WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

MULTI MODAL INTERACTION FRAMEWORK USING DEEP-LEARNING

Angad Jain[1], Anushka Sahu[2], Gauransh Sahu[3], Anshika Kashiv[4], Charu Vaibhav Verma[5]

Prestige Institute of Engineering Management and Research, Indore

ABSTRACT :

The passage discusses advancements in facial emotion recognition using deep learning, particularly convolutional neural networks (CNNs). It emphasizes the challenges of teaching computers to recognize emotions like humans do. The proposed method involves using CNNs for feature extraction and edge detection to preserve texture information. Several datasets are explored for training models aimed at improving emotion detection accuracy.

Furthermore, it introduces a novel approach combining hand gesture and facial emotion recognition for music recommendation systems. This hybrid system utilizes a facial expression recognizer (FER) algorithm alongside MediaPipe for hand detection and TensorFlow for gesture recognition. The system prioritizes hand gestures over facial emotions for music playback, comparing its accuracy with existing methods.

Additionally, the passage addresses attendance management systems, highlighting their importance in various domains and discussing technological advancements in automatic identification methods. It includes a structured review of attendance system technologies, summarizing relevant research and outlining future directions for improvement.

Overall, the passage underscores the intersection of AI, IoT, and machine learning in enhancing emotion detection, music recommendation systems, and attendance management technologies.

Keywords—Mood and sentiment analysis, Artificial intelligence, Internet of things, Bag of words, Attendance system, Image segmentation, Motion segmentation, Neural networks, Motion detection, Gesture recognition, Music playback, Hand gestures.

Introduction :

Over the past two decades, there has been a significant rise in multimodal approaches within linguistics and language sciences, building on early empirical efforts from the late 19th and early 20th centuries. This evolution culminated approximately a century later in a robust adoption of multimodal methodologies. Despite challenges in accurately tracing academic historical trajectories, understanding this evolution requires acknowledging pioneers like Charles Goodwin.

The proliferation of social media has become a major data source, with daily outputs such as 500 million tweets globally, over 2 billion active Facebook users monthly, and Instagram generating 9 million images and posts daily. This abundance underscores the prominence of text mining and sentiment analysis in extracting insights from online social networks. Recent advancements include predicting emojis in real-time text and identifying mental health trends from user activity over time.

Research efforts have focused on mood detection and prediction using classifiers, examining archival data and real-time streams via an Android app called Citizen Sense. This app collects user opinions across various topics, categorizing them into 25 mood options and employing nine conventional learning techniques for classification, including Decision Trees and Support Vector Machines. Studies also analyse mood variations over time using tweet datasets.

Facial Expression Recognition (FER) and Gesture Recognition represent key applications in computer vision and language technology. FER involves classifying human faces into emotion categories, while Gesture Recognition enables tasks like controlling music through hand movements(fig.no.1.2). These technologies, popularized by devices like Microsoft Kinect, aim to enhancecommunication, particularly for sign language users, by translating hand gestures into meaningful messages.

Gesture recognition allows users to interact with devices or applications using hand movements. In the context of music control, it enables users to adjust volume, skip tracks, or play/pause music without physical buttons(fig.no.1.1).

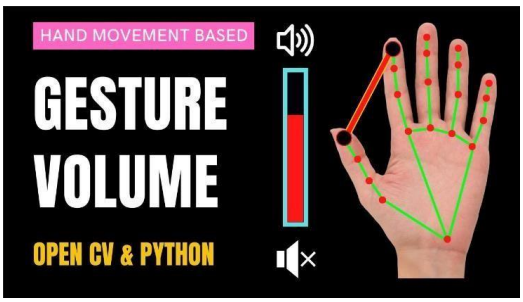


Fig. no. 1.1: Hand Gesture music control

Fig. no. 1.2: Mood analyzer

Additionally, a web-based Attendance Tracker System in PHP automates student attendance management using face recognition(shown in fig.no.1.3). This system records daily attendance subject-wise, initiated by scheduled class times, capturing and recognizing

Fig.no.1.3: Face Recognition Attendance System

students in real-time using deep learning techniques such as histogram of oriented gradients (HOG) for face detection.

Overall, these advancements illustrate the integration of advanced technologies across diverse domains, from social media analytics and emotion detection to gesture-based interfaces and educational management systems.

LITERATURE REVIEW :

Hasan [1] applied multivariate Gaussian distribution to recognize hand gestures using non-geometric features. The input hand image is segmented using two different methods [2]; skin color based segmentation by applying HSV color model and clustering based thresholding techniques [2]. Kulkarni [3] recognize static posture of American Sign Language using neural networks algorithm. Two methods are used for extraction the features; firstly by using the edge mages, and secondly by using normalized features where only the brightness values of pixels are calculated and other black pixels are neglected to reduce the length of the feature vector [4]. One notable study is the "Muni-Muni: Mood Analyzer and Screening Tests for the National Center of Mental Health using Sentiment Analysis"1. In this project, researchers developed a web and mobile application to promote mental health. The system utilized Sentiment Analysis, analyzing user text input to determine polarity (Positive, Neutral, or Negative). It was based on Cognitive Behavioral Therapy (CBT) and Dialectical Behavior Therapy (DBT), commonly used by behavioral specialists. The effectiveness of the software was tested through surveys with 40 students and 10 behavioral field professionals. [5]. Automatic face recognition (AFR) innovations have seen sensational enhancements in execution over the previous years, and such systems are presently generally utilized for security and business applications. A system for human face recognition for an association to stamp the attendance of the employees is been executed. So Smart Attendance utilizing Real Time Face Recognition is a genuine arrangement which accompanies everyday exercises of dealing with employees. The errand is exceptionally troublesome as the ongoing foundation subtraction in a picture is as yet a test. To identify ongoing human face are utilized and a basic quick Principal Component Analysis has used to perceive the faces identified with a high exactness rate. The coordinated face is utilized to stamp attendance of the representative. Our system keeps up the attendance records of employees consequentlly[6]. Face recognition system is one of the biometric method of taking attendance, it is easier to use and working range is larger than others like, fingerprint, iris scanning, signature etc. Face recognition system is able to automatically detect a face from a live image. This involves extracts some features and then recognize it. Regardless of lighting, expression, illumination, ageing, transformations and pose, face recognition systems can detect any face without facing any problem. This paper describes a method of taking attendance using face recognition here we use Haar Cascade algorithm for face detection and Local Binary Pattern Histogram for face recognition. After recognition the system is able to put attendance automatically into a C.S.V with accurate time and date without any human interaction.[7]. This project proposes an automated attendance system using Raspberry Pi 3B+ with OpenCV/Python libraries

and a recognizer algorithm to assist faculty in taking attendance without any disruptions or time wastage. The system utilizes face recognition technology to save time and accurately identify and eliminate the chances of proxy attendance. This proposed system has the potential to be deployed in various domains where attendance tracking is crucial and servesas a key component of project objectives and design criteria. Upon meeting the project objectives and design criteria, it can be concluded that this project provides an engineering solution for universities and colleges to effectively track and manage attendance. Additionally, the proposed system ensures accuracy, eliminates manual errors, and saves valuable time. (Source: Ghalib Al-Muhaidhri [8] Javeed Hussain[9] "Smart Attendance System using Face Recognition" in the 2019 International Journal of Engineering Research & Technology (IJERT)). Face detection and recognition application is implemented using an open source computer vision called OpenCV. The paper doesn't venture into real time application though it has a future in real time detection. The essential algorithms that have been extensively used are Convolution Neural Network and VGG16 architecture along with Haar features. Moreover, significant changes are to be integrated in order to bring
an upheaval in the major image analysis modules such that the framework incorporates accuracy and robustness even when other variations on poses are imposed. Our framework has been evolved to detect and recognise faces in fixed time periods and mark the student's attendance according to the average value of the faces detected and recognised within those fixed time periods such that the online attendance system can't be evaded or cheated.[9].

In paper [9] propose employing a number of machine learning algorithms while taking factors including age, hypertension, body mass index, heart disease, average blood sugar, smoking status, and prior strokes into account in order to predict early stroke disease. These high features attributes have been used to train ten different classifiers, including LR, Stochastic Gradient Descent, DT, AdaBoost, Gaussian, Quadratic Discriminant Analysis, Multi Layer Perceptron, KNN, Gradient Boosting, and XGBoost, to predict strokes.For the paper cited in [10], data on patients with acute thalamic ischemic stroke (melanoma), atypical nevi (cancer risk), and common nevi (no cancer risk) were collected from five renowned hospitals in Bangladesh. In this research, we propose a Modified Boot-strap Aggregating (Bagging) method for categorizing patterns that is ensemble- based. Duncan et al. completed yet another investigation on the recognition of live facial expressions of emotion. In their "cite"b6 study, they proposed transfer learning from VGGs and achieved accuracy of 90.0 percent for training steps and 57 percent for testing. They used an unstandardized set of handcrafted data that was later prepossessed with the case's intended shape and size. F. Y. Shih et al.[11] presented image pattern recognition techniques to extract facial features from color images. They used mathematical morphology and PCA (Personal Component Analysis) technique to accomplish it. Three feature extraction technique approaches were suggested by Urvashi Bakshi and Rohit Singhal in [12]. Their feature-based method examines geometric correlations between regional characteristics like the nose and eyes. Numerous techniques, including eigenfaces, fisher faces, support vector machines, and hidden Markov models (HMM), are included in a holistic approach. Combining local and global features is what makes up a hybrid approach. They covered a variety of face detection and facial feature ex-traction techniques in their article. Emotion classification is the most challengingtask to implement as various people have various kinds of expressions. Reeshad khan and Omar sharif [13] in their review paper proposed various deep learning techniques to implement a more effective and improved emotion recognition.Tanjim Mahmud et al.system. They used a combination of LSTM-RNN approach to achieve better accuracy than the previous related work on emotion recognition. Dipika Raval and Mukesh Sakle [14] proposed methods to recognize emotion from various facial expression. They explicitly describe about the system overview which contains face detection and prepossessing, feature extraction and classification and alsothe techniques used for emotion recognition. Chuan-Yu Chang et al. [15] pro-posed a similar approach to achieve this. They calculated symmetry indexes area between left and right eye and mouth along with Local Ternary pattern (LTP) and Gabor filter to extract the ROI (Region of Interest). They classify stroke face using SVM, Random Forest and Bayes and showed the system accuracy was 100% for svm,95.45% for random forest and 100% for bayes to predict the stroke face from normal face. A drooping mouth recognition model was suggested by O. Foong, K. Hong, and S. Yong in [16] using facial landmarks like the corners of the mouth to determine the coordinates of the key points. They created a mobile platform software to recognize and detect mouth drooping.

Methodology :

The multi-modal interaction system can be explained by fuse-case model in fig. 4.01.The hand gesture recognition project aims to provide a natural and dynamic way to use hand gestures to adjust the volume on a device. For hand gesture recognition and volume control, the project effectively utilizes modules such as Pycaw, MediaPipe, and NumPy. The MediaPipe Hands library, which is used to infer 21 crucial 3D hand properties from a single frame, allows for the precise identification of hand landmarks. Using the camera to detect hand landmarks and calculate the distance between the tips of the index and palm fingers is the fundamental concept behind this project. After that, the volume range is connected to this distance, allowing for dynamic volume control in reaction to hand movements. The code uses MediaPipe Hands to recognize hand landmarks, Pycaw to control the system volume, and OpenCV to capture webcam video. MediaPipe Hands is used to locate hand landmarks in the collected frame after processing and RGB conversion. The mapping of volume levels is carried out by taking the measured distance between specific hand markers. The processed video frame is displayed along with visual feedback features like hand landmarks, circles on finger tips, a box around the index finger, and a volume meter. The project includes an exit command ('Spacebar') for user convenience. All things considered, the project provides a novel and interesting way to control a device's volume with hand movements. With the available Python version, controlling the system volume with hand gestures is made interesting and engaging. The code must execute numerous pip install instructions in order to build up the necessary libraries, which include OpenCV (cv2), MediaPipe, Pycaw, and NumPy. These libraries are necessary for processing video frames, identifying hand landmarks, and controlling system volume. The code starts by using the webcam to take a photo, converting it to RGB, and then transferring it to the MediaPipe Hands library for processing. Next, hand motions are identified using the landmark list returned by the MediaPipe Hands library for each hand in the frame. In conclusion, anyone interested in learning more about gesture detection technologies can start with this code. It is easy to use, flexible to fit different needs and scenarios, and well-documented. The study additionally demonstrates how the integration of diverse libraries and technologies may yield innovative solutions that enhance user engagement and experience.

Process of mood analysis takes place in following steps as described in fig 4.02.

To use the face recognition system, every student in the class needs to register first by providing necessary details. During the registration process, their images will be captured and stored in the system's dataset. In each class session, the system will detect faces from the live streaming video of the classroom. These detected faces will then be compared with the images stored in the dataset. If a match is found between the detected face and the stored images, the attendance will be marked for the respective student. The system architecture of the proposed system is given in fig. no. 4.04.

Figures:

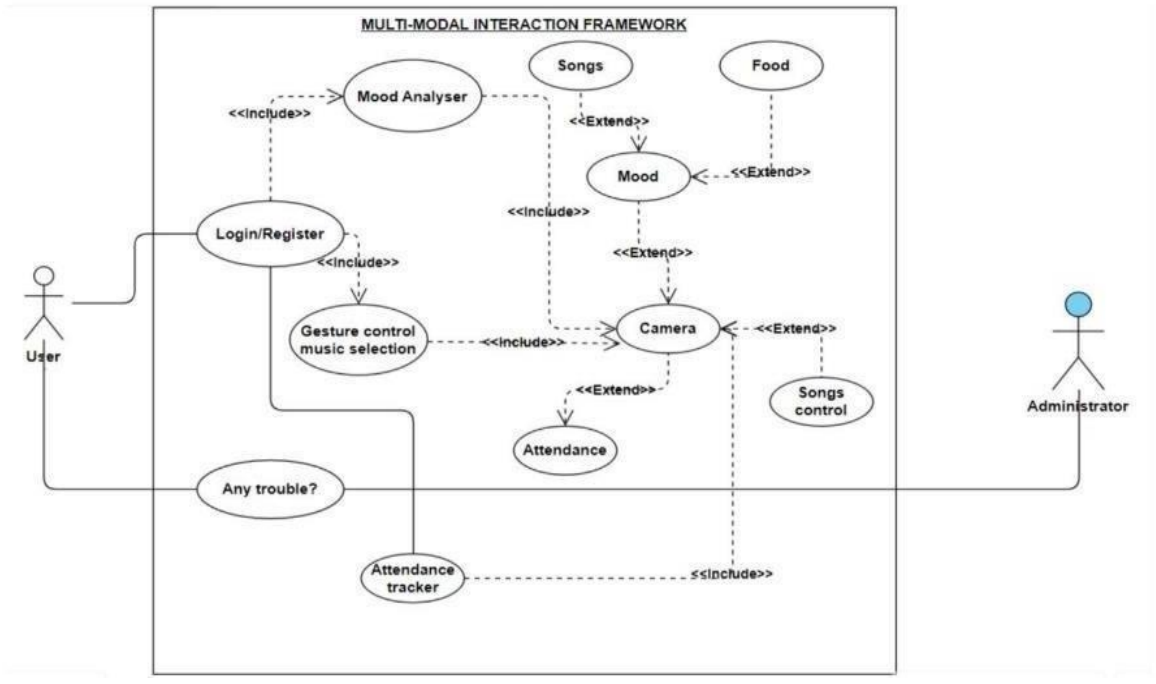


Fig. no. 4.1 Use-Case Diagram

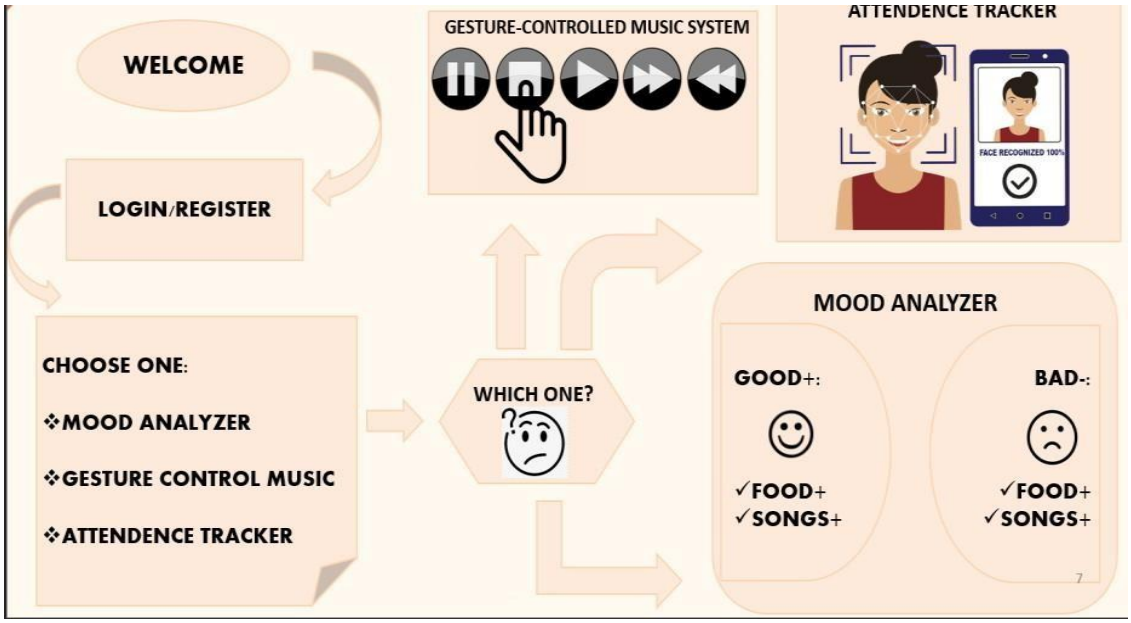


Fig. no. 4.2 Workflow Diagram

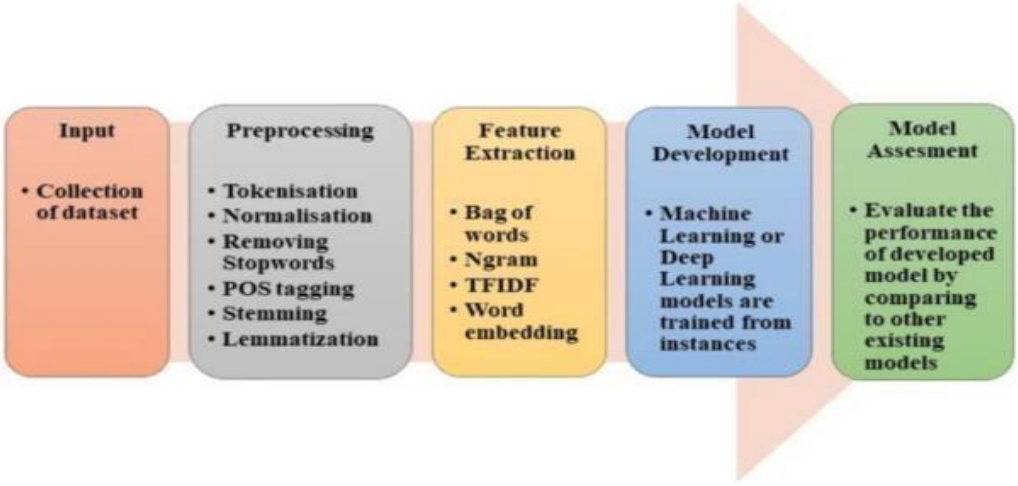


Fig. no. 4.03 Sentiment Ananlysis

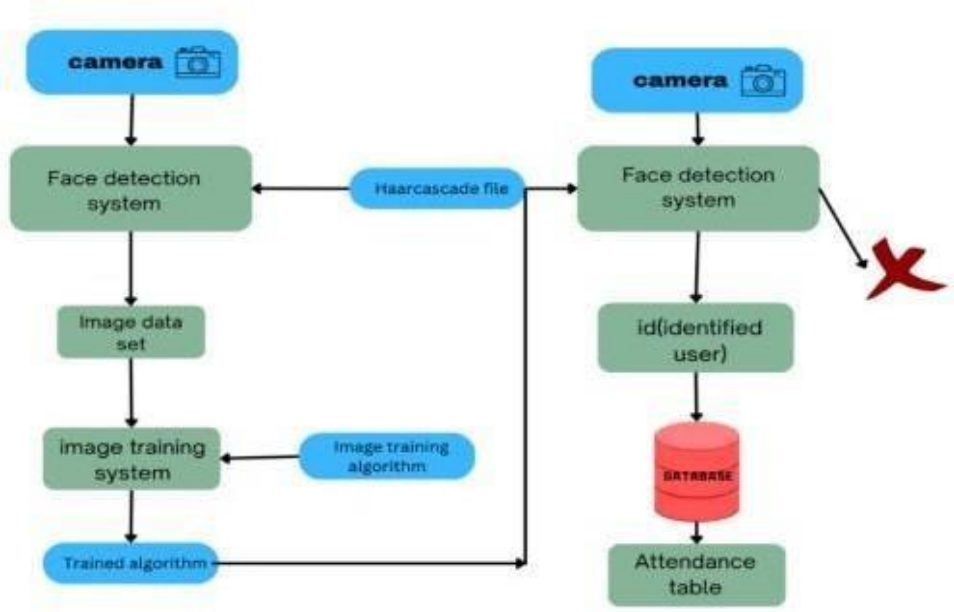


Fig. no. 4.04 Architecture of Face recognition Attendence system

TECHNOLOGY :

For creating website for the minor project we use, the trio of HTML (Hypertext Markup Language), CSS (Cascading Style Sheets), and JavaScript stands as the foundation for creating interactive and visually appealing websites. This report delves into the roles, interactions, and best practices associated with these three technologies.

1. **HTML:**

a) Structural backbone: HTML provides the structural framework for web content. Elements like , <p>, and <ul> define the document's structure, forming the basis for presenting information on the web.

b) HTML5 Features: HTML5 introduces advanced features such as the <canvas> element for graphics, <audio> and <video> elements for multimedia, and semantic tags that enhance both structure and accessibility.

2. **CSS:**

Styling and Layout: CSS is responsible for styling HTML elements, controlling layout, color, and typography.

Responsive Design: CSS enables responsive design through media queries and flexible grid layouts, ensuring a seamless user experience across various devices and screen sizes.

3. JavaScript:

Dynamic Interactivity: JavaScript brings interactivity to web pages, allowing developers to create dynamic content, handle user input, and communicate with servers.

4. PHP:

PHP is open- source, meaning it's free to use and has a large community of developers contributing to its improvement.

PHP can be embedded within HTML, making it easy to add functionality to web pages without needing to separate the code entirely.

PHP has built-in support for various databases, such as MySQL, PostgreSQL, and SQLite, allowing for easy database management and interactions.

5. OpenCV:

An open-source computer vision library that provides tools for image processing, feature extraction, and object tracking.

6. MediaPipe:

A framework by Google that simplifies building real-time applications with machine learning and computer vision components.

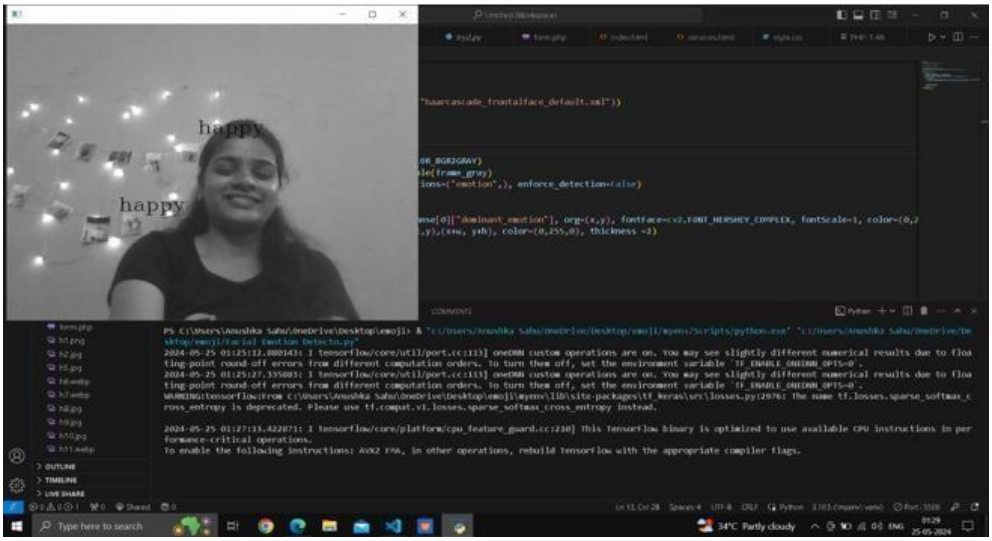
RESULTS AND DECLARATION :

Multi-modal Interaction Framework:

A multi-modal interaction framework combines different input modes (such as eye tracking, hand gestures, voice, etc.) to create a more versatile and intuitive user interface. Researchers have proposed methods that use wearable cameras to calculate gaze points or fixed remote cameras for eye tracking. These methods address issues like low accuracy, involuntary blinking, and pupil tremors1.The proposed multi-modal method combines feature tracking for global navigation and hand gesture recognition for detailed interactions.

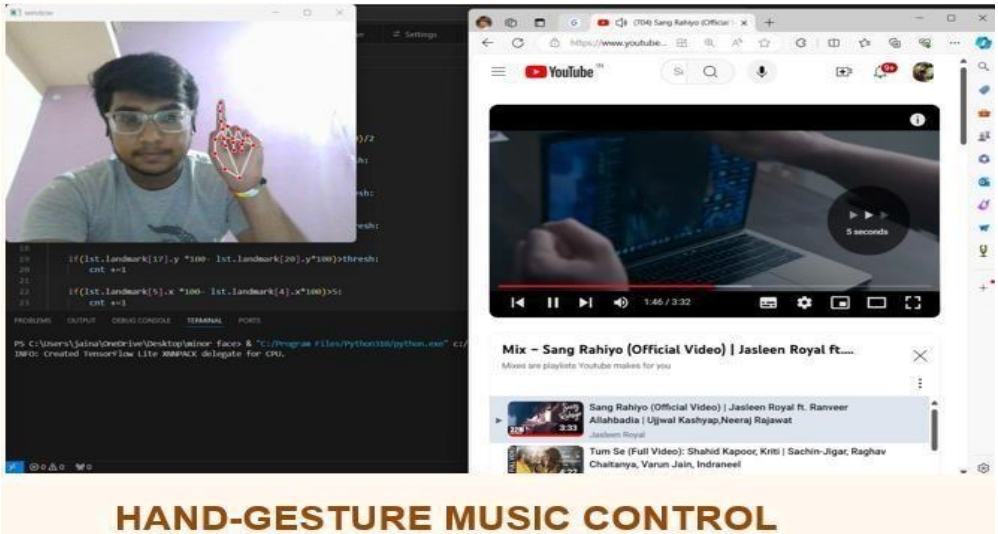
1. Mood Analyzer:

Mood analyzers typically use various cues (such as facial expressions, or physiological signals) to infer a person's emotional state.



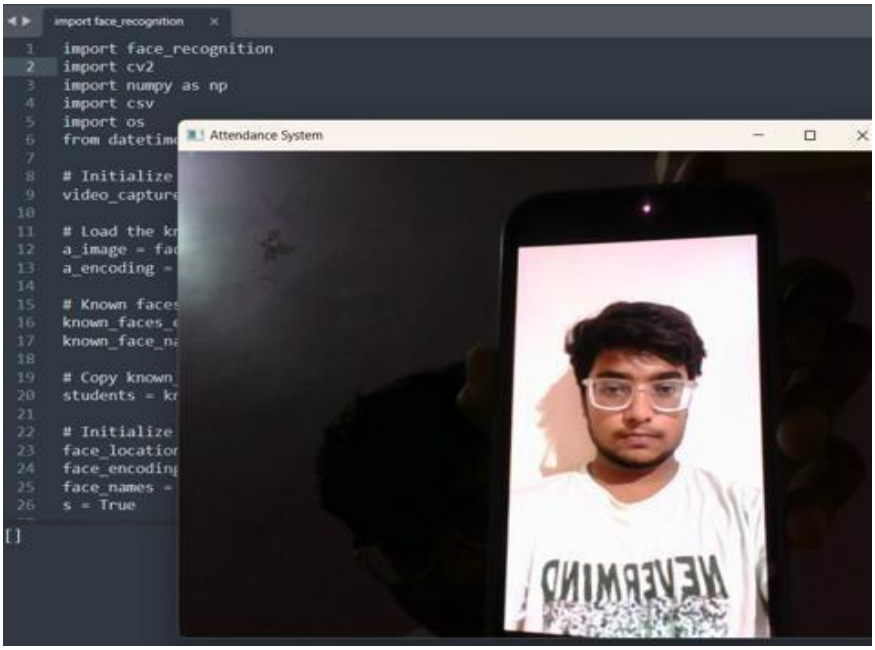
2. Gesture-Controlled Selection:

Gesture-controlled selection refers to using hand movements or gestures to interact with a system (e.g., selecting items, clicking buttons, etc.).



3. Attendance Tracking:

Attendance tracking systems are commonly used in educational institutions or workplaces to monitor attendance.



CONCLUSION :

Our aim with this article is to highlight multimodal interaction analysis (MIA) as a methodological approach that enables an illumination of the diverse array of communicative, interactional, social, and material resources students draw upon while engaging in science practices. When computers look at an image, what they 'see' is simply a matrix of pixel values. In order to classify an image, the computer has to discover and classify numerical patterns within the image .These patterns can be variable, and hard to pin down for multiple reasons. Several human emotions can be distinguished only by subtle differences in facial patterns, with emotions like anger and disgust often expressed in very similar ways. Each person's expressions of emotions can be highly idiosyncratic, with particular quirks and facial cues. There can be a wide variety of divergent orientations and positions of people's heads in the photographs to be classified. For these types of reasons, FER is more difficult than most other Image Classification tasks. However, well-designed systems can achieve accurate results when constraints are taken into account during development. The paper presents multimodal human-computer interaction using

speech and gesture recognition to develop a system for mouse movement and operation. The approach allows users to perform mouse navigation and various mouse operations without the need for physical contact with the system, we also crafted a powerful Face Recognition Attendance System using Python and the face recognition library. Imagine a world where your webcam transforms into a digital gatekeeper, effortlessly tracking and identifying faces. Our attendance system not only captures video frames but also performs a seamless face detection, comparing each face with a database of pre-stored faces.

FUTURE SCOPE :

Multimodal interaction systems aim to support the recognition of naturally occurring forms of human language and behaviour through the use of recognition-based technologies. The approach to analysis we present in this paper draws from theoretical perspectives grounded in the work and supports a dialogic view of human interaction in general Most of the multimodal interaction technologies are still immature for universal use. All the current approaches for multimodal interaction have their strengths and weaknesses and We aim to comprehensively and dynamically model and simulate the dimension, extent, and modes of multi-technology interactions. Based on a system view, we take a technology and propose a multi-modal and multi-dimensional technology interaction framework.

ACKNOWLEDGEMENT

We extend our heartfelt gratitude to our Director, (PIEMR, Indore) Dr. Manoj Kumar Deshpande, and our HOD (Dept. of CSE) Dr. Piyush Choudhary, for their unwavering support. We are thankful to Dr. Charu Vaibhav Verma (Project Guide), Mr. Atul Barve , and Professors, for their guidance. Finally, our thanks to our parents, group members, and all who supported us.

REFERENCES :

1. Mokhar M. Hasan, Pramod K. Mishra, (2012) "Features Fitting using Multivariate Gaussian. Distribution for Hand Gesture Recognition", International Journal of Computer Science & Emerging Technologies IJCSET. https://www.researchgate.net/publication/284626785 .
2. Mokhar M. Hasan, Pramod K. Mishra, (2012). "Robust Gesture Recognition Using Gaussian. Distribution for Features Fitting', International Journal of Machine Learning and Computing, Serrano, M., Nigay, L., Lawson, J.Y.L., Ramsay, A., Murray-Smith, R., Denef, S.: "The openinterface framework: a tool for multimodal interaction." In: CHI'08 Extended Abstracts on Human Factors in Computing Systems, pp. 3501–3506 (2008). https://www.researchgate.net/publication/284626785 .
3. V. S. Kulkarni, S.D.Lokhande, (2010) "Appearance Based Recognition of American Sign Language. Using Gesture Segmentation", International Journal on Computer Science and Engineering (IJCSE), Vol. 2(3), pp. 560-565. https://www.researchgate.net/publication/284626785 .
4. Mokhtar M. Hasan, Pramoud K. Misra, (2011). "Brightness Factor Matching For Gesture Recognition System Using Scaled Normalization", International Journal of Computer Science & Information Technology (IJCSIT). https://www.researchgate.net/publication/284626785.
5. Piccardi, M. (2004). "Background subtraction techniques: a review". IEEE International Conference on Systems, Man and Cybernetics, 4, 3099-3104. https://www.stepacademic.net/ijcsr/article/download/295/121.
6. T. Kamble and N. Mankar (2018) " Automated Human Resource and Attendance Management System Based On Real Time Face Recognition",IEEE. https://www.semanticscholar.org/paper/Automated-Human-Resource-and-Attendance-Management-Kamble-Mankar.
7. Shivangi Awasthi, 2Shubhangi Awasthi Facial Recognition Attendance System Using Python at 2022 International Journal of Research Publication and Reviews. https://www.researchgate.net/publication/370412970 .
8. Divya Pandey, 2Priyanka Pitale, 3Kusum Sharma Face Recognition Based Attendance System using Python at JETIR October 2020, Volume 7. https://www.researchgate.net/publication/370412970.
9. A Goyal, A Dalvi, A Guin, A Gite, A. Thengade "Online Attendance Management System Based on Face Recognition Using CNN" 2nd International Conference on IoT Based Control Networks and Intelligent System (ICICNIS 2021) (2021) https://www.researchgate.net/publication/379406929 .
10. Emon, M.U., Keya, M.S., Meghla, T.I., Rahman, M.M., Al Mamun, M.S., Kaiser, M.S.: Performance analysis of machine learning approaches in stroke prediction. In: 2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA). pp. 1464–1469. IEEE (2020)
11. https://www.researchgate.net/publication/379406929 .
12. Hakim, M.A., Hasan, M.Z., Alam, M.M., Hasan, M.M., Huda, M.N.: An efficient modified bagging method for early prediction of brain stroke. In: 2019 International Conference on Computer, Communication, Chemical, Materials and Electronic Engineering (IC4ME2). pp. 1–4. IEEE (2019)
13. https://www.researchgate.net/publication/379406929.
14. Shih, F.Y., Cheng, S., Chuang, C.F., Wang, P.S.: Extracting faces and facial features from color images. International Journal of Pattern Recognition and Artificial Intelligence 22(03), 515–534 (2008) https://www.researchgate.net/publication/379406929 .
15. Bakshi, U., Singhal, R.: A survey on face detection methods and feature extraction techniques of face recognition. International Journal of Emerging Trends & Technology in Computer Science (IJETTCS) 3(3), 233–237 (2014) . https://www.researchgate.net/publication/379406929 .
16. Khan, R., Sharif, O.: A literature review on emotion recognition using various methods. Global journal of computer science and technology 17(F1), 25– 27 (2017). https://www.researchgate.net/publication/379406929 .
17. Raval, D., Sakle, M.: A literature review on emotion recognition system using various facial expression. IJARIIE 1, 326–329 (2015) https://www.researchgate.net/publication/379406929 .
18. Chang, C.Y., Cheng, M.J., Ma, M.H.M.: Application of machine learning for facial stroke detection. In: 2018 IEEE 23rd International Conference on Digital Signal Processing (DSP). pp. 1–5. IEEE (2018). https://www.researchgate.net/publication/379406929 .
19. Foong, O.M., Hong, K.W., Yong, S.P.: Droopy mouth detection model in stroke warning. In: 2016 3rd International Conference on Computer and Information Sciences (ICCOINS). pp. 616–621. IEEE (2016). https://www.researchgate.net/publication/379406929 .