THAKUR DEGREE COLLEGE OF SCIENCE AND COMMERCE KANDIVALI (EAST), MUMBAI

**PROJECT REPORT ON**

**“WEB SCRAPPER”**

## DESIGNED AND DEVELOPED BY JAGDISH ANESHWAR

**SUBMITTED IN PARTIAL FULFILMENT OF BACHELORS OF SCIENCE (COMPUTER SCIENCE)**

## [UNIVERSITY OF MUMBAI]

**THAKUR DEGREE COLLEGE OF SCIENCE AND COMMERCE KANDIVALI (EAST), MUMBAI, MAHARASHTRA 400 101.**

## ACADEMIC YEAR 2022 – 2023



**A PROJECT ON YOURFITNESS APPLICATION**



## A PROJECT SUBMITTED TO

**THE UNIVERSITY OF MUMBAI FOR PARTIAL COMPLETION OF THE DEGREE OF BACHELOR OF SCIENCE IN COMPUTER SCIENCE**

## UNDER THE FACULTY OF SCIENCE

**PROJECT REPORT BY JAGDISH ANESHWAR**

## UNDER THE GUIDANCE OF MRS. NEHA KAMBLE

**THAKUR COLLEGE OF SCIENCE AND COMMERCE, KANDIVALI (EAST), MUMBAI, MAHARASHTRA 400 101.**

## MARCH 2022

CERTIFICATE

This is to certify that **MR. JAGDISH ASHOK ANESHWAR** has worked and duly completed her Project Work for the Degree of Bachelor in Management Studies under the Faculty of Science in the subject of Computer and the project is entitled, “Web Scrapper” under my supervision.

### I further certify that the entire work has been done by the learner under my guidance and that no part of it has been submitted previously for any Degree or Diploma of any University.

It is her own work and facts reported by her personal findings and investigations.

|  |  |  |
| --- | --- | --- |
| Dr. C.T. CHAKRABORTY | MR.ASHISH TRIVEDI | PROF. NEHA KAMBLE |
| Principal | Head Of Department | Project Guide |

### INTERNAL EXAMINER EXTERNAL EXAMINER

# Date of submission:

# Table of Contents

[Abstract 3](#_TOC_250000)

1. [Introduction](#_bookmark0) 4
   1. [Project Definition](#_bookmark1) 4
   2. [Project Objectives](#_bookmark2) 4
   3. [Project Specifications](#_bookmark3) 4
   4. [Product Architecture and Components](#_bookmark4) 5
      1. [Functional Diagram](#_bookmark5) 5

[1.4.1 User Interface Image](#_bookmark6) 6

* 1. [Applications](#_bookmark7) 6

1. [Literature Review](#_bookmark8) 7
   1. [Project background](#_bookmark9) 7
   2. [Previous Work](#_bookmark10) 7
   3. [Comparative Study 13](#_bookmark11)3
2. [System Design 13](#_bookmark12)3
   1. [Design Constraints](#_bookmark13) 13
      1. [Design Constraint: Engineering Standards 14](#_bookmark14)4
      2. [Design Constraint: Environmental 14](#_bookmark15)4
      3. [Design Constraint: Ethical 14](#_bookmark16)4
   2. [Design Methodology 14](#_bookmark17)4
   3. [Product Subsystems and Components 14](#_bookmark18)4
      1. [Product Subsystem1: Vision Acquisition 14](#_bookmark19)4
      2. [Product Subsystem2: Grayscale Conversion 16](#_bookmark20)5
      3. [Product Subsystem3: Vision Assistant 16](#_bookmark21)5
      4. [Product Subsystem4: Writing in a Measurement File 17](#_bookmark22)6
   4. [Implementation 18](#_bookmark23)8
3. [System Testing and Analysis 21](#_bookmark24)1

[4.2 Subsystem 2: Vision Acquisition 21](#_bookmark25)1

* 1. Subsystem 2: Vision Assistant 221
  2. [Overall Results, Analysis and Discussion 24](#_bookmark26)2

1. [Project Management](#_bookmark27) [26](#_bookmark27)6
   1. [Project Plan](#_bookmark28) [27](#_bookmark28)6
   2. [Contribution of Team Members](#_bookmark28) [27](#_bookmark28)7
   3. [**Project Execution Monitoring**](#_bookmark28)[27](#_bookmark28)8
   4. [**Limitation and Challenges**](#_bookmark28)[27](#_bookmark28)8
   5. **Project Bill of Materials and Budget** 299
2. **Project Analysis** 299
   1. **Life-long Learning** 299
   2. **Impact of Engineering Solutions** 299
   3. [**Contemporary Issues Addressed**](#_bookmark29)30
3. [**Conclusions and Future Recommendations**](#_bookmark30)30
   1. [**Conclusions**](#_bookmark31)30
   2. **Future Work and Expected Final Prototype/Results** 31
   3. **Future Recommendations** 31
4. **References** 31

### Abstract

Face detection and face recognition are very important technologies these days, furthermore we noticed that they got have a variety of uses such as cellphones, army uses, and some high risk information offices. We decided to make a device that detects and recognize the face as a student attendance system and can be a substitute for the regular paper attendance system and finger print attendance system. The main function in our project is going to be done using LabVIEW because, LabVIEW is a very helpful programming tool in regards of facial uses and very helpful in other uses. Our project is based on a main program in LabVIEW that detects and recognize faces with giving scores and parameters, furthermore the subsystems are an Excel sheet that is integrated with the program, and a messaging device that is for either a message for absent students or to the student’s parents. Components of our project are LabVIEW program as the main system and subsystems, Office Excel sheet to include students names, and a computer (or laptop) to integrate the programs together.

# Introduction

## Project Definition

Design of an automatic class attendance system using face detection algorithm of LabVIEW software. The system requires a video capture device and the running LabVIEW algorithm to be implemented successfully. It detects the faces and mark attendance accordingly. This system will prevent unnecessary wastage of time of classes that is usually wasted in form of class roll calls.

## Project Objectives

1. Reducing time wastage during conventional class attendance.
2. Utilizing latest trends in machine vision to implement a feasible solution for class attendance system.
3. Automating the whole process so that we have digital environment.
4. Preventing fake roll calls as one to one attendance marking is possible only.
5. Encouraging the use of technology in daily lives.

## Project Specifications

* 1. Uses Pattern Matching algorithm for face detection.
  2. Score of minimum 600 required to perfectly match a face.
  3. Metric: Camera Resolution.
  4. For prototype fixed to 10 users only but scalable design.
  5. Requires good lighting condition for better camera capture capability.
  6. Attendance sheet is .xlsx format and can be digitally distributed and maintained.

## Product Architecture and Components

##### Functional Diagram

**FIGURE 1.1: BLOCK DIAGRAM OF THE SYSTEM**

The subsystem description is as follows:

**Camera:** The camera is the only hardware component required to capture live video feed of class.

**Vision Acquisition:** This module allows image to be captured by camera into LabVIEW for programming. It includes IMAQ submodules such as IMAQ Create, IMAQdx Open, IMAQdx Grab. They all combine to provide Continuous Acquisition of video feed from camera module.

**Image to Grayscale:** This process is performed using IMAQ ExtractSingleColorPlane VI to convert a 32/16bit image to 8bit image. This is a requirement for our pattern matching algorithm to work completely.

**Pattern Extraction:** This is included in Vision Assistant VI which deals with our face recognition algorithm. Pattern Extraction is feature in which the image inputted features are compared using Pattern Matching Algorithm.

**Feature Extraction:** This feature is used to extract important features out of image. It compares them with templates, saves in database and provides a score of comparison.

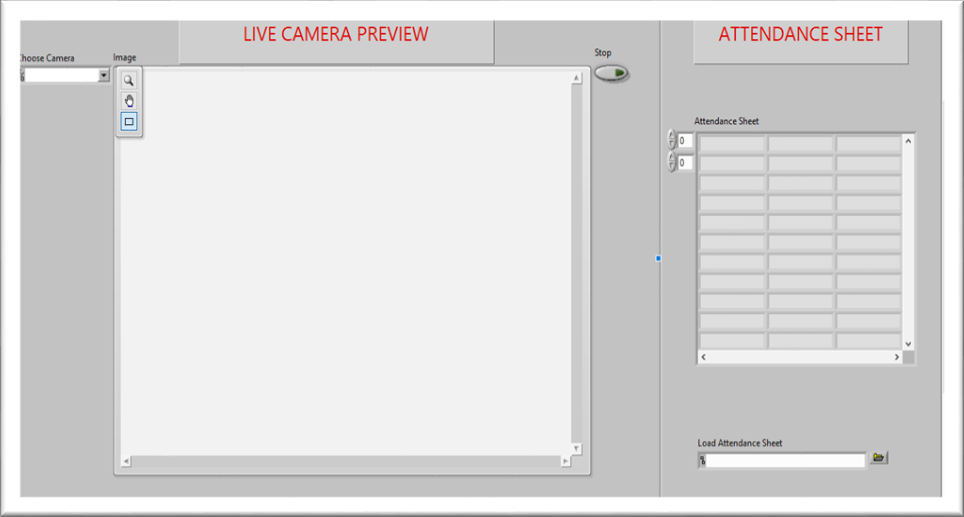
**Find Match in database:** Our database has preserved templates or images of students which we aim to recognize and mark attendance. This database can be updated or appended according to

requirement. This database is used for comparison with extracted feature of image to confirm a successful hit.

**Update Attendance Sheet.xlsx:** If match is found our algorithm updates the attendance of user corresponding to his/her name in excel file of format .xlsx. If not, the system marks absent in front of his/her name in the same excel file.

##### 1.4.1 User Interface Image

This is the front panel of LabVIEW program that the user is going to be using. It shows the attendance sheet with the names of the students, and a live camera of the user in front of the camera.



**FIGURE 1.2: FRONT PANEL OF LABVIEW PROGRAM**

## Applications

1. Large application in institute attendance system where multiple attendances are carried out for different classes. The attendance will be short timed and reduce manual errors.
2. Large application of computer vision in field of Communication, Biomedical, Automatic Product Inspection.

# Literature Review

## Project background

In the face detection and recognition system, the process flow is initiated by being able to detect the facial features from a camera or a picture store in a memory. The algorithm processes the image captured and identifies the number of faces in the image by analyzing from the learned pattern and compare them to filter out the rest. This image processing uses multiple algorithm that takes facial features and compare them with known database.

The motivation behind this project is to simplify the means by which attendance is taken during lectures and how much time it takes. The use of ID cards or manually calling out attendance and writing it down on sheets is not productive and efficient. This system will detect the number of faces on the class and will also identify them from the store database. With the face detection and recognition system in place, it will be easy to tell if a student is actually present in the classroom or not.

## Previous Work

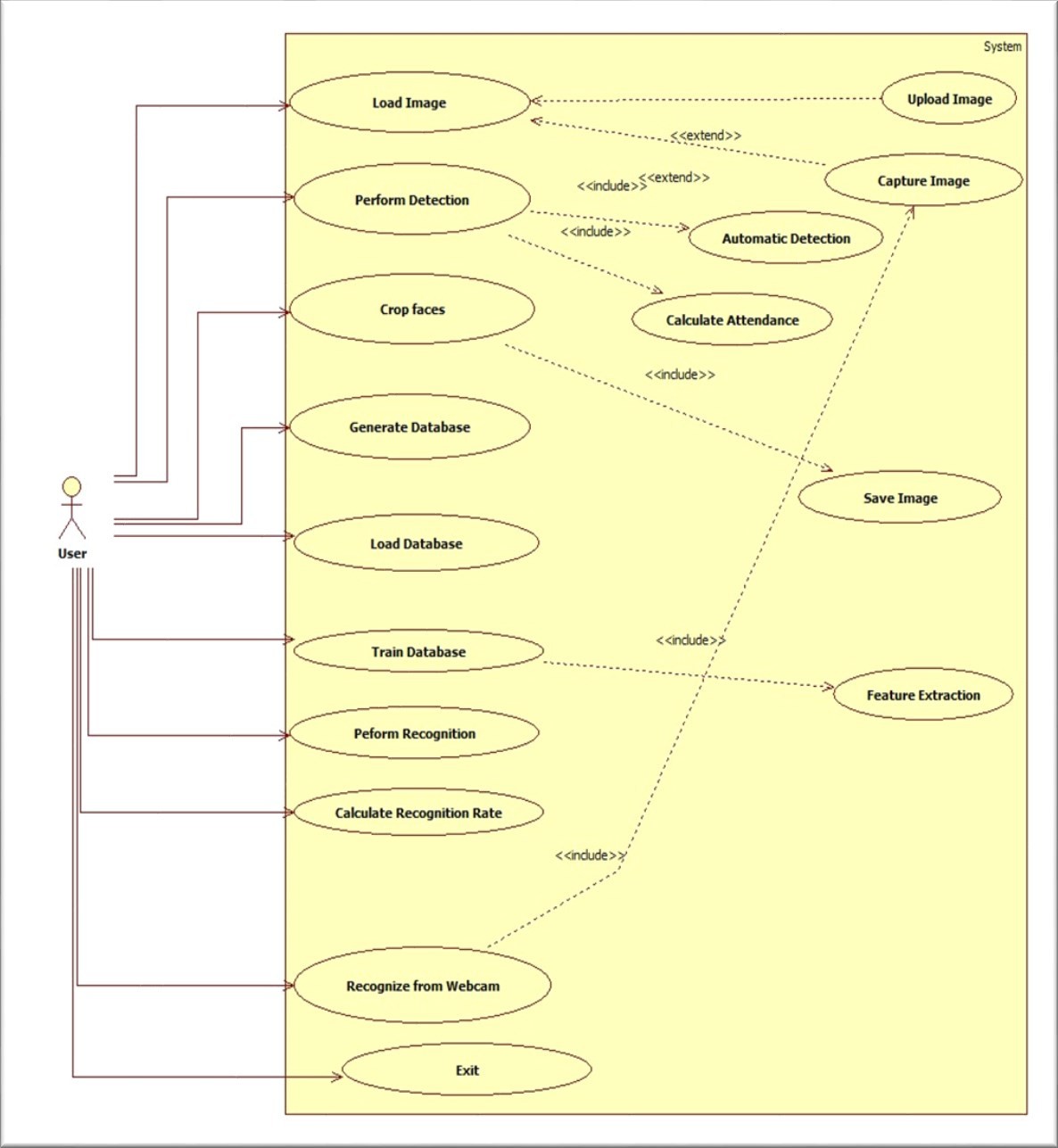
#### PROJECT # 1

This is a project done by students as a final year project at Kingston University London in 2018.

The system will be presented an image either via camera or from memory and it must detect the number of faces on it automatically. After identifying faces, the system should crop the faces from the image and store them in memory for image recognition which will be done in the second step. The system should be able to automatically count the number of faces detected on the image.

The second step will be the recognition part where the system will be able to match faces from the stored dataset and compare it to the input data from the first step. A software will be used for this system which automatically sorts out the faces. The software will be inter-active so to facilitate interaction between multiple tasks as required. Because the system has two steps, the second phase of the system will involve the training of images on a dataset that are to be used for recognition.

The system behavior has been explained in the following flowchart



**FIGURE 2.1: BLOCK DIAGRAM OF PREVIOUS PROJECT #1**

#### Technology Used

The key algorithms are Viola-Jones for face detection and Hidden Markov Model with SVD.

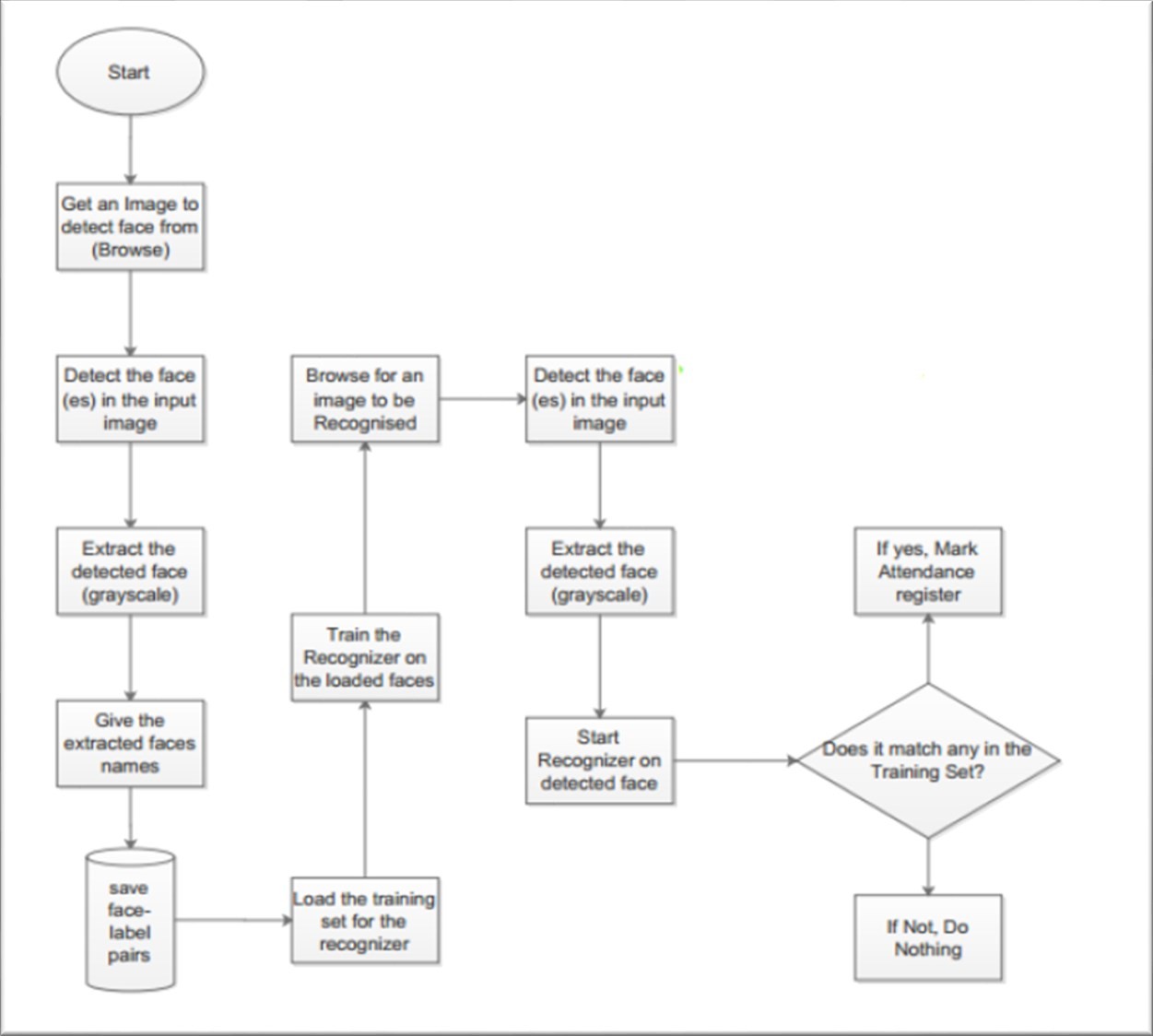
* The implementation of The Viola-Jones algorithm is available on softwares like MATLAB, OpenCV and Web Browsers (using adobe flash).
* The existing implementation of the Hidden Markov Model with SVD for face recognition are available on MATLAB, C++ and OpenCV libraries.

#### PROJECT # 2

This is a project done by students as a final year project at University of Nairobi in 2012.

The system will comprise of two modules. The first module a.k.a face detector is a mobile component, which is basically a camera that captures student faces and stores them in a file using computer vision face detection algorithms and face extraction techniques. The second module is a desktop application that does face recognition of the captured images (faces) in the file, marks the students register and then stores the results in a database for future analysis.

Following flowchart explains the process of the flow of information throughout the process.



**FIGURE 2.2: BLOCK DIAGRAM OF PREVIOUS PROJECT #2**

#### Technology Used

The following tools will be used in the implementation of the designed system. They’ve been divided in to two categories; Mobile and Desktop tools.

#### Mobile Tools

The face detection module will use OpenCV library for implementation by use of the frontal Haar Cascade face detector in either Android studio.

**OpenCV for Android Library -** (Open Source Computer Vision) is a library of programming functions mainly aimed at real-time computer vision.

**Android Studio/ Eclipse IDE -** Android Studio is the official IDE for Android application development, based on IntelliJ IDEA.

#### Desktop Tools

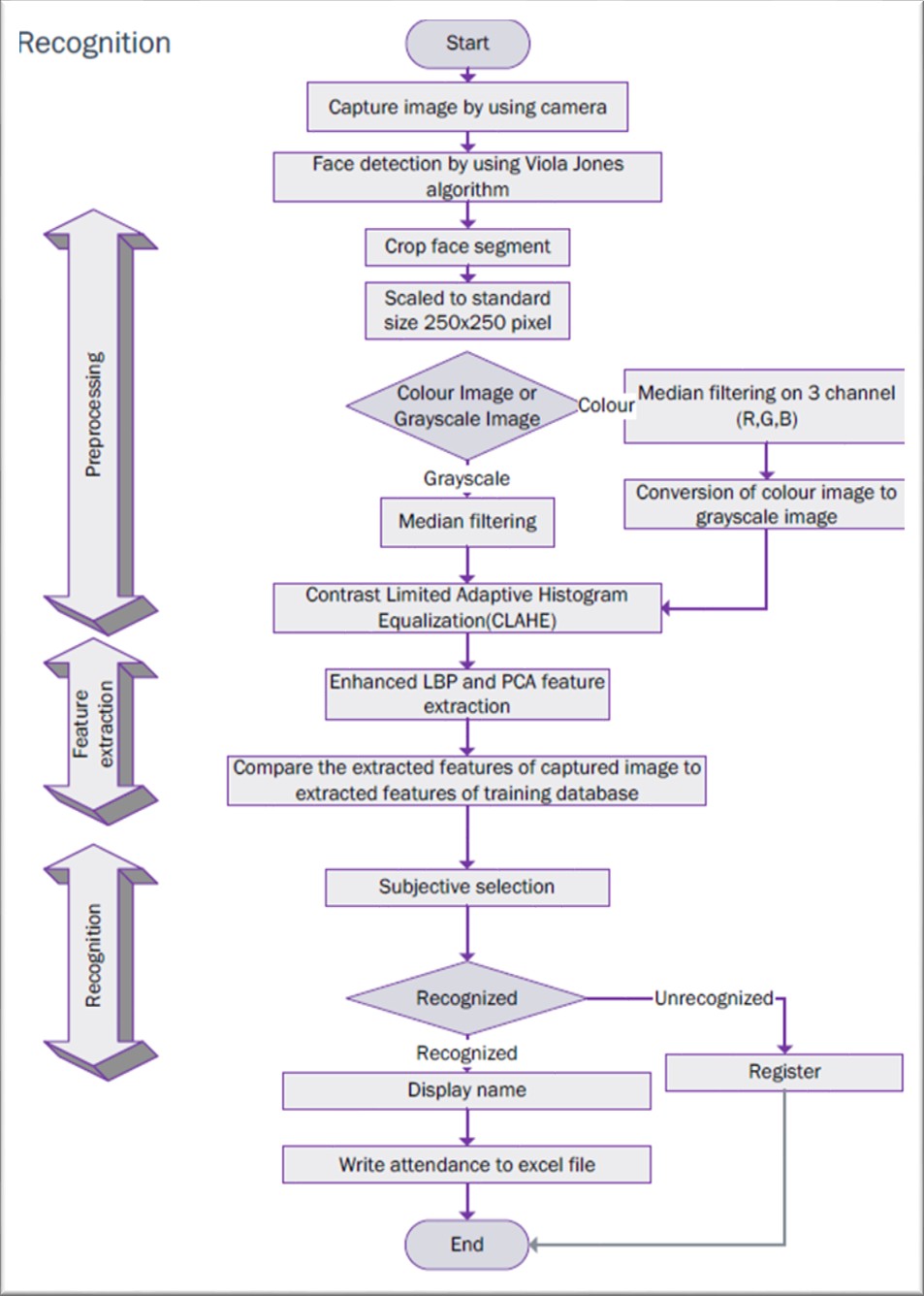
**EmguCV Library -** EmguCV is a cross platform .Net wrapper to the OpenCV image processing library. OpenCV/EmguCV uses a type of face detector called a Haar Cascade. The Haar Cascade is a classifier (detector) trained on thousands of human faces.

**Visual Studio** - Visual Studio is able to build and run the solution examples after a proper configuration of EmguCV. The desktop software will implement the two sub-systems (Training set manager and Face recognizer) together with face detector in windows form.

#### PROJECT # 3

This is a project done by students as a final year project at Universiti Tunku in 2018

The approach performs face recognition-based student attendance system. This method is also similar to others and begins with the input of an image either loaded from memory or from camera. Then it pre-processes the facial features and extracts it followed by subjective selecting and then the recognition of the facial images from known database. Both LBP and PCA feature extraction methods are studied in detail and computed in this approach in order to make comparisons. LBP is enhanced in this approach to reduce the illumination effect. An algorithm to combine enhanced LBP and PCA is also designed for subjective selection in order to increase the accuracy.



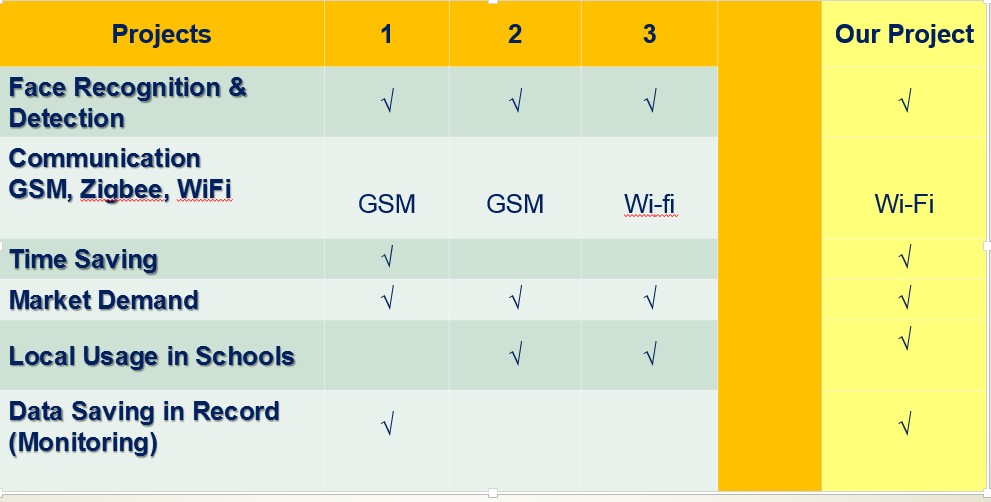
**FIGURE 2.3: BLOCK DIAGRAM OF PREVIOUS PROJECT #3**

The project is completely built in MATLAB with OpenCV libraries implemented in it.

## Comparative Study

Our project is different than all the previous projects made and mentioned above. They have purely used the core of machine vision to implement a face detection mechanism. None of the above- mentioned projects have realized the power of LabVIEW programming and LabVIEW Vision modules in which not only pattern matching but other machine vision algorithms like edge tracking, geometric matching can be implemented with ease.

Though the general mechanism and flow of events is similar in above projects and our current project however, the mechanism of face detection is completely unique and different.



**FIGURE 2.4: COMPARESSON BETWEEN ALL PROJECTS**

# System Design

## Design Constraints

The constraints which were considered while designing on project are following.

##### Design Constraint: Engineering Standards

The samples for database should be increase, as to increase the efficiency of detection. Also, the more the expensive the camera, the easier its algorithm is likely detecting the person.

##### Design Constraint: Environmental

The camera should capture all the students present in the class. Each student present should be seated such that it is visible to camera, so that his/her attendance gets marked easily.

##### Design Constraint: Ethical

The second limitation which is faced include the person appearance by face, which a person changes his/her look and looks different from the picture in the database of the attendance system, then it may be difficult for his/her attendance to be marked.

## Design Methodology

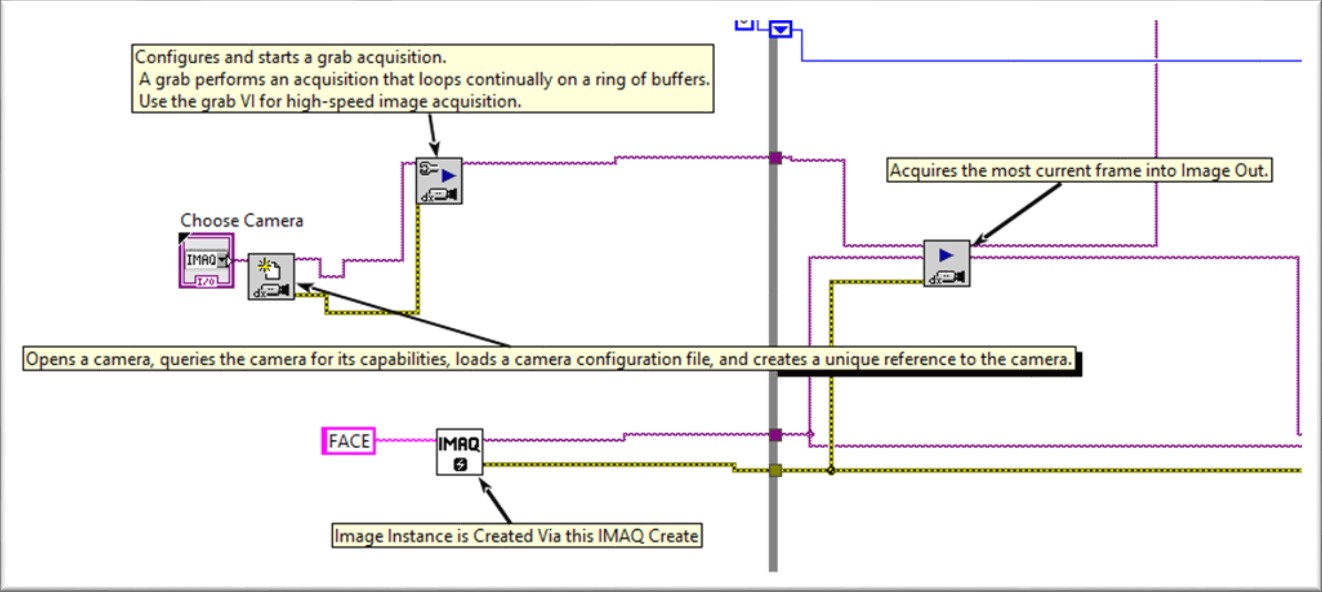
As we mentioned before in (Figure 1.1). The project process is:

* A camera will take continuous stream.
* In LABVIEW, IMAQ library for vision will be used.
* Convert the RGB image to grayscale image.
* Then perform Machine Vision Algorithm and match with patterns stored in our database.
* If pattern matches based on the score of how successful, decide to mark attendance or not.
* Update the marked attendance in a measurement file.

## Product Subsystems and Components

##### Product Subsystem1: Vision Acquisition

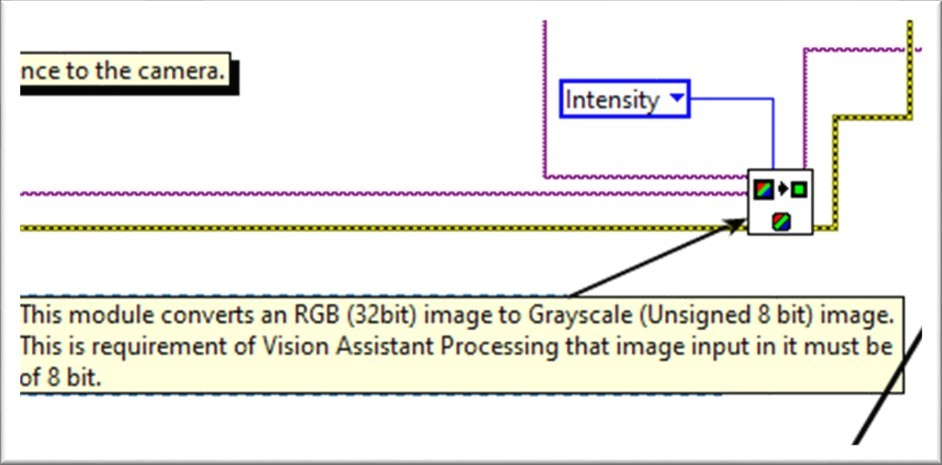
This subsystem is used to acquire continuous stream of video from attached camera. It starts a camera session from desired camera and transmits its image feed to further processing. The feed captured is inline processed and then the next feed is captured as shown in (Figure 3.1).



**FIGURE 3.1: VISION ACQUISITION PROGRAMING ICONS**

##### Product Subsystem2: Grayscale Conversion

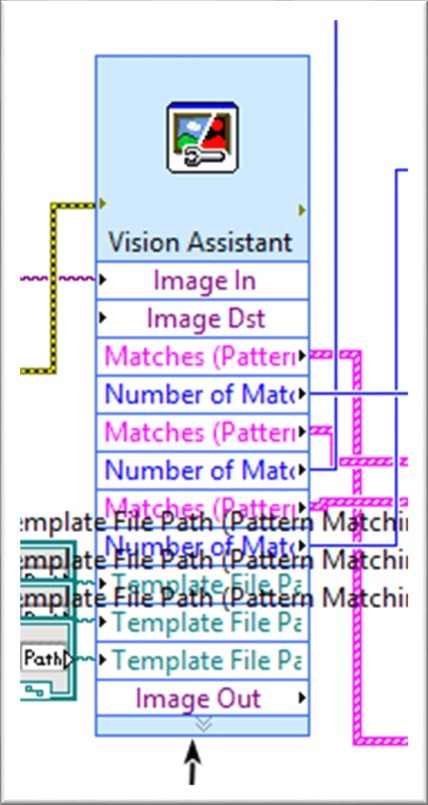
This module converts an RGB (32bit) image to Grayscale (Unsigned 8 bit) image. This is requirement of Vision Assistant Processing whose image input must be 8 bits as shown in (Figure3.2).



**FIGURE 3.2: GRAYSCALE VISION CONVERSION ICON**

##### Product Subsystem3: Vision Assistant

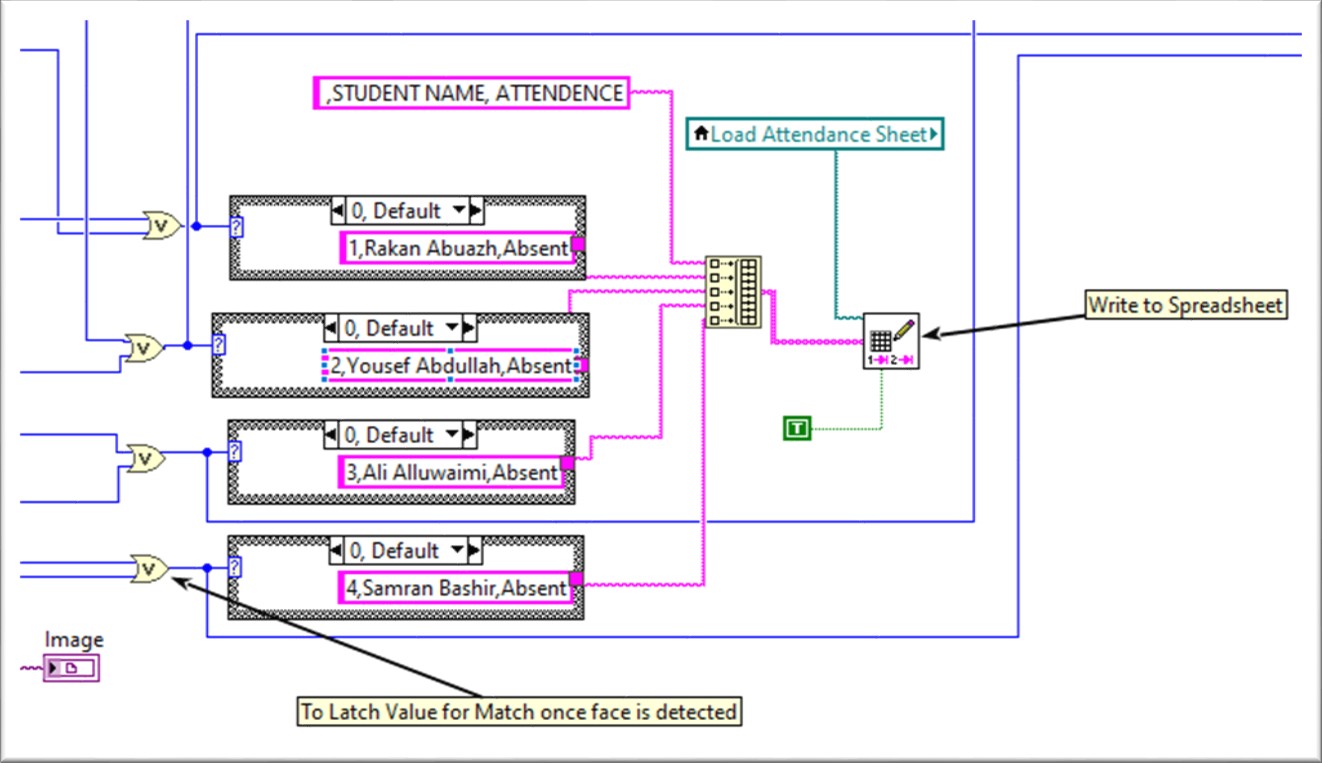
In (Figure 3.3) Vision Assistant helps us to perform Machine Vision Algorithm Pattern Matching on our image. This allows us to detect faces of student in a group of class. First one must add student faces as template in this program to create a database of images using reference images. The result is outputed which includes the information of score 0-1000 to tell how successful a match was, position of match occurred in image, angle of match occurred in image. This information together with number of matches for each user will be used to mark attendance of user in future progresses.



**FIGURE 3.3: VISION ASSISTANT ICON**

##### Product Subsystem4: Writing in a Measurement File

Each Student name and his/her attendance is marked if the case is true, else the attendance marked absent. The file path to the spread sheet delimiter is provided and the delimiter used to separate each student data is new line (new row). The system also caters a latch in form of a OR gate in which once a student is detected the system will maintain its value no matter if he/she gets out of frame and is undetected as demonstrated in (Figure 3.4).

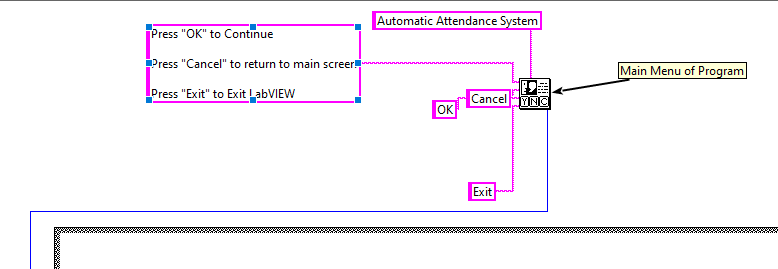


**FIGURE 3.4: WRITING ON SPREADSHEET NAMES**

## Implementation

#### THE MAIN MENU:

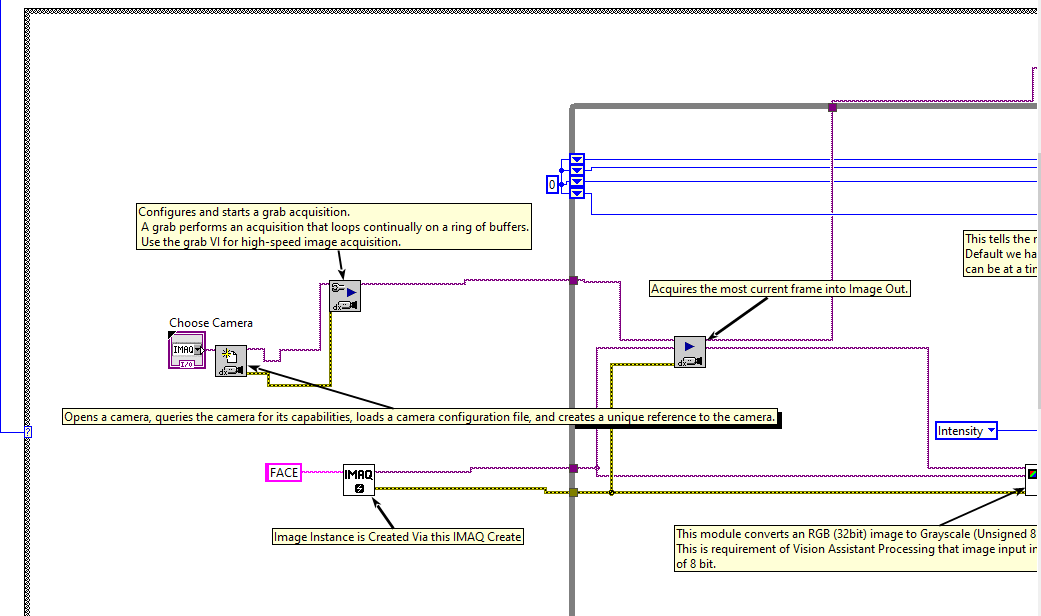
This is the first prompt to appear when program is run. It will ask user to either proceed further for face detection or exit the program as demonstrated in (Figure 3.5).



**FIGURE 3.5: MAIN MENU ICONS**

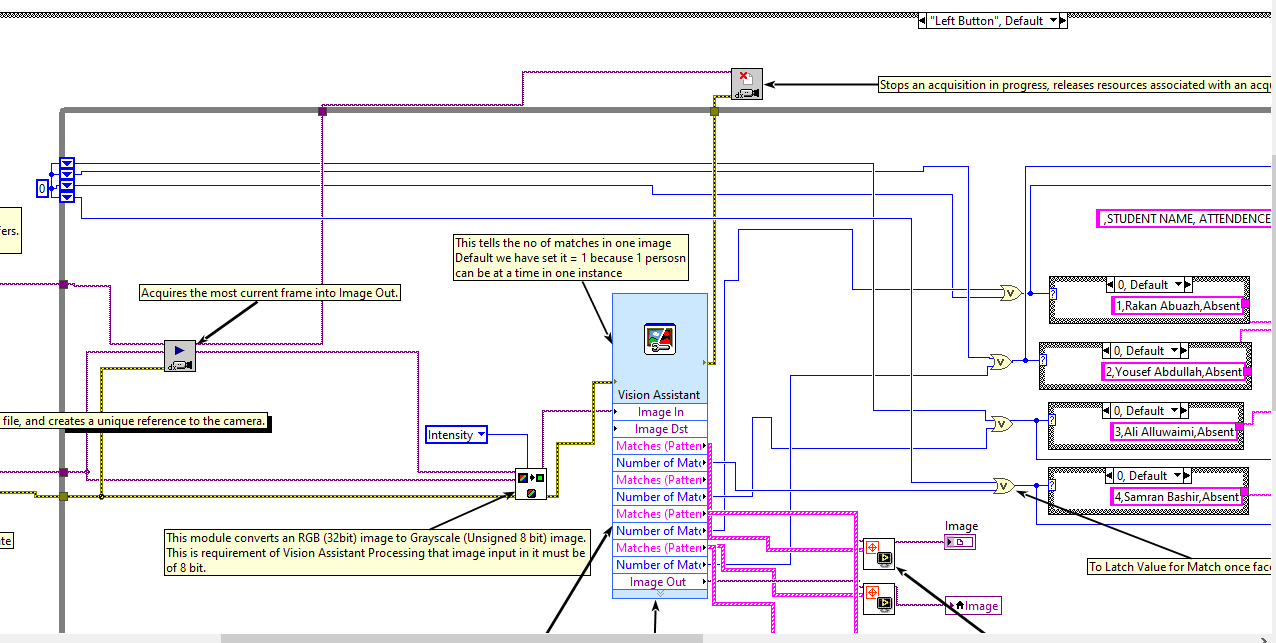
#### THE VISION ACQUISITON MODULES:

Vision acquisition is responsible of the live camera in front panel, changing the image to greyscale so the Vision Assistant can accept it, and the camera choice as shown in (Figure 3.6).

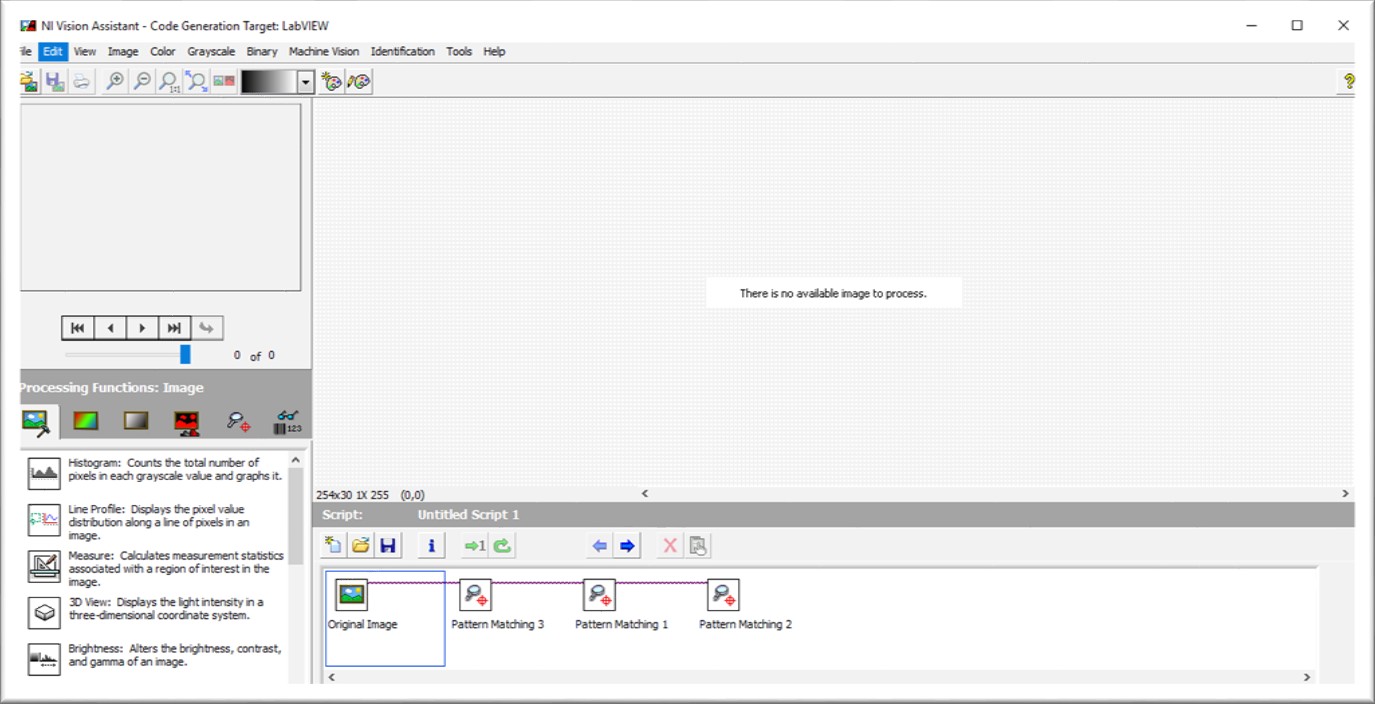


**FIGURE3.6: VISION ACQUISITION MODULES**

#### THE VISION ASSISTANT MODULE:

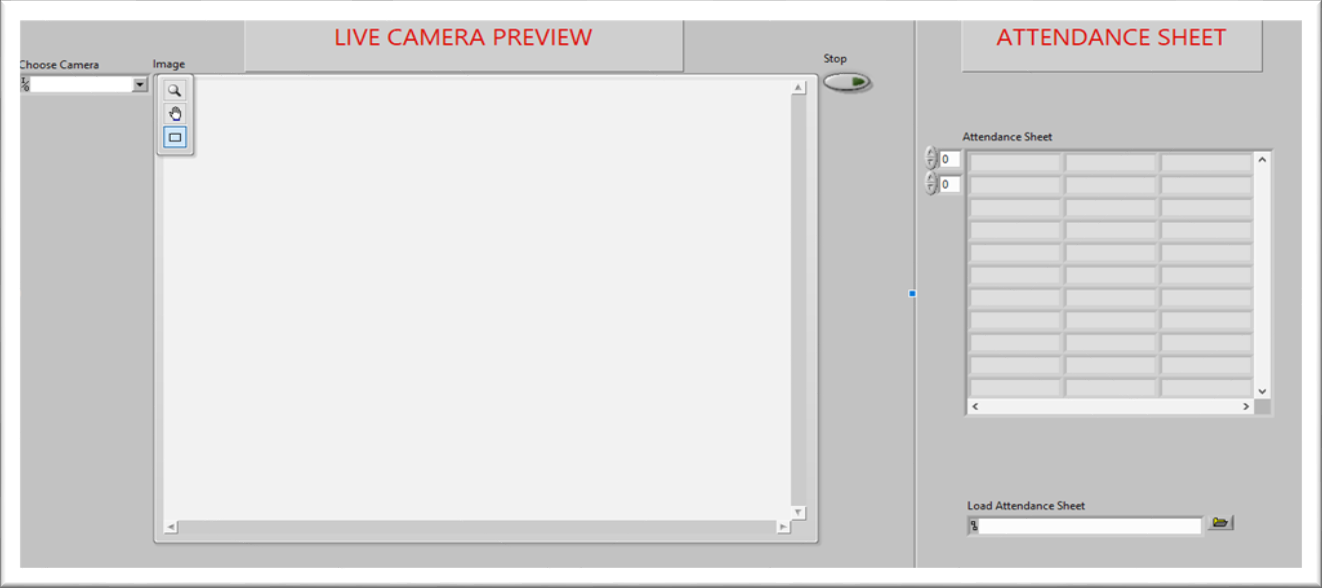


**FIGURE 3.7: VISION ASSISTANT MODULES**

This is the Vision Assistant where processing is performed. But it is the requirement of Vision Assistant to first include the picture of each student for matching. We can add as many patterns as we want.

**FIGURE3.8: INSIDE OF VISION ASSISTANT**

This is the front panel where the capture picture of the class is shown in image out. We have to add path for the save attendance file where it needs to be stored and also the path of the pattern of each student, three cases are generated. You can increase the pattern from Vision Assistant block as already stated above. The marked attendance is updated in the file and also shown in the indicator in front panel as shown in (Figure 3.9).



**FIGURE3.9: FRONT PANEL**

# System Testing and Analysis

## Subsystem 2: Vision Acquisition

##### Objective

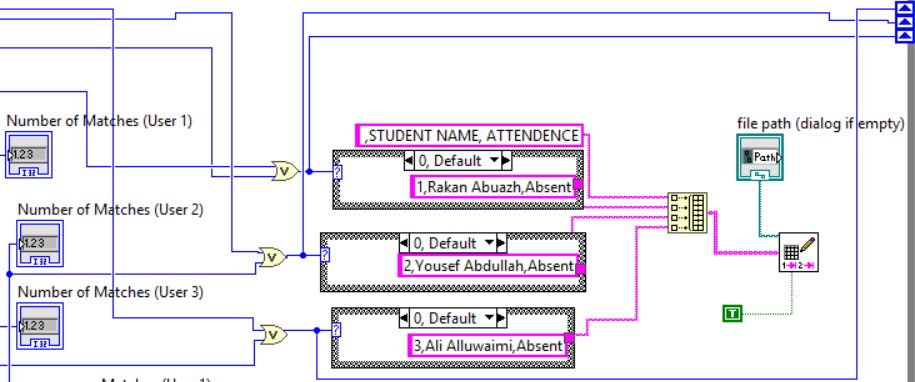
Verify that the camera can capture a live video feed and is able to convert image into grayscale.

##### Setup

We ran the main VI (Working Detection.vi) and checked whether the image of camera feed is being displayed in front panel or not.

##### Results

We were successfully able to capture live feed of video as shown in (Figure 4.1). Also, the image captured is in grayscale which passed our second test as well.



**FIGURE 4.2: CONDITIONS OF ATTENDANCE**

## Overall Results, Analysis and Discussion

##### Objective

Verify that the spreadsheet is updated to corresponding present as soon as image is detected. The “Present” should be marked against the person detected only.

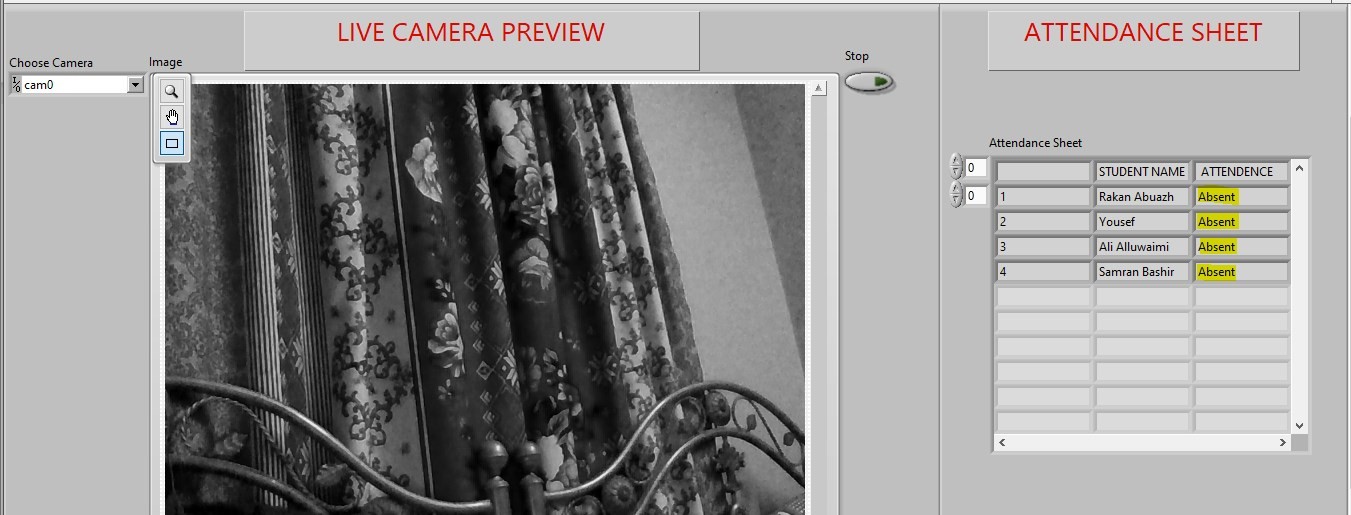
##### Setup

We ran the main VI (Working Detection.vi) and checked Attendance Sheet Section to change from “Absent” to “Present” on detection of person.

##### Results

We were successfully able to detect face and change the value of corresponding face from “Absent” Before to “Present” After as shown in (Figure 4.4).

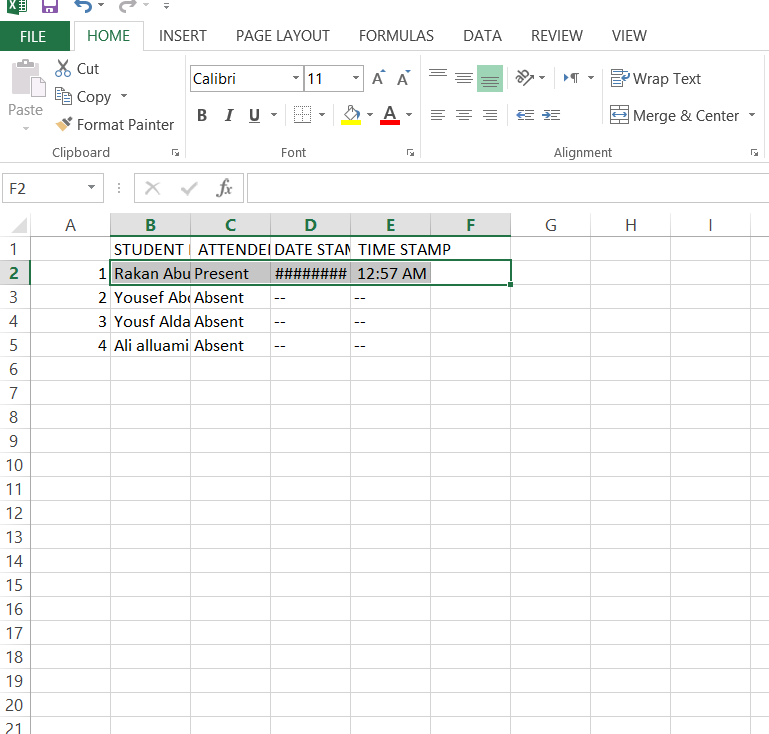
**BEFORE:**



**FIGURE 4.3: ATTENDANCE BEFORE FACE DETECTION**

#### AFTER:

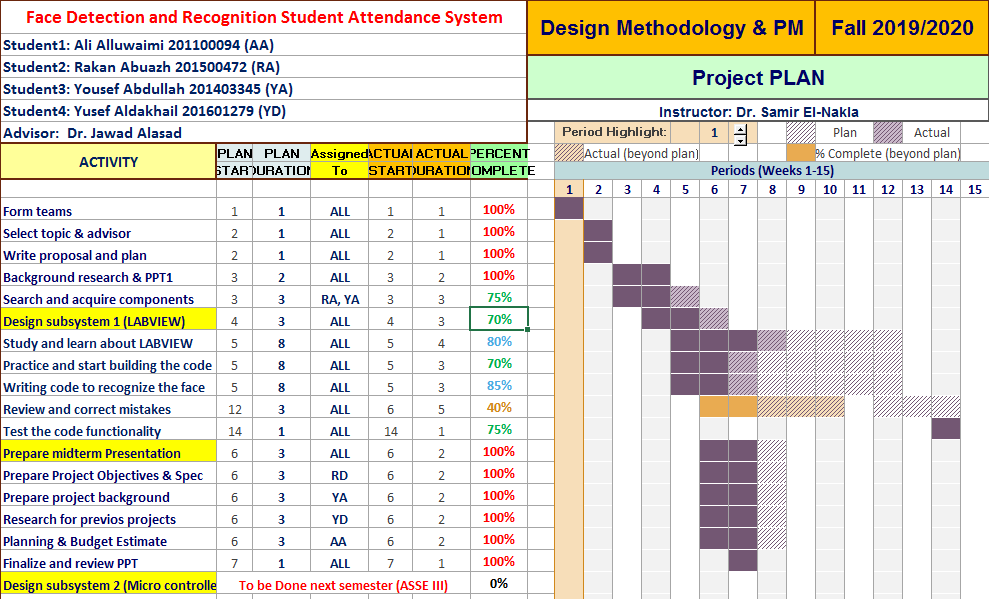
When we opened the attendance sheet on excel, we saw the file was updated correctly and completely as shown in (Figure 4.4).



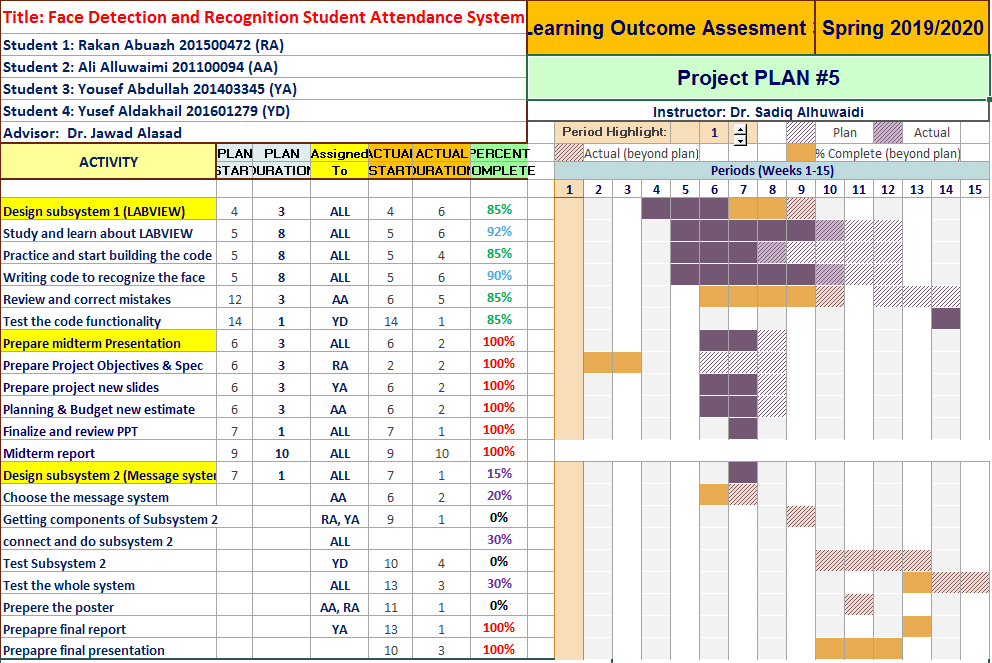
**FIGURE4.4: ATTENDANCE AFTER FACE DETECTION**

# Project Management

## Project Plan



**FIGURE 5.1: PROJECT PLAN FOR SEMESTER #1**



**FIGURE5.2: PROJECT PLAN FOR SEMESTER #2**

## Contribution of Team Members

As group, we managed to divide the work between us as shown in (Table 1). We were collaborating and meeting with each other before the COVID-19, however after the pandemic we were doing virtual meetings in order to try our best to work on the project despite what is happening.

**TABLE 1: CONTRIBUTION OF TEAM MEMBERS**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Task** | **Rakan** | **Qadir** | **Wazir** | **Task Total** |
| **Search & acquire components** | **30%** | **35%** | **35%** | **100%** |
| **Design**  **Subsystems** | **25%** | **30%** | **20%** | **100%** |
| **Test**  **Subsystems** | **30%** | **25%** | **15%** | **100%** |
| **Write**  **Reports & Presentations** | **25%** | **25%** | **20%** | **100%** |

## Project Execution Monitoring

##### Meeting With Advisor

Proper meeting with advisor was held biweekly to discuss the issue that became halt in the progress of the project. Advisor helped with the necessary solution for the required problem.

##### Team Meeting

All members discuss their respective progress in their part in daily meeting and prepare a final sheet for the meeting with the advisor.

##### Testing

After completion of each part mentioned in project plan, proofreading and testing was done for successful verification of the part.

##### Progress Discussion Meeting

An overall progress discussion meeting was held once a month where current standing of the project is presented to the advisor and everyone including members share their opinion and discuss them, and what amendments are necessary to be added.

## Limitation and Challenges

* **COVID-19** played a crucial role while we tried to merge the subsystems, however it prevented us to do so due to the quarantine. **COVID-19** did not allows us to meet as a group, and we were not able to get LABVIEW license from campus due to **COVID-19** pandemic.
* The identical twin issue. We were having a challenge of getting different recognitions for twins.
* Interfacing excel sheet with LABVIEW gave us hard time, since sometimes it worked and was synchronized with it and sometimes not.
* The first time setup for camera was difficult, but after guidance manual and some helped from the manufacture, successful camera setup was built.

will be easy to sort the attendance according to the student roll no. and name while updating the student report for mentioning his/her attendance record. With the help of simple processing, we can estimate the student attendance record with an algorithm and if their attendance is less than 75%, send a notice mail or message to their parents.

## 6.3 Contemporary Issues Addressed

Before face recognition was even possible and available for larger use by just state agencies and airport, people suggest that we all have distinct fingerprint and that same thought could be contributed to our facial features. With the help of face detection, we are moving towards safer network security for the society and it is the most effective way to protect your information. Face detection technology is vast and is now being used in every department and firm in Saudi Arabia.

# Conclusions and Future Recommendations

## Conclusions

Number of modules are available on LabVIEW to achieve incredible number of tasks. The best thing about LabVIEW is that you can view the flow of data from one block to other and have more freedom to make changes according to your requirements. The Automatic Class Attendance System implemented in this project would be much more difficult if it was not implemented on LabVIEW. The objective of class attendance system is to automate the time consuming and error prone attendance system.

There are always limitations of every system. One can only have fixed number of students and provide less freedom to have interclass attendance system. This means the attendance system for one class can’t be used for attendance system of other class. One must change programming to do this.

The Project experience was tremendous as we learned the core of vision algorithms and different programming techniques of LabVIEW. We learned how can a problem be simplified into smaller tasks and can be achieved successfully. It is the reason why we are able to complete our project in 1st of the two semesters.