# Tribhuwan University

## Institute of Engineering

## Pulchowk Campus


A Project Proposal on

Object Recognition and Image Enhancement for Night Vision Surveillance

Prepared for

Department of Electronics and Computer Engineering

Pulchowk Campus, IOE

Prepared by

"AASHISH BHANDARI, 070BEX401"

"AAYUSH KAFLE, 070BEX403"

"PRANJAL DHAKAL, 070BEX427"

"PRATEEK RAJ JOSHI, 070BEX429"

"17TH POUSH 2073"

# COVER LETTER

To,
The Head of Department
Department of Electronics and Computer Engineering,
Pulchowk Campus, IOE
Pulchowk, Lalitpur

Dear Sir,

We the students of Electronics and Communication Engineering intend to develop "Object Recognition and Image Enhancement for Night Vision Surveillance" for our Major Project.
The project "Object Recognition and Image Enhancement for Night Vision Surveillance" is an application oriented project that uses the concept of image processing and machine learning. The project aims to develop a surveillance system capable of operating in both well-lit and dark scenarios.
Please review the proposal attached and we are eagerly anticipating your approval and necessary arrangements.

Sincerely yours,
Aashish Bhandari (070bex401)
Aayush Kafle (070bex403)
Pranjal Dhakal (070bex427)
Prateek Raj Joshi (070bex429)

## Acknowledgement

We would like to express our humble gratitude to everyone who has made contribution in finalizing our major project. We are very grateful for having "Object Recognition and Image Enhancement for Night Vision Surveillance" as our major project. This project encompasses all the key disciplines, we are keenly interested in.

We would like to heartily thank **Mr. Dinesh Baniya Kshatri** sir for his wonderful insight and assistance for helping us to refine and finalize the topic of our major project. We are very thankful of him for providing us with necessary resource materials for the topic.

Finally, we would like to sincerely thank the Department of Electronics and Computer Engineering for elucidating us with the concepts and constraints in choosing our major project.

## Abstract

The project “Object Recognition and Image Enhancement for Night Vision Surveillance” is an application oriented project that uses the concept of image processing and machine learning. The project aims to develop a surveillance system capable of operating in both well-lit and dark scenarios.

The proposed system is capable of automatizing surveillance by recognizing objects and alerting the user upon discovering foreign objects. We plan to learn and implement the concept of client-server approach in our project to devise a lightweight front-end system.

Surveillance systems are highly capable during well-lit situations like daytime but most of these systems rely on manual intervention during night time. Many approaches have been tried to build surveillance systems capable of operating automatically in dark situations. Our plan is to continue this effort using image processing and machine learning technique.

# Table of Contents

Problem Statement ..... 1
Objectives: ..... 2
Introduction: ..... 3
1. Infrared Imaging ..... 3
2. Digital Image Processing ..... 4
3. Machine Learning using Neural Networks ..... 4
4. Computer Vision ..... 5
5. GPU Processing ..... 5
Description of proposed system ..... 6
1. Block Diagram ..... 6
2. System Description ..... 6
A. Client Side ..... 6
B. Server Side ..... 7
Methodology ..... 9
A. Client Interface Operation ..... 9
B. Server Operations ..... 10
Expected Outcome ..... 11
Scope ..... 12
Cost Estimation ..... 13
Time Estimation ..... 14
Appendix ..... 16
A. List of figures ..... 16
B. List of Tables ..... 16
C. References ..... 17

## Problem Statement

Object recognition is critical part of any surveillance system. It is the matter of utmost concern to identify intruders and foreign objects in the area where surveillance is done. In the process of automatized detection of unauthorized objects using camera, the image is first processed to isolate object from background. This is accomplished by working on the fact that there is notable difference in contrast between the object and background.

The performance of surveillance system using traditional camera in daylight is vastly superior as compared to night. The main problem for surveillance during night is the objects captured by traditional cameras have low contrast against the background i.e. there is an increase in cluster of pixel intensity towards left direction of x-axis in image histogram. So, object isolation by setting threshold observing the image histogram becomes impractical. The severity of problem increases with the decrement of light available during image capture.

So, the use of traditional camera for night surveillance poses a number of problems. So, there is a great need for assistive tools and technology for surveillance during low light situations.

## Objectives:

1. To become familiarized with image processing techniques and computer vision used for manipulating images captured in dark situations or with IR cameras.
2. To become acquainted with machine learning approach for image recognition.
3. To implement a Graphics Processing Unit (GPU) accelerated Machine Learning model.
4. To implement a basic client-server model.

## Introduction:

The evolution of Computer Vision and Machine Learning(ML) techniques have unlocked myriad possibilities in today's world. The success of deep neural networks and other ML algorithms in applications like autonomous vehicle driving, speech recognition, image recognitions etc. attests to the value of using such methods to tackle complex real-world problems.

One of the prominent feature of Computer Vision includes image recognition. In this, presence of any object in an image is confirmed and then attempt is made for recognition. Our project, "**Object Recognition and Image Enhancement for Night Vision Surveillance"** is a prominent example of Computer Vision application.

"**Object Recognition**" here means the ability of the system to isolate specific features from the image and then classify them in some predefined categories such that the presence of the object is indicated. "**Night Vision**" signifies that the images will be either underexposed or will be taken from special "Night Vision Camera" in complete darkness which will needs some form of enhancement such that the system is able to further process it.

A brief introduction to every key aspect of the project are discussed below:

### 1. Infrared Imaging

Infrared Cameras deal with imaging using Infrared Radiation Band (700 nm to 14 $\mu$m) rather than the Visible Band in Electromagnetic Spectrum (400 to 700 nm). The basic principle of an infrared camera is the detection of the infrared energy (which is the function of temperature) being emitted by the bodies in the surrounding and its conversion into electrical form. A brighter spot in the image taken from an IR camera implies the point or region with high temperature irrespective of degree of visibility of the surrounding. The images taken from a IR camera is monochromatic. Primarily, infrared imaging is used in night vision applications and building inspections.


*Figure 1: IR Camera Module*

## 2. Digital Image Processing

Image Processing refers to mathematical operations of signal processing where the input may be image, series of images, video or video frames. It basically deals with pixel-wise operation of images without extraction of meaningful information. The output of image processing may be either an image or a set of characteristics or parameters related to the image.

Digital Image processing is the use of computer algorithms to perform image processing on digital images.

## 3. Machine Learning using Neural Networks

Machine Learning enables computer programs to learn without being explicitly programmed. It evolved from the study of pattern recognition and computational learning theory in artificial learning.

In a broad sense, machine learning can be classified into two categories; supervised learning and unsupervised learning. Supervised Learning involves specific set of mathematical operations to minimize a cost function (or loss factor) determined by comparison of expected output to the actual output. Unsupervised Learning on the other hand, deals with extraction of pattern or meaningful information from a cluster of data.

The project aims to implement an appropriate neural network using machine learning for the detection (and also recognition) of a small set of object.

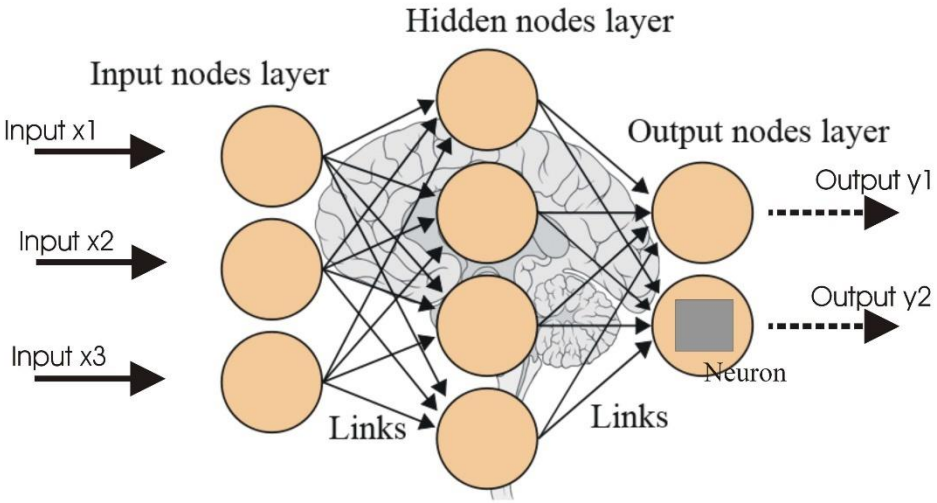


*Figure 2: A 3-layered Neural Network*

A neural network is an architecture of interconnected artificial neural nodes called "neurons" that exchange data between each other. The connections have numeric weights that are adjusted during the training process, so that a properly trained network will respond correctly when presented with an image or pattern to recognize. The network consists of multiple layers of feature-detecting "neurons".

### 4. Computer Vision

Computer Vision (image understanding) refers to the techniques that mainly involves segmentation, recognition and reconstruction (3D models) which work together to give the ability of scene understanding to the computer. It deals with extraction of high-dimensional data from the real world in order to produce numerical or symbolic information. Therefore, meaningful information is extracted from image or image related parameters.

Computer Vision integrates both Image Processing and Machine Learning concepts. This enables it to accomplish many real-world applications with applicable and efficient methods which otherwise would involve myriad complications if needed to be hard coded.

After the Computer Vision techniques are applied, the end result will be detection of presence and if present, classification of the object. This classified data will be used to generate appropriate control signals which will be sent to the client interface. The interface will be designed to operate in accordance to the received control signal performing desired action.

### 5. GPU Processing

Implementation of machine learning algorithms in conventional serial processors is plagued by large number of challenges. The serial processors are extremely slow at processing multidimensional data such as matrix. For the efficient and real time application using machine learning algorithm, there is a need for parallel processing of large chunks of data. Graphics processors are very well suited for this operation.

Graphics cards are used in machine learning applications as there is increase in: the availability of massive amounts of training data, and powerful and efficient parallel computing provided by GPU computing. GPUs are used to train these deep neural networks using far larger training sets, in an order of magnitude less time, using far less datacenter infrastructure. GPUs are also being used to run these trained machine learning models to do classification and prediction in the cloud, supporting far more data volume and throughput with less power and infrastructure.


*Figure 3: GPU Chip*

# Description of proposed system

## 1. Block Diagram

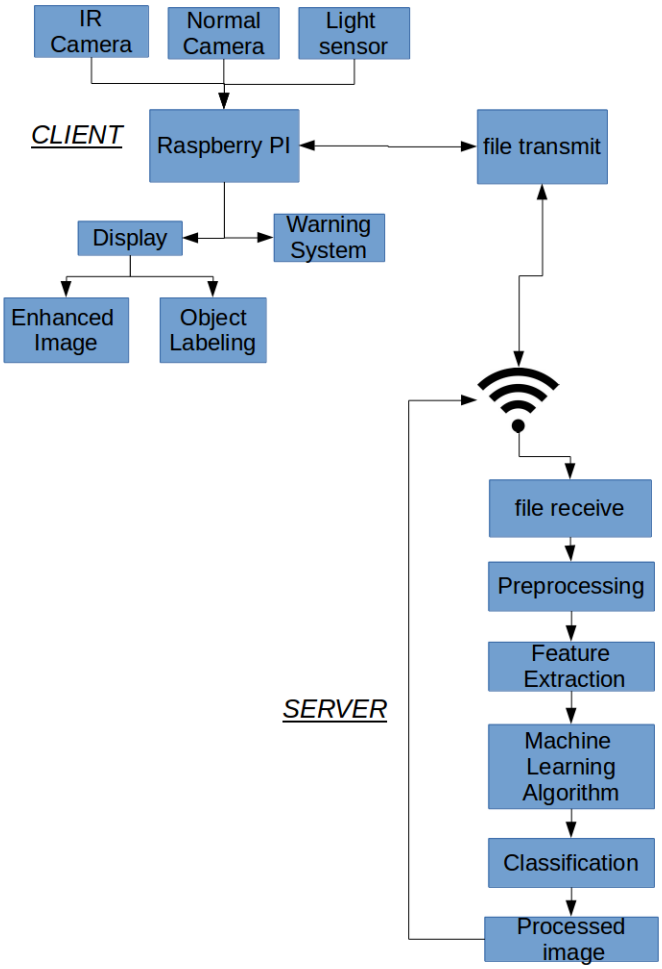


*Figure 4: Block Diagram of Proposed System*

## 2. System Description

### A. Client Side

#### I. Sensor Array:

The client device will be equipped with optical sensors namely IR camera, normal camera and LDR (Light Dependent Resistor). These sensors are used to detect environment lighting condition and take engage appropriate camera to take images of environment. Normal camera will be used in the daylit conditions while the IR camera will be used for night time or appropriate low light condition.

#### II. Raspberry pi:

The main processing module in client unit is Raspberry pi. This unit controls and coordinates all the activities in client unit. It is responsible for taking information of the environment and send the data to server for processing. It uses wireless transmission protocols to send data to server. It receives the processed result and generates appropriate control signal to warn users.

### III. Warning and display module:

The features of surveillance is implemented using this module. It is a collection of warning devices such as alarm or other similar units. It is responsible for providing warning to the user and also display monitoring feed.

A server will also be set-up with high processing power which involves use of Graphics Processing Unit. The Raspberry Pi will transfer the captured frame of specific durations to the server through wireless connections. In the server, Computer Vision techniques in conjunction with Machine Learning approaches will be used to detect the presence of a small set of objects in the obtained footage. Once the presence of object is detected, the server will reply back with certain control signals which will trigger some form of "Warning" or "Alert" to the end user. This way a surveillance system will be established for both day and night.

## B. Server Side

### I. File Receive

The file receive segment will function in the background. The server will receive information from the client interface through wireless network.

### II. Preprocessing

The file received will be an archive of video frames. After extraction of the archive, the video frames will be subjected to image processing. The processed image will be converted to appropriate data format for subsequent operations.

### III. Feature Extraction, Machine Learning Algorithm, Classification and Processed Image

There are several conventional methods of Feature Extraction from images. These involve some "Hand-Engineered" methods like extraction of Histogram of Oriented Gradients (HOG). These are often used in conjunction classifiers like Support Vector Machine (SVM) to get the desired classification. Alternatively, a deep learning approach could be applied like Convolutional Neural Network (CNN) which takes care of both Feature Extraction and Classification and has better performance. Also, the extracted HOG could be used in a simple neural network which would also require the HOG of the training images.

## IV. Signal Generation

The action of the client interface will be based on the control signals from the server. The server will be responsible to issue the necessary control signals from the first step of the system operation.

The client interface will begin capturing video frames once it receives an initiation signal from the server. Also, after server finished its operation, the classified image will be used to generate control signals for the client interface. Based on these signals, the client will take appropriate actions like setting off alarms, lighting up some warning lights or even database update for remote monitoring.

## Methodology

A well-planned approach is needed for the successful completion of any projects. With reference to many papers, books and online tutorials, we have found some crucial concepts for effective project development.

### A. Client Interface Operation

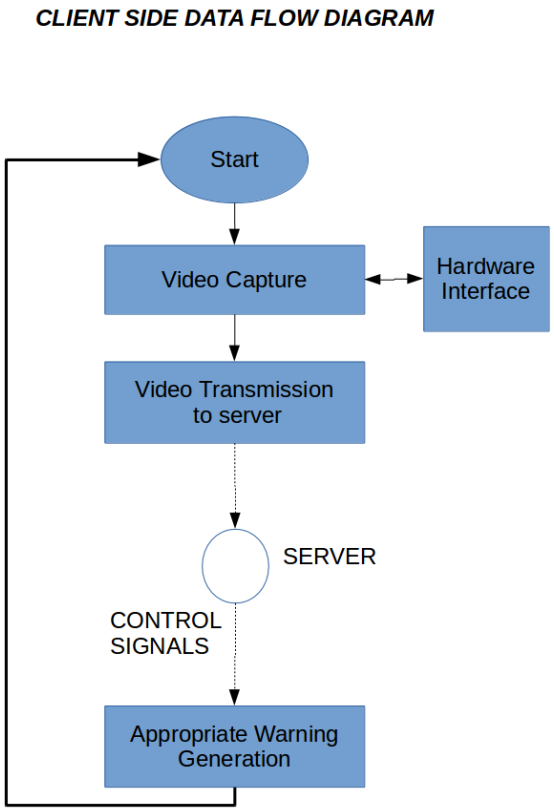


*Figure 5: Client Side Data Flow Diagram*

The first step of the system operation involves video frame capturing of the real world. These frames will be taken either in daylight or night-time. The selection of camera will be done on the basis of exposure of the image. A Light Dependent Resistor (LDR) will be used to detect the lighting condition of the environment and appropriate camera will be used.

An archive of video frames will be sent to the server through wireless means using HTTP or FTP. The server process this archive and then generate appropriate control signals which will be sent back to the server, again through wireless means. Based on these control signals, the client interface will then take appropriate action.

## B. Server Operations

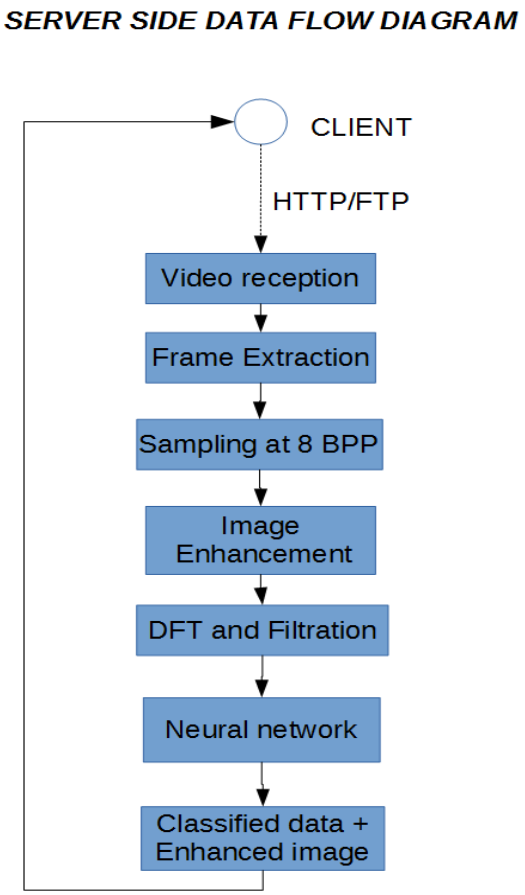


*Figure 6: Server Side Data Flow Diagram*

As server receives frame archive from the client (using FTP or HTTP), the frames are then extracted from it. The image is then sampled at 8 bpp in spatial domain. Adjusting operations like brightness, contrast and saturation adjustments is then performed for the image.

The adjusted image in the spatial domain is then subjected to Discrete Fourier Transform (DFT) where the image is further filtered and fed into the neural network for the object recognition of objects within the image.

The classified data and enhanced image is the transmitted to the client and the output is shown in the respective output units (display and alarm).

## Expected Outcome

We aim to develop a system that is user friendly, oriented towards problem-solving and readily applicable in real world scenarios.

The completed system will function as an assistive tool for surveillance in conditions where ample of lighting is available or lack thereof. The system will be able to detect and correctly classify certain number of objects into specified categories. It will be capable of alerting the user upon detecting anomaly by generating appropriate control signals.

Through this project, we shall learn the concept of machine learning and image processing for object recognition and image enhancement and apply them to solve real world problems like surveillance. We shall also learn and implement the concept of client-server networking model.

## Scope

Automating surveillance is of acute concern in many industries and home. There are number of areas where the proposed system may be used:

1. In industries, for providing surveillance without providing dedicated manpower.
2. In military application, for monitoring movement in hostile areas during low-lit situations.
3. In border security, for monitoring access across the country's border.
4. In personal offices or home, for security and monitoring purposes.

## Cost Estimation

The price estimate for the proposed system is presented below:

*Table 1: Cost Estimation*

| S. No. | Items | Cost* (Rupees) |
|---|---|---|
| 1 | Graphics card | 7600 |
| 2 | Raspberry Pi | 4450 |
| 3 | Infrared Camera | 2000 |
| 4 | Normal Camera | 2600 |
| 5 | Sensors, hardware components and assembly | 4000 |
| 6 | Research & Development | 5000 |
| 7 | Transportation | 3000 |
| 8 | Display Module | 4500 |
| | **Total** | 33,150 |

*(The price mentioned in the table are taken from present price of items in amazon.com)

## Time Estimation

We have divided the time estimation into two sections, one from the period between proposal submission and mid-term defense. The second section being the period between mid-term defense and final presentation.

**Phase 1:**

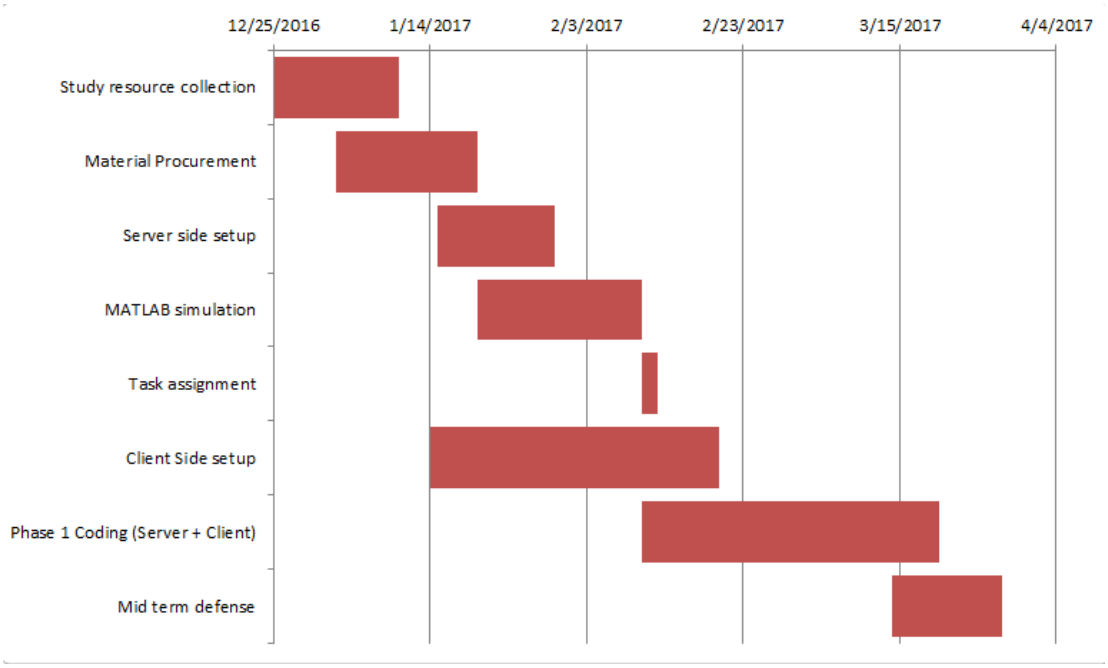


*Figure 7: Time Estimation (1)*

A working prototype of the proposed model able to classify few objects is expected in the 1st part. Resources necessary for the project will have been decided. The prototype shall contain basic server setup. A basic Object Detection Model based on Neural Networks will be built and simulated using MATLAB (or other appropriate programming tools). Client Interface might not be fully designed in this part. Training shall be done using freely available datasets. Immediate testing will be done using web-cams of PC. This part mainly involves creation of the software model and a basic server setup. Also, necessary documentations for the mid term defense sessions will be made.

**Phase 2:**

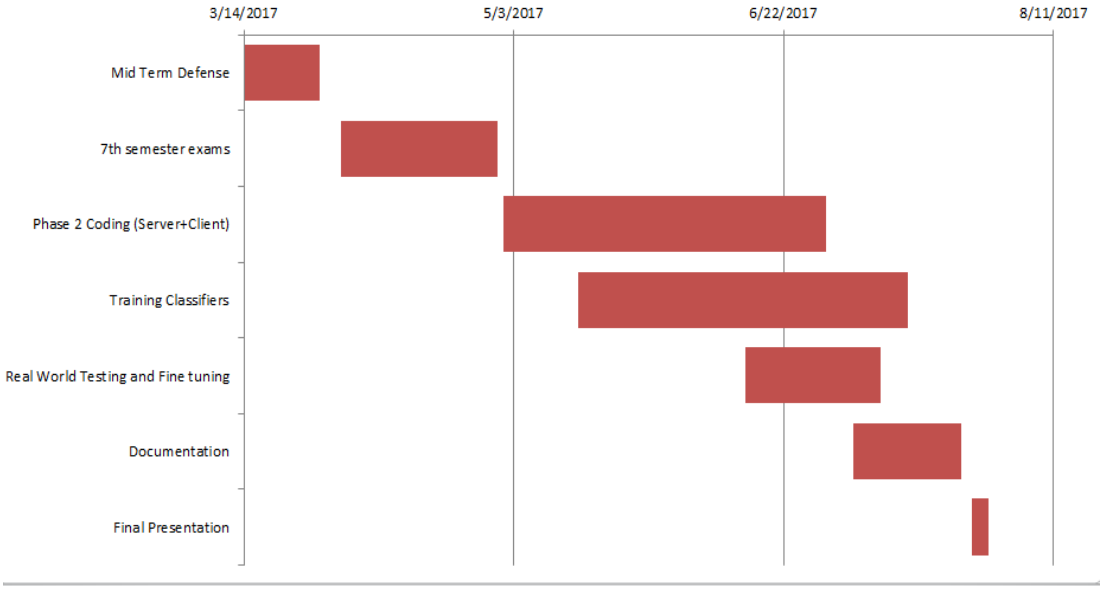


*Figure 8: Time Estimation (2)*

Further refining of algorithms, rigorous training, finalizing hardware and testing/debugging is planned to be done in this part. A viable product will be designed involving Raspberry Pi, cameras and sensors. By the end of the 2nd part, a fully trained CNN able to classify common objects is expected. A complete classification of the underexposed and night vision images is also expected. The system will be put through rigorous training, real world testing and debugging. “Warning” and “Alert” mechanisms will have been set up. Necessary documentations will also be made and presented in the final presentation.

# Appendix

## A. List of figures

Figure 1: IR Camera Module ..... 3
Figure 2: A 3-layered Neural Network ..... 4
Figure 3: GPU Chip ..... 5
Figure 4: Block Diagram of Proposed System ..... 6
Figure 5: Client Side Data Flow Diagram ..... 9
Figure 6: Server Side Data Flow Diagram ..... 10
Figure 7: Time Estimation (1) ..... 14
Figure 8: Time Estimation (2) ..... 15

## B. List of Tables

Table 1: Cost Estimation ……13

## C. References

[1] Digital Image Processing Tutorial. (n.d.). Retrieved December 25, 2016, from https://www.tutorialspoint.com/dip/index.htm

[2] Hijazi, S., Kumar, R., & Rowen, C. (n.d.). Using Convolutional Neural Networks for Image Recognition. www.cadence.com

[3] Intensity Histogram. (n.d.). Retrieved December 27, 2016, from http://homepages.inf.ed.ac.uk/rbf/HIPR2/histgram.htm

[4] Jones, S. (2016). How the GPU Is Revolutionizing Machine Learning | NVIDIA Blog. Retrieved January 01, 2017, from https://blogs.nvidia.com/blog/2015/12/10/machine-learning-gpu-facebook/

[5] S, T., & Ye, Y. (n.d.). Object Detection based on Convolutional Neural Network. *Stanford University*. Retrieved from http://cs231n.stanford.edu/reports/CS231n_final_writeup_sjtang.pdf

[6] Thermographic camera. (n.d.). Retrieved December 26, 2016, from https://en.wikipedia.org/wiki/Thermographic_camera