

# CHAPTER ONE

---

# 1

## Introduction

**1.1 Overview**

**1.2 Project Motivation**

**1.3 Project Aims**

**1.4 Project idea and Importance**

**1.5 Literature Review**

**1.6 List of Abbreviation**

**1.7 Estimated Cost**

**1.8 Scheduling Table**

## 1.1 Overview

The aims of project of Blind assistance is promoting a widely challenge in computer vision such as recognition of objects of the surrounding objects practiced by the blind on a daily basis. the camera placed on blind person's glasses, MS COCO is a large-scale object detection, segmentation, are employed to provide the necessary information about the surrounding environment. A dataset of objects gathered from daily scenes is created to apply the required recognition. Objects detection is used to find objects in the real world from an image of the world such as bicycles, chairs, doors, or tables that are common in the scenes of a blind. based on their locations, and The camera is used to detect any objects. The proposed method for the blind aims at expanding possibilities to people with vision loss to achieve their full potential. The main object of the project is to design and implement a real time object recognition using blind glass.

## 1.2 List of Abbreviation

**Table (1.1):** List of Abbreviation

Abbreviation	Full Meaning
GPS	Global Positioning System
TF	Tensor flow
HOG	Histogram of Oriented Gradients
SURF	Speeded Up Robust Features
MS COCO	Common Objects in Context
API	Application Programming Interface
PASCAL	pattern analysis statically modeling and computational learning
VPU	Visual processing Unit

## 1.3 Project Motivation

Percentage of persons with disabilities in Palestinian society. Especially those with visual disabilities (blind) which is at estimated [0.6 %] It is not simple. From here the idea of our project begins where it aims. the aims of project of Blind assistance is promoting a widely challenge in computer vision such as **recognition of objects** of the surrounding objects practiced by the blind on a daily basis.

## 1.4 Project Aims

The main objective of the project is to design and implement a smart glass for blind people using special mini camera.

- Connect the mini camera with a Raspberry Pi.
- Programing the Raspberry Pi using Python language, its powerful for processing.
- Process and analyze the camera records using Raspberry Pi in real time.
- Detect and recognize objects in front of the blind.
- Design and build an alarm system to notify the user about the recognized objects using voice messages.

## 1.5 Project idea and Importance

- ❖ This project is mainly aimed at helping people who are blind and who suffer from a total lack of vision.
- ❖ Due to the development of technology, we must be tapped to help blind people.
- ❖ Due to the large number of blind people in Palestine.
- ❖ The next future and the future of technology is to serve people and help them in life.

## 1.6 Literature Review

### 1- Real-Time Objects Recognition Approach for Assisting Blind People.

[Jamal S. Zraqou Wissam M. Alkhadour and Mohammad Z. Siam, Multimedia Systems Department, Electrical Engineering Department, Isra University, Amman-Jordan Accepted 30 Jan 2017, Available online 31 Jan 2017, Vol.7, No.1]

Blind assistance is promoting a widely challenge in computer vision such as navigation and path finding. In this paper, two cameras placed on blind person's glasses, GPS free service, and ultra-sonic sensor are employed to provide. the necessary information about the surrounding environment. A dataset of objects gathered from daily scenes is created to apply the required recognition. Objects detection is used to find objects in the real world from an image of the world such as faces, bicycles, chairs, doors, or tables that are common in the scenes of a blind.

The two cameras are necessary to generate the depth by creating the disparity map of the scene, GPS service is used to create groups of objects based on their locations, and the sensor is used to detect any obstacle at a medium to long distance.

The descriptor of the Speeded-Up Robust Features method is optimized to perform the recognition. The proposed method for the blind aims at expanding possibilities to people with vision loss to achieve their full potential. The experimental results reveal the performance of the proposed work in about real time system.

## **2-Voscal vision for visually impaired**

[The International Journal Of Issn: 2319 – 1813 Isbn: 2319 – 1805 Engineering And Science(Ijes)-01-07||2013|| Shrilekha Banger , Preetam Narkhede , Rajashree Parajape.]

This project is a vision substitute system designed to assist blind people for autonomous navigation. Its working concept is based on ‘image to sound’ conversion. The vision sensor captures the image in front of blind user. This image is then fed to MATLAB for processing. Process intuit processes the captured image and enhances the significant vision data. This processed image is then compared with the data base kept in microcontroller. The processed information is then presented as a structured form of acoustic signal and it is conveyed to the blind user using set of ear phones. Color information from the interested objects evaluated to determine the color of the object. The color output is informed to the blind user through headphones.

## **3-Object Detection Combining Recognition and Segmentation**

[Fudan University, Shanghai, PRC, yfshen@fudan.edu.cn University of Pennsylvania,3330WalnutStreet, Philadelphia, PA19104 Liming Wang1, Jianbo Shi2, Gang Song2, and I-fan Shen.]

We develop an object detection method combining top-down recognition with bottom-up image segmentation. There are two main steps in this method: a hypothesis generation step and a verification step. In the top-down hypothesis generation step, we design an improved Shape Context feature, which is more robust to object deformation and background clutter. The improved Shape Context is used to generate a set of hypotheses of object locations and figure ground masks, which have high recall and low precision rate. In the verification step, we first compute a set of feasible segmentations that are consistent with top-down object hypotheses, then we propose a False Positive Pruning(FPP) procedure to prune out false positives. We exploit the fact that false positive regions typically do not align with any feasible image segmentation. Experiments show that this simple framework is capable of achieving both high recall and high precision with only a few positive training examples and that this method can be generalized to many object classes.

## **4 - Microsoft COCO Common Objects in Context**

[Tsung-Yi Lin, Michael Maire, Serge Belongie, Lubomir Bourdev, Ross Girshick, James Hays, Pietro Perona, Deva Ramanan, C. Lawrence Zitnick, Piotr Dollár(Submitted on 1 May 2014 (v1), last revised 21 Feb 2015 (this version, v3))]

We present a new dataset with the goal of advancing the state-of-the-art in object recognition by placing the question of object recognition in the context of the broader question of scene understanding. This is achieved by gathering images of complex everyday scenes containing common objects in their natural context. Objects are labeled using per-instance

segmentations to aid in precise object localization. Our dataset contains photos of 91 objects types that would be easily recognizable by a 4-year-old. With a total of 2.5 million labeled instances in 328k images, the creation of our dataset drew upon extensive crowd worker involvement via novel user interfaces for category detection, instance spotting and instance segmentation. We present a detailed statistical analysis of the dataset in comparison to PASCAL, ImageNet, and SUN. Finally, we provide baseline performance analysis for bounding box and segmentation detection results using a Deformable Parts Model.

## 1.7 Project Cost

**Table (1.2):** Project Cost

Component	Cost \$
Mini Camera	20 \$
Raspberry pi 3 model b	60 \$
Rubber Glasses	10 \$
Ear phone	10 \$
Rechargeable Battery	30 \$
Total	130 \$

## 1.8 Scheduling Table

**Table (1.3):** shows the activities that done in the project, and the time of each one.

Weeks Activities	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
System Definition																
System Analysis																
System Design																
Presentation Preparing																
Documentation																



# CHAPTER TWO

---

# 2

## **Human Eye: Anatomy & Physiology**

### **2.1 Introduction**

### **2.2 Human Eye Anatomy**

### **2.3 Visual Processing**

### **2.4 Visual Impairment and blindness**

#### **2.4.1 Causes of Visual Impairment and Blindness**

## 2.1 Introduction

The human eye is the organ which gives us the sense of sight, allowing to observe and learn more about the surrounding world than we do with any of the other from sense. We use our eye in almost every activity we perform, whether reading, working, watching television, writing a letter, driving a car, and in countless other ways. Most people probably would agree that sight is the sense they value more than all the rest.

## 2.2 Human Eye Anatomy

The human eye is very nearly spherical, with a diameter of approximately 24 millimeters (nearly 1 inch), or slightly smaller than a Ping-Pong ball. It consists of three concentric layers, each with its own characteristic appearance, structure, and functions [1].

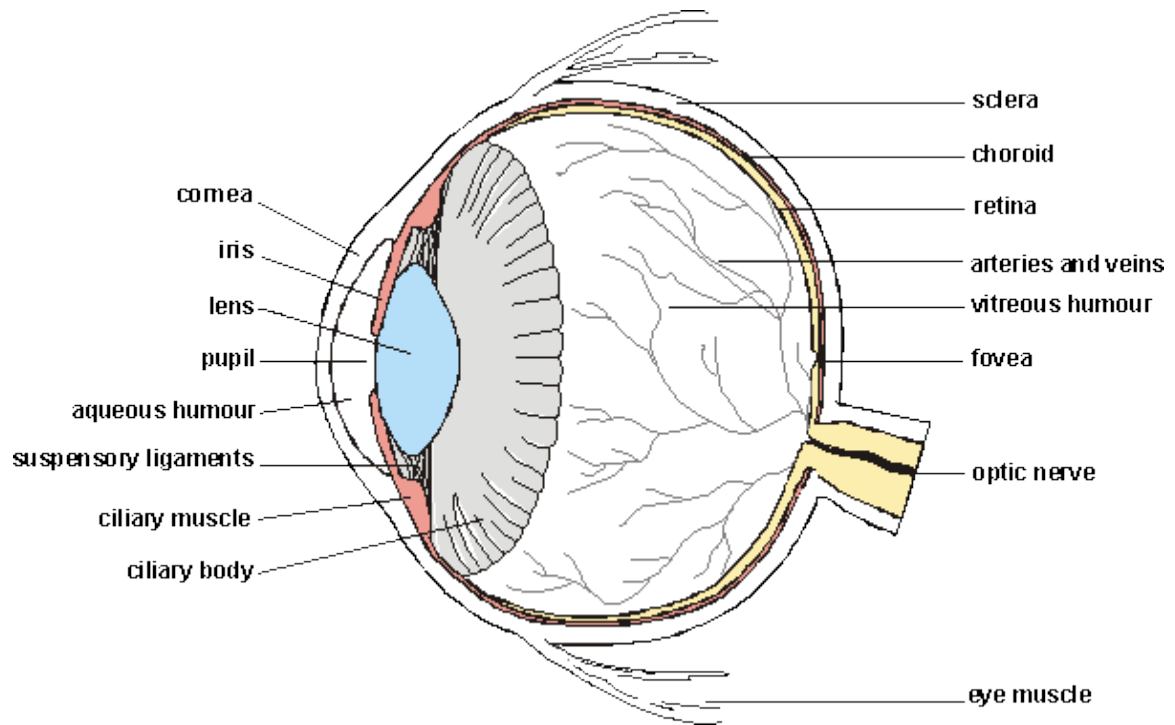
From outermost to innermost, the three layers are the fibrous tunic, which protects the eyeball; and the retina which detects light and initiates neural messages bound for the brain [2].

The eye is made up of three coats, enclosing three transparent structure. the outermost layer, known as the fibrous tunic, is composed of the cornea and sclera. The middle layer known as the vascular tunic or uvea consists of the choroid, ciliary body, and iris. The innermost is the retina, which gets its circulation from the vessels of the choroid as well as the retinal vessels, which can be seen in an ophthalmoscope [3].

Within these coats are the aqueous humor, the vitreous body, and the flexible lens. the aqueous humor is a clear fluid that is contained in two areas: the anterior chamber between the cornea and the iris, and the posterior chamber between the iris and the lens. The lens is suspended to the ciliary body by the suspensory ligament (Zonule of Zinn), made up of fine transparent fibers the vitreous body is a clear jelly that is much larger than the aqueous humor present behind the lens, and the rest is bordered by the sclera, zonula and lens. they are connected via the pupil [4]. Figure (2.1) illustrate the main components of human eye [5].

- Conjunctiva: is a thin protective covering of epithelial cells. It protects the cornea against damage by friction (tears from the tear glands help this process by lubricating the surface of the conjunctiva).
- Cornea: is the transparent, curved front of the eye which helps to converge the light rays which enter the eye.
- Sclera: is an opaque, fibrous, protective outer structure. It is soft connective tissue, and the spherical shape of the eye is maintained by the pressure of the liquid inside. it provides attachment surface for eye muscle.





**Figure (2.1):** Components of eye [5].

- **Choroid:** has a network of blood vessels to supply nutrients to the cells and remove waste products. It is pigmented that makes the retina appear black, thus preventing reflection of light within the eyeball.
- **Ciliary body:** has suspensory ligaments that hold the lens in place. It secretes the aqueous humor, and contains ciliary muscles that enable the lens to change shape, during accommodation (focusing on near and distant objects).
- **Iris:** is a pigmented muscular structure consisting of an inner ring circular muscle and outer layer of radial muscle.
- **Pupil:** is a hole in the middle of the iris where light is allowed to continue its passage. In bright light it is constricted and in dim light is dilated.
- **Lens:** is a transparent, flexible, curved structure. Its function is to focus incoming light rays onto the retina using its refractive properties.
- **Retina:** is a layer of sensory neurons, the key structure being photoreceptors (rod and cone cells) which respond to light. Contains relay neurons and sensory neurons that pass impulses along the optic nerve to the part of the brain that controls vision.
- **Fovea (yellow spot):** a part of the retina that is directly opposite the pupil and contains only cone cells. It is responsible for good visual acuity (good resolution).
- **Blind Spot:** is where the bundle of sensory fibers from the optic nerve, it contains no light-sensitive receptors.

- Vitreous Humor: is a transparent, jelly-like mass located behind the lens. It acts as a ‘suspension’ for the lens so that delicate lens is not damaged. It helps to maintain the shape of the posterior chamber of the eyeball.
- Aqueous Humor: helps to maintain the shape of the anterior chamber of the eyeball.

## 2.3 Visual Processing

The ability to see clearly depends on how well these parts work together. Light rays bounce off all objects. If a person is looking at a particular object, such as a tree, light is reflected off the tree to the person’s eye and enters the eye through the cornea (clear, transparent portion of the coating that surrounds the eyeball) [6].

Next, light rays pass through an opening in the iris (colored part of the eye), called the pupil. The iris controls the amount of light entering the eye by dilating or constricting the pupil. In bright light, for example, the pupil shrinks to the size of a pinhead to prevent too much light from entering. In dim light, the pupil enlarges to allow more light to enter the eye [7].

Light then reaches the crystalline lens. The lens focusses light rays onto the retina by bending (refracting) them. The cornea does most of the refraction and the crystalline lens fine-tunes the focus. In a healthy eye, the lens can change its shape (accommodate) to provide clear vision at various distances. If an object is close, the ciliary muscle of the eye contracts and the lens becomes rounder. To see a distant object, the same muscle relaxes and the lens flattens [8].

Behind the lens and in front of the retina is a chamber called vitreous body, which contains a clear, gelatinous fluid called vitreous humor. Light rays pass through the vitreous before reaching the retina. The retina lines the back two-thirds of the eye and is responsible for the wide field of vision that most people experience. For clear vision, light rays must focus directly on the retina. When light focuses in front or behind the retina, the result is blurry vision [9].

The retina contains millions of specialized photoreceptor cells called rods and cones that convert light rays into electrical signals that are transmitted to the brain through the optic nerve. Rods and cones provide the ability to see in dim light and see in color, respectively [10].

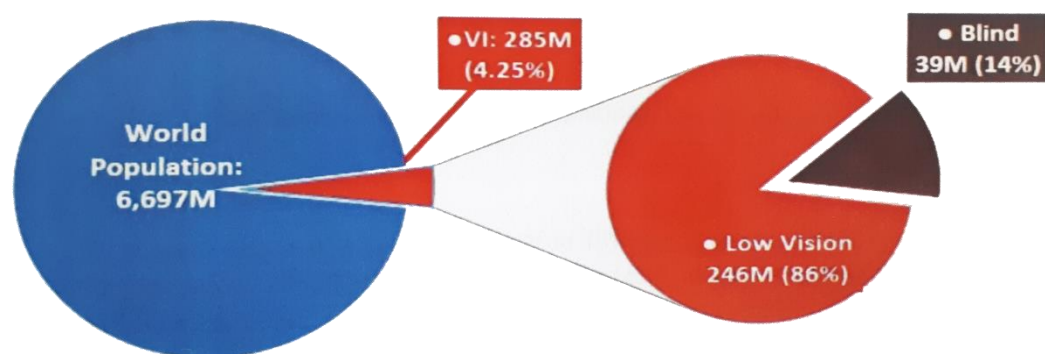
The macula, located in the center of the retina, is where most of the cone cells are located. The fovea, a small depression in the center of the macula, has the highest concentration of cone cells. The macula is responsible for central vision, seeing color, and distinguishing fine detail. The outer portion (peripheral retina) is the primary location of rod cells and allows for night vision and seeing movement and objects to the side (i.e., peripheral vision) [11].

The optic nerve, located behind the retina, transmits signals from the photoreceptor cells to the brain. Each eye transmits signals of a slightly different image that are inverted. Once they reach the brain, they are corrected and combined into one image. This complex process of analyzing data transmitted through the optic nerve is called visual processing [12].

## 2.4 Visual Impairment and Blindness

The World Health Organization (WHO) defines **Visual impairment** decrease or severe reduction in vision that cannot be corrected with standard glasses or contact lenses and reduce an individual's ability to function at a specific or all tasks [13].

**Blindness** as severe sight loss, where a person is unable to see clearly how many fingers are being held up at a distance of 3m (9.8 feet) or less, even when they are wearing glasses or contact lenses. However, someone who is blind may still have some degree of vision [13].



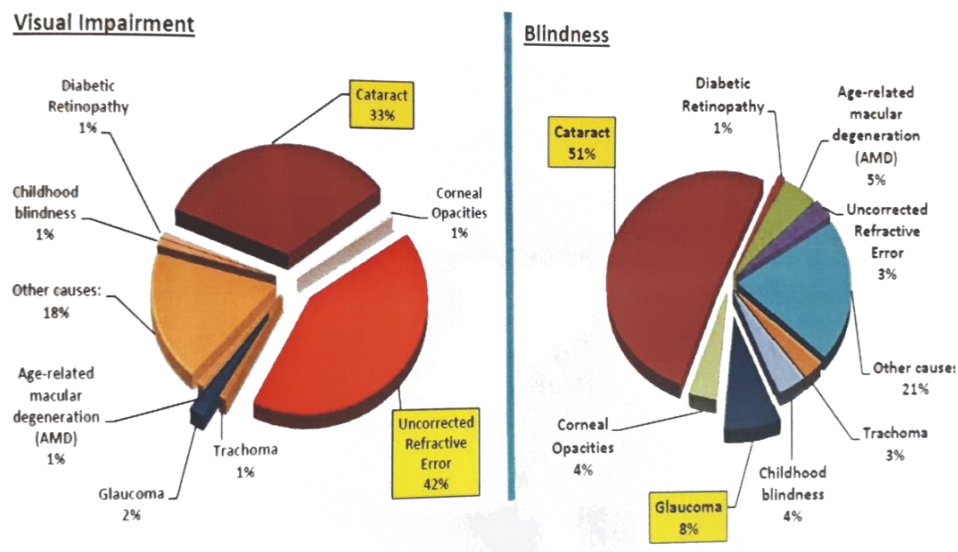
**Figure (2.2):** Global estimate of visual impairment [13].

According to WHO criteria, global estimate predicts that there are 285 million people with visual impairment approximately are blind.

Most people (87%) who are visually impaired live in developing countries. In developing countries, cataracts (a cloudy area that forms in the lens of the eye) are responsible for most cases of blindness (48%). visual impairment usually affects older people. Globally, women are more at risk than men. With the right treatment, about 85% of visual impairment cases are avoidable, and approximately 75% of all blindness can be treated or prevented [13].

### 2.4.1 Causes of Visual Impairment and Blindness

In spite of the progress made in surgical techniques in many countries during the last ten years, cataract (47.9%) remains the leading cause of visual impairment in all areas of the world, except for developed countries.



**Figure (2.3):** Causes of Visual Impairment and Blindness [13]

Other main causes of visual impairment on 2010 are glaucoma (2%), age-related macular degeneration (AMD) (1%), corneal opacities (1%), diabetic retinopathy (4.8%), childhood blindness (1%), trachoma (1%), and onchocerciasis (0.8%). The causes of avoidable visual impairment worldwide are all the above except for AMD. In the least-developed countries, and in particular Sub-Saharan Africa, the causes of avoidable blindness are primarily, cataract (51%), glaucoma (8%), corneal opacities (4%), trachoma (3%), childhood blindness (4%) and onchocerciasis (1%) [13].

Looking at the global distribution of the avoidable blindness based on the population in each of the WHO regions, we see the following: South Asian 28%, Western Pacific 26%, African 16.6%, Eastern Mediterranean 10%, the American 9.6%, and European 9.6% [13].

In addition to uncorrected refractive errors, these six diseases or groups of diseases which have effective known strategies for their elimination, make up the targets of the WHO Global Initiative to Eliminate Avoidable Blindness, “VISION 2020: The Right to Sight”, which aims to eliminate these causes as a public health problem by the year 2020. cataract, onchocerciasis, and trachoma are the principle diseases for which world strategies and programmers have been developed. For glaucoma, diabetic retinopathy, uncorrected refractive errors, and childhood blindness (except for exophthalmia), the development of screening and management strategies for use at the primary care level is ongoing at WHO[13].

# CHAPTER THREE

# 3

## Computer vision and Image Processing

### 3.1 Human vision

### 3.2 Computer vision

#### 3.2.1 Human Vision VS Computer Vision

#### 3.2.2 Main goal of computer vision

#### 3.2.3 Advantages and Disadvantages of computer vision

#### 3.2.4 Applications of Computer Vision

### 3.3 Levels of Computer vision

### 3.4 Fundamental steps in digital image processing

#### 3.4.1 Image Acquisition

#### 3.4.2 Enhancement Image Processing

#### 3.4.3 Restoration Image Processing

#### 3.4.4 Color Image Processing

#### 3.4.5 Wavelets and multiresolution processing

#### 3.4.6 Compression Image Processing

#### 3.4.7 Morphological Image Processing

#### 3.4.8 Segmentation Image Processing

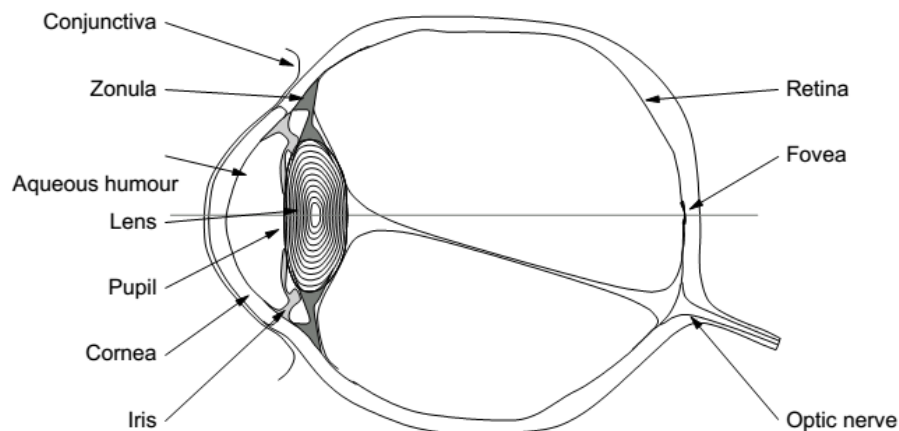
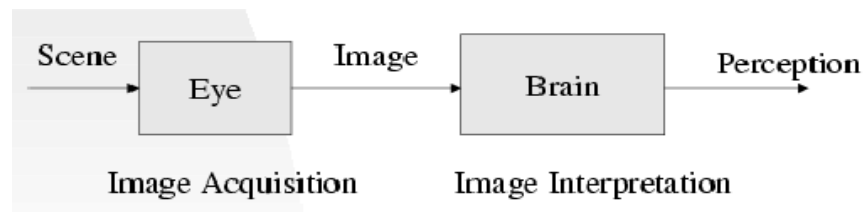
#### 3.4.9 Representation & description

#### 3.4.10 Object Recognition

### 3.1 human vision

Vision is the process of discovering what is present in the world and where it is by looking. The human visual system can be regarded as consisting of two parts. The eyes act as image receptors which capture light and convert it into signals which are then transmitted to image processing centers in the brain. These centers process the signals received from the eyes and build an internal “picture” of the scene being viewed. Processing by the brain consists of partly of simple image processing and partly of higher functions which build and manipulate an internal model of the outside world. Although the division of function between the eyes and the brain is not clear-cut, it is useful to consider each of the components separately [24].

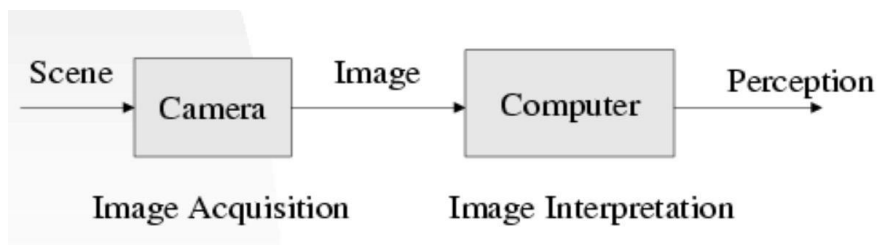
- ❖ Limitations of Human Vision [30]:
  - limited memory-cannot remember a quickly flashed image
  - limited to visible spectrum
  - illusion



**Figure (3.1):** A cross-section of the right human eye, viewed from above [24].

### 3.2 Computer vision

Computer vision is an interdisciplinary field that deals with how computers can be made for gaining high-level understanding from digital images or videos. From the perspective of engineering, it seeks to automate tasks that the human visual system can do[14][15][16].



**Figure (3.2):** Computer vision similar to those as by humans [24].

Computer vision tasks include methods for acquiring, processing, analyzing and understanding digital images, and extraction of high-dimensional data from the real world in order to produce numerical or symbolic information, e.g., in the forms of decisions[17][18][19][20].

Understanding in this context means the transformation of visual images (the input of the retina) into descriptions of the world that can interface with other thought processes and elicit appropriate action. This image understanding can be seen as the disentangling of symbolic information from image data using models constructed with the aid of geometry, physics, statistics, and learning theory[21]

### 3.2.1 Human Vision VS Computer Vision

In object classification, it was known that the human brain tactics visible know-how in a semantic area traditionally, i.e. Extracting the semantically significant elements equivalent to line segments, shape, boundaries etc. In any case, by late data handling methods, these sorts of components can't be recognized by PCs heartily so that in PC vision it's still hard to prepare visual data as people do. PCs need to prepare visual data in information space framed by the vigorously distinguishable yet less important components, for example, hues, surfaces, and so on. In this way, the handling philosophy in PCs is entirely not the same as that in individuals [29].



0	3	2	5	4	7	6	9	8
3	0	1	2	3	4	5	6	7
2	1	0	3	2	5	4	7	6
5	2	3	0	1	2	3	4	5
4	3	2	1	0	3	2	5	4
7	4	5	2	3	0	1	2	3
6	5	4	3	2	1	0	3	2
9	6	7	4	5	2	3	0	1
8	7	6	5	4	3	2	1	0

**Figure (3.3):** What we see what computer sees [31].

Computer vision is an interdisciplinary field that deals with how computers can be made for gaining high-level understanding from digital images or videos. From the perspective of engineering, it seeks to automate tasks that the human visual system can do[14][14][16]. "Computer vision is concerned with the automatic extraction, analysis and understanding of useful information from a single image or a sequence of images. It involves the development of a theoretical and algorithmic basis to achieve automatic visual understanding." [22] As a scientific discipline, computer vision is concerned with the theory behind artificial systems that extract information from images. The image data can take many forms, such as video sequences, views from multiple cameras, or multi-dimensional data from a medical scanner.[23] As a technological discipline, computer vision seeks to apply its theories and models for the construction of computer vision systems.

### **3.2.2 Main goal of computer vision:**

Computer Vision has a dual goal. From the biological science point of view, computer vision aims to come up with computational models of the human visual system. From the engineering point of view, computer vision aims to build autonomous systems which could perform some of the tasks which the human visual system can perform (and even surpass it in many cases). Many vision tasks are related to the extraction of 3D and temporal information from time-varying 2D data such as obtained by one or more television cameras, and more generally the understanding of such dynamic scenes. Of course, the two goals are intimately related. The properties and characteristics of the human visual system often give inspiration to engineers who are designing computer vision systems. Conversely, computer vision algorithms can offer insights into how the human visual system works. In this paper we shall adopt the engineering point of view [28].

### **3.2.3 Advantages and Disadvantages of computer vision [29]**

#### **Advantages of computer vision:**

- Simple & Faster processes: - Speedy computer systems replace lengthy visual tests.
- Reliability: - Cameras and computers, as opposed to a human eye, in no way get tired.
- Accuracy: - By using utilizing computer imaginative and prescient the completing of the tip product increases to a large extent.
- A wide variety of use: - It has an awfully huge range of applications.
- Price discount: - Time is saved on people and devices, therefore misguided merchandise is eliminated.
- No boundaries like human perception.



- Do not must have instruments embedded, bodily printed or externally hooked up to objects specified for detection.
- Picture shooting gadgets are convenient to mount, do away with, substitute and upgrade.
- Upgrading photo sensors doesn't require upgrading tags, identifiers or transponder devices.

#### **Disadvantages of computer vision:**

- Data processing and analytics is intensive and requires a lot of computation resources and memory
- Main technical barriers are its robustness in the face of fixing the environment.
- Illumination variation further complicates the design of effective algorithms on account that of changes in shadows being cast.

### **3.2.4 Applications of Computer Vision**

The computer vision is being utilized today as a part of a wide assortment of true applications, as:

1. Optical character realization (OCR): Studying handwritten postal codes on letters and Automatic Number Plate Recognition (ANPR).
2. Retail: Object acknowledgment for mechanized checkout paths.
3. 3D model building (Photogrammetry): Fully mechanized development of 3D models from airborne photos utilized as a part of frameworks, for example, Bing Maps.
4. Medical Imaging: Registering pre-agent and intra-agent symbolism or performing long haul investigations of individuals' cerebrum morphology as they age.
5. Automobile defense: Detecting startling deterrents like people on foot in the city, under conditions where dynamic vision strategies, for example, radar don't function admirably.
6. Movement seizes (mobcap): Using retro –reflective markers saw from different cameras or other vision-based systems to catch on-screen characters for Computer animation.
7. Surveillance: Monitoring for gate crashers, investigating thruway activity, and observing pools for suffocating casualties.
8. Fingerprint attention and biometrics: For programmed get to validation and additionally scientific applications.

### **3.3 levels of Computer vision**

#### **1. Low-level Vision:**

Based on low-level image processing, low-level vision tasks could be performed, such as image matching, optical flow computation and motion analysis. Image matching basically is to find correspondences between two or more images. These images could be the same scene taken from different view points, or a moving scene taken by a fixed camera, or both. Constructing image correspondences is a fundamentally important problem in vision for both geometry

recovery and motion recovery. Without exaggeration, image matching is part of the base for vision [24].

Optical flow is a kind of image observation of motion, but it is not the true motion. Since it only measures the optical changes in images, an aperture problem is unavoidable. But based on optical flows, camera motion or object motion could be estimated [34].

## **2. Middle-level Vision:**

There are two major aspects in middle-level vision: (1) inferring the geometry and (2) inferring the motion. These two aspects are not independent but highly related. A simple question is “can we estimate geometry based on just one image?”. The answer is obvious. We need at least two images. They could be taken from two cameras or come from the motion of the scene [34].

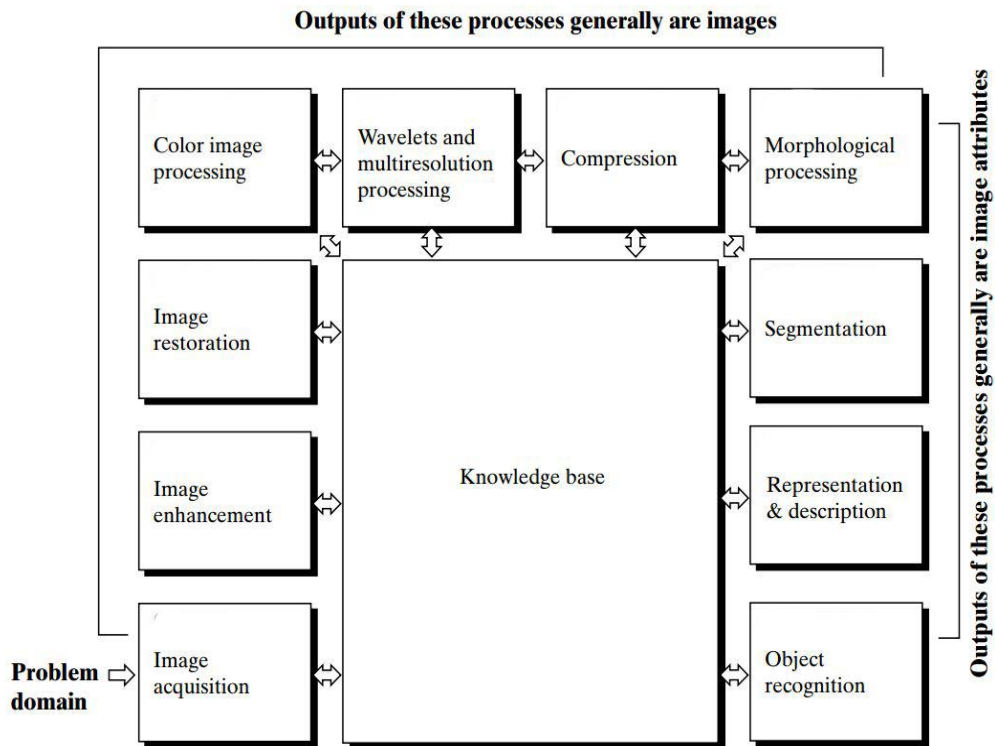
Some fundamental parts of geometric vision include multi view geometry, stereo and structure from motion (SFM), which fulfill the step of from 2D to 3D by inferring 3D scene information from 2D images. Based on that, geometric modelling is to construct 3D models for 6 objects and scenes, such that 3D reconstruction and image-based rendering could be made possible[29].

Another task of middle-level vision is to answer the question “how the object moves”. Firstly, we should know which areas in the images belong to the object, which is the task of image segmentation. Image segmentation has been a challenging fundamental problem in computer vision for decades. Segmentation could be based on spatial similarities and continuities. However, uncertainty cannot be overcome for static image. When considering motion continuities, we hope the uncertainty of segmentation could be alleviated. On top of that is visual tracking and visual motion capturing, which estimate 2D and 3D motions, including deformable motions and articulated motions [34].

## **3. High-level Vision:**

High-level vision is to infer the semantics, for example, object recognition and scene understanding. A challenging question in many decades is that how to achieve invariant recognition, i.e., recognize 3D object from different view directions. There have been two approaches for recognition: model-based recognition and learning-based recognition. It is noticed that there was a spiral development of these two approaches in history. Even higher level vision is image understanding and video understanding. We are interested in answering questions like “Is there a car in the image? Or Is this video a drama or an action? or is the person in the video jumping? Based on the answers of these questions, we should be able to fulfill different tasks in intelligent human-computer interaction, intelligent robots, smart environment and content-based multimedia [34].

## **3.4 Fundamental steps in digital image processing**



**Figure (3.4):** Fundamental steps in digital image processing [40].

### 3.4.1 Image Acquisition

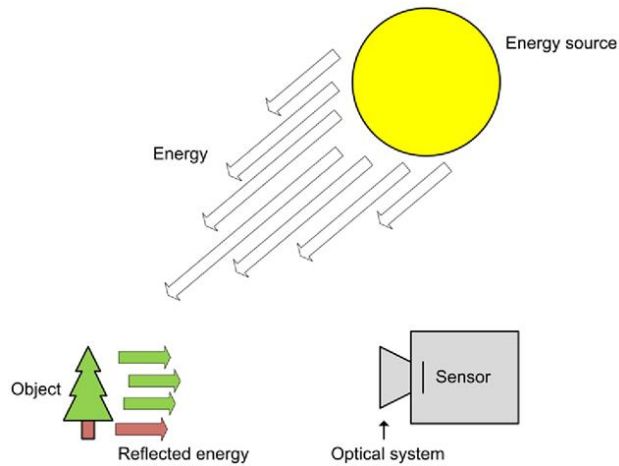
Before any video or image processing can commence an image must be captured by a camera and converted into a manageable entity. This is the process known acquisition. The image acquisition process consists of two steps; energy reflected from the object of interest, an optical system which focuses the energy and finally a sensor which measures the amount of energy. In Fig. 3.1 the three steps are shown for the case of an ordinary camera with the sun as the energy source.

Energy in order to capture an image a camera requires some sort of measurable energy. The energy of interest in this context is light or more generally electromagnetic waves. An electromagnetic (EM) wave can be described as massless entity, a photon, whose electric and magnetic fields vary sinusoidal, hence the name wave. The photon belongs to the group of fundamental particles and can be described in three different ways [35].

$$\lambda = c / f, \quad E = h \cdot f \quad \Rightarrow \quad E = h \cdot c \lambda \dots\dots\dots (3.1)$$

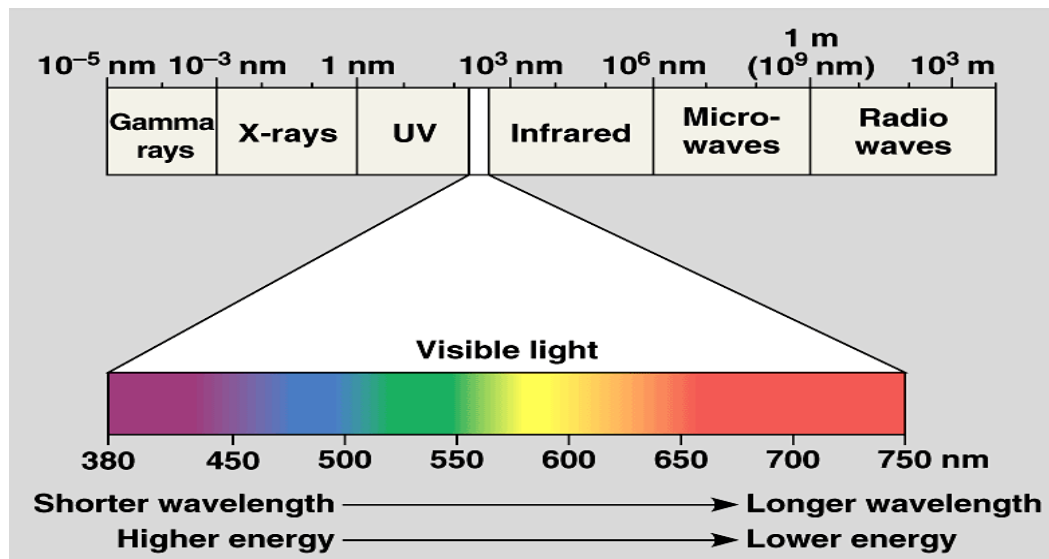
- A photon can be described by its energy E, which is measured in electron volts [eV]
- A photon can be described by its frequency f, which is measured in Hertz [Hz]. A frequency is the number of cycles or wave-tops in one second
- A photon can be described by its wavelength  $\lambda$ , which is measured in meters [m].

A wavelength is the distance between two wave-tops The three different notations are connected through the speed of light c and Planck's constant h.



**Figure (3.5):** Overview of the typical image acquisition process, with the sun as light source, a tree as object and a digital camera to capture the image [35].

The range from approximately 400–700 nm (nm = nanometer =  $10^{-9}$ ) is denoted the visual spectrum. The EM waves within this range are those your eye (and most cameras) can detect. This means that the light from the sun (or a lamp) in principle is the same as the signal used for transmitting TV, radio or for mobile phones etc. The only difference, in this context, is the fact that the human eye can sense EM waves in this range and not the waves used for e.g., radio. Or in other words, if our eyes were sensitive to EM waves with a frequency around  $2 \cdot 10^9$  Hz, then your mobile phone would work as a flash light, and big antennas would be perceived as “small suns”. Evolution has (of course) not made the human eye sensitive to such frequencies but rather to the frequencies of the waves coming from the sun, hence visible light [35].



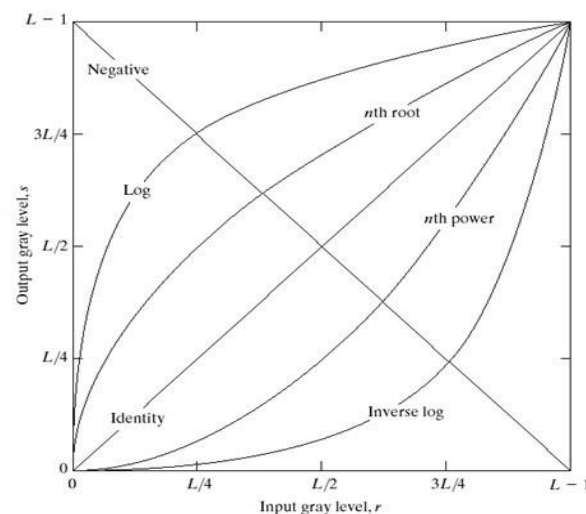
**Figure (3.6):** The visual spectrum [35].

### 3.4.2 Enhancement Image Processing

The principal objective of enhancement is to process an image so that the result is more suitable than the original image for a specific application. Regardless of the method used, however, image enhancement is one of the most interesting and visually appealing areas of image processing.

Image enhancement approaches fall into two broad categories: The term spatial domain refers to the image plane itself, and approaches in this category are based on direct manipulation of pixels in an image. Frequency domain processing techniques are based on modifying the Fourier transform of an image

There is no general theory of image enhancement. When an image is processed for visual interpretation, the viewer is the ultimate judge of how well a particular method works. Visual evaluation of image quality is a highly subjective process, thus making the definition of a “good image” an elusive standard by which to compare algorithm performance. When the problem is one of processing images for machine perception, the evaluation task is somewhat easier [35].



**Figure (3.7):** Some basic gray-level transformation functions used for image enhancement [35]

We begin the study of image enhancement techniques by discussing gray-level transformation functions. These are among the simplest of all image enhancement techniques. The values of pixels, before and after processing, will be denoted by  $r$  and  $s$ , respectively. As indicated in the previous section, these values are related by an expression of the form:  $s=T(r)$ , where  $T$  is a transformation that maps a pixel value  $r$  into a pixel value  $s$ . Since we are dealing with digital quantities, values of the transformation function typically are stored in a one-dimensional array and the mappings from  $r$  to  $s$  are implemented via table lookups. For an 8-bit environment, a lookup table containing the values of  $T$  will have 256 entries. As an introduction to gray-level transformations, consider Fig. 3.3, which shows three basic types of functions used frequently for image enhancement: linear (negative and identity transformations), logarithmic (log and inverse-log transformations), and power-law ( $n$ th power and  $n$ th root transformations) [35].

### 3.4.3 Restoration Image Processing

Image Restoration techniques aim at modelling a degradation corrupting the image and inverting this degradation to correct the image so that it is as close as possible to the original.

- Image restoration attempts to restore images that have been degraded.
  - Identify the degradation process and attempt to reverse it.
  - Similar to image enhancement, but more objective.
- The sources of noise in digital images arise during image acquisition (digitization) and transmission.
  - Imaging sensors can be affected by ambient conditions.
  - Interference can be added to an image during transmission.

We can consider a noisy image to be modelled as follows:

$$g(x, y) = f(x, y) + \eta(x, y) \dots \dots \dots (3.5)$$

where  $f(x, y)$  is the original image pixel,  $\eta(x, y)$  is the noise term and  $g(x, y)$  is the resulting noisy pixel. If we can estimate the model of the noise in an image, this will help us to figure out how to restore the image [36] [37] [38].

#### The Noise Sources

where  $f(x, y)$  is the original image pixel,  $\eta(x, y)$  is the noise term and  $g(x, y)$  is the resulting noisy pixel. If we can estimate the model of the noise in an image, this will help us to figure out how to restore the image.

The principal sources of noise in digital images arise during image acquisition and/or transmission.

- Image acquisition e.g., light levels, sensor temperature, etc.
- Transmission e.g., lightning or other atmospheric disturbance in wireless network.

#### Filtering to Remove Noise

We can use spatial filters of different kinds to remove different kinds of noise. The arithmetic mean filter is a very simple one and is calculated as follows:

$$\hat{f}(x, y) = 1/mn \sum g(s, t) \dots \dots \dots (3.6)$$

This is implemented as the simple smoothing filter. It blurs the image to remove noise.

### 3.4.4 Color Image Processing

**Color Image Processing is divided into two major areas:**

1) Full-color processing

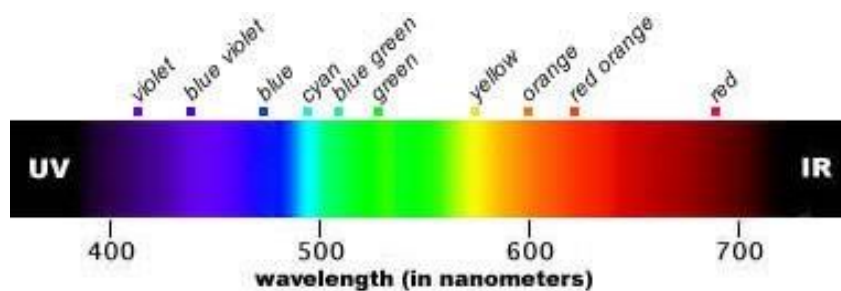
- Images are acquired with a full-color sensor, such as a color TV camera or color scanner.
- Used in publishing, visualization, and the Internet

2) Pseudo color processing

- Assigning a color to a particular monochrome intensity or range of intensities.

Visible light as a narrow band of frequencies in EM

- A body that reflects light that is balanced in all visible wavelengths appears white
- However, a body that favors reflectance in a limited range of the visible spectrum exhibits some shades of color
- Green objects reflect wavelength in the 500 nm to 570 nm range while absorbing most of the energy at other wavelengths [40].



**Figure (3.8):** wavelength (in nanometers)[40].

### 3.4.5 Wavelets and multiresolution processing

Wavelets are mathematical functions that splits up data into different frequency components, and then study each component with a resolution matched to its scale.

Wavelet transform decomposes a signal into a set of basic functions. These basis functions are called as “wavelets” [41]. use wavelet for:

- Good approximation properties.
  - Efficient way to compress the smooth data except in localized-region.
- Easy to control wavelet properties.  
( Example : Smoothness, better accuracy near sharp gradients).

Wavelets are a powerful statistical tool which can be used for a wide range of applications:

Signal processing.- Image processing.-Smoothing and image denoising.- Speech recognition.

The advantage of wavelet compression is that, in contrast to JPEG, wavelet algorithm does not divide image into blocks, but analyze the whole image. Wavelet transform is applied to sub images, so it produces no blocking artifacts. Wavelets have the great advantage of being able to separate the fine details in a signal. Wavelet allows getting best compression ratio, while maintaining the quality of the images [41].

Image compression using wavelet transforms results in an improved compression ratio as well as image quality. Wavelet transform is the only method that provides both spatial and frequency domain information. These properties of wavelet transform greatly help in identification and selection of significant and no significant coefficient. Wavelet transform techniques currently provide the most promising approach to high quality image compression [41].

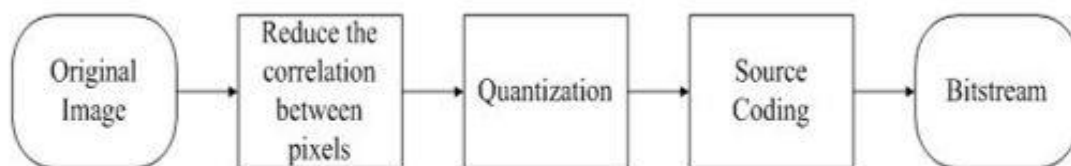
### 3.4.6 Compression Image Processing

In terms of storage, the capacity of a storage device can be effectively increased with methods that compress a body of data on its way to a storage. device and decompresses it when it is retrieved. In terms of communications, the bandwidth of a digital communication link can be effectively increased by compressing data at the sending end and decompressing data at the receiving end. At any given time, the ability of the Internet to transfer data is fixed. Thus, if data can effectively be compressed wherever possible, significant improvements of data throughput can be achieved. Many files can become binned into one compressed document making sending easier [38].

Image Compression is the art and science of reducing amount data required to represent an image. A technique used to reduce the volume of information to be transmitted about an image.

#### The Flow of Image Compression Coding

What is the so-called image compression coding? Image compression coding is to store the image into bit-stream as compact as possible and to display the decoded image in the monitor as exact as possible. Now consider an encoder and a decoder as shown in Fig. 3.11 When the encoder receives the original image file, the image file will be converted into a series of binary data, which is called the bit-stream. The decoder then receives the encoded bit-stream and decodes it to form the decoded image. If the total data quantity of the bit-stream is less than the total data quantity of the original image, then this is called image The full compression flow is as



shown in Fig. 3.9

**Figure (3.9):** The general encoding flow of image compression [38].



The compression ratio is defined as follows:

$$Cr = n_1/n_2 \quad \dots\dots\dots (3.6)$$

where  $n_1$  is the data rate of original image and  $n_2$  is that of the encoded bit-stream.

### 3.4.7 Morphological Image Processing

Binary images may contain numerous imperfections. In particular, the binary regions produced by simple thresholding are distorted by noise and texture. Morphological image processing pursues the goals of removing these imperfections by accounting for the form and structure of the image. These techniques can be extended to grey scale images [43].

#### Basic concepts

Morphology a branch in biology that deals with the form and structure of animals and plants. [40]

Mathematical Morphology as a tool for extracting image components, that are useful in the representation and description of region shape, and The language of mathematical morphology is Set theory [40].

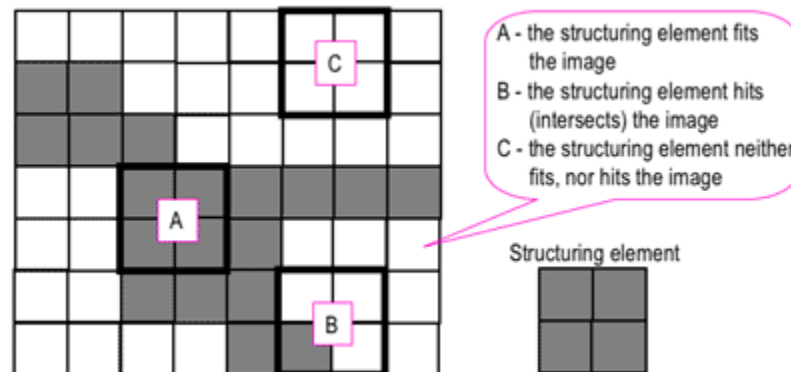
**Table (3.1):** Set Theory [40].

Subset	$A \subseteq B$
Union	$A \cup B$
Intersection	$A \cap B$
Disjoint / mutually exclusive	$A \cap B = \emptyset$
Complement	$A^c \equiv \{w   w \notin A\}$
Difference	$A - B \equiv \{w   w \in A, w \notin B = A \cap B^c\}$
Reflection	$B \equiv \{w   w = -b, \forall b \in B\}$
Translation	$(A)_z \equiv \{c   c = a + z \forall a \in A\}$

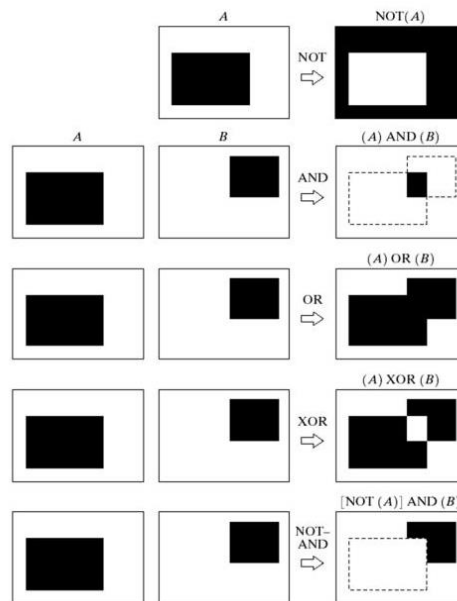
In binary images, the set elements are members of the 2-D integer space, where each element (x, y) is a coordinate of a black (or white) pixel in the image. [40]

Morphological techniques probe an image with a small shape or template called a **structuring element**, the structuring element is positioned at all possible locations in the image and it is compared with the corresponding neighborhoods of pixels. Some operations test

whether the element "fits" within the neighborhoods, while others test whether it "hits" or intersects the neighborhoods:



**Figure (3.10):** Probing of an image with a structuring element[40].

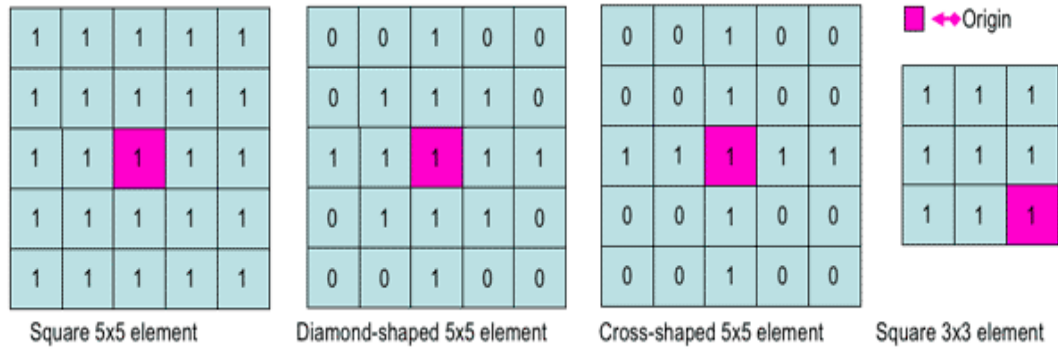


**Figure (3.11):** Some logic operations between binary image. Black represents binary 1s and white binary 0s in this example. [40]

A morphological operation on a binary image creates a new binary image in which the pixel has a non-zero value only if the test is successful at that location in the input image.

The **structuring element** is a small binary image, i.e. a small matrix of pixels, each with a value of zero or one: [43]

- The matrix dimensions specify the *size* of the structuring element.
- The pattern of ones and zeros specifies the *shape* of the structuring element.
- An *origin* of the structuring element is usually one of its pixels, although generally the origin can be outside the structuring element.



**Figure (3.12):** Examples of simple structuring elements. [43]

A common practice is to have odd dimensions of the structuring matrix and the origin defined as the center of the matrix. Structuring elements play in morphological image processing the same role as convolution kernels in linear image filtering [43].

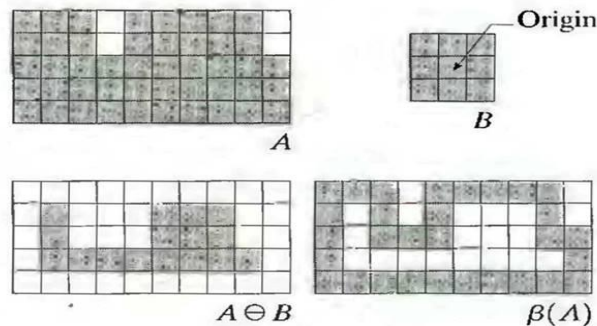
### Morphological filtering

of a binary image is conducted by considering compound operations like opening and closing as filters. They may act as filters of shape. For example, opening with a disc structuring element smooths corners from the inside, and closing with a disc smooths corners from the outside. But also these operations can filter out from an image any details that are smaller in size than the structuring element, e.g. opening is filtering the binary image at a scale defined by the size of the structuring element. Only those portions of the image that fit the structuring element are passed by the filter; smaller structures are blocked and excluded from the output image. The size of the structuring element is most important to eliminate noisy details but not to damage objects of interest [40].

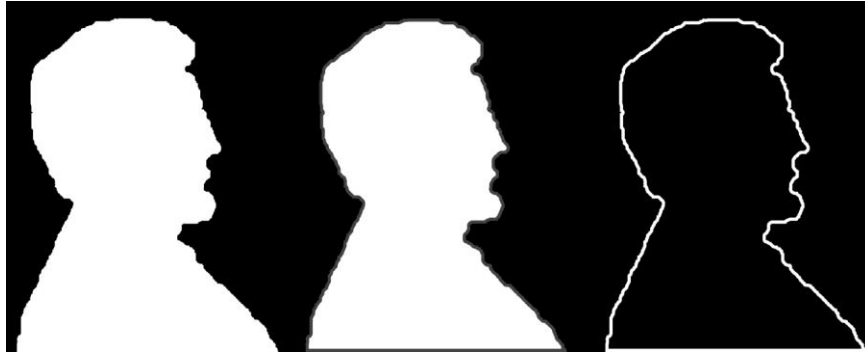
### Boundary Extraction

First, erode  $A$  by  $B$ , then make set difference between  $A$  and the erosion. The thickness of the contour depends on the size of constructing object –  $\beta$ :

$$\beta(A) = A - (A \ominus B) \dots \dots \dots (3.7)$$



**Figure (3.13):** Boundary Extraction using logic Theory [40].



**Figure (3.14):** Example of Boundary Extraction [40].

### 3.4.8 Segmentation Image Processing

#### Image analysis:

Segmentation, i.e. subdivision of the image into its constituent parts or objects. Autonomous segmentation is one of the most difficult tasks in image processing, Segmentation algorithms are based on two basic properties of gray level values:

•**Discontinuity:** the image is partitioned based on abrupt changes in gray level. Main approach is edge detection.

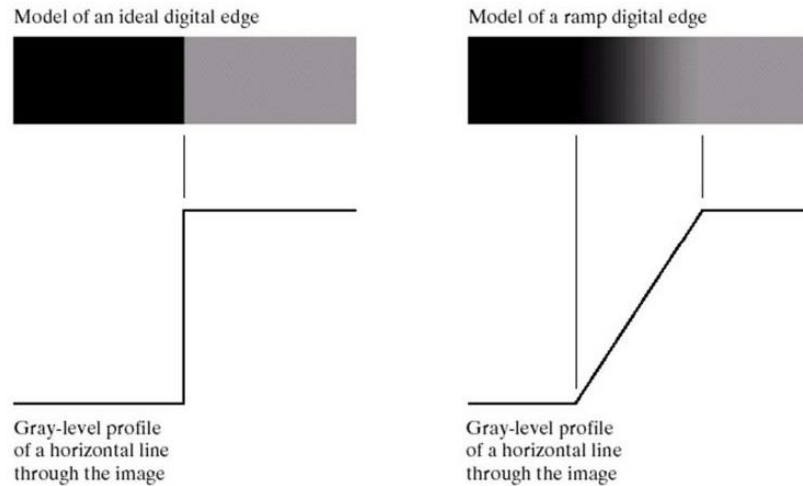
•**Similarity:** partition an image into regions that are similar. Main approaches are thresholding, region growing, and region splitting and merging. [40]

Three basic types of discontinuities in digital images: Points, Lines, Edges.

- Point detection: detect an isolated point (dark pixel within the bright zone below) [40]
- Line detection: If a mask is moved around an image it will respond more strongly to lines in the corresponding direction.
- Edges detection show in figure 3.13

-1	-1	-1	-1	-1	2	-1	2	-1	2	-1	-1
2	2	2	-1	2	-1	-1	2	-1	-1	2	-1
-1	-1	-1	2	-1	-1	-1	2	-1	-1	-1	2
Horizontal			+45°			Vertical			-45°		

**Figure (3.15):** The corresponding direction [40], Note: zero-sum masks



**Figure (3.16):** (a)Model of an ideal digital edge. (b)Model of ramp edge. The slope of the ramp is proportional to the degree of blurring in the edge [40].

## Thresholding

Selecting features within a scene or image is an important prerequisite for most kinds of measurement or analysis of the scene. Traditionally, one simple way this selection has been accomplished is to define a range of brightness values in the original image, select the pixels within

this range as belonging to the foreground, and reject all of the other pixels to the background. Such an image is then usually displayed as a binary or two-level image, using black and white (or sometimes other colors) to distinguish the regions. There is no standard convention on whether the features of interest are white or black; the choice depends on the particular display hardware in use and the designer's preference; in the examples shown here, the features are black and the background is white, which matches most modern computer displays and printing that show black text on a white background[40].

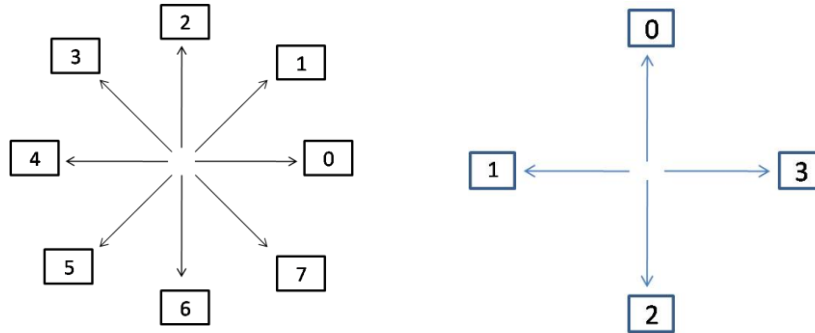
### 3.4.9 Representation & description Image Processing

The segmentation techniques usually consider the pixel along a boundary and pixel contained in the region. And an approach to obtain the descriptor that is compact the data into representation.

The results of segmentation are a set of regions. Regions have then to be represented and described. Two main ways of representing a region, external characteristics (its boundary): focus on shape, internal characteristics (its internal pixels), focus on color, textures. [32].

## Chain Code

The chain code is used to represent a boundary by the length and the direction of straight line segments. Typically, this representation is based on 4 or 8 connectivity of the segments. [32]



**Figure (3.17):** (a) 4-directional chain code (b) 8-directional chain code.

## Merging Techniques

Merging techniques based on average error or other criteria have been applied to the problem of polygonal approximation. The approach is to merge points along a boundary until the least square error line fit of the points merged so far exceeds a preset threshold.

## Descriptor

In general, descriptors are some set of numbers that are produced to describe a given shape. The shape may not be entirely reconstructing able from the descriptors, but the descriptors for different shapes should be different enough that the shapes can be discriminated. [32].

Simple descriptors:

- Length
- Number of pixels
- Number of vertical and horizontal components +  $\sqrt{2}$  times the number of diagonal components.
- Diameter (length of the major axis).

Basic rectangle (formed by the major and the minor, axis encloses the boundary) and its eccentricity (major/minor axis).

### 1) Shape Numbers.

Order of a shape: the number of digits Shape numbers, the first difference of a chain-coded boundary depends on the starting point. The shape number of such a boundary, based on the 4-directional code is defined as the first difference of smallest magnitude. For a desired shape order, we find the rectangle of order  $n$  whose eccentricity best approximates that of the basic rectangle and use this new rectangle to establish the grid size. [32]

## 2) Fourier Descriptors.

The Fourier descriptors are starting at an arbitrary point  $(x, y)$ . Each coordinate pair can be treated as a complex number so that:

$$s(k) = x(k) + jy(k) \dots\dots\dots (3.8)$$

Fourier descriptors are not insensitive to translation, but effects on the transform coefficients are known.

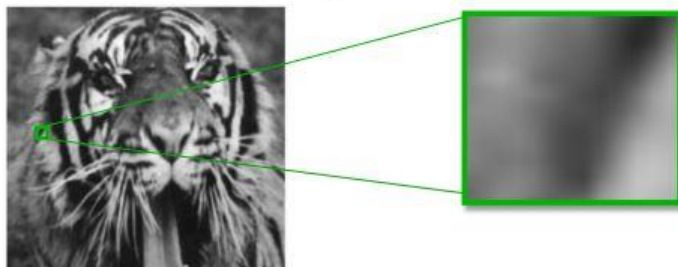
### Typical Region Features:

- Color
- Mean RGB value.
- color histograms in R, G, and color histogram in (R, G, B).
- Shape
- Number of pixels.
- Width and height attributes.
- Boundary smoothness attributes.
- Adjacent region labels.

## 3.4.10 Object Recognition

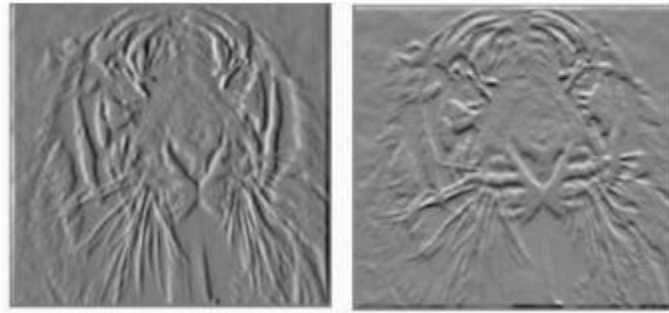
### Common Image Descriptors for Detection

Descriptors encode local neighboring window around key points



**Figure (3.18):** Example neighboring window around key points [44].

- Commonly descriptors in object detection try to capture gradient information [44].
  - Human Perception is sensitive to gradient orientation
  - Invariant to changes in lighting and small deformations



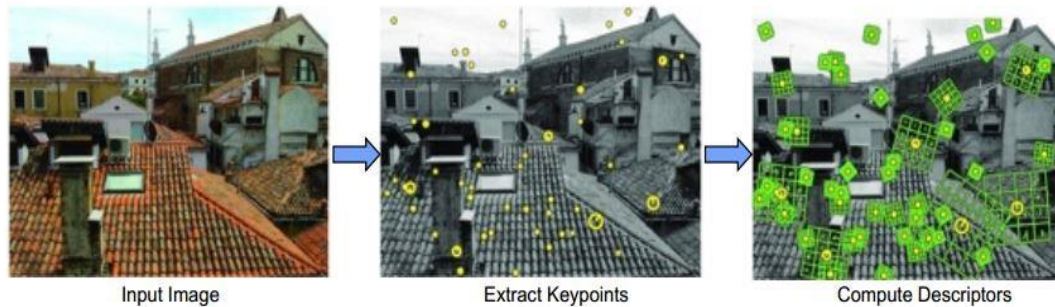
**Figure (3.19):** Capture gradient information [44].

**Most common image descriptors currently used in object detection** [44].

- Scale Invariant Feature Transform
- Histogram of Oriented Gradients
- many variants of these

**Scale Invariant Feature Transform(SIFT)** [44].

- Input an Image
- Extract Keypoints
  - Finds “corners”
  - Determines scale and orientation of the keypoint
- Compute Descriptor for each Keypoint
  - Histogram of gradients in Gaussian window around keypoint

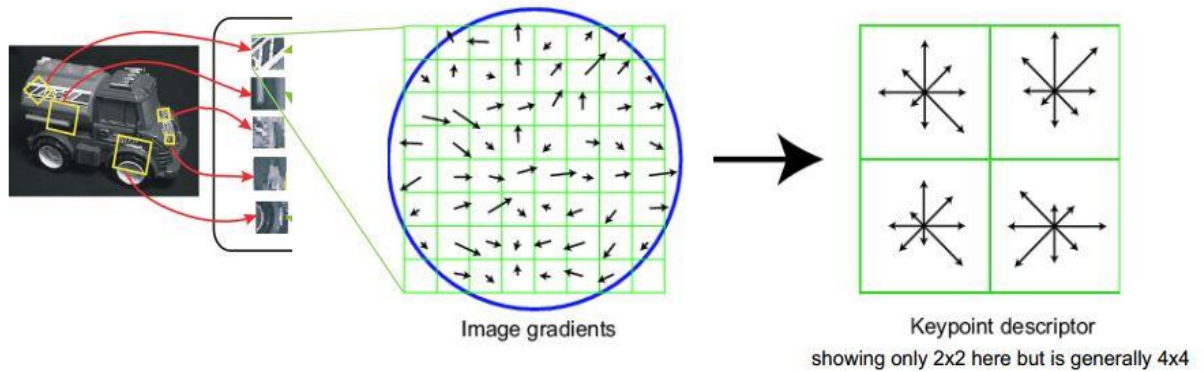


**Figure (3.20):** Scale Invariant Feature Transform [44].

- Compute the gradient for each pixel in local neighboring window
  - Typically 8 gradient directions
  - Neighboring window is determined by scale of the keypoint
- Pool Gradients into a 4x4 histogram
  - Weight each magnitude by a Gaussian window centered around the keypoint



- $8 \times 4 \times 4 = 128$  dimensional output vector normalized to 1

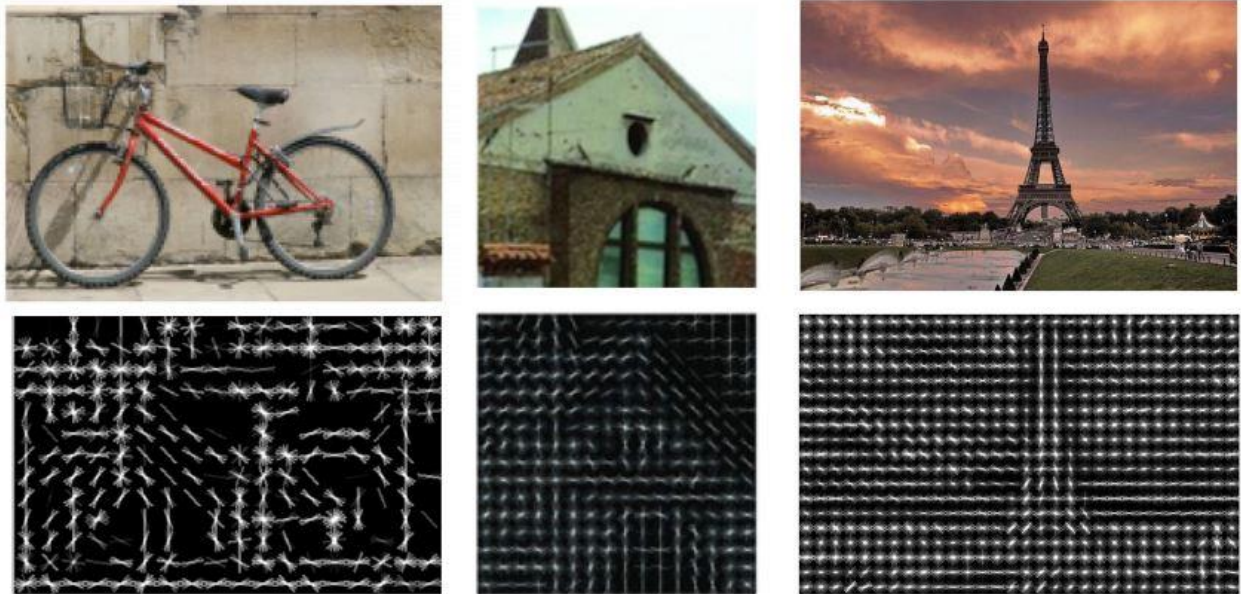


**Figure (3.21):** Key point descriptor Scale Invariant Feature Transform [44].

- Match groups of keypoints across images
  - Invariant to scale and some changes in lighting and orientation
- Great for finding the same instance of an object!
- Not good at finding different instances of an object [44].

## Histogram of Oriented Gradients(HOG) [44]

- Input an Image
- Normalize Gamma and Color
- Compute Gradients
- Accumulate weighted votes for gradient orientation over spatial bins
- Normalize contrast within overlapping blocks of cells
- Collect HOGs for all blocks over image



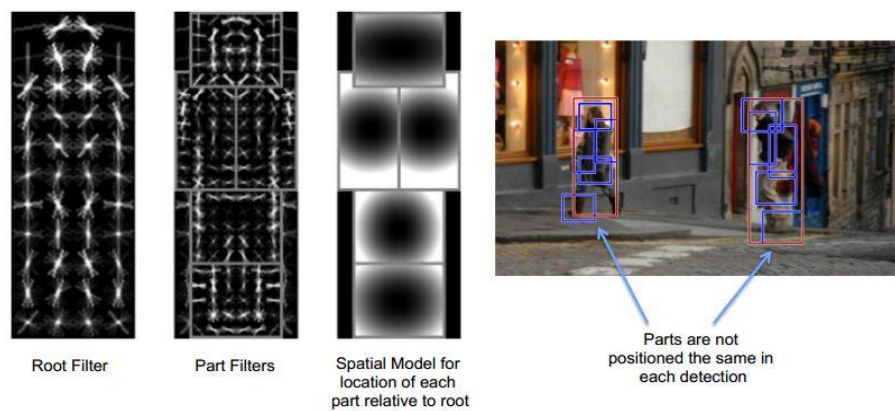
**Figure (3.22):** Example Histogram of Oriented Gradients [44].

## Structure Model

- Part models.
- Voting models.

Models an object as a number of smaller parts that are allowed to deviate slightly from average appearance. [44].

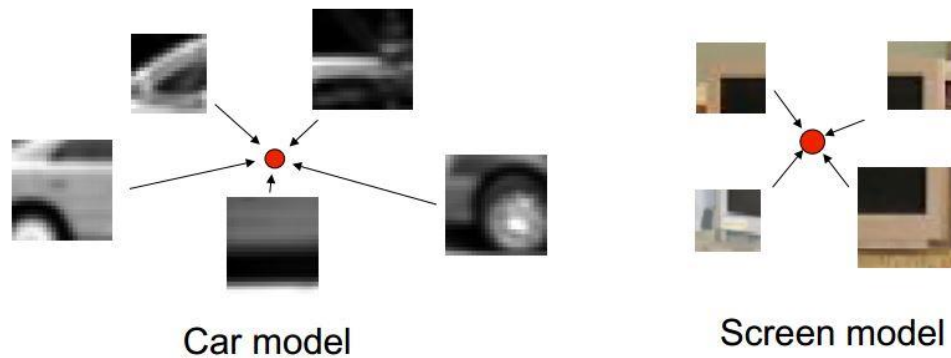
- Star model - coarse root and higher resolution part filters



**Figure (3.23):** Part Based Model [44].

## Voting Models

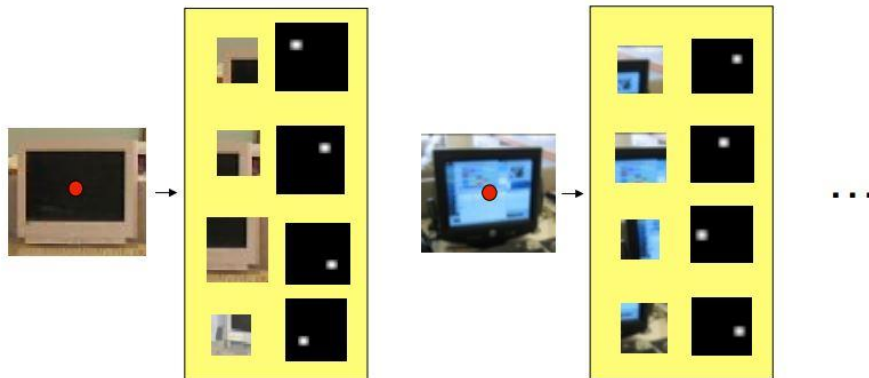
- Create weak detectors by using parts and voting for the object's center location



**Figure (3.24):** Voting Models [44].

## Collecting Parts

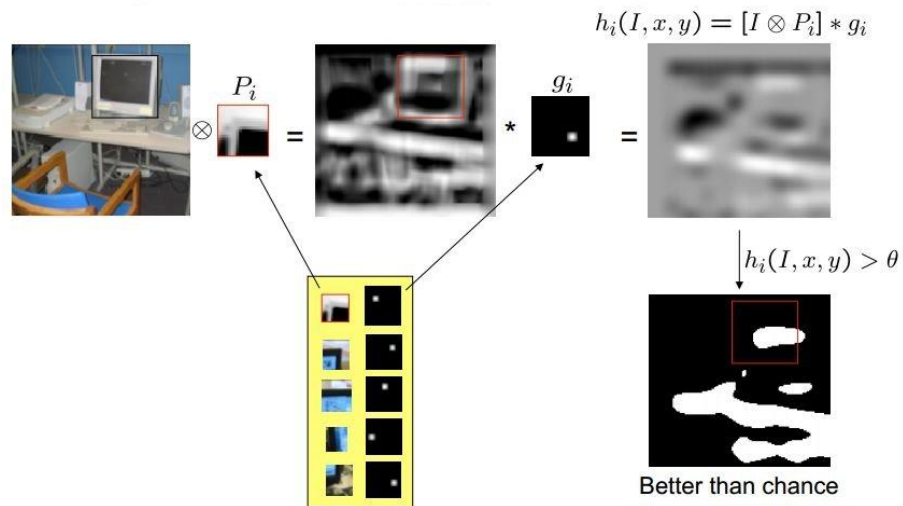
First we collect a set of part templates from a set of training objects.



**Figure (3.25):** Example Collecting Parts [44].

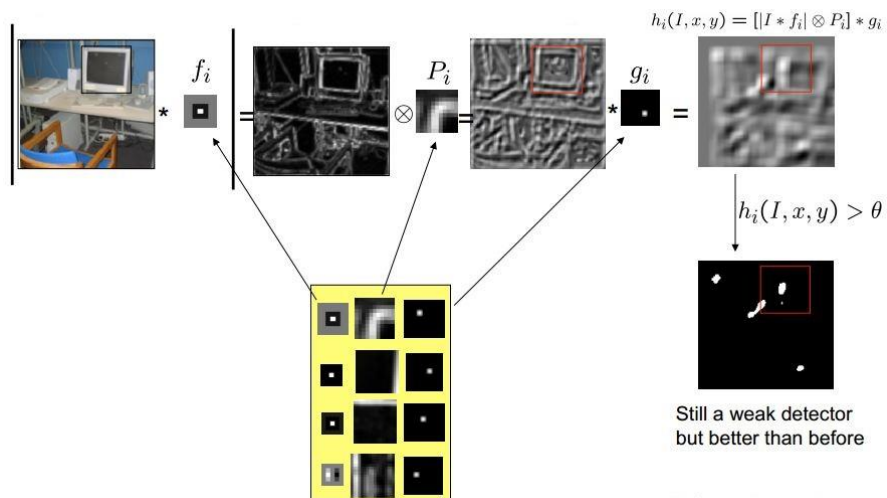
## Weak Part Detectors

-We now define a family of “weak detectors” as:



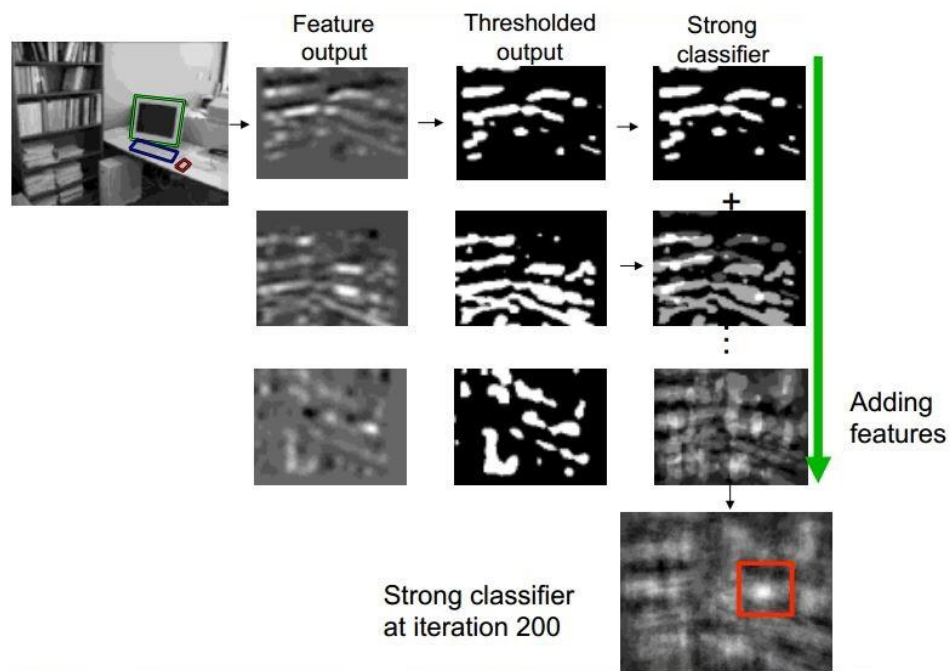
**Figure (3.26):** Weak Part Detectors [44].

-We can do a better job using filtered images



**Figure (3.27):** Weak Part Detectors using filtered images [44].

## Voting Model



**Figure (3.28):** Example of Screen Detection [44].

## Datasets for Object Classification Detection

- Caltech101
- Caltech256
- PASCAL
- ImageNET
- LabelMe

# CHAPTER FOUR

## CHALLENGE LOOK

# 4

---

### System Design

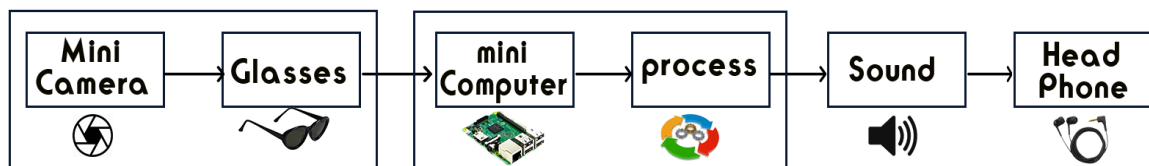
- 4.1 Introduction**
- 4.2 System Block Diagram**
- 4.3 System Flow Chart**
- 4.4 Special mini camera**
- 4.5 Programming Language using Python**
- 4.6 Raspberry pi 3**
- 4.7 Dataset of image**
- 4.8 Recognition**
- 4.9 Power Supply**

## 4.1 Introduction

One of the primary goals of computer vision is the understanding of visual scenes. Scene understanding involves numerous tasks including recognizing what objects are present, localizing the objects in 2D and 3D, determining the objects' and scene's attributes, characterizing relationships between objects and providing a semantic description of the scene.

## 4.2 System Block Diagram

The main idea of system block diagram is work as similar part eye in human of The person who is not blind and help blind people for understand the object around his life by detection and recognition using camera and tell the result on earphone what's camera can see. This block can describe it.



**Figure (4.1):** Block diagram of project.

the camera is a box that controls the amount of light which reaches a light-sensitive surface inside (either film, a digital sensor, or another surface). The original cameras did not even have a glass lens, though today we can say that most cameras include: a light-tight box, a glass lens, and a surface that captures light.

The camera has come a long way from its humble beginnings, but it is still just a box that controls the amount of light that reaches a piece of film (or sensor). The camera has different types of body and size and shape in this project we use mini special camera that use in surgical it very small and can but it easily on glasses and it is very good in low dark and low current it has 6 LEDS inside camera work in dark and high sensitivity and high pixel of image to get high quality then is better for fast recognition and connect it USB cable to minicomputer .

in real time camera as Visual multimedia source that combines a sequence of images to form a moving picture. The video transmits a signal to a screen and processes the order in which the screen captures should be shown.

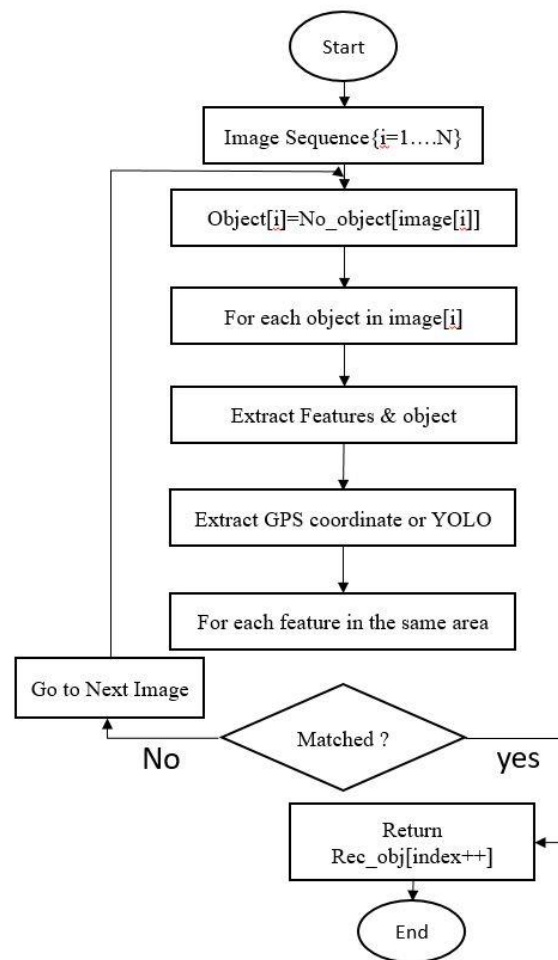
The main processing in my project to work as similar part of brain to understand the object on what camera can see by matching into database of images we use minicomputer is called Raspberry pi 3 model b when get input sequence of image by camera then do method of image processing we talk it on chapter 3 and when finish the processing the minicomputer matching all sequence of image with image net or database is stored in minicomputer and it detected using python programming language the result of output sound of object by connect earphone to tell



what camera can see and what's can understanding after complete processing then can blind people know what's on around life for fast time and feeling more comfortable without need any help from another person.

### 4.3 System Flow Chart

In this work, object recognition approach is presented by applying several steps: Object detection, creating unique descriptor for each object, retrieving from model database, and matching. A model database is a prerequisite stage required to be built in order to apply the matching process. It contains features for all common objects in the environments of the blind [58].



**Figure (4.2):** System Flow chart of the proposed method [58].

All objects that exist in a blind's environment are manually extracted and identified by the user to apply machine learning. The model database saves the features for each object, which is used later to apply the matching process. The extracted features are then saved in the database as shown in Table 4.1. In order to reduce mismatches and computational time, GPS service is



used to determine the location for each object. Hence the comparisons are only applied to the objects that exist in an involved area [58].

**Table (4.1):** Database structure for instances in the real world [58].

Object ID	Feature vector	GPS Coordinate
Chair	V-object1	(Latitude1,Longitude1)
Door	V-object1	(Latitude2,Longitude2)
....	.....	.....
Object{N}	V-objectN	(LatitudeN,LongitudeN)

The database of models is created by applying the following: [58]

- 1) For each object in the input image: Do the next steps.
- 2) Extract the SURF features descriptor.
- 3) Get the GPS coordinate.
- 4) Identify the object by the user.
- 5) Save the extracted info.

Based on the models database, fast indexing is performed using the sign of the Laplacian for the underlying interest point and the GPS service to specify the involved areas. Typically, as performed in (Bay et al, 2006), the interest points are found at blob-type structures and the sign of Laplacian differentiates bright blobs on dark backgrounds from the reverse situation. This feature and the GPS area-based service are utilized to apply the proposed work at no extra computational cost. It should be noted that we only compare features if they have the same type of contrast and in same location. Therefore, this information allows faster matching and provides a slight increase in the performance as shown in the flow chart. [58]

## 4.4 Special Mini Camera

MISUMI is specialized in making customized design and modification of our products to meet your specific needs. Misumi R&D team is equipped with “Rapid Prototyping (RP) Machine” to custom-make your sample in as quickly as 1 day (excluding shipping time). We are also equipped with PADS software & Printed Circuit Board Plotter, which are instrumental in accelerating the process of designing, engineering, producing, and testing our ongoing new products. In addition, a T.Q.M. programmer has been implemented to ensure the highest quality standard at all times. [56]

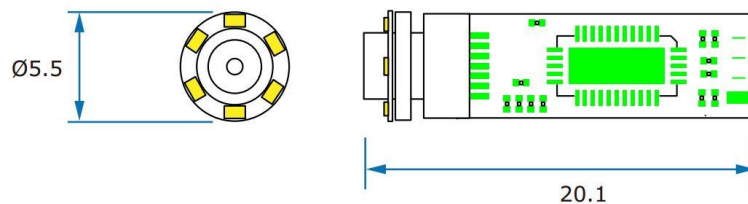


**Figure (4.3):** Special mini camera [57].

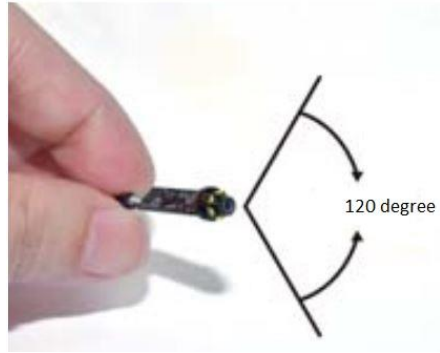
Properties of mini camera:

- Low dark current for low-light.
- Conditions.
- High sensitivity.
- High performance.
- Full HD image pixel.

Video cameras are used primarily in two modes. The first, characteristic of much early broadcasting, is live television, where the camera feeds real time images directly to a screen for immediate observation. A few cameras still serve live television production, but most live connections are for security, military/tactical, and industrial operations where surreptitious or remote viewing is required. In the second mode the images are recorded to a storage device for archiving or further processing and we use this camera in figure it suitable for processing and progress scan. [56]



**Figure (4.4):** The size of mini camera [App. A]



**Figure (4.5):** Angle of viewing camera [App. A]

Now we design the size of camera to be comfortable and small size to place on glasses, the diameter of lens is 5.5 mm and angle can see around 120 degrees, to make the space wider the eyes of the camera and the possibility of the person to recognize the more object.

## 4.5 Programming Language using Python

Being a very high level language, Python reads like English, which takes a lot of syntax-learning stress off coding beginners. Python handles a lot of complexity for you, so it is very beginner-friendly in that it allows beginners to focus on learning programming concepts and not have to worry about too much details. [46]

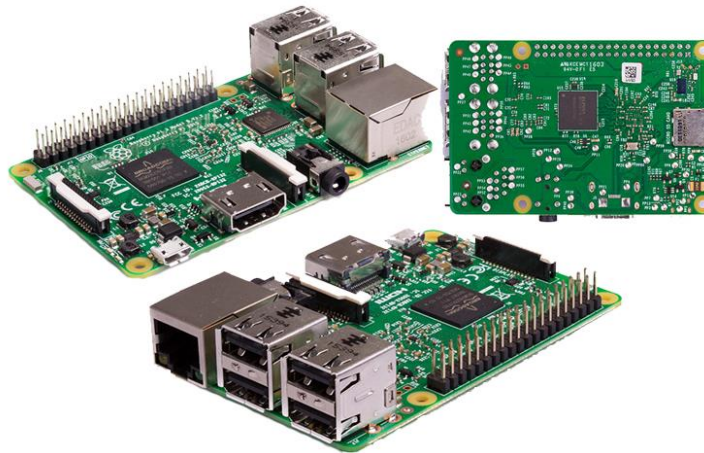
As a dynamically typed language, Python is really flexible. This means there are no hard rules on how to build features, and you'll have more flexibility solving problems using different methods (though the Python philosophy encourages using the obvious way to solve things). Furthermore, Python is also more forgiving of errors, so you'll still be able to compile and run your program until you hit the problematic part. [46]

As you step into the programming world, you'll soon understand how vital support is, as the developer community is all about giving and receiving help. The larger a community, the more likely you'd get help and the more people will be building useful tools to ease the process of development. [46]

Stack Overflow is a programming Q&A site you will no doubt become intimate with as a coding beginner. Python has 85.9k followers, with over 500k Python questions. Python questions are also the 3rd most likely to be answered when compared to other popular programming languages. [46]

## 4.6 Raspberry pi 3

Computers are being developed in a short time with increased speed, hardware, software, lower cost, and the increased availability of access technology. Hence, several works on assistive technologies are created to enable localization, navigation and object recognition. The best interface then can be customized based on a user request whether that is vibrations, sounds or the spoken word [58].



**Figure (4.6):** Raspberry pi 3 model B [59].

The **Raspberry Pi** is a series of small single-board computers developed in the United Kingdom by the Raspberry Pi Foundation to promote the teaching of basic computer science in schools and in developing countries. The original model became far more popular than anticipated, selling outside of its target market for uses such as robotics. Peripherals (including keyboards, mice and cases) are not included with the Raspberry Pi. Some accessories however have been included in several official and unofficial bundles [59].

Several generations of Raspberry Pi is have been released. All models feature a Broadcom system on a chip (SoC) with an integrated ARM compatible central processing unit (CPU) and on-chip graphics processing unit (GPU). Processor speed ranges from 700 MHz to 1.2 GHz for the Pi 3 and on-board memory range from 256 MB to 1 GB RAM. Secure Digital (SD) cards are used to store the operating system and program memory in either SDHC or MicroSDHC sizes. Depending on the model; The boards have either a single USB port or up to four USB ports. For video output, HDMI and composite video are supported, with a standard 3.5 mm phono jack for audio output. Lower level output is provided by a number of GPIO pins which support common protocols like I<sup>2</sup>C. The B-models have an 8P8CEthernet port and the Pi 3 and Pi Zero W have on-board Wi-Fi 802.11n and Bluetooth [59].

The organization behind the Raspberry Pi now consists of two arms. Originally developed under the auspices of the Raspberry Pi Foundation, the success of the Pi Model B prompted the Foundation to set up Raspberry Pi Trading, with Dr Eben Upton as CEO, to develop the third model, the B+. Raspberry Pi Trading is responsible for developing the technology while the

Foundation is an educational charity that exists to get that message out to schools. Raspberry Pi Trading reinvests about a third of its profit in R&D, and the rest goes to the foundation [59].

The Foundation provides Raspbian, a Debian-based Linux distribution for download, as well as third-party Ubuntu, Windows 10 IOT Core, RISC OS, and specialized media center distributions. It promotes Python and Scratch as the main programming language, with support for many other languages. The default firmware is closed source, while an unofficial open source is available [59].

**Table (4.2):** Specification of Raspberry Pi 3 [59].

<b>Processor</b>	Broadcom BCM2387 chipset. 1.2GHz Quad-Core ARM Cortex-A53 802.11 b/g/n Wireless LAN and Bluetooth 4.1 (Bluetooth Classic and LE)
<b>GPU</b>	Dual Core Video core IV® Multimedia Co-Processor. Provides Open GL ES 2.0, hardware-accelerated OpenVG, and 1080p30 H.264 high-profile decode. Capable of 1Gpixel/s, 1.5Gtexel/s or 24GFLOPs with texture filtering and DMA infrastructure.
<b>Memory</b>	1GB LPDDR2
<b>Operating System</b>	Boots from Micro SD card, running a version of the Linux operating system
<b>Dimensions</b>	85 x 56 x 17mm
<b>Power</b>	Micro USB socket 5V1, 2.5A

**Table (4.3):** Connectors in Raspberry Pi [59].

<b>Ethernet</b>	10/100 Base t Ethernet socket
<b>Video Output</b>	HDMI (rev 1.3 & 1.4 Composite RCA (PAL and NTSC)
<b>Audio Output</b>	Audio Output 3.5mm jack, HDMI USB 4 x USB 2.0 Connector
<b>GPIO</b>	Connector 40-pin 2.54 mm (100 mil) expansion header: 2x20 strip Providing 27 GPIO pins as well as +3.3 V, +5 V and GND supply lines
<b>Camera Connector</b>	15-pin MIPI Camera Serial Interface (CSI-2)
<b>Display Connector</b>	Display Serial Interface (DSI) 15 way flat flex cable connector with two data lanes and a clock lane
<b>Memory Card</b>	Slot Push/pull Micro SDIO

## 4.7 Dataset of image

The ImageNet dataset [47], which contains an unprecedented number of images, has recently enabled breakthroughs in both object classification and detection research [48], [50], [51]. The community has also created datasets containing object attributes [51], scene attributes [52], key points [53], and 3D scene information [54] the goal of advancing the state-of-the-art in object recognition by placing the question of object recognition in the context of the broader question of scene understanding. [55]

the properties of the Microsoft Common Objects in Context (MS COCO) dataset in comparison to several other popular datasets. These include ImageNet [47], PASCAL VOC 2012 [48], and SUN [49]. Each of these datasets varies significantly in size, list of labeled categories and types of images. ImageNet was created to capture a large number of object categories, many of which are fine-grained. SUN focuses on labeling scene types and the objects that commonly occur in them. Finally, PASCAL VOC's primary application is object detection in natural images. MS COCO is designed for the detection and segmentation of objects occurring in their natural context [55].

The Microsoft Common Objects in Context (MS COCO) dataset contains 91 common object categories with 82 of them having more than 5,000 labeled instances, Figure (4.7) In total the dataset has 2,500,000 labeled instances in 328,000 images. In contrast to the popular ImageNet dataset [1], COCO has fewer categories but more instances per category. This can aid in learning detailed object models capable of precise 2D localization. The dataset is also significantly larger in number of instances per category than the PASCAL VOC [48] and SUN [49] datasets. Additionally, a critical distinction between our dataset and others is the number of labeled instances per image which may aid in learning contextual information [55].



**Figure (4.7):** Samples of images in the MS COCO dataset [55].


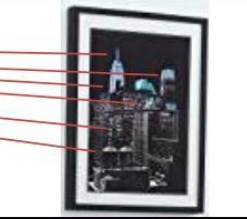




## 4.8 Recognition

The results are presented based on our own dataset. The dataset includes objects captured from images of the real life. Initially, the objects are gathered and identified by ourselves. The dataset consists of 300 images of 25 objects. The tested images comprise 180 images that were taken from the right camera. The resolution of images is about 600 x 500 pixels. Objects recognition from the model database proceeds as follows:

The GPS coordinate is extracted, and it retrieves all objects that are associated with it. The images in the test set are compared to all objects in the database having the same location coordinates. The objects that have acknowledged features under the location term from the database are then chosen as recognized objects as shown in Table (4.4) The matching is applied by calculating the Euclidean distance between the descriptor vectors of the input object and all objects having the same location in the database. If the distance is closer than 0.8 times the distance of the second nearest neighbor, then matching pair is considered to be detected. This threshold value was adapted based on the best result that has been achieved. The output is composed of concatenated strings for both English and Arabic languages. The API of Google cloud speech was used to convert text to audio. This tool was chosen because it supports over 80 languages. Hence, the proposed approach can be globally used based on the supported languages by Google cloud.

**Table (4.4):** The features of objects for the involved scene (GPS based location) only are extracted and matched with the reference image.

Image of real scene	Only the identified objects in the matched area are extracted	Features descriptor in models database	GPS	
			Latitude	Longitude
		Features of paint	$X_1$	$Y_1$
		Features of couch	$X_2$	$Y_2$
		Features of flower vase	$X_3$	$Y_3$

## **4.9 Power Supply**

Capacity is a 4000mAh lithium ion battery, a charging circuit you charge it via the USB cable attached, and a boost converter that provides 5 Volt DC up to 1 Amp via a USB A port and it still work for 7 hours.



# CHAPTER FIVE

# 5

---

## **Object Detection & Recognition Using Tensor Flow**

### **5.1 Introduction**

### **5.2 Tensor Flow**

### **5.3 Why Tensor Flow?**

### **5.4 Neural Network**

### **5.5 Object Detection with Tensor Flow**

#### **5.5.1 Computations are done in Two steps**

#### **5.5.2 Convert labels to the TF Record format**

### **5.6 Detection Models**

#### **5.6.1 Single Shot Detector (SSD)**

#### **5.6.2 RCNN**

#### **5.6.3 Fast RCNN**

### **5.7 Recognition**

#### **5.7.1 Three Steps Recognition**

## 5.1 Introduction

Computer vision is an interdisciplinary field that deals with how computers can be made for gaining high-level understanding from digital images or videos. From the perspective of engineering, it seeks to automate tasks that the human visual system can do. [61] Computer vision is concerned with the automatic extraction, analysis and understanding of useful information from a single image or a sequence of images. It involves the development of a theoretical and algorithmic basis to achieve automatic visual understanding.

One of the primary goals of computer vision is the understanding of visual scenes. Scene understanding involves numerous tasks including recognizing what objects are present, localizing the objects in 2D and 3D, determining the objects' and scene's attributes, characterizing relationships between objects and providing a semantic description of the scene. [61] [62].

## 5.2 Tensor Flow

Tensor Flow is an open source software library for high performance numerical computation. Its flexible architecture allows easy deployment of computation across a variety of platforms (CPUs, GPUs, TPUs), and from desktops to clusters of servers to mobile and edge devices. Originally developed by researchers and engineers from the Google Brain team within Google's AI organization, it comes with strong support for machine learning and deep learning and the flexible numerical computation core is used across many other scientific domains [63].

## 5.3 Why Tensor Flow

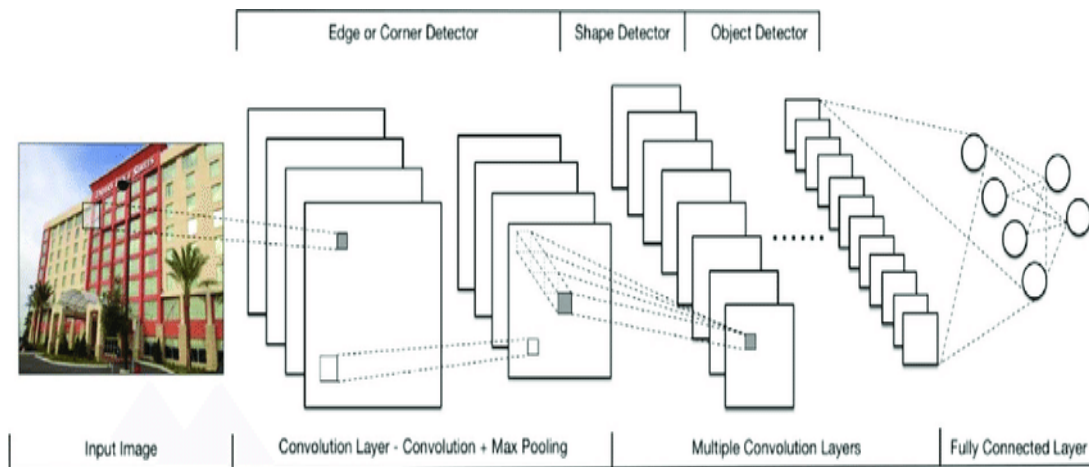
- Python API
- Portability: deploy computation to one or more CPUs or GPUs in a desktop, server, or mobile device with a single API.
- Flexibility: from Raspberry Pi, Android, Windows, IOS, Linux to server farms.
- Visualization.
- Checkpoints (for managing experiments).
- Auto-differentiation (no more taking derivatives by hand)
- Large community (> 10,000 commits and > 3000 TF-related repos in 1 year).
- Awesome projects powerful using Tensor Flow [63].

## 5.4 Neural Network

Neural Network commonly referred to as “Neural Networks” has been motivated right from its inception by the recognition that the human brain computes in an entirely different way from the conventional digital computer. The brain is a highly complex, nonlinear, and parallel computer (information-processing system). It has the capability to organize its structural constituents, known as neurons, so as to perform certain computations (e.g., pattern recognition, perception, and motor control) many times faster than the fastest digital computer in existence today. Consider, for example, human vision, which is an information-processing task. It is the function of the visual system to provide a representation of the environment around us and more important, to supply the information we need to interact with the environment. To be specific, the brain routinely accomplishes perceptual recognition tasks (e.g., recognizing) [65].

Neural Networks help us cluster and classify. You can think of them as a clustering and classification layer on top of the data you store and manage. They help to group unlabeled data according to similarities among the example inputs, and they classify data when they have a labeled dataset to train on. (Neural networks can also extract features that are fed to other algorithms for clustering and classification; so you can think of deep neural networks as components of larger machine-learning applications involving algorithms for reinforcement learning, classification and regression.)

The Convolutional Neural Networks (CNNs), an important and powerful kind of learning architecture widely diffused especially for Computer Vision applications. They currently represent state of the art algorithm for image classification tasks and constitute the main architecture used in Deep Learning [65].



**Figure (5.1):** Convolutional Neural Networks(CNN) [65].

## 5.5 Object detection with Tensor Flow

### 5.5.1 Computations are done in two steps:

- **First:** Build the graph.
- **Second:** Execute the graph. Both steps can be done in many languages (python, C++) Best supported so far is python [64].

we will walk through all the steps for building a custom object classification model using Tensor Flow's API:

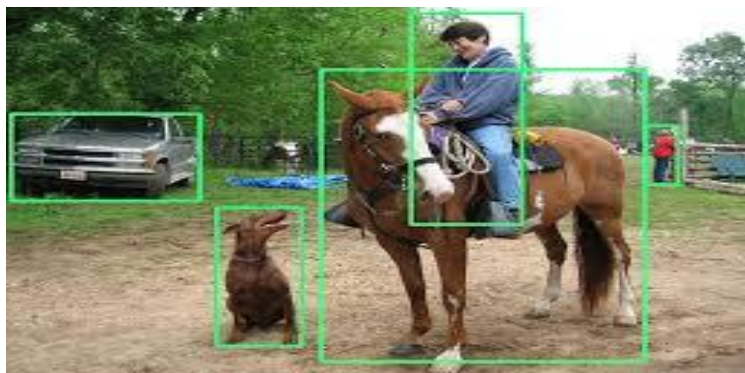
#### ✓ **Gathering a data set:**

Some very large detection data sets, such as MS-COCO, exist already.

#### ✓ **Creating bounding boxes:**

In order to train our object detection model, for each image we will need the image's width, height, and each class with their respective xmin, xmax, ymin, and ymax bounding box. Simply put, our bounding box is the frame that captures exactly where our class is in the image.

Creating these labels can be a huge ordeal, but thankfully there are programs that help create bounding boxes. Labeling is an excellent open source free software that makes the labeling process much easier. It will save individual xml labels for each image, which we will convert into a csv table for training. The labels for all the images used in the pawn detector we are building are included in the Get Hub repository[64].



**Figure (5.2):** Train our object detection model [64].

✓ **Install the object detection API:**

Before getting started, we have to clone and install the object detection API into our Get Hub repository. Installing the object detection API is extremely simple; you just need to clone the Tensor Flow Models directory and add some things to your Python path.

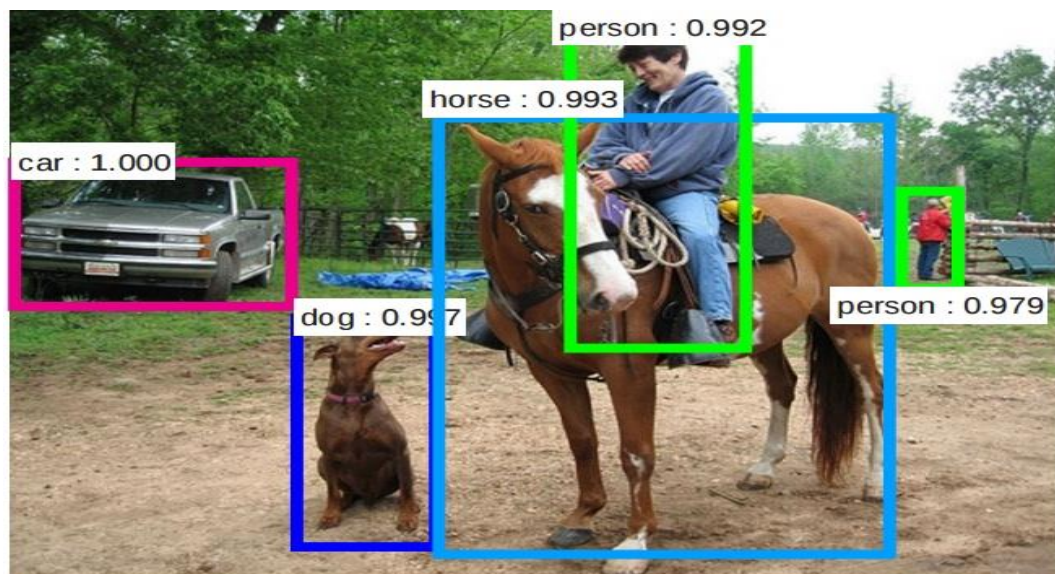
### 5.5.2 Convert labels to The Tensor Flow Record format:

When training models with Tensor Flow using Tensor Flow Record, files help optimize your data feed. We can generate a Tensor Flow Record file using code adapted from this raccoon detector.

❖ **Choose a model:**

There are models in the Tensor Flow API you can use depending on your needs. If you want a high-speed model that can work on detecting video feed at high fps, the single shot detection (SSD) network works best. Some other object detection networks detect objects by sliding different sized boxes across the image and running the classifier many times on different sections of the image, this can be very resource consuming. As its name suggests, the SSD network determines all bounding box probabilities in one go; hence, it is a vastly faster model [64].

❖ **Retrain the model with coco:** we simply run the `train.py` file in the object detection API directory.



**Figure (5.3):** Object Detection [64].

## 5.6 Detection Models

### 5.6.1 Single Shot Detector SSD:

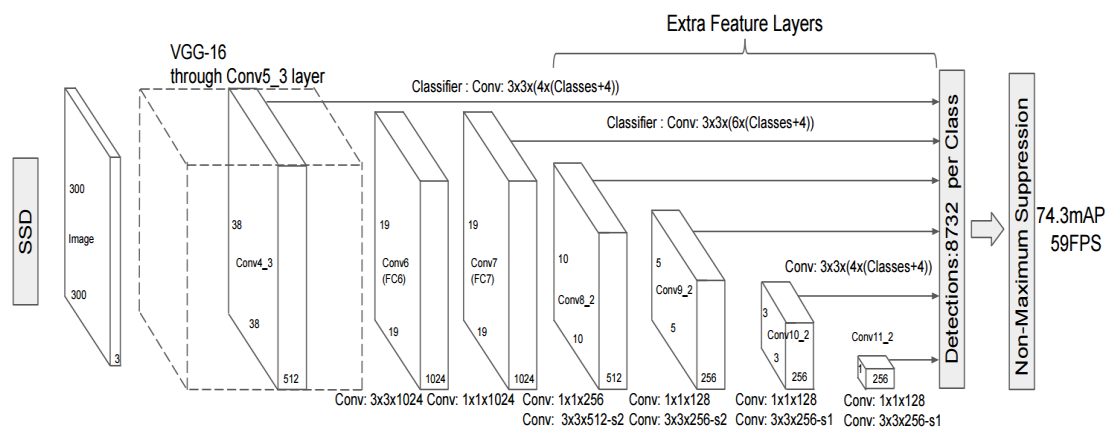
We present a method for detecting objects in images using a single deep neural network. Our approach, Named **SSD**.

**Single Shot:** this means that the tasks of object localization and classification are done in a single forward pass of the network.

**Multi Box:** this is the name of a technique for bounding box regression.

**Detector:** The network is an object detector that also classifies those detected objects [65].

Detectors are convolutional filters, Each detector outputs a single value. discretizes the output space of bounding boxes into a set of default boxes over different aspect ratios and scales per feature map location. At prediction time, the network generates scores for the presence of each object category in each default box and produces adjustments to the box to better match the object shape. Additionally, the network combines predictions from multiple feature maps with different resolutions to naturally handle objects of various sizes. Our SSD model is simple relative to methods that require object proposals because it completely eliminates proposal generation and subsequent pixel or feature resampling stage and encapsulates all computation in a single network. This makes SSD easy to train and straightforward to integrate into systems that require a detection component [65].



**Figure (5.4):** Single Shot Detector SSD [65].

### 5.6.2 RCNN (Region Proposal + CNN)

The Region-based Convolutional Network method (RCNN) achieves excellent object detection accuracy by using a deep ConvNet to classify object proposals. R-CNN [65].

Use selective search to come up with regional proposal First object detection method using CNN.

Training RCNN:

**Step1:** train your own CNN model for classification using Image Net dataset.

**Step2:** focus on 20 classes + 1 background. Remove the last FC layer and replace it with a smaller layer and fine-tune the model using PASCAL VOC dataset.

**Step3:** extract feature. Store all the features.

**Step4:** train SVM for each class: -Crop /Warp image.

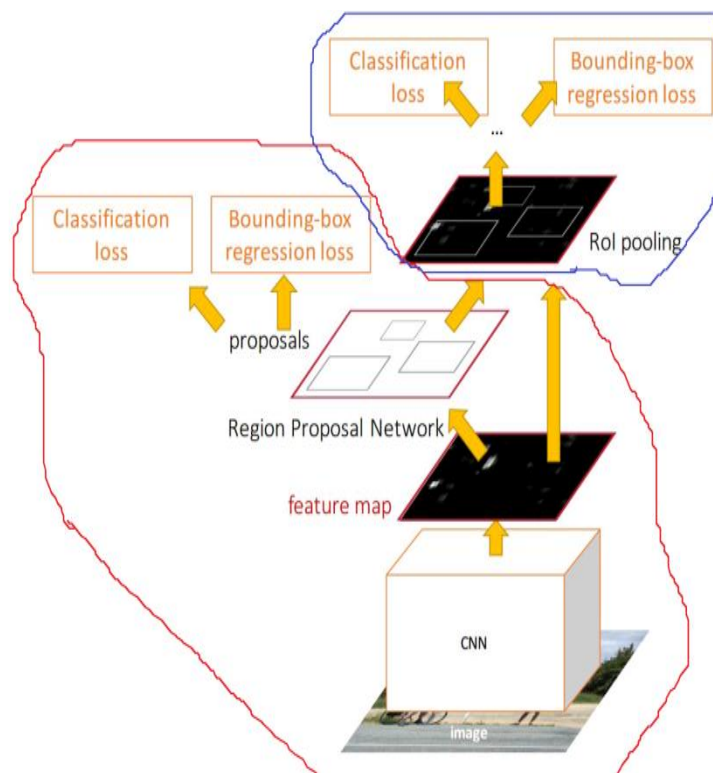


Figure (5.5): RCNN [65].

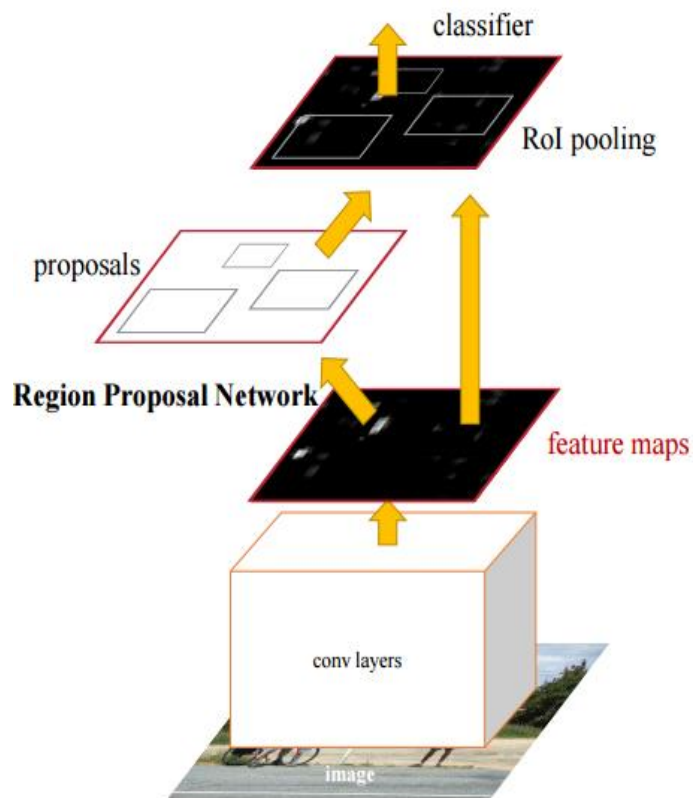


### 5.6.3 Fast RCNN

Faster RCNN is the follow on to Fast RCNN and RCNN. Faster RCNN starts with a CNN adds a Region Proposal Network (RPN) to create proposals (bounding boxes) from the features given by the CNN. Then ROI pooling and a classifier is used to classify and score each bounding box [65].

Once a Fast R-CNN network is fine-tuned, detection amounts to little more than running a forward pass (assuming object proposals are pre-computed). The network takes as input an image and a list of R object proposals to score.

Share convolution layers for proposals from the same image Faster and More accurate than RCNN.



**Figure (5.6):** Fast RCNN [65].



## 5.7 Object Recognition with Tensor Flow

A recognition algorithm (image classifier) takes an image as input and outputs what the image contains. In other words, the output is a class label (e.g. “cat”, “dog”, “table” etc.) [66].

### 5.7.1 Three Steps Recognition:

#### Step 1: Preprocessing

Often an input image is pre-processed to normalize contrast and brightness effects. A very common preprocessing step is to subtract the mean of image intensities and divide by the standard deviation. Sometimes, gamma correction produces slightly better results. While dealing with color images, a color space transformation (e.g. RGB to LAB color space) may help get better results [66].

#### Step 2: Feature Extraction

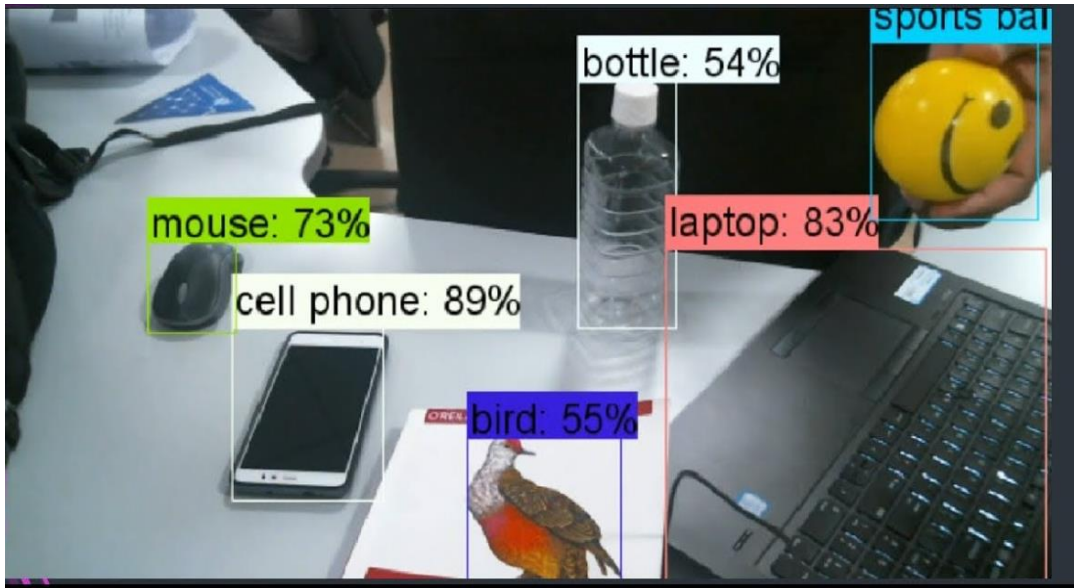
The input image has too much extra information that is not necessary for classification. Therefore, the first step in image classification is to simplify the image by extracting the important information contained in the image and leaving out the rest. For example, if you want to find shirt and coat buttons in images, you will notice a significant variation in RGB pixel values. However, by running an edge detector on an image we can simplify the image. You can still easily discern the circular shape of the buttons in these edge images and so we can conclude that edge detection retains the essential information while throwing away non-essential information. The step is called feature extraction. In traditional computer vision approaches designing these features are crucial to the performance of the algorithm.

Turns out we can do much better than simple edge detection and find features that are much more reliable. In our example of shirt and coat buttons, a good feature detector will not only capture the circular shape of the buttons but also information about how buttons are different from other circular objects like car tires [66].

#### Step 3: Learning Algorithm for Classification

In the previous section, we learned how to convert an image to a feature vector. In this section, we will learn how a classification algorithm takes this feature vector as input and outputs a class label (e.g. cat or background).

Before a classification algorithm can do its magic, we need to train it by showing thousands of examples of cats and backgrounds. Different learning algorithms learn differently, but the general principle is that learning algorithms treat feature vectors as points in higher dimensional space, and try to find planes / surfaces that partition the higher dimensional space in such a way that all examples belonging to the same class are on one side of the plane / surface [66].



**Figure (5.7):** Result of Objects Recognition [66].

# CHAPTER SIX

## CHAPTER SIX

# 6

---

### **SIMULATION & RESULTS**

#### **6.1 Simulation**

##### **6.1.1 Connecting the Camera**

##### **6.1.2 Camera Setup and Configuration**

##### **6.1.3 Understanding Training process**

#### **6.2 Results**

#### **6.3 challenges**

#### **6.4 conclusion & future work**

## 6.1 Simulation:

### 6.1.1 Connecting the Camera

The system starts by connecting the Raspberry Pi camera module with the raspberry pi microcomputer through a cable, the cable connects between the fast camera Serial Interface bus and the system-on-chip processor, the camera is connected to the raspberry pi.

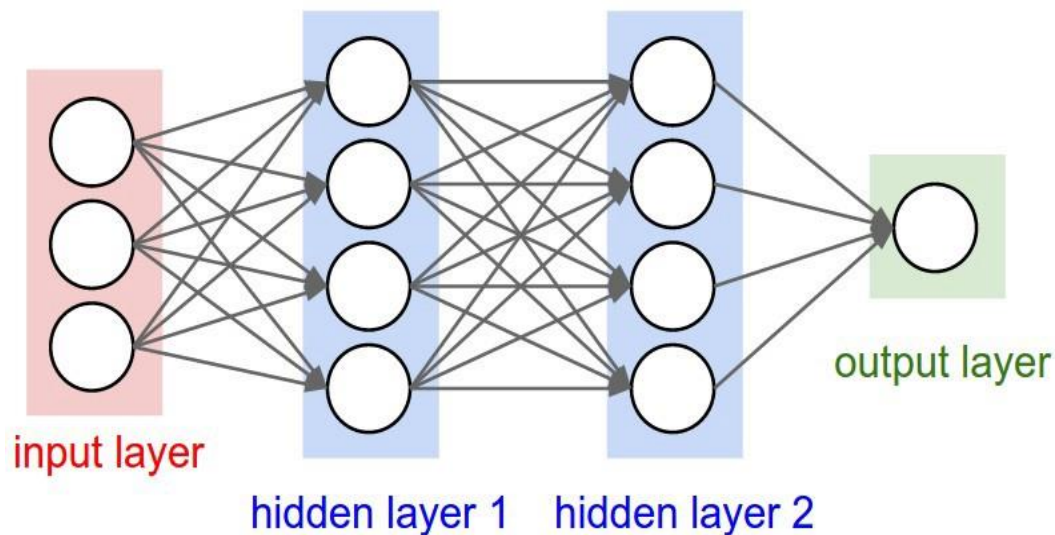
### 6.1.2 Camera Setup and Configuration

This stage assumes that Raspberry pi an Operating System is already Installed on Raspberry Pi Microcomputer, Programming codes After the system is ready on side with the correct programs, all required programs that will do the job was written in python.

Here, we will employ a dataset which includes objects captured from images of the real life, the dataset consists of more than 300 images of 90 objects. The resolution of images is about 600 x 500 pixels.

### 6.1.3 Understanding Training process:

Deep neural networks are nothing but mathematical models of intelligence which to a certain extent mimic human brains, Deep learning recognizes objects in images by using three or more layers of artificial neural networks in which each layer is responsible for extracting one or more features of the image [69].



**Figure (6.1):** Layers in Neural Networks [69].

A neural network is a computational model that is analogous to the arrangement of neurons in the human brain. Each neuron takes an input, performs an operation, and then sends output to one or more adjacent neurons.

## Train A Neural Network

Training a Neural Network is very similar to training a little child. You show the child a ball and tell her that it is a “ball”. When you do that many times with different kinds of balls, the child figures out that it is the shape of the ball that makes it a ball and not the color, texture or size. You then show the child an egg and ask, “What is this?” She responds “Ball.” You correct them that it is not a ball, but an egg. When this process is repeated several times, the child is able to tell the difference between a ball and an egg [68] [69].

To train a Neural Network, you show it several thousand examples of the classes (e.g. table, cup, Other) you want it to learn. This kind of training is called **Supervised Learning** because you are providing the Neural Network an image of a class and explicitly telling it that it is an image from that class [68] [69].

**To train a Neural Network, we need three things:**

- 1-**Training data:** Thousands of images of each class and the expected output.
- 2- **Cost function:** We need to know if the current setting is better than the previous knob setting. A cost function sums up the errors made by the neural network over all images in the training set.
- 3- **How to update the knob settings:** Finally, we need a way to update the knob settings based on the error we observe over all training images [68] [69].

## Steps Category Labeling in Image:

- The first task in annotating our dataset is determining which object categories are present in each image.
- In the next stage all instances of the object categories in an image were labeled.
- final stage is the laborious task of segmenting each object instance, this stage for image segmentation.
- Finally, PASCAL VOC’s primary application is object detection in natural images. MS COCO is designed for the detection and segmentation of objects occurring in their natural context.

## Time Testing

Our graduation project is able to process and match each training or recognition image in about two seconds on a computer, when using raspberry pi its take 3second to 4 second but is enough and good for blind people to know the object.

## 6.2 Results:

The view detection and recognition approaches have proven to work well in practice, after testing the project inside the office in front of the blind person. It was detection and recognition correctly and in short time not exceeding two second.

**Table 6.1:** Test of Objects

Object Name	Number of Tries	Detection Ratio	Pass	Failure	Percentage Error
person	50	91	49	1	2%
backpack	50	92	48	2	4%
bottle	50	94	48	2	4%
cup	50	91	49	1	2%
banana	50	93	49	1	2%
apple	50	93	48	2	5%
spoon	50	94	49	1	2%
bowl	50	91	47	3	6%
chair	20	91	47	3	6%
laptop	25	92	48	2	4%
tv	20	93	47	3	7%
mouse	50	92	49	1	2%
keyboard	50	92	49	1	2%
cell phone	50	91	49	1	2%
book	50	94	48	2	4%
clock	50	93	47	3	7%
scissors	50	92	49	1	2%
remote	50	93	48	2	4%
toothbrush	50	92	49	1	2%

In this table, we show you 20 object of the total 90, where we calculated the error rate in the object selection and recognition it. A range of errors was set between 2% and 7%. This value is based on the object and the accuracy of the camera, And the project is providing more features for an accurate of object detection like the Possibility to detect different object of the same type, detection from different angle, detection multiple object together.



**Figure (6.2):** Test of result detection and recognition by camera of project.

## 6.3 Challenges

- While building the system, there are many challenge was faced, such as:
- Not all the required component for the project are available
- in the Palestinian market; as a result, some of the main components were purchased from outside.
- Some of the project components are expensive like Raspberry pi & Mini Camera.
- Some problems in dealing and Understand Python programming language also with some project components like Raspberry pi.

## 6.4 Conclusion & Future work

we designed and implemented a smart glass for blind people using special mini camera.

Objects detection is used to find objects in the real world from an image of the world, that are common in the scenes of a blind. based on their locations, and The camera is used to detect any objects.

We expect further improvements in the future as we develop new feature types including color, distance and other features.

We also recommend using this component Movidius Neural Compute Stick (NCS) is a deep learning USB drive. The NCS is powered by the low-power high-performance Movidius Visual Processing Unit (VPU). run multiple devices on the same platform to scale performance.