Continuous Face Authentication and Hand Gesturesfor OS Control using MobileNet V2 CNN

R Sreyas B. tech StudentDept.of CSE FISAT Angamaly

Rosmi JohnsonB. tech Student Dept.of CSE FISAT Angamaly

Sanjay Johny B.tech Student Dept.of CSE FISAT Angamaly

Soumya S Raj Assistant Professor Dept.of CSE FISAT Angamaly

Vignesh S KumarB.tech Student Dept.of CSE FISAT Angamaly

Abstract—**With the aid of a computer interface system that detects hand gestures, a person can control a computer through various hand movements. The system's camera captures these gestures, and they are subsequently analyzed and interpreted by the software using image processing methods. Additionally, facial recognition is routinely utilized as a security measure.**

Index Terms—**computer inteface system, detects hand gestures, image processing, facial recognition**

I. Introduction

A gesture refers to a bodily movement, typically made with the hands, that conveys an emotion or conveys information. Employing hand gestures can aid in the description of topics, indicating objects and people in the vicinity, and providing emphasis and structure to speech, while also providing insights into one's emotional state. A typical hand gesture recognition system comprises three stages: detection, tracking, and recognition. Although it remains an unresolved issue, there are numerous advantages to communicating with an electronic device via non-manual means. The two primary benefits of such a system are the convenience it provides and the preservation of hygiene standards. Users are able to communicate with one another via these devices without the need for a keyboard or mouse, which are major modes of disease transmission. By utilizing hand gesture recognition, the frequency of such interactions is reduced, ultimately decreasing the likelihood of contracting illnesses. The suggested system allows users to operate a computer through a hand gesture-based interface, utilizing various hand movements. The system's camera captures these gestures, which are subsequently interpreted by the software through the use of image processing techniques. In addition, facial recognition is regularly employed as a security measure. The user-friendly interface enables individuals to accomplish common tasks like menu navigation and selection without the need for conventional input devices such as a mouse or keyboard. This can be particularly advantageous for individuals with physical limitations or in virtual reality scenarios.

II. LITERATURE SURVEY

In 2021 Truong Quang Vinh and Nguven Tran Ngoe Anli, proposed a real-time face mask detection system using the YOLOv3 algorithm. To improve the processing speed and accuracy of the detector, the Haar cascade detector is used to detect the face region in the input images, and the region of interest (ROI) is then passed through the YOLOv3 to detect the face mask. The YOLOv3 algorithm employs a Darknet-53 as a feature detection backbone and contains 53 convolutional layers in its feature extractor. The YOLO algorithm is able to detect objects in a single input image with just one pass through the network. In this study, YOLOv3 was chosen for its high accuracy and real-time processing capabilities. The deep learning model was trained on a dataset of 7,000 samples.The face mask detection algorithm proposed in this study consists of three steps: preprocessing, face detection, and mask detection. The preprocessing step involves improving the quality of the input image by applying auto white balance and an unsharp filter for edge enhancement. The auto white balance is used to ensure consistent color in the input image frames across a range of color temperatures, while the unsharp filter enhances the edges in the input images.The face detection step involves identifying the face region. The Haar cascade classifier, proposed by Viola-Jones, is used for this purpose. This classifier extracts features using the Haar Wavelet technique with a 24x24 window size, removes redundant features using AdaBoost, and applies cascade classifiers to detect objects. The face regions detected by the Haar cascade classifier are then passed as input to the YOLOv3 algorithm to detect regions of face masks.The final step in the process is to use the YOLOv3 algorithm to determine whether the person is wearing a mask or not. The prediction process is configured at three scales: 13x13, 26x26, and 52x52. The first detection is made by the 82^{nd} layer and produces a feature map of size 13x13x27. The second detection is made by the 94^{th} layer and has a size of 26x26x27. The last detection is made by the 106^{th} layer and has a size of 52x52x27.The training process was completed in 75 epochs, and the accuracy of the proposed algorithm can reach up to 90.1%. The use of the YOLOv3 network allows the system to operate in real-time at 30fps.

In 2021 Min Wang, Hussein A. Abbass, Jiankun HuA designed a multi-modal biometric system that continuously verifies the presence of a logged-in user using both face images and electroencephalography (EEG) signals. The information from each modality is fused at the matching score level. For the face modality, matching scores are calculated based on the distances between eigenface coefficients. For the EEG signals, an event-related potential (ERP) modality is established using a simple ERP elicitation protocol and the calculation of cross-correlation similarities. The scores from the two modalities are normalized and fused using three schemes: the sum-score, max-score, and min-score schemes. The experiments showed that the individual variations found in the ERPs are detectable and can be used for continuous authentication, proving that ERP biometrics are feasible for user authentication and worthy of further research. The results also showed that combining ERP biometrics with face biometrics using a sum-score scheme performs better than either modality used in isolation, indicating the potential of integrating ERP into multimodal authentication systems.In this paper, the use of EEG as a biometric for continuous authentication in trusted autonomous systems is explored. EEG data, which is collected in many human-machine augmented trusted autonomous systems, is a natural continuous data source and has the potential to be used for authentication. A specific EEG biometric method based on P300 potential is designed and tested, and it is found to have an accuracy of above 90% when using a simple oddball paradigm. Additionally, the effectiveness of combining EEG with another biometric modality (face recognition) through the use of three fusion schemes (sum, max, and min) is examined. Results show that the integration of EEG with another biometric modality can reduce system vulnerability and improve performance.

In 2021 Tellaeche Iglesias, A. Fidalgo Astorquia, I.; V azquez G omez, J.I.; Saikia proposed a hand gesture recognition system based on convolutional neural networks and color images has been proposed. This system is robust against environmental variations and has real-time performance on embedded systems. It also addresses the main issues with previous approaches. A new CNN network with a small number of layers and neurons has been specifically designed for use on computationally limited devices. The system was found to have an average success rate of 96.92%, which is higher than the scores obtained by previous algorithms in the field.In the first stage of this system, an image is captured from a video sequence. In the second stage, the region containing the gestural interlocutor's hand is identified through segmentation. In the third stage, a thinning process is applied to this region of interest to reduce the amount of information to be processed and facilitate successful recognition. Points of interest are then identified on the thinned image, and the position of these points relative to the center of mass of the hand is used to represent the hand in a vector form, where the dimensions correspond to the number of fingers. Each component of this vector holds information on the inclination of a finger relative to the inclination of the forearm. Finally, the system compares this vector to a set of "base vectors" that were established during a training stage using the mean square error to determine if the vector is similar enough to any of the base vectors.The networks mentioned have demonstrated exceptional performance in a variety of challenges and competitions, including ILSVRC, the Kaggle platform, and The Low-Power Image Recognition Challenge.One of the main drawbacks of this technology is that it requires high computational power for most problems involving CNN. This limits the potential use of modern neural computation to systems with sufficient computational resources, such as embedded systems in industrial equipment.The proposed system has a gesture detection accuracy of 96.92% and is resistant to variations that can cause failure in other algorithms. For instance, the system is able to handle lighting variations and different skin tones, as long as the training dataset is properly generalized. These types of variations can often make skin segmentation approaches for gesture detection difficult. Additionally, the system has a small CNN with a Faster-RCNN architecture and is able to process a 1 Mpix image in an average time of 0.14 seconds using GPU computation, which can be considered real-time for systems that use gestures as a form of interaction.

In 2021 Mohammad Alsaffar, Abdullah Alshammari, Gharbi Alshammari, Tariq S Al- murayziq, Saud Aljaloud, Dhahi Al-shammari and Assaye Belay r described an electronic system that is able to recognize 12 different hand gestures made by a person in real time, under conditions with controlled lighting and background. The system can handle hand rotations, trans-lations, and changes in scale within the camera's field of view. It requires the use of an Analog Devices ADSP BF-533Ez-Kit Lite evaluation card. As a final step in the development process, it is recommended to display a letter corresponding to the recognized gesture. Alternatively, a visual representation of the proposed algorithm can be found in a personal computer's visual toolbox. In this work, a new alphabet is proposed that is based on the number of fingers and their position on the hand. This approach allows for a relatively large set of gestures with distinct structural differences. The fingers of the hand are modeled as binary inputs to the recognition system, with the thumb treated as the most significant bit. The system was developed using the ADV7183 video encoder, the parallel peripheral interface (PPI), the DMA controller, and the asynchronous memory SDRAM. The use of PPI and DMA allows for hardware-based subsampling of the image, which does not significantly impact the performance of the application but does improve processing speed by reducing memory accesses. The DMA is configured to generate an interrupt once the entire image has been stored in memory, interrupting the data transfer. This process results in a black and white image of the captured scene being stored in memory. The DMA interrupt routine handles image processing, after which the DMA is enabled again to transfer another image and the process is repeated.This work has resulted in an efficient tool that enables communication between a user and a machine, allowing for remote and real-time control of the machine. It also allows for the management of ports and other personal computer peripherals, which could enable future developments aimed at helping deaf and mute people participate in teleconferences. This tool has the potential to help this population reduce isolation by enabling them to interact with others who do not understand their language through a machine that can synthesize sound or generate text.

Nawaf O. Alsrehin ,Mu'tasem A. Al Taamneh in 2021 did a comparative study of statistical-based face recognition techniques including Eigen-faces, Fisher-faces, and Local Binary Patterns Histograms (LBPH) is presented. The study aims to evaluate the effectiveness and efficiency of these techniques using real database images. Face recognition has various applications and is expected to see significant technological advancements in the future. The study found that statistical techniques are effective in terms of accuracy and performance, but there is a noticeable difference in execution time between the techniques. Additionally, the paper offers a comprehensive comparison of 17 face recognition techniques that use artificial neural networks and categorizes them into understandable categories. It also provides suggestions for addressing the challenges faced by face recognition, such as pose variations, illumination, and low resolution or blurry images. The paper notes that there is currently no widely accepted face recognition method that can effectively address all of these issues. The study is significant for researchers in the field of recognition, software companies, and government security officials, as it provides insights that can inform the development of new face recognition approaches. The paper's comprehensive review of various techniques is noteworthy. The main focus of this study is to evaluate the accuracy and execution time of selected face recognition algorithms. The analysis reveals that the Fisherfaces and Eigenfaces techniques perform well in terms of accuracy, while the LBPH recognizer has an acceptable accuracy of 90% . Among the evaluated techniques, the LBPH recognizer produces the best results in terms of accuracy and execution time. It is also noted that the performance of Fisherfaces and Eigenfaces is similar due to their shared implementation aspects

In 2021 Rajkumar Janakiraman, Sandeep Kumar, Sheng Zhang, Terence Sim,developed a face verification system to constantly check if the current user of a computer console is still present. It employs a Bayesian framework, which uses a sliding window of verification data points (approximately ten seconds) as input to calculate the likelihood of the user's presence. If the system determines that the probability of unauthorized activity has fallen below a particular threshold, it may temporarily stop or slow down specific processes of the current user. This measure safeguards the computer and its resources against misuse by unauthorized users. The system may choose to halt or slow down processes without any conditions or based on specific types of system requests, particularly those that are security-related. The face verification system runs on the Linux 2.4.26 kernel and the Redhat 9.0 distribution. Upon login, the monitor system retrieves the user ID and loads a profile that outlines the appropriate actions to take if unauthorized activity is detected. The system uses a video box to record periodic images and run face detection and verification algorithms. The monitor acts as the central coordinating entity, keeping track of the user's presence probability and communicating with the kernel level driver (drv) to manage user processes. The changes to the Linux kernel to allow process preemption are minimal and limited to specific areas. Optimizations have demonstrated that modern Pentium processors (2.4GHz with 512MB RAM) can achieve a frame rate of 25 fps during face tracking.

In 2021 Jie Zhu, Zhiqian Chen Facial detection is crucial for interactive user interfaces and is a significant research focus in computer vision. To achieve a fully automated facial image analysis system, robust and efficient facial detection algorithms are necessary. This research paper proposes a real-time facial detection system that utilizes the AdaBoost algorithm and Haar-like features. Haar-like features are one of the most effective and efficient methods for facial recognition, and they can be learned through the AdaBoost algorithm. They are particularly advantageous because they can be rapidly calculated using integral images. The AdaBoost algorithm selects a suitable set of weak learners to construct a robust classifier. The accuracy of the facial detector was evaluated on the challenging CMU face dataset, which contains various faces and lighting conditions. The research demonstrated the effectiveness of the facial detector on this difficult dataset. The facial detection system is designed as follows: (1) gathering a dataset of faces and non-faces for training; (2) creating a collection of weak classifiers based on Haar-like features; (3) constructing a robust classifier (facial detector) using boosting methods; (4) installing a digital camera to capture input images and developing software to link the hardware. The system's performance is demonstrated through several experiments, indicating both high accuracy and speed.

In 2021 Luo Jiang , Juyong Zhang , Bailin Deng introduced a fresh approach to face recognition that utilizes facial attributes in the training process by using a convolutional neural network (CNN). The method involves the use of an attribute-aware loss function to guide the CNN learning process and promote the development of more distinguishable features that correspond to the attributes. An attribute-aware loss function is a loss function in machine learning that takes into account the inherent characteristics or attributes of the training data. This innovative loss function has the potential to resolve the problem of uneven sampling in the training dataset and enhance the accuracy of face recognition. A face recognition model was trained using RGB-D and a vast dataset with more than 100,000 identities. The experimental outcomes demonstrated the efficacy of the new attribute-aware loss function and the model's good generalization capacity. This research is the first to use non-facial attributes that are invariant to the capture environment to regulate the feature mapping of face recognition. The attribute-aware loss term is added to the classification loss, such as softmax. Despite only using gender, age, and ethnicity attributes for regulation in this study, significant enhancements were still observed in the experimental results.

In 2022 JUN XU HANCHEN WANG,JIANRONG ZHANG,LINQIN CAI presented a robust method for recognizing static and dynamic hand gestures using RGB-D data. For static hand gesture recognition, the method begins by extracting the hand gesture contour and identifying the palm center using the Distance Transform algorithm. The fingertips are then located using the K-Curvature-Convex Defects Detection algorithm. The distances of the pixels on the hand gesture contour to the palm center and the angles between the fingertips are used as auxiliary features to create a multimodal feature vector. A recognition algorithm is then applied to robustly identify the static hand gestures based on this feature vector. For dynamic hand gesture recognition, this method combines the Euclidean distance between hand joints and the shoulder center joint with the modulus ratios of skeleton features to create a unifying feature descriptor for each dynamic hand gesture. An improved dynamic time warping algorithm (IDTW) is then used to recognize the dynamic hand gestures based on these descriptors. Finally, extensive experiments were conducted to test and validate the static and dynamic hand gesture recognition algorithms and demonstrate the potential for a low-cost, real-time application of natural interaction with virtual environments using hand gestures.After evaluating the system, it was found to have an average performance of 97.4%, with 96% accuracy for static and dynamic gestures, respectively. However, the current interactive gestures are relatively simple and mainly focus on the scene and vision of the virtual environment. Additionally, this method may not be suitable for people with hand deformities. Alternative HCI methods, such as speech and brain-computer interface technology, may be more appropriate for these individuals. The system also currently uses Kinect to obtain RGB-D data, and future work will involve testing other depth sensors, such as Intel RealSense, Leap Motion Controller, and ASUS Xtion, with the proposed algorithms for static and dynamic hand gesture recognition.

In 2021 Harshala Gammulle,Sridha Sridharan,Clinton Fookes introduced a gesture recognition framework called Temporal Multi-Modal Fusion (TMMF) which can be used to detect and classify multiple gestures in a video using a single model. This approach is able to learn the transitions between gestures and non-gestures without the need for a pre-processing step to identify individual gestures. This is achieved through the use of a multi-modal fusion mechanism that allows for the integration of information from multiple inputs and is scalable to any number of modes.In addition to the TMMF framework, two models were proposed for mapping features: Unimodal Feature Mapping (UFM) for uni-modal features and Multi-modal Feature Mapping (MFM) for fused multi-modal features. To further improve performance, a mid-point based loss function was introduced that helps the model learn natural gesture transitions by encouraging smooth alignment between the ground truth and the prediction. The proposed framework was shown to be effective at handling variable-length input videos and outperforming the state-of-the-art on three challenging datasets: EgoGesture, IPN hand, and ChaLearn LAP Continuous Gesture Dataset (ConGD).Ablation experiments were conducted to demonstrate the importance of the various components of the proposed method, which outperformed state-of-the-art systems on all three datasets by a significant margin. The model has potential applications in various real-world domains that require classification of continuous data, and the fusion model can be used in other fusion problems involving video or signal inputs, with or without the UFM or MFM blocks.

In 2022 Gibran Benitez-Garcia, Yoshiyuki Tsuda, Norimichi Ukita proposed a Continuous Finger Gesture Spotting and Recognition Based on Similarities Between Start and End Frames. Recognizing finger gestures can be difficult because (i) gesture and non-gesture frames may be similar, (ii) it can be challenging to identify the temporal boundaries of continuous gestures, and (iii) there is significant intraclass variability in the duration of gestures, which can make it difficult to use finger gestures to control in-car devices.To tackle the challenges of (i) distinguishing gestures from non-gestures and (ii) identifying the temporal boundaries of continuous gestures, a gesture spotting method was proposed that segments continuous gestures by detecting boundary frames and evaluating the similarity of the hand between the start and end boundaries of each gesture. To address the issue of (iii) intraclass variability in gesture duration, a gesture recognition method was introduced that is based on temporal normalization of features extracted from the set of spotted frames. The proposed gesture recognition method normalizes the features of any gesture to enable representation with a fixed number of features. To ensure real-time performance, the method is based on compact deep neural networks. Additionally, the effectiveness of the method was demonstrated through an alternate approach that uses hand-crafted features, which can perform in real-time without requiring a GPU.To further evaluate the proposed approach, a realistic driving setup was used to capture a dataset of continuous finger gestures, including over 2,800 instances in untrimmed videos covering safety driving requirements. The proposed approaches were able to run at 53 fps and 28 fps on a GPU and CPU, respectively, when tested on this dataset, which is around 13 fps faster than previous works while maintaining better performance (at least 5% higher mean tIoU).

In 2021 Wenjin Zhang,Jiacun Wang,Fangping Lan proposed a new way for Dynamic Hand Gesture Recognition Based on Short-Term Sampling Neural Networks. Hand gestures are a natural way for human-robot interaction. Vision based dynamic hand gesture recognition has become a hot research topic due to its various applications. The paper presents a novel deep learning network for hand gesture recognition. The network integrates several well-proved modules together to learn both short-term and long-term features from video inputs and meanwhile avoid intensive computation. The study proposed a short-term sampling neural network for dynamic hand gesture recognition. Each hand gesture was captured as a video input. Each video input was divided into a fixed number of groups of frames. A sample frame was taken randomly from each group of color and optical flow frames. The samples are fed into ConvNets for feature extraction and the features are fused and passed to an LSTM for prediction of the class of a hand gesture input.The developed system based on the new model was trained and evaluated on the Jester dataset. To test the robustness of the new approach, the Jester dataset was zoomed out by coping with the boundary of original images. An average accuracy of 95.73%, 95.69% We achieved on the Jester dataset and the “zoomed-out” Jester dataset, respectively. This model was also tested on Nvidia dataset and achieved a great performance with a classification accuracy of 85.13%. The results of these experiments show that the short-term sampling neural network model is effective for hand gesture recognition.

III. COMPARITIVE STUDY

Table 1 and Table 2 shows the comparative study of the above discussed papers,based on their methodology,advantages and disadvantages.

IV. PROBLEM STATEMENT

To implement a system that minimises the hardware input devices to provide a more user-friendly computer interface that make use of human hand gestures and also provide a continuous face verification system that will only allow the registered user to control the system.

Current methods for authenticating users and controlling operating systems (OS) often rely on passwords, PINs, or physical buttons, which can be inconvenient and prone to security breaches. There is a need for more intuitive and secure methods of user authentication and OS control. This project aims to develop and evaluate a system that utilizes continuous face authentication and hand gestures as a means of identifying users and controlling an OS, using the MobileNet V2 CNN model. The goal is to determine the feasibility and effectiveness of this approach, and to identify any potential security risks or usability issues

There are many applications where hand gesture can be used for interaction with systems like, videogames, controlling UAV’s, medical equipment’s, etc. These hand gestures can also be used by handicapped people to interact with the systems. Classical interactions tools like keyboard, mouse, touchscreen, etc. may limit the way we use the system. All these systems require physical contact, in order to interact with system. Gestures can interpret same functionality without physically interacting with the interfacing devices. The problem lies in understanding these gestures, as for different people, the same gesture may look different for performing the same task.

V. OBJECTIVE

The objective of this project is to develop a system, which facilitates the user to interact with the Operating System Windows 10 using Hand Gestures. The users can freely use various hand gestures to do certain actions in the OS like navigation , selection etc. This will reduce the use of various hardware devices like mouse to operate the computer. The gestures will provide a more user friendly way to operate the computer.The touchless way of interaction with the computer also provides a hiegenic and more immersive way for computer interaction.

The system will accurately identify a user’s identity in real-time and control an operating system (OS) using hand gestures, with the goal of providing an intuitive and convenient way for users to interact with their devices.It will also investigate the potential security risks and vulnerabilities of using continuous face authentication and hand gestures as a means of controlling an OS. and examine the usability and user experience of the proposed system, and gather feedback from test users to inform future improvements. It can also identify any potential real-world applications of the continuous face authentication and hand gesture control system, such as in the fields of home automation, gaming, or accessibility.

The proposed system will also provide a security feature that will continuously verify the user and authenticate him. When the user goes away from the screen the system will be locked and no other person can use the system.

VI. COMPARISON TABLE

TABLE 1 Comparison Table Based on Face Authentication and Human Computer Interaction

Based on Face Authentication			Based on Human Computer Interaction		
Title	Real-Time Face Mask Detector Using YOLOv3 Algorithm and Haar Cascade Classifier	Continuous Authentication Using EEG and Face Images for Trusted Autonomous Systems	Title	Gesture-Based Human Machine Interaction Using RCNNs in Limited Computation Power Devices	Human-Computer Interaction Using Manual Hand Gestures in Real Time
Face Detection	Haar cascade detection	Cascade object detector	Journal	MDPI,2021	Hindawi Computational Intelligence and Neuroscience,2021
Methodology	YOLOv3	EEG Biometric method	Methodology	Region based CNN Approach	PPI together with the DMA
Accuracy	90.1% Accuracy	90% Accuracy	Accuracy	More Accurate	Less Accurate
Preprocessing	Auto white balance, edge enhancement	Brightness, contrast	Efficiency	More efficient	Less efficient

TABLE 2 Comparison Table Based on Face Recognition

Title	Face Recognition Techniques using Statistical and Artificial Neural Network: A Comparative Study	Using Continuous Face Verification to Improve Desktop Security	Real Time Face detection System Using Adaboost and Haar-like Features	Robust RGB-D Face Recognition Using Attribute-Aware Loss
Methodology	Eigen-faces, Fisher-faces, and Local Binary Patterns Histograms (LBPH)	Bayes' Classifier	Adaboost and Haar-like Features	Attribute-aware loss function for CNN based face recognition
Advanatges	From the analysis results, the performance accuracy seems very good for Fisherfaces and Eigenfaces techniques and an acceptable result (90%) for the LBPH recognizer.	The ptimizations have shown that even 25 fps is possible on a modern Pentium (2.4GHz with 512MB RAM) with "face tracking".	High performance in both accuracy and speed was experienced in the developed system.	The experimental results demonstrate the effectiveness of our novel attribute-aware loss, and the good generalization ability of our trained RGB-D face recognition model.
Disadvantages	Analysis based on different pose variation, illumination, blurry and low-resolution images wasn't considered.	The usability of the system in a public setting is yet to develop.	May not perform as well as some other methods on very large or small faces	This work only use gender, age, and ethnicity attribute for regularization.

TABLE 3 Comparison Table Based on Gesture Recognition

Title	Robust Hand Gesture Recognition Based on RGB-D Data for Natural Human–Computer Interaction	TMMF: Temporal Multi-Modal Fusion for Single-Stage Continuous Gesture Recognition	Continuous Finger Gesture Spotting and Recognition Based on Similarities Between Start and End Frames	Dynamic Hand Gesture Recognition Based on Short-Term Sampling Neural Networks
Journal	Digital Object Identifier 10.1109/ACCESS.2022.3176717	IEEE Transactions On Image Processing, VOL. 30, 2021	IEEE Transactions On Intelligent Transportation Systems, VOL. 23, NO. 1, JANUARY 2022	IEEE/CAA Journal Of AUTOMATICA SINICA, VOL. 8, NO. 1, JANUARY 2021
Methodology	RGB-D data based recognition method	Temporal Multi-Modal Fusion (TMMF)	Temporal stacking for Deep Network & Temporal Normalization for Hand-crafted Based Approach	Short-term sampling neural network
Advanatges	In the experiment evaluation, the system achieved an average performance of 97.4%, 96% for static and dynamic gestures, respectively.	It can be applied to varying length gesture videos, and is able to perform the gesture detection and classification in a single direct step.	The deep-network approaches presented the smallest model size, while being significantly faster than previous works.	An average accuracy of 95.73%, 95.69% was achieved on the Jester dataset and the “zoomed-out” Jester dataset, respectively.
Disadvantages	Current interactive gestures mainly focus on the scene and vision of virtual environment, which are relatively simple.	The limitation of frame-wise metrics is that they do not capture the segmentation behaviour of continuous data.	Few detection errors are attributed to issues of motion registration by the optical flow approach.	According to confusion matrix, model tends to make mistakes in identifying “turning hand clockwise” and “turning hand counterclockwise”.

VII. Methodology and Design

The system is designed to be intuitive and easy to use, allowing users to perform common tasks, such as navigating through menus and selecting options, without the need for a traditional input device, such as a mouse or keyboard

A. *Proposed System*

The proposed system allows a user to control a computer by using various hand gestures. These gestures are captured by using camera, and are then interpreted by the system software.

Project Flow

- Input video feed from camera
- Detect hand from video feed using media pipeline
- Pass the ROI to model (mobilenetv2)
- Classify the gesture based on model output
- Change controls based on the hand gesture

Requirements

Software Requirements:

- windows 10 (Operating System)
- Programming language - python
- Python Anaconda
- Spyder

Hardware Requirements:

- Intel i3 or above (processor)
- 4GB or above RAM memory
- Hard Disk : 500 GB or above
- webcam (internal/external)

B. *Architectural Diagram*

In this architecture, the media pipeline captures video frames from a camera and processes them using the Haar cascade and LBPH algorithms to detect and track faces and continuously verify users. The detected ROI of hand by Media Pipeline are then passed to the MobileNet V2 CNN model for further processing, which is trained to perform gesture recognition tasks. The output of the MobileNet V2 CNN model is used by the gesture recognition modules to control the OS accordingly.

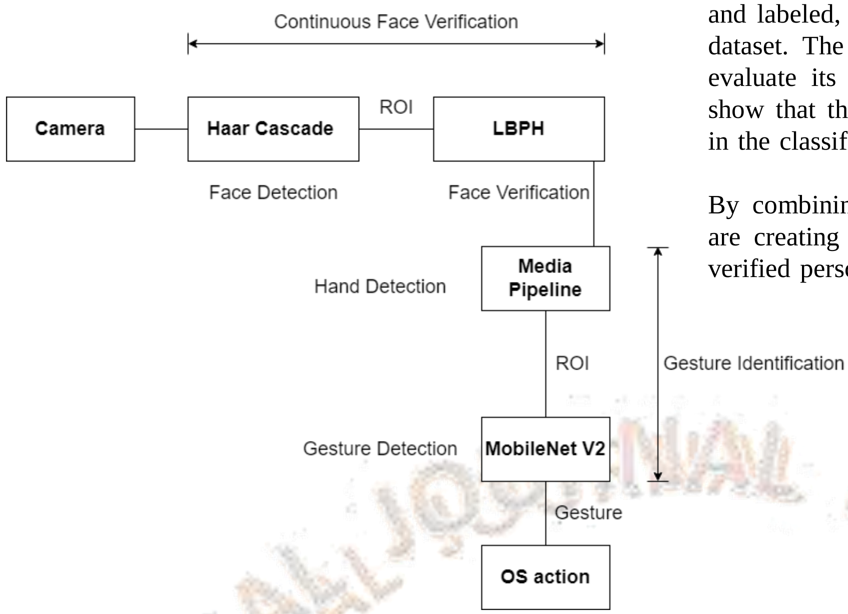

Fig. 1. Schematic Diagram of Proposed System

C. Logical Design

Haar cascades are machine learning classifiers that are trained to identify features in an image. They are used to detect objects in images by training on positive and negative samples and using the resulting classifier to detect objects in new images.
LBPH, on the other hand, is a texture-based method that encodes local spatial relationships between pixels in an image. It converts an image into a collection of local binary patterns and uses them to create a histogram, which is then compared to histograms of known faces to determine the identity of an unknown face. Together, Haar cascades and LBPH can be used to create a robust and reliable face recognition system.

We are using media pipeline library for detecting hand from from video. The system utilizes a combination of color and texture features to identify hand regions, which are then verified using a support vector machine classifier. The system was tested on a dataset of videos containing various hand gestures and achieved an good accuracy. The use of the Media Pipeline Library allows for real-time processing of the video stream, making the system suitable for use in interactive applications.

The classification of hand gestures using MobileNet V2 is a deep learning approach to recognizing and identifying different hand gestures in real-time. This method utilizes the MobileNet V2 convolutional neural network (CNN) to extract features from hand images and classify them into predefined categories. The mobile net architecture is chosen due to its efficient computation and ability to run on mobile devices. In this approach, a dataset of hand images is first collected and labeled, and the MobileNet V2 model is trained on this dataset. The model is then tested on a separate dataset to evaluate its accuracy in classifying hand gestures. Results show that the MobileNet V2 model achieves high accuracy in the classification of hand gestures.

By combining face recognition and hand gesture we are creating a secure system that can only be operated by verified person

Fig. 2. Logical Diagram

VIII. CONCLUSION

The spreading of Covid-19 virus have alarmingly taught us the importance of minimizing the touch interactions with the various material things around us. Computers are devices that involves constant touch interaction with the hardware devices. This is in turn can be a cause of concern for the spreading of infectious diseases or viruses as seen in the case of Coivd-19. Also personal computers nowadays provide only a single authentication system that verifies the user only in the starting phase. So if someone outside the organization has managed to crack that initial verification phase it could lead to various problems like potential data theft and further loss. A continuous monitoring system that could verify the user continuously will prevent such break-ins into the computer.

Another aspect to be looked is the ease of use for a computer illiterate person. At the initial stage such a person will be confused when he has to use a couple of hardware devices to operate the computer. If a system which is more user friendly to use is developed they could forget about the hassles of the hardware devices and use the computer system seamlessly.

A gesture Based system that can act according to certain hand gestures can actually smoothen the process of human computer interaction. Also the virtual reality and virtual environments like metaverse are gaining popularity and in such an environment the use of traditional hardware devices will be an outdated approach and gesture based interaction will provide more immersion.

The proposed system accommodate every problem raised above by enabling the user to control the OS namely a Windows 10. They can use their hands to perform various operations like moving the mouse around , navigate through the menus and select different options without the use of any hardware devices like mouse. Moreover the system uses a continuous face verification that will be ongoing the entire time and the system will lock itself when any unauthorised user comes in front of the system.

The system uses Media Piepline Library for the purpose of identifying the Region of Interest(ROI) that is hand from the video input and this is fed to pre trained MobileNetV2 model whiich uses CNN for the classification of hand gestures. For the continuous face verification Haar cascades and Local Binary Patterns Histograms(LBPH) are used. The Haar cascades will identify the ROI that is face and the LBPH authenticates the face. The Haar cascade is a rapid detection system and that will ensure the continuous face verification happens without any time delay.Together, Haar cascades and LBPH can be used to create a secure and fast continuous face verification system.

IX. REFERENCES

[1] Truong Quang Vinh and Nguven Tran Ngoe Anli, "Real-Time Face Mask Detector Using YOLOv3 Algorithm and Haar Cascade Classifier,"

[2] Min Wang, Hussein A. Abbass, Jiankun Hu, "Continuous Authentication Using EEG and Face Images for Trusted Autonomous Systems"

[3] Tellaeche Iglesias, A. Fidalgo Astorquia, I.; V´azquez G´omez, J.I.; Saikia, S , "Gesture-Based Human Machine Interaction Using RCNNs in Limited Computation Power Devices" , Sensors 2021, 21, 8202

[4] Mohammad Alsaffar, Abdullah Alshammari, Gharbi Alshammari, Tariq S Almurayziq, Saud Aljaloud, Dhahi Alshammari and Assaye Belay, "Human-Computer Interaction Using Manual Hand Gestures in Real Time ",Hindawi Computational Intelligence and Neuroscience Volume 2021, Article ID 6972192

[5] Rajkumar Janakiraman, Sandeep Kumar, Sheng Zhang, Terence Sim, "Using Continuous Face Verification to Improve Desktop Security"

[6] Harshala Gammulle,Sridha Sridharan,Clinton Fookes , "TMMF: Temporal Multi-Modal Fusion for Single-Stage Continuous Gesture Recognition"

[7] Gibran Benitez-Garcia, Yoshiyuki Tsuda, Norimichi Ukita,"Continuous Finger Gesture Spotting and Recognition Based on Similarities Between Start and End Frames"IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS, VOL. 23, NO. 1, JANUARY 2022

[8] Wenjin Zhang,Jiacun Wang,Fangping Lan , "Dynamic Hand Gesture Recognition Based on Short- Term Sampling Neural Networks" IEEE/CAA JOURNAL OF AUTOMATICA SINICA, VOL. 8, NO. 1, JANUARY 2021

[9] Nawaf O. Alsrehin ,Mu'tasem A. Al Taamneh, "Face Recognition Techniques using Statistical and Artifi- cial Neural Network: A Comparative Study"

[10] Rajkumar Janakiraman, Sandeep Kumar, Sheng Zhang, Terence Sim , "Using Continuous Face Verification to Improve Desktop Security"

[11] Jie Zhu, Zhiqian Chen, "Real Time Face detection System Using Adaboost and Haar-like Features"

[12] Luo Jiang , Juyong Zhang , Bailin Deng , "Robust RGB-D Face Recognition Using Attribute-Aware Loss"