15EC401M-MULTIDISPLINARY DESIGN

On

STUDENT ATTENDANCE MONITORING USING FACUAL RECOGNITION IN A CLASSROOM

Submitted for partial fulfilment for the award of the degree

BACHELOR OF TECHNOLOGY

In

ELECTRONICS AND COMMUNICATION ENGINEERING

Submitted By

K. CHANDRASEKARAN RA11711004040028
S. AATHITH RA1711004040030

Under the guidance of
MR. DINESH BABU A
(Assistant Professor, Department of Electronics and Communication Engineering)

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING
FACULTY OF ENGINEERING AND TECHNOLOGY SRM INSTITUTE OF SCIENCE & TECHNOLOGYY VADAPALANI CAMPUS, CHENNAI-600 026

MAY 2020

BONAFIDE CERTIFICATE

Certified that this project report titled "**STUDENT ATTENDANCE MONITORING USING FACIAL RECOGNITION IN A CLASSROOM**" is the bonafide work of **K. CHANDRASEKARAN - RA1711004040028** and **S.AATHITH - RA1711004040030** who carried out the project work under my supervision as a batch. Certified further, that to the best of my knowledge the work reported herein is not from any other project on the basis of which a degree or award was conferred on an earlier occasion for this or any other candidate.

In-charge	Head of the Department
Mr. A. Dinesh Babu	Dr. C. Gomathy,
Assistant Professor,	Professor and Head,
Department of ECE,	Department of ECE,
SRM IST, Vadapalani	SRM IST, Vadapalani

Date:

Submitted for University Examination held in MAY 2020 at SRM INSTITUTE OF SCIENCE AND TECHNOLOGY, Vadapalani, Chennai-600 026.

Date: Internal Examiner External Examiner

ACKNOWLEDGEMENT

This work would not have been completed without motivation and help from our **SRM IST MANAGEMENT**. We place our deep sense of gratitude to our beloved chancellor, **Dr. T.R.PACHAMUTHU, SRM IST** for providing us with the requisite infrastructure throughout the course, We would like to thank our respected **Dr. DURAIVELU**, Dean (E&T), SRM IST, Vadapalani rendering all facilities to do our project. We also convey our immense gratitude **to Dr. C.GOMATHY**, Head of the Department, Department of ECE for her constant support and encouragement rendered to us for completion of our project.

We would like to have a special mention of our class in-charge and guide **Mr. A.DINESH BABU**, Assistant professor, Department of ECE for his efficient and excellent guidance, inspiring discussions insightful and timely encouragements for the successful completion of our project.

We also extend our heartfelt thanks to all the teaching and non-teaching staff members of ECE Department who have rendered their valuable help in making for successful completion of the project. We could not have finished this work without support from our family and friends. They are dearest in our heart and a great source of motivation. Our hearty appreciation and thanks to our friends whose support and encouragement made all this possible.

INDEX

LIST OF FIGURES

LIST OF ABBREVIATIONS

ACRONYM	ABBREVIATIONS
AI	Artificial Intelligence
SOC	System on Chips
ML	Machine Learning
CV	Computer Vision

ABSTRACT

Student’s attendance in the classroom is a very an important task. The existing method of taking attendance consumes a finite of amount of time which can be used for other productive action. We found this problem in many universities and schools. We want a reliable system which identifies which student is present in the classroom and gives the attendance based on his presence. We found that facial recognition overcomes some of the problems that exist in the attendance systems like Fingerprint, RFID or IRIS based recognition. Face recognition-based attendance system is a process of recognizing the students face for taking attendance by using face biometrics based on high - definition monitor video and other information technology. In my face recognition project, a computer system will be able to find and recognize human faces fast and precisely in images or videos that are being captured through a surveillance camera. Numerous algorithms and techniques have been developed for improving the performance of face recognition but the concept to be implemented here is Deep Learning. The user-friendly graphical user interface provides flexibility and ease in running these powerful face recognition algorithms powered by deep learning.

Chapter 1

Introduction

1.1 Computer Vision

Computer vision is an interdisciplinary field that deals with how computers can be made to gain 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. "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. 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. As a technological discipline, computer vision seeks to apply its theories and models for the construction of computer vision systems.

Computer vision is an interdisciplinary scientific field that deals with how computers can gain high-level understanding from digital images or videos. From the perspective of engineering, it seeks to understand and automate tasks that the human visual system can do.

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. Understanding in this context means the transformation of visual images (the input of the retina) into descriptions of the world that make sense to thought processes and can 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.

The scientific discipline of 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, multi-dimensional data from a 3D scanner, or medical scanning device. The technological discipline of computer vision seeks to apply its theories and models to the construction of computer vision systems.

Sub-domains of computer vision include scene reconstruction, event detection, video tracking, object recognition, 3D pose estimation, learning, indexing, motion estimation, visual serving, 3D scene modeling, and image restoration.

1.2 Digital image processing

An image may be defined as a two-dimensional function, f (x, y), where x and y are spatial coordinates and the amplitude of "f" at any pair of coordinates (x, y) is called the intensity or gray level of the image at that point. When x, y, and gray level of "f" are all finite discrete quantities, the image is then digital image.

The processing of digital images by means of a digital computer is called digital image processing. Digital image is composed of a finite number of elements, each of which has a particular value and location, called as pixels. The different stages for image processing are low-level, mid-level and higher-level processing. It contains preprocessing to reduce noise, contrast enhancement, image sharpening, segmentation, classification (recognition) and finally ―making sense of recognized objects performing the cognitive functions normally associated with vision.

Today, there is almost no area of technical that are not impacted in some way by digital image processing such as in medical space and day to day life. Images based on radiation from the electromagnetic spectrum are most familiar such as gamma ray, X-ray, microwave imaging. Image compression, facial recognition iris detection and finger print detection are more common as security point of view.

With the rapid development of mathematical theory on multivariate statistics and multi-media technology especially image processing, facial expressions recognition researchers have achieved many useful results.

The face recognition algorithm needs to compare a given face with a given template and verify their equivalence. Such a setup (one-to-one matching) can occur when biometric technology is used to secure financial transactions, for example, in an automatic teller machine (ATM).

1.3 Machine learning/ Deep learning

Deep learning is a machine learning technique that teaches computers to do what comes naturally to humans: learn by example. Deep learning is a key technology behind driverless cars, enabling them to recognize a stop sign, or to distinguish a pedestrian from a lamppost. It is the key to voice control in consumer devices like phones, tablets, TVs, and hands-free speakers. Deep learning is getting lots of attention lately and for good reason. It's achieving results that were not possible before.

In deep learning, a computer model learns to perform classification tasks directly from images, text, or sound. Deep learning models can achieve state-of-the-art accuracy, sometimes exceeding human-level performance. Models are trained by using a large set of labeled data and neural network architectures that contain many layers.

Most deep learning methods use neural network architectures, which is why deep learning models are often referred to as deep neural networks. The term "deep" usually refers to the number of hidden layers in the neural network. Traditional neural networks only contain 2-3 hidden layers, while deep networks can have as many as 150.

Deep learning models are trained by using large sets of labeled data and neural network architectures that learn features directly from the data without the need for manual feature extraction.

A neural network is a sort of computer software, inspired by biological neurons. Similarly, a neural network is made up of cells that work together to produce a desired result, although each individual cell is only responsible for solving a small part of the problem.

The "signal" at a connection is a real number, and the output of each neuron is computed by some non-linear function of the sum of its inputs. The connections are called edges. Neurons and edges typically have a weight that adjusts as learning proceeds. The weight increases or decreases the strength of the signal at a connection. Neurons may have a threshold such that a signal is sent only if the aggregate signal crosses that threshold. Typically, neurons are aggregated into layers. Different layers may perform different transformations on their inputs. Signals travel from the first layer (the input layer), to the last layer (the output layer), possibly after traversing the layers multiple times.

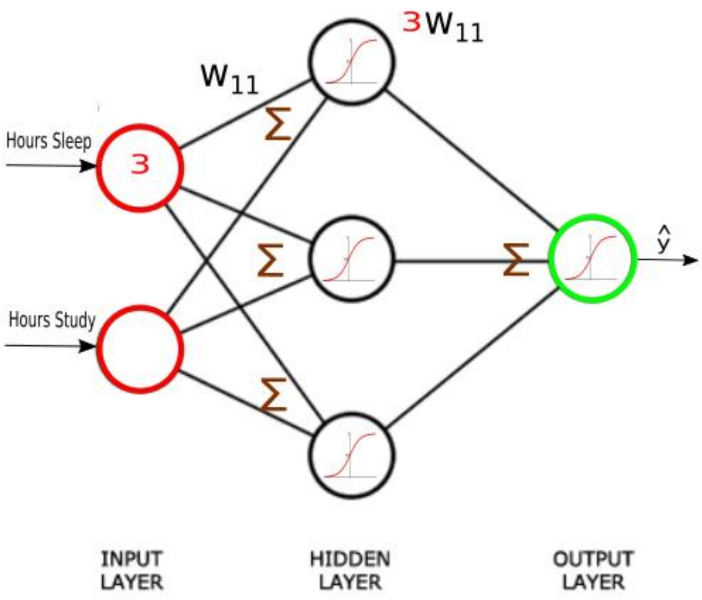

Fig 1: A representation of neural network

One of the most popular types of deep neural networks is known as convolutional neural networks (CNN or ConvNet). A CNN convolves learned features with input data, and uses 2D convolutional layers, making this architecture well suited to processing 2D data, such as images.

CNNs eliminate the need for manual feature extraction, so you do not need to identify features used to classify images. The CNN works by extracting features directly from images. The relevant features are not pretrained; they are learned while the network trains on a collection of images. This automated feature extraction makes deep learning models highly accurate for computer vision tasks such as object classification.

1.4 CHAPTER ORGANISATION

The rest of this project report is structured as follows. chapter 2 deals with the literature survey. Next, chapter 3 gives information about the sub systems present in the overall design. Chapter 4 covers the mathematical modelling of some of the systems explained in chapter 3. Objectives of the project is given in the chapter 5; this makes sure that we are on the correct track. Chapter 6 is about the constraints of the project. The implementation specific details are provided in the chapter 7. The discussion about the future works in chapter 8.

Chapter 2

Literature Survey

Existing Recognition systems

There are now plenty of systems which differ in many aspects: core technology they are based on, way of use, cost, reliability, security and etc. Many of those depend on students having to carry specific identification devices. One of the common types of the attendance systems is Radio Frequency Identification (RFID) where students have to carry appropriate RFID cards. There are also location-based attendance tracking systems. The location of an employee can be determined via Global Positioning System (GPS). The presence is determined by calculating the proximity between a student's and the school's location. Both of the above-mentioned types of the attendance systems have weaknesses. Students could forget the RFID card or the location device, or someone else could check instead of them. This could also be a potential security issue. Therefore, there are systems that exclude the usage of external devices for attendance purposes by exploiting the individual attributes: fingerprints, iris, voice, face and etc. These types of systems are heavily based on computer vision and machine learning algorithms. Recent advances in these areas, especially in deep learning, provide possibilities to use these methods searching for practical solutions. These solutions could be more flexible and could reduce human errors.

Facial Recognition

Face Recognition algorithms have many limitations in a real-world scenario like lighting conditions and proxy via pictures and low-quality image processing. In this era of Machine learning and Artificial Intelligence computers have evolved to process huge chunks of data at very high rates with maximum process efficiency and image processing and computer vision powered by Deep learning algorithms have geared up to make face recognition a reality, solving or reducing most of the previous limitations. Vision is a fully isolated system setup on a Raspberry Pi and is modular. The system can be controlled via a network communication to the PI server and also, all the reports and well maintained in the attendance server onboard and online. This module is equipped with a user-friendly GUI (Graphical User Interface) for ease of use. The GUI offer various options for the administrator to operate the system like choose between Realtime surveillance or single snap mode, to train and update the classifier with new face data Fig.4. The backbone of the software is driven by python code with Open-source computer vision library and Tkinter library for Graphical user interface.

Chapter 3

Sub Systems

Our proposed solution should have a camera and it should a facial recognition algorithm. Keeping in mind these 2 points, we designed the major sub systems in our solution. To control and interface the camera we needed a system and to run the facial recognition we needed another system. So, 2 major sub systems were designed to addressing these 2 specifications.

The major sub systems involved in our solutions include “System to take photo” and “System that identifies the students in the photo”.

We wanted these 2 systems to be separate as when we design them keeping in mind, they are separate and can be interconnected by another system.

2.1 **System that takes the photo**

The system that takes photo, it is evident from the name itself what action is it going to perform. It contains the components such as

- camera,
- a system to signal the camera
- a system to transfer the taken photo.

We kept another system to explicitly signal the camera and there is a need for it. As one of our motivation is that we wanted to give attendance to those students who are present in the class for its full duration. To tackle this problem, we have simply divided the whole class duration into separate chunks of 5-minute duration. So, every 5 minutes the camera is signalled to take photo and the photo taken is then sent to the next major sub system for facial recognition.

2.1.1 Camera

A camera is an optical instrument used to record images. At their most basic, cameras are sealed boxes (the camera body) with a small hole (the aperture) that allow light in to capture an image on a light-sensitive surface (usually photographic film or a digital sensor). Now days everybody is familiar with the camera thanks to the smartphone revolution. The camera is going to be the important part and the decision of choosing which camera to use if very important. One of the obvious parameters of the camera which is wide known is the megapixels of the camera. A megapixel (MP) is a million pixels; the term is used not only for the number of pixels in an image but also to express the number of image sensor elements of digital cameras or the number of display elements of digital displays. For example, a camera that makes a 2048 × 1536-pixel image (3,145,728 finished image pixels) typically uses a few extra rows and columns of sensor elements and is commonly said to have "3.2 megapixels" or "3.4 megapixels", depending on whether the number reported is the "effective" or the "total" pixel count. Sensor Size – If you had to guess camera quality based only on one parameter, size of the sensor, will be your best bet for most accurate response. The sensor size determines how much light it can use to create the image. This is also a reason why DSLR with same MP count often perform better than Smartphone Rear cameras.

2.1.2 A system to signal the camera

This system actually takes care of signalling the camera. Now what is the need for explicitly signalling the camera. Now lets us first consider another scenario where we get all the images from the camera and we will run the face recognition algorithm for all the images. For example, let’s say our camera outputs 60 frames per second, so that means just for one class we will get 60 * 60 which is astonishing 3600 frames per second and each image has to transported and processed. Again, the complexity increases when we consider multiple cameras for better detection inside the classroom. This will again increase the burden on the facial recognition sub system this will force us to use a high-performance device for facial recognition. This invariably increase the cost and complexity of the system that we proposed. To implement this, we can use a microcontroller that is interfaced with the camera, the micro controller can be programmed to signal the camera every 5 minutes. Now the reason for 5 minutes is very simple, we already specified that every 5 minutes the camera takes photo so every 5 minutes the micro controller signals the camera and the camera takes the picture.

A microcontroller (MCU for microcontroller unit) is a small computer on a single metal-oxide-semiconductor (MOS) integrated circuit (IC) chip. A microcontroller contains one or more CPUs (processor cores) along with memory and programmable input/output peripherals. Program memory in the form of ferroelectric RAM, NOR flash or OTP ROM is also often included on chip, as well as a small amount of RAM. Microcontrollers are designed for embedded applications, in contrast to the microprocessors used in personal computers or other general-purpose applications consisting of various discrete chips.

In modern terminology, a microcontroller is similar to, but less sophisticated than, a system on a chip (SoC). SoC may include a microcontroller as one of its components, but usually integrates it with advanced peripherals like graphics processing unit (GPU), Wi-Fi module, or one or more coprocessors. Microcontrollers are used in automatically controlled products and devices, such as automobile engine control systems, implantable medical devices, remote controls, office machines, appliances, power tools, toys and other embedded systems.

By reducing the size and cost compared to a design that uses a separate microprocessor, memory, and input/output devices, microcontrollers make it economical to digitally control even more devices and processes. Mixed signal microcontrollers are common, integrating analog components needed to control non-digital electronic systems. In the context of the internet of things, microcontrollers are an economical and popular means of data collection, sensing and actuating the physical world as edge devices. Some microcontrollers may use four-bit words and operate at frequencies as low as 4 kHz for low power consumption (single-digit milliwatts or microwatts). They generally have the ability to retain functionality while waiting for an event such as a button press or other interrupt; power consumption while sleeping (CPU clock and most peripherals off) may be just nanowatts, making many of them well suited for long lasting battery applications.

2.1.3 A system to transfer the taken photo

This part can also be programmed in the same micro controller that was discussed before for signalling the camera. Now here the main aim is to transfer the photo that was taken by the camera after it was signalled by the specified micro controller. The microcontroller should be interfaced with a transmitter so that it can send the photo inside a channel. Now the choice transmitter and the interfacing of Arduino with the transmitter all depends on which channel we select. These are the following channels that are popularly used.

2.1.3.1 Short-range wireless

Bluetooth mesh networking – Specification providing a mesh networking variant to Bluetooth low energy (BLE) with increased number of nodes and standardized application layer (Models).

Light-Fidelity (Li-Fi) – Wireless communication technology similar to the Wi-Fi standard, but using visible light communication for increased bandwidth.

Near-field communication (NFC) – Communication protocols enabling two electronic devices to communicate within a 4 cm range.

Radio-frequency identification (RFID) – Technology using electromagnetic fields to read data stored in tags embedded in other items.

Wi-Fi – Technology for local area networking based on the IEEE 802.11 standard, where devices may communicate through a shared access point or directly between individual devices.

ZigBee – Communication protocols for personal area networking based on the IEEE 802.15.4 standard, providing low power consumption, low data rate, low cost, and high throughput.

Z-Wave – Wireless communications protocol used primarily for home automation and security applications.

2.1.3.2 Medium-range wireless

LTE-Advanced – High-speed communication specification for mobile networks. Provides enhancements to the LTE standard with extended coverage, higher throughput, and lower latency.

5G - 5G wireless networks can be used to achieve the high communication requirements of the IoT and connect a large number of IoT devices, even when they are on the move.

2.1.3.3 Long-range wireless

Low-power wide-area networking (LPWAN) – Wireless networks designed to allow long-range communication at a low data rate, reducing power and cost for transmission. Available LPWAN technologies and protocols: LoRaWan, Sigfox, NB-IoT, Weightless, RPMA.

Very small aperture terminal (VSAT) – Satellite communication technology using small dish antennas for narrowband and broadband data.

2.1.3.4 Wired

Ethernet – General purpose networking standard using twisted pair and fibre optic links in conjunction with hubs or switches.

Power-line communication (PLC) – Communication technology using electrical wiring to carry power and data. Specifications such as Home Plug or G.hn utilize PLC for networking IoT devices.

After choosing the channel, we will go on to select the transmitter that is appropriate for the channel so that we can send information properly on the selected channel. The choice of which channel it all depends on how are we going to implement the two sub systems and the distance between them.

2.2 System that marks the attendance

Coming on to the next major sub system, this is the part that is going to do the facial recognition. This contains a lot of sub system as it is going to do the important part of identifying somebody from the photo. The flow of operation is as follows

- receive the photo
- identify the faces in the photo,
- For each face identified give the unique descriptor of that face
- using the descriptor marking the attendance

A short explanation about unique descriptor is that it can be considered as a unique number generated for unique face that the machine can understand. Ideally the facial recognition algorithm should output the same number if you input the same face even if is the same person in a different outfit or in a different makeup, i.e. even when the physical characteristics changes. To be more technical we can divide this sub system into a number of components as follows, these components are divided in the same way as the flow is described above.

- System that receives the photo

- System that runs the face detection algorithm that outputs “m” detected faces
- System that outputs a descriptor for “m” faces
- System that iterates through “m” descriptor and marks the attendance

2.2.1 System that receives the photo

The action that it is going to perform is already prevalent from its name. It actually is responsible for receiving the photo that is transmitted from the first major sub system that was present in the class

2.2.2 System that runs the face detection algorithm that outputs “m” detected faces

System that runs the face detection algorithm that outputs “m” detected faces. The face detection algorithm, finds the faces in the photos, essentially to visualize we can draw bounding boxes that contain the face. The face detection algorithm outputs a list of co-ordinates with which we can draw the bounding box

2.2.3 System that outputs a descriptor for “m” faces

“The system that outputs a descriptor for all the faces in the photo”. After we find “m” faces in photo we are going to iterate through each individual face. Now each face is fed into the facial recognition algorithm that gives us the unique descriptor for each face so at last we will get “m” descriptors for a photo and we will store that.

This can be realized for example using a Convolutional neural network which can be trained for outputting a different descriptor for different unique images fed as the input.

2.2.4 System that iterates through “m” descriptor and marks the attendance

“The system that iterates “m” descriptors and marks the attendance”. We are having “m” descriptors that the machine can understand, now how to mark the attendance. Well at first think we are having a data base that contains the student’s register number and the descriptor of the student’s face stored. Then our work becomes easy just, iterate through the “m” descriptors and check whether they are in the database of student, if it is available mark the attendance against that student.

Chapter 4

Mathematical Modelling

In this section we will see the mathematical modelling of some of the previously mentioned systems in the “Subsystems” chapter.

3.1 System that signals the camera

This system is one of the components of the major sub system “System that takes photo”, this system should signal the camera every 5 minutes, a mathematical mode of this system is given below

```
output = signal to camera
begin:
    if (5 minutes elapsed):
        signal to camera
        recieve the photo
        store it
        hand over to next sys
```

Fig 2: Signaling the camera

3.2 System that outputs a descriptor for one face

This system is one of the components of the major sub system “System that marks the attendance”, this system is actually the main part, which actually performs the face recognition it is the algorithm.

The facial recognition algorithm that takes the input as an image containing one face and then gives the descriptor of that face. We have already given a basic concept of what is descriptor and what is its function. To be more precise the algorithm actually outputs a vector of fixed size that contains floating numbers. The values of the number changes for different faces.

A mathematical mode of this system is given below

```
input = photo-of-face
output = vector(size = d x 1)
begin:
    pre_processing(photo-of-face)
    vector = CNN(photo-of-face-processed)
end
```

Fig 3: Facial Recognition

3.3 System that checks whether two descriptors are same

This system belongs to the major sub system "System that marks the attendance", essentially this system compares whether two faces are same, it is kind of what when we do when we are identifying an unknown person in a crowd with the help of a photo graph of the unknown photo graph.

In the explanation of "system that takes m descriptor and marks the attendance" there was mention about data base and how we are checking the descriptors whether they are equal to mark the attendance.

But as algorithm here does not return a number but a vector or simply an array of values. To compare vectors for the similarity we need to use cosine similarity. Essentially what we are doing is telling how much similar these 2 faces are, one that we take in the class, one that was previously taken.

```
n[i] descriptor of size (1 x d) from the database
m[j] descriptor of size (d x 1) from the detected face

input = n[i] and m[j]
output = True or False
begin:
    M = m[j]
    N = n[i]
    theta = MN #cosine similarity
    if (theta > 0.5):
        output = True
    else
        output = False
end
```

Fig 4: Cosine similarity

Chapter 5

Objectives

We have 5 objectives broadly which are explained below

5.1 "1st Objective"

One camera and one person to stand before the camera. Detect the face in the image and run the face recognition.

5.2 "2nd Objective"

One person and a group of people to stand before the camera. Run a face detection model to find the faces in the picture and face recognition algorithm against it.

5.3 "3rd Objective"

Try different face detection algorithms. Select the algorithm which has lowest false negatives. Note down other metrics like the least dimension of the facial image that is required for the face detection algorithm.

5.4 "4th Objective"

Design the camera position, the angle etc. The algorithm that was decided on 3rd objective to be used here to test the camera angle and other metrics like finding how many persons can be covered.

5.5 "5th objective"

Find the best camera angle and algorithm (both face detection and face recognition) that has reduced false negatives, is fast etc. Iterate through 3 and 4 objectives to find the best combination of camera and algorithm

Chapter 6

Constraints

We have constraints for Camera and Facial detection & Recognition algorithms.

6.1 Camera

Proper lighting should be there. The position of camera should not be changed as it is fixed (depends on the algorithm). As when the distance increases the image resolution of detected face decreases.

6.2 Facial Detection & Recognition

The face or faces should not be covered. There is a maximum distance till which a student can sit for the camera to detect the face beyond that facial recognition won't work. When the photo is taken if he/she covers their face then the algorithm can miss the attendance of the person in that time frame. If physical characteristics of people change over time – If they gain weight, loss their hair, grow a beard or start wearing glasses, it is tough for face recognition.

6.3 Overcoming Constraints

We can overcome the constraints related to camera by ensuring proper environment of the classroom. The low-resolution images can be fed to a super resolution algorithm which then increases the resolution thereby increasing the quality of the image. We can update the database containing the descriptors for each student as the days go by, so the change in the face is accounted in the database. By using this approach, we can overcome the complications present in identifying faces of changed physical characteristics.

Chapter 7

Implementation

7. 1 Comparison of different face detection algorithm

This is for deciding which algorithm to use

7.1.1 Haar Cascade Face Detector in OpenCV

Haar Cascade based Face Detector was the state-of-the-art in Face Detection for many years since 2001, when it was introduced by Viola and Jones. There have been many improvements in the recent years.

Pros

- Works almost real-time on CPU
- Simple Architecture
- Detects faces at different scales

Cons

- The major drawback of this method is that it gives a lot of False predictions.
- Doesn’t work on non-frontal images.
- Doesn’t work under occlusion

7.1.2 HoG Face Detector in Dlib

This is a widely used face detection model, based on HoG features and SVM. You can read more about HoG in our post. The model is built out of 5 HOG filters – front looking, left looking, right looking, front looking but rotated left, and a front looking but rotated right. The model comes embedded in the header file itself.

The dataset used for training, consists of 2825 images which are obtained from LFW dataset and manually annotated by Davis King, the author of Dlib.

Pros

- Fastest method on CPU
- Works very well for frontal and slightly non-frontal faces
- Light-weight model as compared to the other three.
- Works under small occlusion

Cons

- The major drawback is that it does not detect small faces as it is trained for minimum face size of 80×80. Thus, you need to make sure that the face size should be more than that in your application. You can however, train your own face detector for smaller sized faces.
- The bounding box often excludes part of forehead and even part of chin sometimes.
- Does not work very well under substantial occlusion
- Does not work for side face and extreme non-frontal faces, like looking down or up.

7.1.2 DNN Face Detector in OpenCV

This model was included in OpenCV from version 3.3. It is based on Single-Shot-Multibox detector and uses ResNet-10 Architecture as backbone. The model was trained using images available from the web, but the source is not disclosed. OpenCV provides 2 models for this face detector.

1) Floating point 16 version of the original caffe implementation (5.4 MB)
2) 8-bit quantized version using TensorFlow (2.7 MB)

The method has the following merits:

Most accurate out of the four methods

- Runs at real-time on CPU
- Works for different face orientations – up, down, left, right, side-face etc.
- Works even under substantial occlusion
- Detects faces across various scales (detects big as well as tiny faces)

This is the slowest among the face detectors

7.1.3 CNN Face Detector in Dlib

This method uses a Maximum-Margin Object Detector (MMOD) with CNN based features. The training process for this method is very simple and you don’t need a large amount of data to train a custom object detector.

Pros

- Works for different face orientations
- Robust to occlusion
- Works very fast on GPU
- Very easy training process

Cons

- Very slow on CPU
- Does not detect small faces as it is trained for minimum face size of 80×80. Thus, we need to make sure that the face size should be more than that in our application.

For medium to large image sizes - Dlib HoG is the fastest method on CPU. But it does not detect small sized faces (< 70x70). So, if you know that your application will not be dealing with very small sized faces (for example a selfie app), then HoG based Face detector is a better option. Also, if you can use a GPU, then MMOD face detector is the best option as it is very fast on GPU and also provides detection at various angles.

We have stated the comparison of face detection algorithms as we can see that the first 2 algorithms are the most popular in case of algorithms that does not use CNN.

The Face detection algorithm using CNN actually gives the advantage over these algorithms but at the cost of performance. Based on our environment considerations and our implementation we have decided to not to use CNN based face detection.

7.2 Raspberry Pi 4

Raspberry Pi 4 Model B was released in June 2019 with a 1.5 GHz 64-bit quad core ARM Cortex-A72 processor, on-board 802.11ac Wi-Fi, Bluetooth 5, full gigabit Ethernet (throughput not limited), two USB 2.0 ports, two USB 3.0 ports, and dual-monitor support via a pair of micro-HDMI (HDMI Type D) ports for up to 4K resolution. The Pi 4 is also powered via a USB-C port, enabling additional power to be provided to downstream peripherals, when used with an appropriate PSU. The Raspberry Pi 4 uses a Broadcom BCM2711 SoC with a 1.5 GHz 64-bit quad-core ARM Cortex-A72 processor, with 1 MiB shared L2 cache. Unlike previous models, which all used a custom interrupt controller poorly suited for virtualization, the interrupt controller on this SoC is compatible with the ARM Generic Interrupt Controller (GIC) architecture 2.0, providing hardware support for interrupt distribution when using ARM virtualization capabilities, The Raspberry Pi 4, with a quad-core ARM Cortex-A72 processor, is described as having three times the performance of a Raspberry Pi 3. The Raspberry Pi 4 is available with 2, 4 or 8 GiB of RAM

7.3 Raspberry pi 8 MP camera

The 8MP Raspberry Pi Camera Module v2 can be used to take high-definition video, as well as stills photographs. It uses high quality 8-megapixel Sony IMX219 image sensor custom designed add-on board for Raspberry Pi, featuring a fixed focus lens. It is capable of 3280 x 2464-pixel static images, and also supports 1080p30, 720p60 and 640x480p60/90 video. It attaches to Pi by way of one of the small sockets on the board upper surface and uses the dedicated CSI interface, designed especially for interfacing to cameras. It is suitable for mobile or other applications where size and weight are important. It connects to Raspberry Pi by way of a short ribbon cable. It’s easy to use for beginners. The camera works with all models of Raspberry Pi 1, 2, and 3. It can be accessed through the MMAL and V4L APIs, and there are numerous third-party libraries built for it, including the Picamera Python library. Specifications of 8MP Raspberry Pi Camera V2: 8-megapixel camera capable of taking photographs of 3280x2464 pixels. Capture video at 1080p30, 720p60 and 640x480p90 resolutions. All software is supported within the latest version of Raspbian Operating System. Supports Raspberry Pi 1,2 and Applications: CCTV security camera, motion detection, time lapse photography.

7.4 Block Diagram

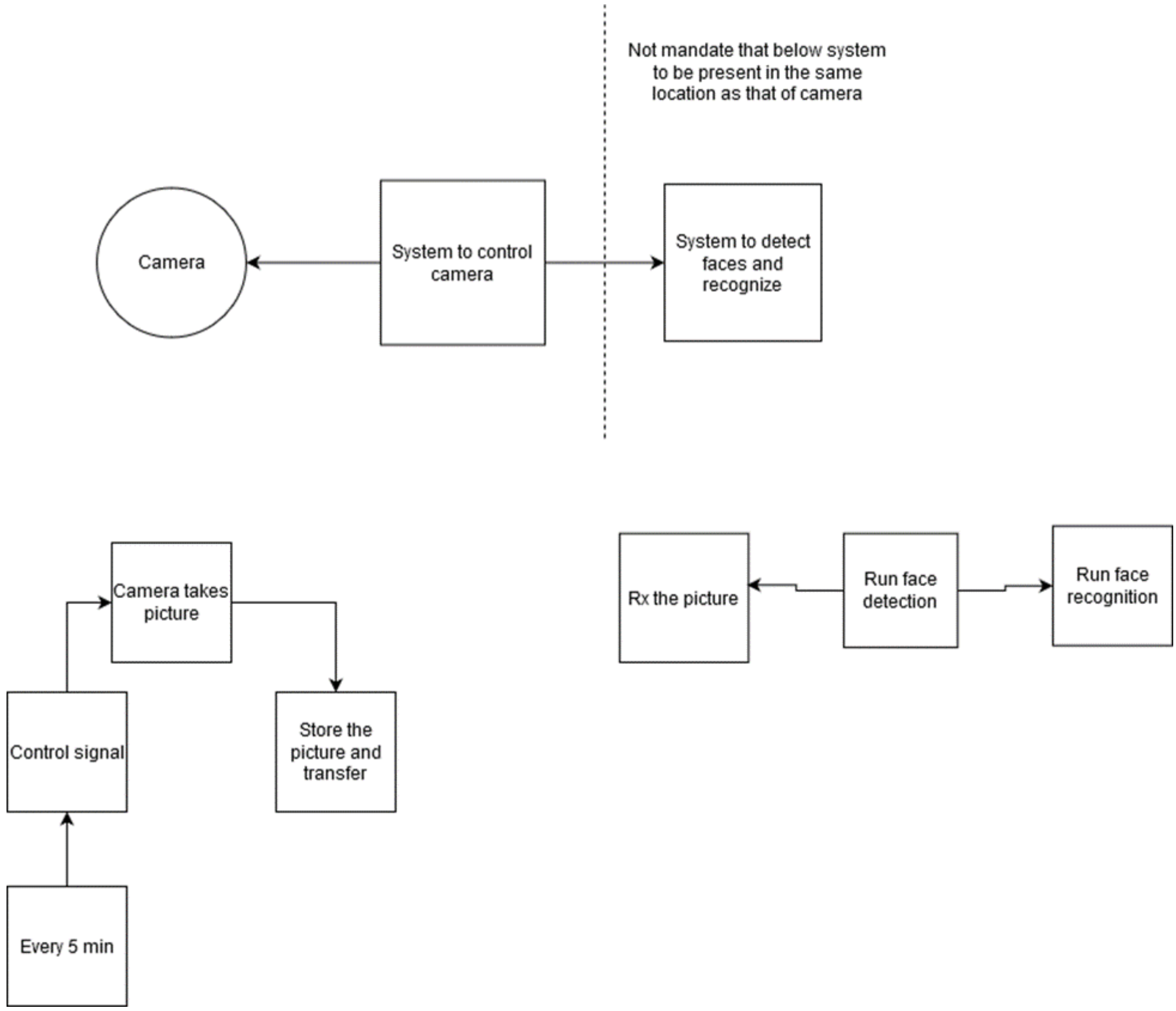

Fig 5: Block diagram

This is the block diagram that represents the connection between various sub systems that was discussed in the previous chapters.

Chapter 8

Future Works

In the future we can implement a separate system for running face recognition, connecting the system in class with the other system wirelessly. Using lidar we can try to improve the facial recognition algorithm which depends on 2D image. An IR lidar will enable us to detect face even in darkness. Also, we will try to integrate surveillance into this system.

REFERENCES

1)C. Ding and D. Tao. Trunk-branch ensemble convolutional neural networks for video-based face recognition. IEEE Trans Pattern Anal Mach Intell, 40(4):1002–1014, apr 2018.

2)Y. Chen, Y. Tai, X. Liu, C. Shen, and J. Yang. FSRNet: End-to-end learning face super-resolution with facial priors. In 2018 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2018.

3) Deepan.P1, Raja Vignesh.R2, Venkateswaran.S3, "Face Recognition based Automated Attendence Management System using Hybrid Classifier", International Research Jouranal of Engineering and Technology(IRJET) e-ISSN: 2395-0056, vol.4 Issue: 5, p-ISSN: 2395-0072, May 2017

4) Qingdong Liang, Wenting Fang, "College Student Attendance System based on Face Recognition", IOP Conf. Ser.: Mater. Sci Eng. 466 012107, 2018

5) Li Cuimei, Qi Zhiliang, Jia Nan, Wu Jianhua, "Human face detection algorithm via Haar classifier combined with three additional classifiers",13th IEEE International Conference on Electronics Measurement & Instruments (ICEMI), 2017

6) Ankita Digambar Patil, Dr. S. M. Mukane, Aishwarya Vijaykumar Patil, Simran Raju Inamdar, " Automatic Attendance system using Webcam", et.al; International Journal of Advance Research, Idea and Innovations in Technology", ISSN: 2454- 132X Impact factor: 4.295, vol. 4, Issue 2

7) Megalingam, R.K. Nair, R.N. Prakhya, "Automated voice based Home navigation system for the elderly and physically challenged", S.M.2011 2nd International Conference on Wireless Communication, Vehicular Technology, Information Theory and Aerospace and Electronics Systems Technology, Wireless VITAE 2011 5940832