCLUSION AND FUTURE SCOPE**

Now coming to the future scope of this project or application, since in this we are taking any image, video or with camera we are detecting humans and getting count of it, along with accuracy. So some of the future scope can be :

* This can be used in various malls and other areas, to analyse the maximum people count, and then providing some restrictions on number of people to have at a time at that place.
* This can replace various mental jobs, and this can be done more efficiently with machines.
* This will ultimately leads to some kind of crowd-ness control in some places or areas when implemented in that area.

#### 7 REFERENCES

1. *A.J. Schofield, P.A. Mehta, T.J. Stonham, A system for counting people in video images using neural networks to identify the background scene, Pattern Recognition, Volume 29, Issue 8,*
2. *Xi Zhao, Emmanuel Dellandrea and Liming Chen, A People Counting System based on Face Detection and Tracking in a Video*
3. *Tsong-Yi Chen, Chao-Ho Chen, Da-Jinn Wang, and Yi-Li Kuo, A People Counting System Based on Face-Detection*
4. *Edeoghon, Isi. (2021). Design and Implementation of a Human Detector and Counting System using MATLAB. 5.*