Emotion Space Online Assistance

MR. MOHIT JAYANT DHAWALE
MR. RAMAN PRADEEP PAWAR
MS. SAMREEN SHAAD SHEIKH
MS. SAVI AJAY WAGHMARE
MR. SHIBAN IFTEKHAR ALISAYYAD

PROF. RAJESHWARI SURYAWANSHI (Guide)

Department of Information Technology

Govindrao Wanjari College of Engineering & Technology Nagpur

ABSTRACT

These Human facial expressions convey a lot of information visually rather than articulately. Facial expression recognition plays a crucial role in the area of human-machine interaction. Automatic facial expression recognition system has many applications including, but not limited to, human behavior understanding, detection of mental disorders, and synthetic human expressions. Recognition of facial expression by computer with high recognition rate is still a challenging task. Two popular methods utilized mostly in the literature for the automatic FER systems are based on geometry and appearance. Facial Expression Recognition usually performed in four-stages consisting of pre-processing, face detection, feature extraction, and expression classification. In this project we applied various deep learning methods (convolutional neural networks) to identify the key seven human emotions: anger, disgust, fear, happiness, sadness, surprise and neutrality. The primary Aim of this project was to develop an emotion recognition web-application that was capable of recognising emotions expressed by people in a video stream but could also be deployed to run natively on embedded devices. Current solutions revolve around deep learning techniques which require a great deal of processing. This means that when embedded devices are used, they function primarily as an input video source for a remote server that carries out the processing. The web-application developed had to be capable of recording a video stream and detecting the emotions expressed in it. The detected emotions had to then be displayable in a meaningful manner in order to give insights into the emotions expressed by people being recorded. This could then be used by marketing professionals to gain truthful information about how people feel when watching an advert or having a product demonstrated to them. With the emotion recognition system, AI can **detect the emotions of a person through**

their facial expressions. Detected emotions can fall into any of the six main data of emotions: happiness, sadness, fear, surprise, disgust, and anger. For example, a smile on a person can be easily identified by the AI as happiness.

INTRODUCTION

"2018 is the year when machines learn to grasp human emotions" --Andrew Moore, the dean of computer science at Carnegie Mellon. With the advent of modern technology our desires went high and it binds no bounds. In the present era a huge research work is going on in the field of digital image and image processing. The way of progression has been exponential and it is ever increasing. Image Processing is a vast area of research in present day world and its applications are very widespread. Image processing is the field of signal processing where both the input and output signals are images. One of the most important application of Image processing is Facial expression recognition. Our emotion is revealed by the expressions in our face. Facial Expressions plays an important role in interpersonal communication. Facial expression is a non verbal scientific gesture which gets expressed in our face as per our emotions. Automatic recognition of facial expression plays an important role in artificial intelligence and robotics and thus it is a need of the generation. Some application related to this include Personal identification and Access control, Videophone and Teleconferencing, Forensic application, Human-Computer Interaction, Automated Surveillance, Cosmetology and so on. The objective of this project is to develop Automatic Facial Expression Recognition System which can take human facial images containing some expression as input and recognize and classify it into seven different expression class such as :

1 Neutral
2 Angry
3 Disgust
4 Fear
5 Happy
6 Sadness
7 Surprise

PROBLEM STATEMENT

- The goal is to use a model based approach for facial emotion recognition of driver in real time environment
- The system should work on a embedded platform
- The system should assist user on basis of emotions.
- The program is so developed to work on pipelined architecture and parallel processing

PROPOSED SYSTEM

Design is the first step in the development phase for any techniques and principle for the purpose of defining a device process or system in sufficient detail to permit in physical realization. System design is the process of defining architecture, components, module and data for system to satisfy specified requirements. System design could be seen as the application of a system theory to product development.

The System design document is a required document for every project. It should include a high level description of why the system design document has been created, provide what the new system is intendedfor or is intended to replace and contain detailed description of the architecture and system components.

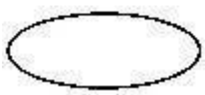
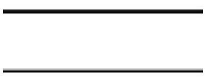
Once the software requirements have been analyzed and specified the software design involves three technical activities – design, coding and implementation and testing that are required to build and verify the software. The design activities are of main importance in this phase, because in this activity decision ultimately the success of the software implementation and its ease of maintenance are made. The decisions havethe final bearing upon reliability and maintainability of the system. Design is the only way to accurately translatethe customers requirement into finished software or a system.

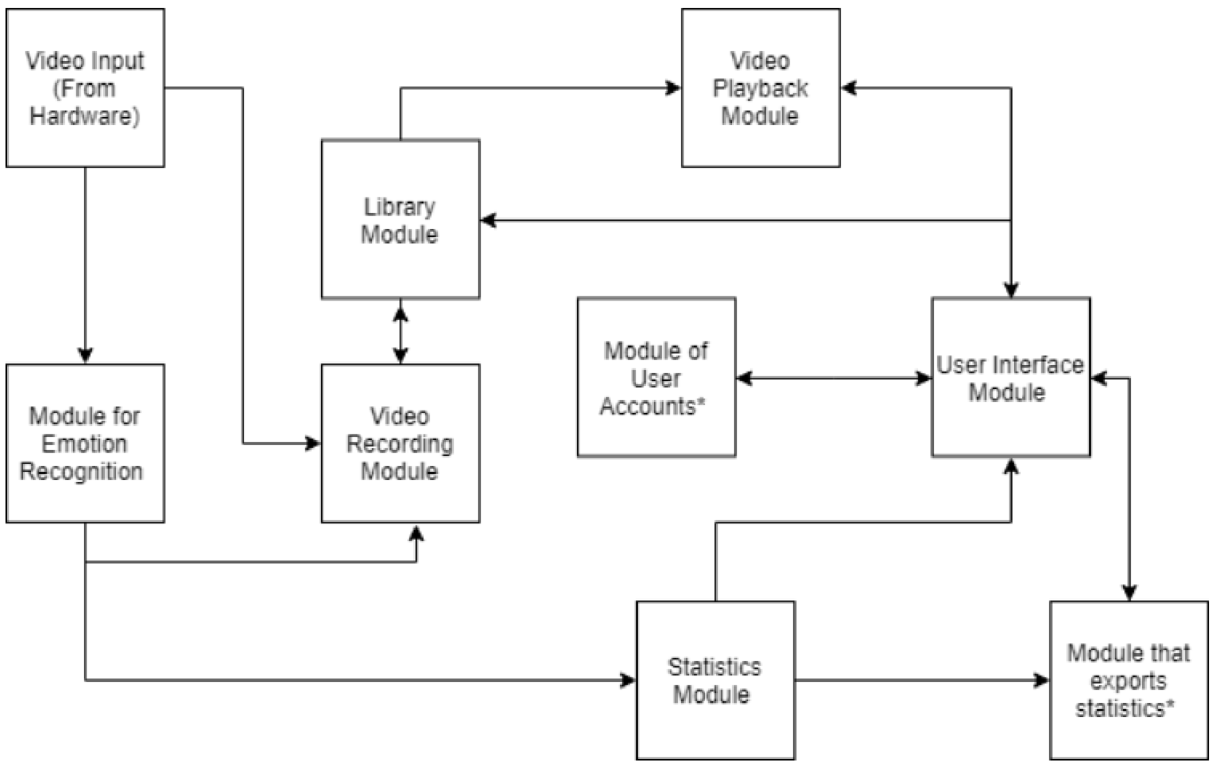
Design is the place where quality is fostered in development. Software design is a process through which requirements are translated into a representation of software. Software design is conducted in two step. Preliminary design is concerned with the transformation of requirements into data.

DATA FLOW DIAGRAM

5.1 Data Flow Diagram

A data flow diagram (DFD) is a graphical tool used to describe and analyse the movement of data through a system by depicting the flow of data, storage of data, source or destination of data and the processes that respond to change in data. The DFD is one of the most important tools used by the system analysts to model system components. The Data flow diagram (DFD) contains some symbol for drawing the data flow diagram.

	**Data Flow** – Data flow are pipelines through the packets of informationflow.
	**Process :** A Process or task performed by the system.
	**Entity :** Entity are object of the system. A source or destination data of asystem.
	**Data store :** A place to be where data is store



5.2 Entity Relationship Diagram

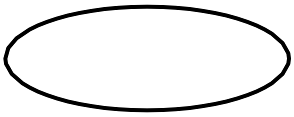
An Entity Relationship Diagram (ERD) is a graphical tools to express the overall structure of a database. Anentity is a place, person, thing or event of interest to the organization and about which data are captured, stored, processed. The attribute are various kinds of data that describe an entity. An association of several entities in an, Entity Relationship model is called relationship.

An ERD consists of the following major components

- Rectangle :

Used for representing entity types

- Ellipse :



Used for representing attribute

- Lines :

Used for linking attribute to entity type

5.1.2 Advantages

- Emotion-sensing technologies can help employees make better decisions, improve their focus and performance in the workplace, manage stress, and also help them adopt healthier and more productive working styles.
- Voice-based emotion analysis in real time opens up many business opportunities by enabling an automated customer service agent to recognise a caller's emotional state and adapt accordingly. Such information can also be helpful in analysing and managing stress levels of human workers.
- Emotion-sensing technology will also enable companies to establish deeper emotional connections with their consumers through virtual assistants. Data collected through such devices can help companies understand how internal and external environmental factors impact their employees. As a result, companies can redesign processes accordingly, to help keep their personnel better engaged and productive.
- Emotion sensing through wearables can help monitor the user's state of mind in terms of mental and other health conditions.

5.1.3 Disadvantages

- Validation of emotion dataset is a challenge in order to have accurate emotion recognition system.
- Finding or detecting emotion is haystack.
- It is a challenge to make emotion available in different languages.
- There are limitations with different types and versions of the softwares such as dataset input is only textual data, image, pattern, video and audio inputs are invalid.
- Performance and results of the emotion sensing system depends on accuracy of the sensors such as cameras, thermal image sensors, facial recognition algorithm used and so on. Highly accurate system will be expensive due to use of costly components.

5.1.4 Applications

- App and product development Emotion recognition can play a huge role in optimizing various software engineering processes which comprises of testing of ease with which a product can be used. It's a longestablished fact that level of comfort with different software products depends hugely upon human emotions. A products overall look and feel can also alter human feelings which in turn makes the person buy the product or not. Thus Emotion Detection using Image Processing in Python researching about different emotional states of a human body and how it is influenced by usage of different products is a matter of prime importance for related industry professionals.

- Improved learning practices Evidence suggests that part of emotional states vouches for better learning practices while the other part try to suppress them. The difference between the two groups possessing different emotional states is not so common to find. For example, positive emotional state is thought to be bad for learning purposes while slightly negative emotional state fosters analytical thinking and is also appropriate for carrying out critical tasks.

- Improvised web development With the mammoth scale at which the internet is expanding, service providers are interested in collecting tons and tons of data which can be extracted from the users. Correspondingly, all the content and advertisements are played based on the users' profile. Subsequently, adding intricate details about the different human emotions can provide much more precise behavioral models of different types of users.

- 4. Immersive gaming Video games constitute a large chunk of entertainment industry. Thus, in order to make these games much more intensive and intrusive, video game creators base their research on different types of human emotions commonly found. In order to allure more and more players, video games are made in such a way that they incorporate human emotions naturally into their game play.

IMPLEMENTATION

Software Environment

Environment refers to the collection of hardware and software tools a system developer uses to build software systems. As technology improves and user expectations grow, an environment's functionality tends to change. Over the last 20 years the set of software tools available to developers has expanded considerably.

We can illustrate this change by observing some distinctions in the terminology. Programming environment and software development environment are often used synonymously, but here we will make a distinction between the two. By "programming environment" we mean an environment that supports only the coding phase of the software development cycle—that is, programming-in-the-small tasks such as editing and compiling. By software development environment we mean an environment that augments or automates the activities comprising the software development cycle, including programming-in-the-large tasks such as configuration management and programming-in-the-many tasks such as project and team management. We also mean an environment that supports largescale, long-term maintenance of software.

The evolution of environments also demands that we distinguish basic operating system facilities—fundamental services such as memory, data, and multiple program management—from the enhanced functionality that characterizes state-of-the-art environments. This enhanced functionality is typically achieved through tools such as browsers, window managers, configuration managers, and task managers. In a sense, environments have evolved in concert with the software engineering community's understanding of the tasks involved in the development of software systems.

To better understand the technological trends that have produced state-of the-art environments, we here present a taxonomy of these trends. We cite examples of research and commercial systems within each class. We intend the taxonomy to show the description of the trends and to suggest where more work needs to be done.

Language:

Python: Python is a high-level, interpreted general purpose programming language. The selection of

Python as the programming languages was made for a few reasons. Python is a language commonly used in Data Analytics, Machine Learning and Computer Vision, meaning that there would be lots of support available if needed. OpenCV has a python interface, so importing OpenCV to the project would be easy to do. Having the Raspberry PI in mind, the language needed to be compatible with this and Python is pre-installed out of the box with Raspbian (Raspberry Pi's operating system). HTML: Hypertext Mark-up Language (HTML) is the standard language used to design how webpages are displayed in a web-browser. HTML is the language that was chosen to develop the user interface as it has cross-browser support, it is relatively simple to understand and interfaces with Flask and Jinja2 templates. Jinja2: Jinja2 is a Python templating engine that is used by flask and is necessary for flask to run. JavaScript: JavaScript is an interpreted object-oriented scripting language used to create effects within webbrowsers. JavaScript was used to interact between different sections of the application. Specifically, for handling requests to start and stop recording the emotion detection. The statistics that have been developed utilised a JavaScript library too. Libraries & Frameworks: OpenCV: OpenCV is an open source computer vision library that was used for the Computer Vision Section of the system. It has been developed over time and is currently in a maturity stage where there is a big enough community so resources can be found with ease. Many of the projects examined in the literature review use OpenCV, and I have used this in the past so I was familiar with this and preferred it over other Machine Learning libraries that could be adapted for Computer Vision tasks. Flask: Flask is a lightweight python web application framework that is designed to get a simple web application running quickly and easily. As the focus of this project was oriented towards emotion detection in real time, this was the ideal framework for the task.

Hardware: PC: For the development of the project although it would have been possible to develop entirely on the Raspberry Pi, I didn't want to rely only on this especially when it came to re-training facial recognition algorithms. This is because although the Raspberry Pi model that I used was computationally powerful, the processing power of it couldn't compare to that of the processor in any modern-day computer. Raspberry Pi 3b+: The Raspberry Pi is a microcomputer that is small enough to fit in a pocket, meaning that a solution implemented on this would be portable. The Raspberry Pi is used by many hobbyists and researchers because of how lightweight and affordable it is. This means that there is a large community offering lots of support when needed. Raspberry Pi camera: The need for video input to the system led to the purchase of a Raspberry Pi camera module. This can easily be installed and set up and is affordable yet also produces video at a high enough quality to be used for the system. The Raspberry Pi camera has an 8 megapixel native resolution sensor, and supports video capture at resolutions of 1080p, 720p and 640x480 at framerates of 30, 60 and 90 FPS respectively. Development Tools: PyCharm: PyCharm is a Python Development Environment used industry wide that I am familiar with and comfortable using, so it made sense to use this as the IDE for developing and debugging any code. System Architecture: Once the high-level design had been completed, I knew what submodules existed within the system and could therefore determine how they interacted between each other.

Figure 1 below shows the interactions between the different modules of the system. This was helpful to understand how internal components of the system would be integrated and how the development of each would be implemented as it showed what dependencies exist between submodules.

Softwares

As the project is developed in python, we have used Anaconda for Python 3.6.5 and Spyder.

Anaconda Itis a free and open source distribution of the Python and R programming languages for data science and machine learning related applications (large-scale data processing, predictive analytics, scientific computing), that aims to simplify package management and deployment. Package versions are managed by the package management system conda. The Anaconda distribution is used by over 6 million users, and it includes more than 250 popular data science packages suitable for Windows, Linux, and MacOS.

Spyder Spyder (formerly Pydee) is an open source cross-platform integrated development environment (IDE) for scientific programming in the Python language. Spyder integrates NumPy, SciPy, Matplotlib and IPython, as well as other open source software. It is released under the MIT license. Spyder is extensible with plugins, includes support for interactive tools for data inspection and embeds Python-specific code quality assurance and introspection instruments, such as Pyflakes, Pylint and Rope. It is available cross-platform through Anaconda, on Windows with WinPython and Python (x,y), on macOS through MacPorts, and on major Linux distributions such as Arch Linux, Debian, Fedora, Gentoo Linux, openSUSE and Ubuntu.

Features include:

- editor with syntax highlighting and introspection for code completion
- support for multiple Python consoles (including IPython)
- the ability to explore and edit variables from a GUI

Available plugins include:

- Static Code Analysis with Pylint
- Code Profiling
- Conda Package Manager with Conda

Requirement Analysis

Software Requirements

- Operating System : Windows 7 or above
- Front End: HTML, CSS, javascript
- Backend: Python
- IDE : Notepad
- Language : Python (Flask)

Hardware Requirements:

- RAM : 1GB and Above
- Keyboard : Standard Windows keyboard
- Mouse : Two or Three button mouse
- Monitor : SVGA

Implementation

1) After successfully installing all the necessary softwares, we must start by creating a Dataset. Here, we can create our own dataset by analyzing group of images so that our result is accurate and there is enough data to extract sufficient information. Or we can use an existing database.

2)The dataset is then organized into two different directories. First directory will contain all the images and the second directory will contain all the information about the different types of emotions.

3)After running the sample images through the python code, all the output images will be stored into another directory, sorted in the order of emotions and its subsequent encoding.

4) Different types of classes can be used in OpenCV for emotion recognition, but we will be mainly using Fisher Face one.

5)Extracting Faces:OpenCV provides four predefined classifiers, so to detect as many faces as possible, we use these classifiers in a sequence.

6)The dataset is split into Training set and Classification set. The training set is used to teach the type of emotions by extracting information from a number of images and the classification set is used to estimate the classifier performance.

7)For best results, the images should be of exact same properties i.e. size.

8)The subject on each image is analyzed, converted to grayscale, cropped and saved to a directory.

9)Finally, we compile training set using 80% of the test data and classify the remaining 20% on the classification set. Repeat the process to improve efficiency .

RESULTS

The completed training implementation uses Viola-Jones's Haar-like feature cascade detector to detect faces as well as eyes and mouths. Detected faces are cropped, resized, and mean subtracted, then PCA is performed. Using the reduceddimensionality training dataset Fisher LDA is performed to extract Fisherfaces on which we can project test data. Also during training, eye and mouths are detected using Haar-like features, or using a Harris corner based approach is Haar-like features fail. The detected eye and mouth regions are then extracted and resized. HOG features are extracted from each region, and a SVM is trained using a combined eye-mouth HOG vector and training labels. The primary reason we use this dual-classifier approach is improving speed with maintaining accuracy. When we use test images from the Extended Cohn-Kanade dataset and project those images onto our Fisherfaces for classification based on our established thresholds, we have an accuracy of 56%. This is a poor result, as it is only marginally better than randomguessing. Upon further investigation, this is due to the Fisherface-approach's inability to effectively detect the expressions corresponding to disgust and contempt. However, when only detecting expressions of test images that correspond to anger, fear, happiness, sadness, and surprise, the Fisherface approach is more than 90% accurate. This leads up to consider that anger, fear, happiness, sadness, and surprise are "easy-to-distinguish" emotions. This is likely attributable to the fact that the "easy-to-distinguish" emotions have very distinct and blatant feature manipulations associated with them. For example, the happiness expression has a very strong turning up of the mouth, which is seen to be strongly emphasized in the Fisherface for discerning happiness. The expression associated with fear has a very strong lateral pulling of the edges of the mouth, also evident in the associated Fisherface. The anger expression involves a downwards furrowing of the brow, the sad expression involves an obvious turning-down of the mouth, and surprise involves a veryobvious open mouth. Contempt and disgust on the other hand, are much more difficult to detect, for potentially different reasons. It is possible that disgust is difficult to detect because it has feature orientations that are similar to those in several other emotions, such as an opening of the mouth that could be confused with happiness, fear, or surprise. The brows during a display of disgust is also furrowed similarly to the anger expression. The most tell-tale sign of disgust is an upward pulling of the nose, leading to wrinkling around the bridge of the nose. However, this detail is much more nuanced than the other more obvious expression characteristics, and can be lost during resolution reduction, mean-subtraction, and image misalignment. Contempt on the other hand, is difficult to detect since its characteristics are very faint in intensity. The expression for contempt is characterized by a neutral expression overall, with a unilateral pulling up of the mouth. This can be very difficult to distinguish as a human, so incorrect labeling of training data, as well as a methodological inability to capture the faint characteristics of the expression make contempt very difficult to detect. The dual-classifier approach works well when the Fisherface cannot effectively determine a prediction. This happens in two cases. First is if a test image is not one of the "easy-to-distinguish" emotions, and second is if the Fisherface classifier cannot decide between two or more predicted emotions. The overall testing approach is to pass a test image through each of the five "easy-to-distinguish" Fisherface classifiers. If only one classifier makes a positive prediction, then that test image is assigned that Fisherface's emotion as the prediction. If no classifier offers a positive prediction, or more than one classifier offers a positive prediction, then the test image moves to phase two of the classification process. The test image first undergoes Haar-like feature detection for the eye region and mouth region. The detailed Harris corner method is used as a backup if Haar detection fails. Then the HOG features are extracted for both regions, concatenated, then passed to the trained SVM for a final prediction. When using the HOG and SVM classifier only, the accuracy for detection is 81%, much better than a Fisherface only approach. When using the dual-classifier method, the accuracy is the same as HOG-only at 81%, but the testing process is 20% faster. This is because not all images must undergo eye and mouth detection, extraction, then undergo HOG feature extraction, but only those test images that are not given a prediction by the much faster Fisherface classifier.

CONCLUSION

7.1 Conclusion :

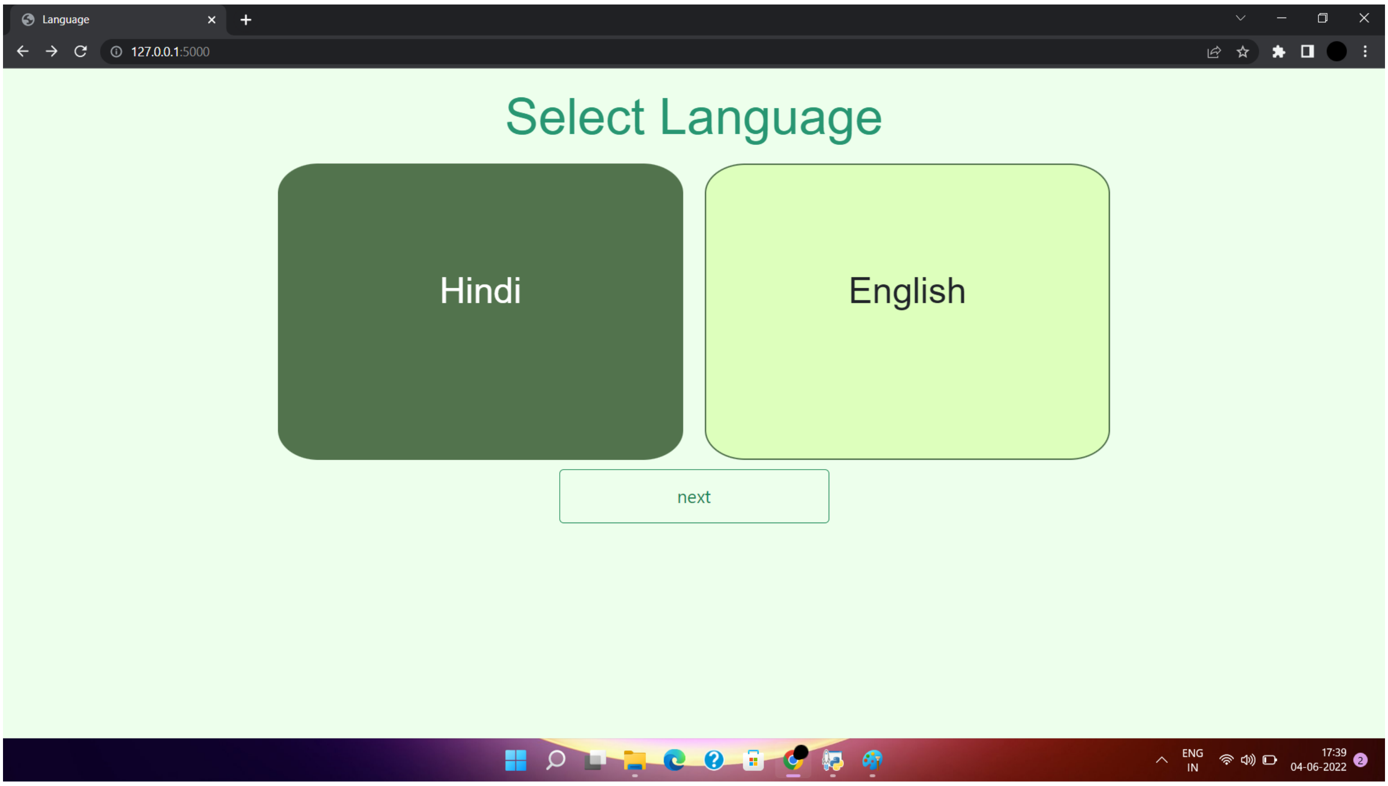
An image processing and classification method has been implemented in which face images are used to train a dualclassifier predictor that predicts the seven basic human emotions given a test image. The predictor is relatively successful at predicting test data from the same dataset used to train the classifiers. However, the predictor is consistently poor at detecting the expression associated with contempt. This is likely due to a combination of lacking training and test images that clearly exhibit contempt, poor pre-training labeling of data, and the intrinsic difficulty at identifying contempt. The classifier is also not successful at predicting emotions for test data that have expressions that do not clearly belong exclusively to one of the seven basic expressions, as it has not been trained for other expressions. Future work should entail improving the robustness of the classifiers by adding more training images from different datasets, investigating more accurate detection methods that still maintain computational efficiency, and considering the classification of more nuanced and sophisticated expressions.

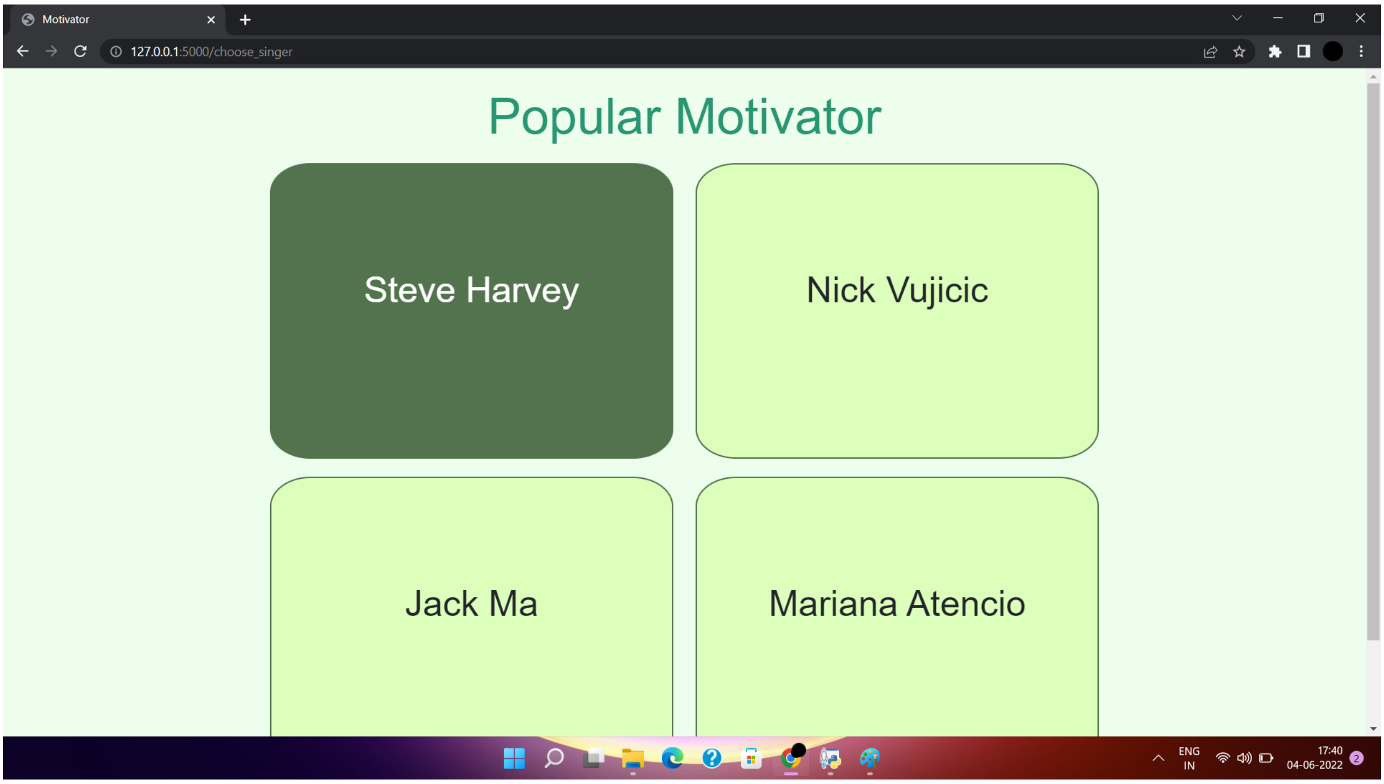
7.2Future Scope

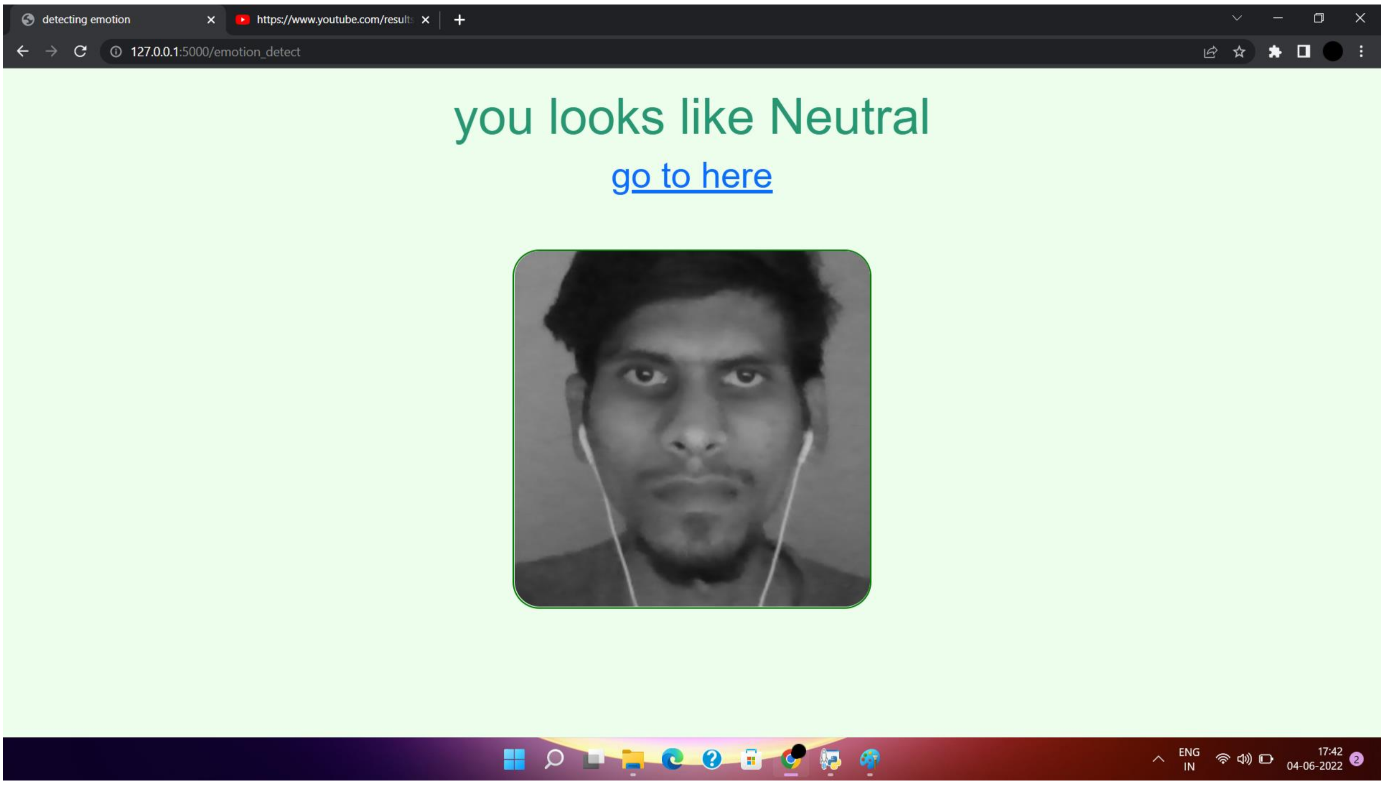
This chapter proposes solutions to challenges brought about by the widespread adoption of online classes during the pandemic using facial recognition and emotion detection technology.
Facial recognition-based attendance registration is performed using HOG to achieve 97.2% accuracy in identifying faces under well-lit optimal conditions. The model is highly effective in the recognition of faces which is essential for users during online classes. This reduces the chances of false negatives and false positives while registering attendance. In distance learning, automating this process would relieve the lecturer of having to worry about attendance. Future scope involves enhancement by applying the framework in real-time attendance monitoring, which would be computationally expensive and require powerful hardware.

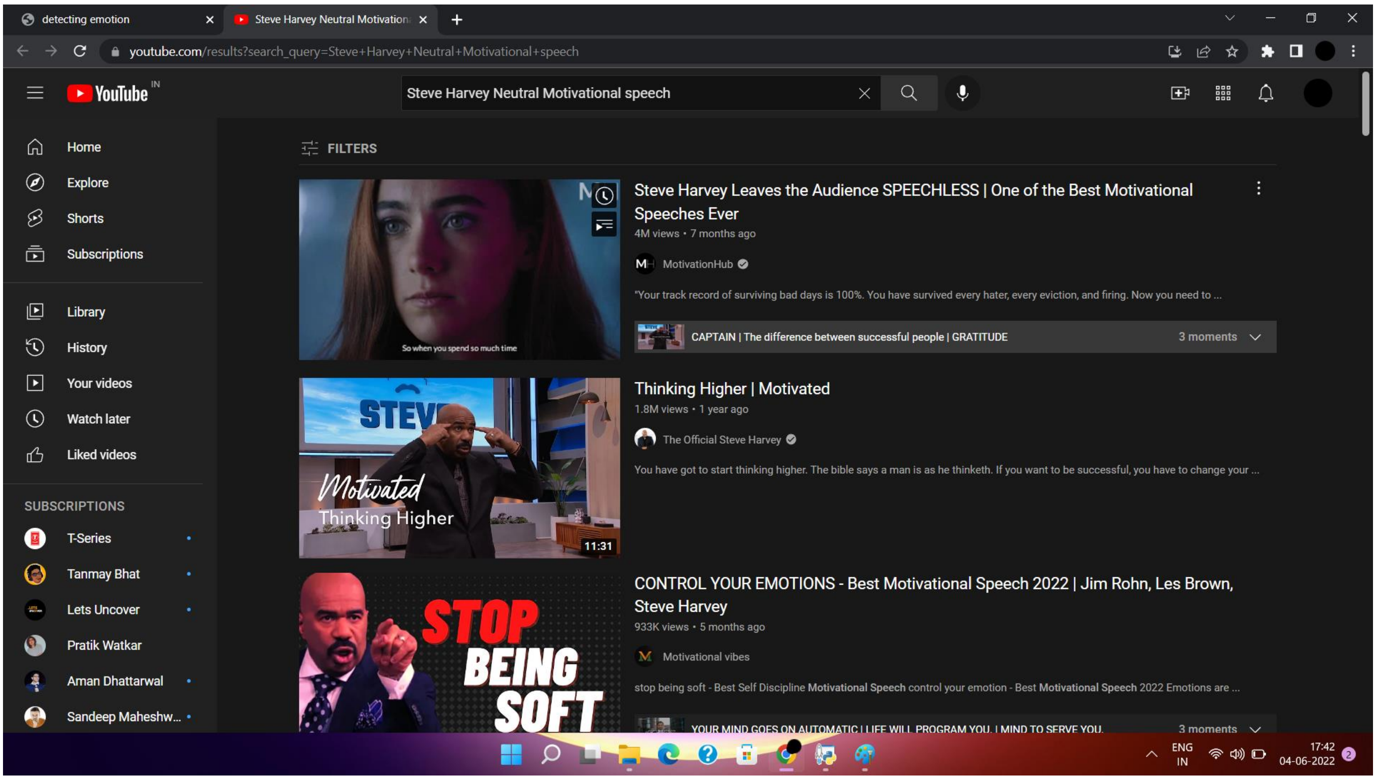
Emotion detection is performed by detecting facial expressions using Haar cascade classifiers, pretrained weights of the VGG-16 neural network, and a custom sequential model. Due to training limitations, a convolutional neural network, VGG-16 with pre-trained weights, is used to create bottleneck features for the images. These are then fed to the sequential model using ReLU activation, dropout, and a batch normalization layer with a learning rate of 0.001. Thus, the model classifies facial expressions from a preprocessed image into one of seven emotion classes—Anger, Disgust, Fear, Happiness, Neutral, Sadness, and Surprise. The model achieves 89.2% accuracy while testing on the test dataset. On performing testing in real-world conditions, the model achieves accuracies around 87%–79%, which is effective for the intent of this framework. In the future, there exists scope for improvement in the performance to increase accuracy to over 90%. Providing the dominant emotion detected from the student to the lecturer serves to have a basic form of “student reaction” to aid the teaching process of the lecturer. However, in the age of distance learning and unreliable Internet connections, optimized models which are not computationally demanding are a challenge. In the future, real-time emotion monitoring systems that can perform similar functions would serve enhanced utility in similar educational situations, e.g., classrooms, and would provide benefits to education. Real-time emotion detection models that can operate on a local machine instead of requiring high-bandwidth Internet and powerful computers can benefit online learning.

Project Screenshots




Motivator
127.0.0.1:5000/choose_singer
Popular Motivator
Steve Harvey
Nick Vujicic
Jack Ma
Mariana Atencio


detecting emotion
127.0.0.1:5000/emotion_detect
you looks like Neutral
go to here



REFERENCES

[1] Ekman, P. & Keltner, D. (1997). Universal facial expressions of emotion: An old controversy and new findings. In Segerstråle, U. C. & Molnár, P. (Eds.), Nonverbal communication: Where nature meets culture (pp. 27-46). Mahwah, NJ: Lawrence Erlbaum Associates.

[2] Matsumoto, D. & Kupperbusch, C. Idiocentric and allocentric differences in emotional expression, experience, and the coherence between expression and experience. Asian Journal of Social Psychology (4), pp. 113-131 (2001).I.S. Jacobs and C.P. Bean, “Fine particles, thin films and exchange anisotropy,” in Magnetism, vol. III, G.T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271-350.

[3] Lucey, P., Cohn, J.F., Kanade, T., Saragih, J. Ambadar, Z. The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression. IEEE Computer Society Conference CVPRW (2010)

[4] Zhang, Z. Feature-based facial expression recognition: Sensitivity analysis and experiments with a multilayer perceptron. International Journal of Patten Recognition and Artificial Intelligence 13 (6):893-911 (1999).

[5] Michael J. Lyons, Shigeru Akemastu, Miyuki Kamachi, Jiro Gyoba. Coding Facial Expressions with Gabor Wavelets, 3rd IEEE International Conference on Automatic Face and Gesture Recognition, pp. 200-205 (1998)