FACE TRACKING WITH OPENCV

CONTENTS

- INTRODUCTION
- OBJECTIVE
- METHODOLOGY
- COMPONENTS
- APPLICATION
- CHALLENGES
- FUTURE SCOPE
- CONCLUSION
- REFERENCE

INTRODUCTION

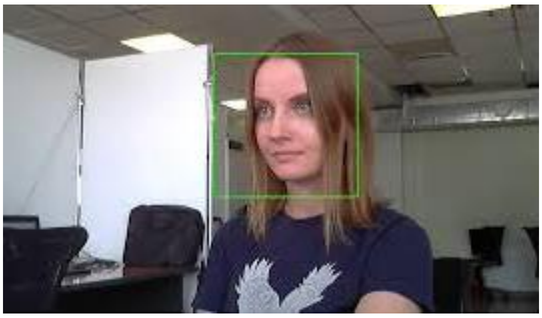


Image processing is a method to perform some operations on an image, in order to get an enhanced image or to extract some useful information from it. It is a type of signal processing in which input is an image and output may be image or characteristics/features associated with that image. Nowadays, image processing is among rapidly growing technologies. It forms core research area within engineering and computer science disciplines too.

Image processing basically includes the following three steps:

- Importing the image via image acquisition tools;
- Analysing and manipulating the image;
- Output in which result can be altered image or report that is based on image analysis.

There are two types of methods used for image processing namely, analogue and digital image processing. Analogue image processing can be used for the hard copies like printouts and photographs. Image analysts use various fundamentals of interpretation while using these visual techniques. Digital image processing techniques help in manipulation of the digital images by using computers. The three general phases that all types of data have to undergo while using digital technique are pre-processing, enhancement, and display, information extraction.

Face detection is a computer vision technology that helps to locate/visualize human faces in digital images. This technique is a specific use case of object detection technology that deals with detecting instances of semantic objects of a certain class (such as humans, buildings or cars) in digital images and videos. With the advent of technology, face detection has gained a lot of importance especially in fields like photography, security, and marketing.

OBJECTIVE

- To use the face detection and tracking system based on open source platforms such as Arduino and OpenCV to detect faces.
- To use the webcam used here to send video frames to IDE running on a PC.
- Face detection algorithm written in OpenCV library detects a face, calculates its center's X, Y coordinates and tracks motion of face.
- To send the coordinates to the COM port which in turn is connected to Arduino board via a USB connection.
- The Arduino then controls the movement of the webcam with the help of two pan/tilt servos to follow the detected face.
- The servos will move until the face is centered on the screen. The servos will only move when a face is detected if it doesn’t the servos will stop where they are.

METHODOLOGY

An image is nothing but a standard NumPy array containing pixels of data points. More the number of pixels in an image, the better is its resolution. You can think of pixels to be tiny blocks of information arranged in the form of a 2 D grid, and the depth of a pixel refers to the colour information present in it. In order to be processed by a computer, an image needs to be converted into a binary form. The colour of an image can be calculated as follows:

Bits/Pixel	Possible colours
1	2^1 = 2
2	2^2 = 4
3	2^3 = 8
4	2^4 = 16
8	2^8 = 256
16	2^16 = 65000

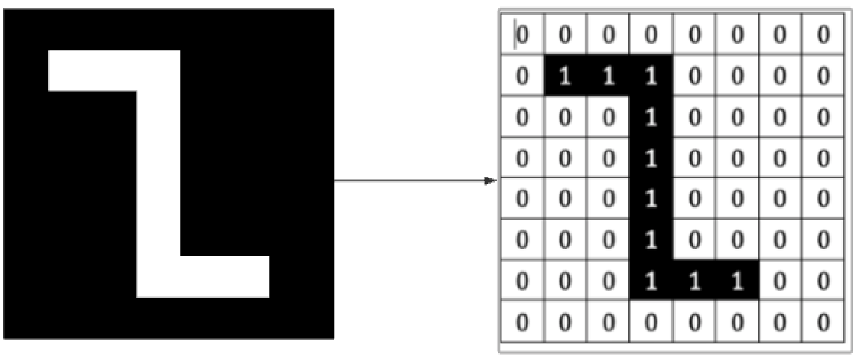
Number of colours/ shades = 2^bpp where bpp represents bits per pixel.

Naturally, more the number of bits/pixels, more possible colours in the images. The following table shows the relationship more clearly.

Let us now have a look at the representation of the different kinds of images:

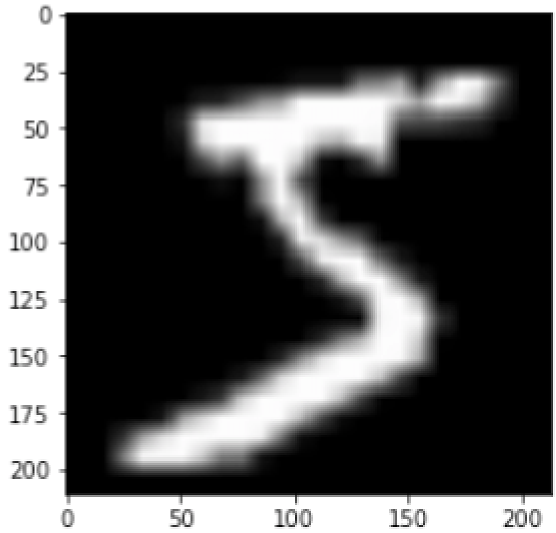
1. Binary Image

A binary image consists of 1 bit/pixel and so can have only two possible colours, i.e., black or white. Black is represented by the value 0 while 1 represents white.



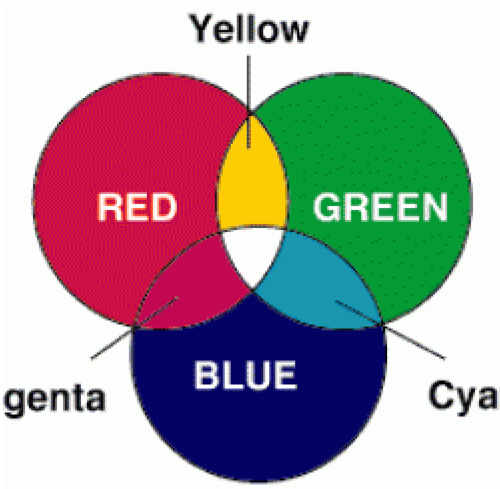
Representation of a black and white image in form of a binary where '1' represents pure white while '0' represents black. Here the image is represented by 1 bit/pixel which means image can be represented by only 2 possible colours since 2^1= 2

2. Grayscale image



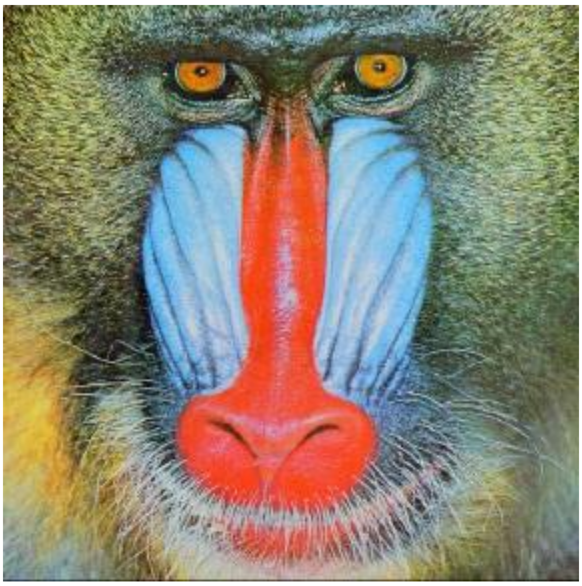
A grayscale image consists of 8 bits per pixel. This means it can have 256 different shades where 0 pixels will represent black colour while 255 denotes white. For example, the image below shows a grayscale image represented in the form of an array. A grayscale image has only 1 channel where the channel represents dimension.

3. Coloured image



Coloured images are represented as a combination of Red, Blue, and Green, and all the other colours can be achieved by mixing these primary colours in correct proportions.

A coloured image also consists of 8 bits per pixel. As a result, 256 different shades of colours can be represented with 0 denoting black and 255 white. Let us look at the famous coloured image of a mandrill which has been cited in many image processing examples.



If we were to check the shape of this image we would get:

Shape

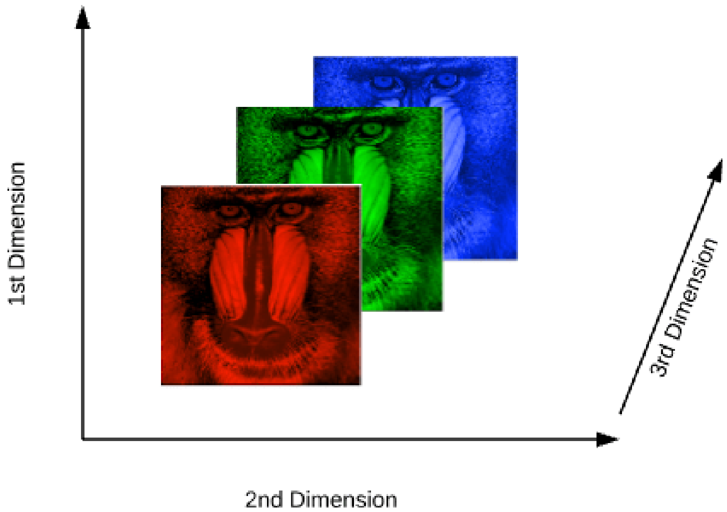
(288, 288, 3)

288: Pixel width

288: Pixel height

3: colour channel

This means we can represent the above image in the form of a three-dimensional array.



Overview on Face Detection

Face detection is a technique that identifies or locates human faces in digital images. A typical example of face detection occurs when we take photographs through our smartphones, and it instantly detects faces in the picture. Face detection is different from Face recognition. Face detection detects merely the presence of faces in an image while facial recognition involves identifying whose face it is. In this article, we shall only be dealing with the former.

Face detection is done using OpenCV (Open Source Computer Vision Library) which is an open source computer vision and machine learning software library. OpenCV was built to provide a common infrastructure for computer vision applications and to accelerate the use of machine perception in the commercial products. Being a BSD-licensed product, OpenCV makes it easy for businesses to utilize and modify the code.

OpenCV

The library has more than 2500 optimized algorithms, which includes a comprehensive set of both classic and state-of-the-art computer vision and machine learning algorithms. These algorithms can be used to detect and recognize faces, identify objects, classify human actions in videos, track camera movements, track moving objects, extract 3D models of objects, produce 3D point clouds from stereo cameras, stitch images together to produce a high resolution image of an entire scene, find similar images from an image database, remove red eyes from images taken using flash, follow eye movements, recognize scenery and establish markers to overlay it with augmented reality, etc. It has C++, Python, Java and MATLAB interfaces and supports Windows, Linux, Android and Mac OS. OpenCV leans mostly towards real-time vision applications and takes advantage of MMX and SSE instructions when available.

Face detection is performed by using classifiers. A classifier is essentially an algorithm that decides whether a given image is positive(face) or negative (not a face). A classifier needs to be trained on thousands of images with and without faces. Fortunately, OpenCV already has two pre-trained face detection classifiers, which can readily be used in a program. The two classifiers are:

- Haar Classifier
- Local Binary Pattern (LBP) classifier

Haar feature-based cascade classifiers

Haar-like features are digital image features used in object recognition. They owe their name to their intuitive similarity with Haar wavelets and were used in the first real-time face detector. Paul Viola and Michael Jones in their paper titled "Rapid Object Detection using a Boosted Cascade of Simple Features" used the idea of Haar-feature classifier based on the Haar wavelets. This classifier is widely used for tasks like face detection in computer vision industry.

Haar cascade classifier employs a machine learning approach for visual object detection which is capable of processing images extremely rapidly and achieving high detection rates. This can be attributed to three main reasons:

Haar classifier employs 'Integral Image' concept which allows the features used by the detector to be computed very quickly.

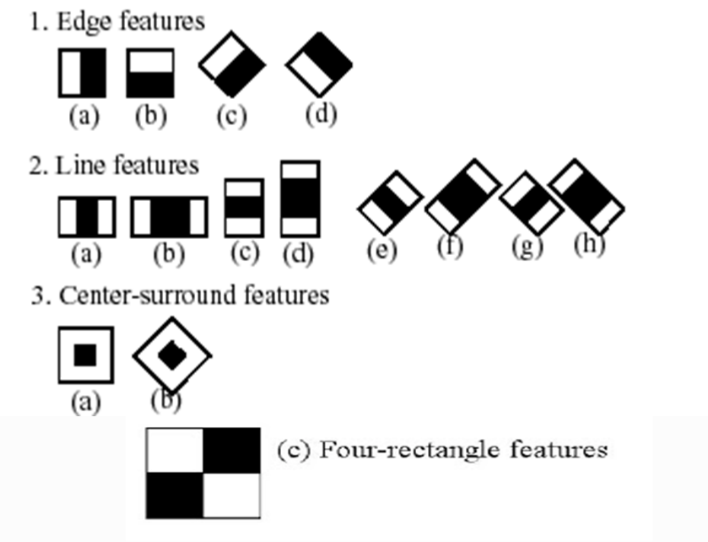
The learning algorithm is based on AdaBoost. It selects a small number of important features from a large set and gives highly efficient classifiers.

More complex classifiers are combined to form a 'cascade' which discard any non-face regions in an image, thereby spending more computation on promising object-like regions.

<u>How the algorithm works on images in steps:</u>

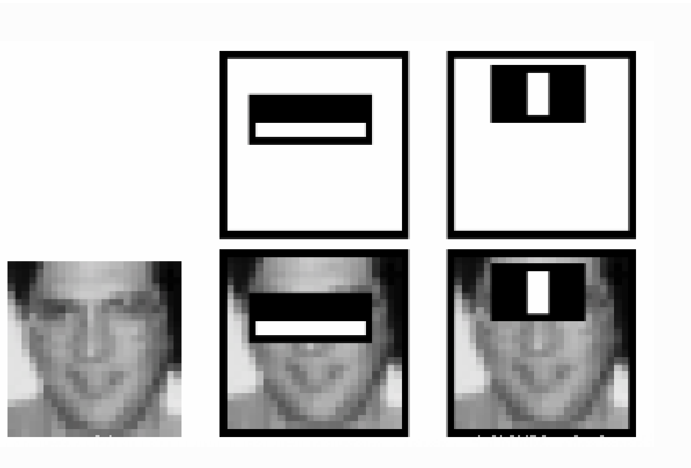
1. <u>'Haar features' extraction</u>

After the tremendous amount of training data (in the form of images) is fed into the system, the classifier begins by extracting Haar features from each image. Haar Features are kind of convolution kernels which primarily detect whether a suitable feature is present on an image or not.

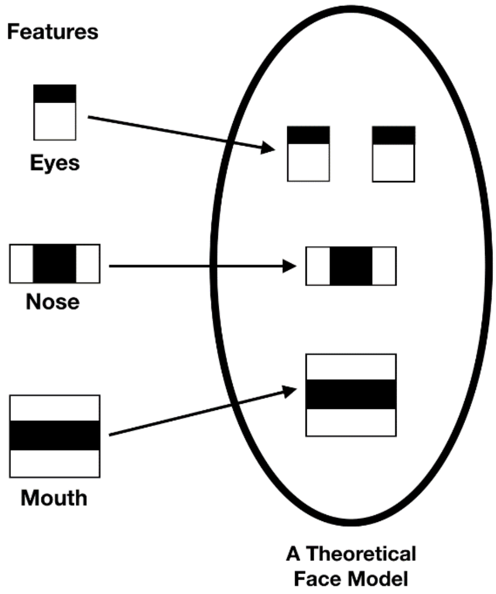


Some examples of Haar features are mentioned below:

These Haar Features are like windows and are placed upon images to compute a single feature. The feature is essentially a single value obtained by subtracting the sum of the pixels under the white region and that under the black. The process can be easily visualized in the example below.

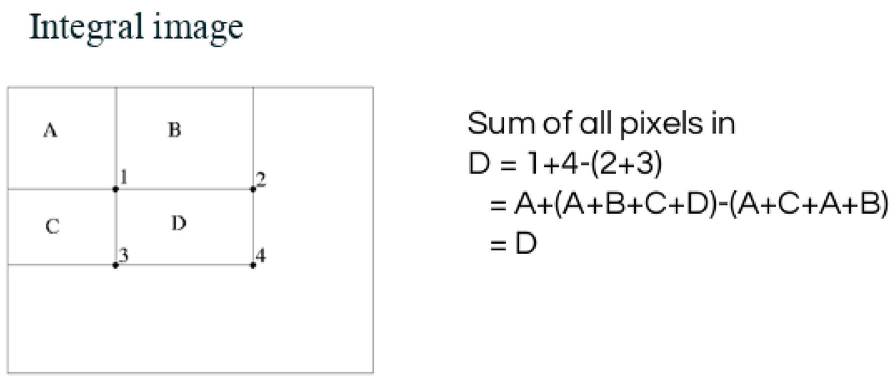


For demonstration purpose, let's say we are only extracting two features, hence we have only two windows here. The first feature relies on the point that the eye region is darker than the adjacent cheeks and nose region. The second feature focuses on the fact that eyes are kind of darker as compared to the bridge of the nose. Thus, when the feature window moves over the eyes, it will calculate a single value. This value will then be compared to some threshold and if it passes that it will conclude that there is an edge here or some positive feature.



2. 'Integral Images' concept

The algorithm proposed by Viola Jones uses a 24X24 base window size, and that would result in more than 180,000 features being calculated in this window. Imagine calculating the pixel difference for all the features? The solution devised for this computationally intensive process is to go for the Integral Image concept. The integral image means that to find the sum of all pixels under any rectangle, we simply need the four corner values.



This means, to calculate the sum of pixels in any feature window, we do not need to sum them up individually. All we need is to calculate the integral image using the 4 corner values. The example below will make the process transparent.

31	2	4	33	5	36
12	26	9	10	29	25
13	17	21	22	20	18
24	23	15	16	14	19
30	8	28	27	11	7
1	35	34	3	32	6

31	33	37	70	75	111
43	71	84	127	161	222
56	101	135	200	254	333
80	148	197	278	346	444
110	186	263	371	450	555
111	222	333	444	555	666

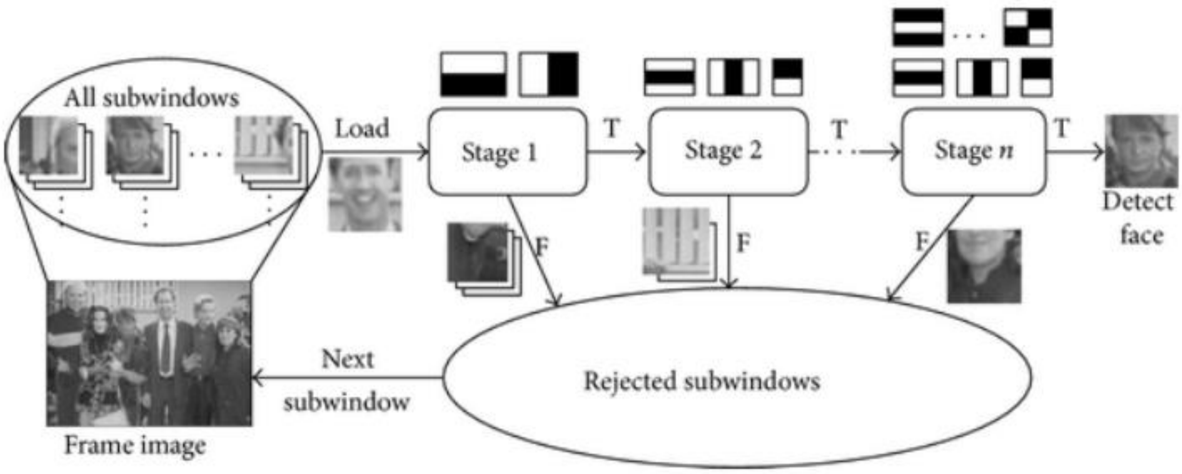
$$\boxed{15 + 16 + 14 + 28 + 27 + 11} =$$
$$101 + 450 - 254 - 186 = 111$$

3.'<u>Adaboost': to improve classifier accuracy</u>

As pointed out above, more than 180,000 features values result within a 24X24 window. However, not all features are useful for identifying a face. To only select the best feature out of the entire chunk, a machine learning algorithm called Adaboost is used. What it essentially does is that it selects only those features that help to improve the classifier accuracy. It does so by constructing a strong classifier which is a linear combination of a number of weak classifiers. This reduces the number of features drastically to around 6000 from around 180,000.

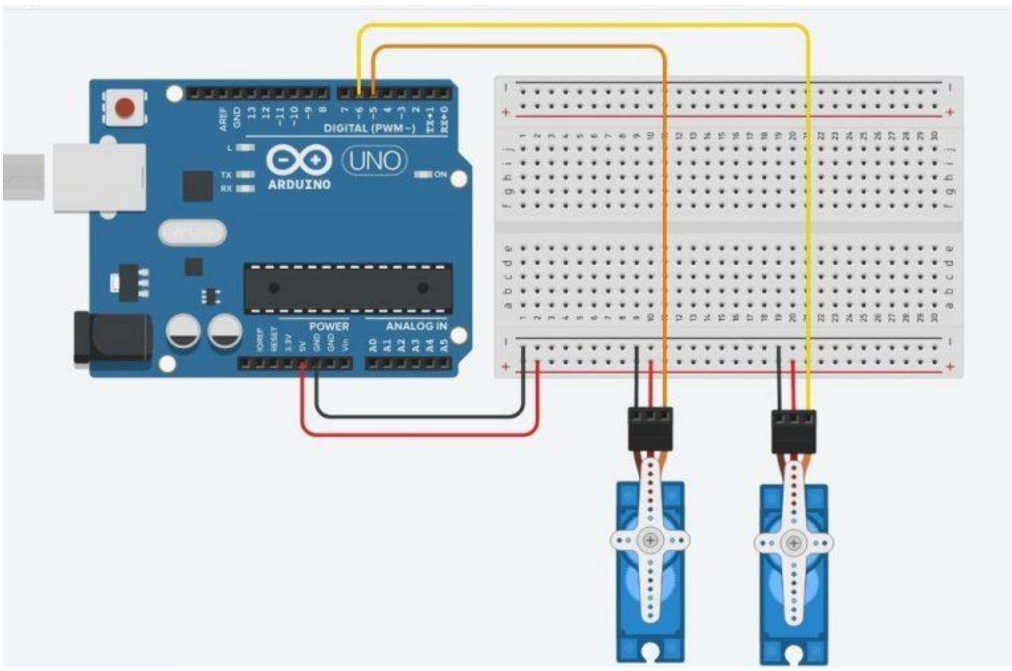
4.<u> Using 'Cascade of Classifiers'</u>

Another way by which Viola Jones ensured that the algorithm performs fast is by employing a cascade of classifiers. The cascade classifier essentially consists of stages where each stage consists of a strong classifier. This is beneficial since it eliminates the need to apply all features at once on a window. Rather, it groups the features into separate sub-windows and the classifier at each stage determines whether or not the sub-window is a face. In case it is not, the sub-window is discarded along with the features in that window. If the sub-window moves past the classifier, it continues to the next stage where the second stage of features is applied. The process can be understood with the help of the diagram below.



<u>Cascade structure for Haar classifiers.</u>

From this process the coordinates of the face are obtained and is sent to the Arduino for appropriate actions.

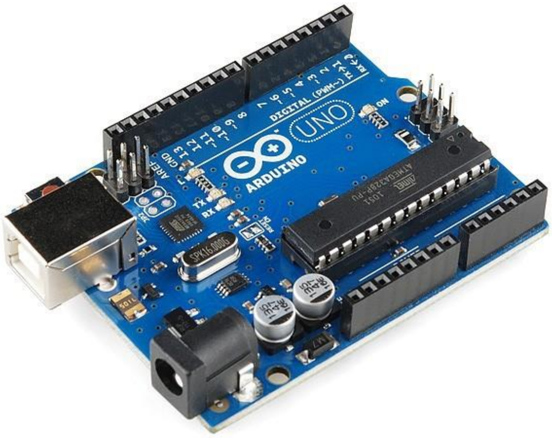


<u>Wiring Diagram for Face Tracking OpenCV</u>

COMPONENTS

Arduino UNO R3:

The Arduino UNO is an open-source microcontroller board based on the Microchip ATmega328P microcontroller and developed by Arduino.cc. The board is equipped with sets of digital and analog input/output (I/O) pins that may be interfaced to various expansion boards (shields) and other circuits. The board has 14 Digital pins,



6 Analog pins, and programmable with the Arduino IDE via a type B USB cable. It can be powered by the USB cable or by an external 9-volt battery, though it accepts voltages between 7 and 20 volts.

Its function is to receive instructions from the laptop which is then relayed to the servos.

Servo Motor SG90:

A servomotor is a rotary actuator or linear actuator that allows for precise control of angular or linear position, velocity and acceleration. It consists of a suitable motor coupled to a sensor for position feedback. Tiny and lightweight with high output power. These can rotate approximately 180 degrees (90 in each direction), and works just like the standard kinds but smaller. You can use any servo code, hardware or library to control these servos.

- Model: SG90
- Weight: 9 gm
- Operating voltage: 3.0V~ 7.2V
- Stall torque @4.8V: 1.2kg-cm
- Stall torque @6.6V: 1.6kg-cm

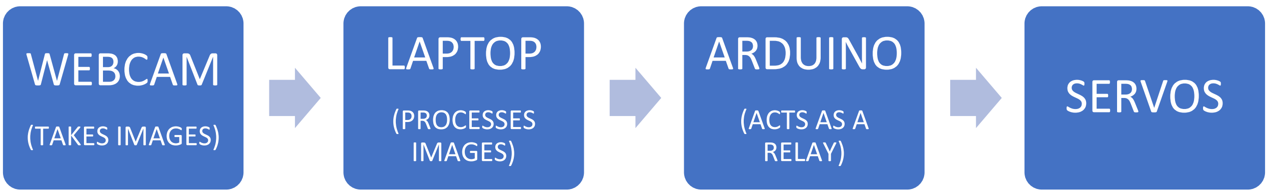
QUANTUM QHM495LM Webcam:

Quantum 25 MP Webcam with 6 Lights & Microphone Possess this Quantum 25 MP Webcam with 6 Lights & Microphone and never go out of touch. This 25 MP camera with 6 lights is a clean sweep as one of the best products that are available in the market. It also comes with a built-in microphone which helps you chat with your loved ones while you are online.

The evolutions in this web camera are the 6 lights that are detailed on this with a potentiometer that automatically switches on these lights when in the dark. This webcam from Quantum comes with 16 special effects and 10 background frames so you can now pep up your images with these frames or special effect and frame these photographs in your living room. Image control like brightness, sharpness, color contrast options are also available, and the same can be adjusted to get the expected image output. The 10 times real time digital zoom ensures that the click still gives you high resolution quality. It comes with high quality 2G glass lens for sharper picture quality. It also comes with an anti-flicker 50 Hz, 60 Hz for outdoor resolution.

Its purpose is to take video of the objects and send it to the laptop where the image is processed.

Data transfer block diagram:



APPLICATION

The applications of Facial Recognition and tracking are the following:

PREVENT RETAIL CRIME

Face recognition is currently being used to instantly identify when known shoplifters, organized retail criminals or people with a history of fraud enter retail establishments. Photographs of individuals can be matched against large databases of criminals so that loss prevention and retail security professionals can be instantly notified when a shopper enters a store that prevents a threat. Face recognition systems are already radically reducing retail crime. According to our data, face recognition reduces external shrink by 34% and, more importantly, reduces violent incidents in retail stores by up to 91%.

FIND MISSING PERSONS

Face recognition can be used to find missing children and victims of human trafficking. As long as missing individuals are added to a database, law enforcement can become alerted as soon as they are recognized by face recognition—be it an airport, retail store or other public space. In fact, 3000 missing children were discovered in just four days using face recognition in India!

HELP THE BLIND

Listerine has developed a groundbreaking facial recognition app that helps the blind using face recognition. The app recognizes when people are smiling and alerts the blind person with a vibration. This can help them better understand social situations.

TRACK SCHOOL ATTENDANCE

In addition to making schools safer, face recognition has the potential to track students' attendance. Traditionally, attendance sheets can allow students to sign another student, who is ditching class, in. But China is already using face recognition to ensure students aren't skipping class. Tablets are being used to scan students' faces and match their photos against a database to validate their identities.

VALIDATE IDENTITY AT ATMS

It seems likely that face scans will eventually replace ATM cards completely since face recognition is such a powerful identity authentication tool. But in the meantime, face recognition can be used to make sure that individuals using ATMs cards are who they say they are. Face recognition is currently being used at ATMs in Macau to protect peoples' identities.

AID FORENSIC INVESTIGATIONS

Facial recognition can aid forensic investigations by automatically recognizing individuals in security footage or other videos. Face recognition software can also be used to identify dead or unconscious individuals at crime scenes.

FACILITATE SECURE TRANSACTIONS

In China, there is a financial services company called Ant Financial that enables customers to pay for meals by scanning their faces. Customers place orders through a digital menu, and then use face scan as a payment option. After providing their telephone number they can then purchase their meal.

CHALLENGES

The Facial Recognition System has come a long way. Its usage is crucial in quite a few applications, for instance - photo retrieval, surveillance, authentication/access control systems etc. But it is yet to completely overcome the challenges which have constantly played with its quality of delivery. Listed below are the challenges which limit the potential of a Facial Recognition System to go that extra mile.

Illumination

For instance, a slight change in lighting conditions has always been known to cause a major impact on its results. If the illumination tends to vary, then; even if the same individual gets captured with the same sensor and with an almost identical facial expression and pose, the results that emerge may appear quite different.

Background

The placement of the subject also serves as a significant contributor to the limitations. A facial recognition system might not produce the same results outdoors compared to what it produces indoors because the factors - impacting its performance - change as soon as the locations change. Additional factors, such as individual expressions, aging etc. contribute significantly to these variations.

Pose

Facial Recognition Systems are highly sensitive to pose variations. The movements of head or differing POV of a camera can invariably cause changes in face appearance and generate intra-class variations making automated face recognition across pose a tough nut to crack.

Occlusion

Occlusions of the face such as beard, moustache, accessories (goggles, caps, mask etc.) also meddle with the evaluation of a face recognition system. Presence of such components make the subject diverse and hence it becomes difficult for the system to operate in a non-simulated environment.

Expressions

Another significant factor which needs to be taken into account is different expressions of the same individual. Macro and micro expressions find their place on someone's face due to changes in one's emotional state and in the wake of such expressions - which are many - the efficient recognition becomes difficult.

Complexity

Existing state-of- the-art methods of facial recognition rely on ‘too-deep’ Convolutional Neural Network (CNN) architecture which are very complex and unsuitable for real-time performance on embedded devices.

Frame Rate

Face Tracking can be difficult when the captured video has a low frame rate. If the frame rate is too low, the software might lose track of the face from one frame to another. When the person moves within the frame, the software may detect the same face as two separate people. To overcome this problem, we recommend a minimum frame rate of 6 frames per second (fps) for accurate Face Tracking.

Extreme head poses, poor lighting conditions, and partial occlusion of the face can affect the accuracy of Face Tracking. An ideal Face recognition system should be tolerant to variations in illumination, expression, pose and occlusion. It should be scalable to large number of users with need for capturing minimal images during registration while doing away with complex architecture at the same time.

FUTURE SCOPE

Learning to learn through deep learning: The feature common to all these disruptive technologies is known as Artificial Intelligence (AI) and more precisely deep learning where a system is capable of learning from data. Artificial neural network algorithms are helping face recognition algorithms to be more accurate.

Health: Significant advances have been made in this area. Thanks to deep learning and face analysis, it is already possible to: track a patient's use of medication more accurately, detect genetic diseases such as DiGeorge syndrome with a success rate of 96.6%, support pain management procedures.

Further together – towards hybridized solutions: The identification and authentication solutions of the future will borrow from all aspects of biometrics. This will lead to "biometrix" or a biometric mix capable of guaranteeing total security and privacy for all stakeholders in the ecosystem. In this solution, geolocation, IP-addresses (the device being used) and keying patterns can create a strong combination to securely authenticate users for on-line banking or egovernment services. This seventh trend belongs to us. It's our job to envisage it together and make it happen through high-added-value biometric projects.

Storing Facial Data: By saving the information on the users face it can be used to control various devices securely and efficiently. For example, smart homes or cars can be used to recognize the users and prevent and theft while providing optimal service.

CONCLUSION

Face recognition technology has come a long way in the last twenty years. Today, machines are able to automatically verify identity information for secure transactions, for surveillance and security tasks, and for access control to buildings etc. These applications usually work in controlled environments and recognition algorithms can take advantage of the environmental constraints to obtain high recognition accuracy. However, next generation face recognition systems are going to have widespread application in smart environments -- where computers and machines are more like helpful assistants.

To achieve this goal computers must be able to reliably identify nearby people in a manner that fits naturally within the pattern of normal human interactions. They must not require special interactions and must conform to human intuitions about when recognition is likely. This implies that future smart environments should use the same modalities as humans, and have approximately the same limitations. These goals now appear in reach -- however, substantial research remains to be done in making person recognition technology work reliably, in widely varying conditions using information from single or multiple modalities.

REFERENCE

- https://www.instructables.com/id/Real-Time-Face-Tracking-Robot-With-Arduino-and-Mat/
- https://www.learnrobotics.org/blog/face-tracking-opencv/
- https://www.opencv-srf.com/2010/09/object-detection-using-color-seperation.html
- https://create.arduino.cc/projecthub/WolfxPac/face-tracking-using-arduino-b35b6b
- https://sisu.ut.ee/imageprocessing/book/1
- https://www.gemalto.com/govt/biometrics/facial-recognition