Virtual Private Voice Assistants with Face Recognition

Raj Gupta, Bivin Joseph | Hussain Ahmed Chowdhury | SCOPE

Introduction

Enhanced security measures, such as two-factor authentication and liveness detection, are being implemented to safeguard WhatsApp messaging services and user privacy. Innovative features like WhatsApp Timed Messaging and improved virtual personal assistant (VPA) functionalities are introduced to enhance user experience. Future developments aim to improve voice recognition technology, strengthen security features, and explore new functionalities to meet evolving user demands.

Motivation

AI-driven virtual personal assistants (VPAs) revolutionize daily tasks but lack strong security and adaptability. Our "Virtual Private Voice Assistants with Face Recognition" initiative focuses on privacy and security through advanced technologies like voice and face recognition. Motivated by security concerns, user experience enhancement, and future development prospects, we aim to implement features such as liveness detection and two-factor authentication while continually improving functionality to meet evolving user needs.

SCOPE of the Project

Our project aims to enhance AI-powered virtual personal assistants (VPAs) by integrating voice and face recognition technologies to improve security and user experience. We will implement features like liveness detection and two-factor authentication to safeguard user data and enhance usability. Continuous refinement of functionality will ensure adaptation to evolving user needs and technological advancements in voice recognition and security.

Methodology

1. Face Recognition

Face Detection: The suggested approach detects faces by using an oriented gradient histogram . HOG is a feature descriptor that is also known as the histogram of oriented gradients. frequently used for object detection in computer vision and image processing. The technique counts the occurrences of gradient orientation in a certain region of a picture. The structure or shape of an entity is highlighted in the HOG description. It computes the features using both the angle and magnitude of the gradient, in contrast to conventional edge descriptors. For every region of the image, it creates a histogram based on the gradient's magnitude and directions. The gradient matrices are split into 8x8 cells to create a block once each pixel's gradient is determined. A 9-point histogram is computed for each block. A 9-point histogram consists of nine bins with a 20-degree angle range between each bin. A histogram with nine bins is assigned values. These nine-point histograms can each be displayed as a histogram with bins that correspond to the gradient strength for that specific bin.

Face Recognition: The suggested method recognises faces using a linear SVM classifier and a deep neural network. Rather than instructing the network to identify objects in images, we train it to produce measures for every face. The stages that the algorithm takes to identify a face are listed below. It first uses the HOG method to encode the image in order to obtain a reduced version of the image. Using this reduced form of the image, it finds the region that most closely fits a generic HOG encoding of a face. It establishes the facial posture by recognising the important facial landmarks. Once we've positioned them so that the mouth and eyes are in the centre, use those markers to distort the image. The 128 measurements are then saved after the centred face image is run through a neural network that has been trained to evaluate facial features. Next, in order to identify the closest match, a linear SVM classifier is trained to compare the 128 measurements of the test image with the measurements of every image in the dataset.

2. Speech Recognition

Algorithms The Web voice API recognises user voice with deep neural networks, Viterbi search, and Perceptual Linear Prediction characteristics. Equal loudness pre-emphasis and cube-root emphasis are two feature extraction techniques used in Perceptual Linear Prediction (PLP), a speech recognition system. An technique known as Viterbi search finds the maximum probability of the most likely sequence of concealed states, leading to a path known as the Viterbi path. It leads to a series of recorded occurrences within the framework of Hidden Markov models (HMM).

3. Features

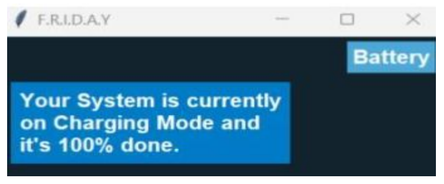
The typical functions that a voice assistant can perform include the following: (i) Internet browsing; (ii) system control (such as volume control, battery life, etc.) (iii) Mathematical functions (factororial, logarithm, trigonometry, and basic operations, among others); (iv) Look up a definition in a dictionary; (V) Exploring YouTube; (vi) reading the news

4. Novel Features of the VPA

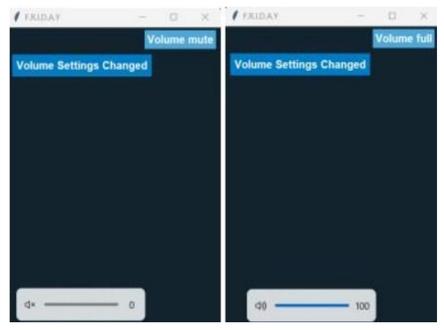
WhatsApp Timed Message

Results

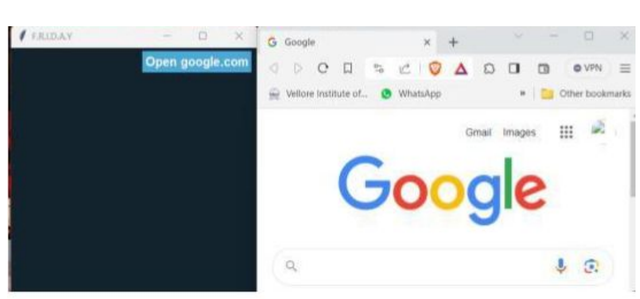
(i)System info



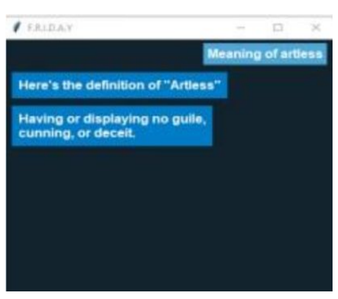
(ii)Battery info



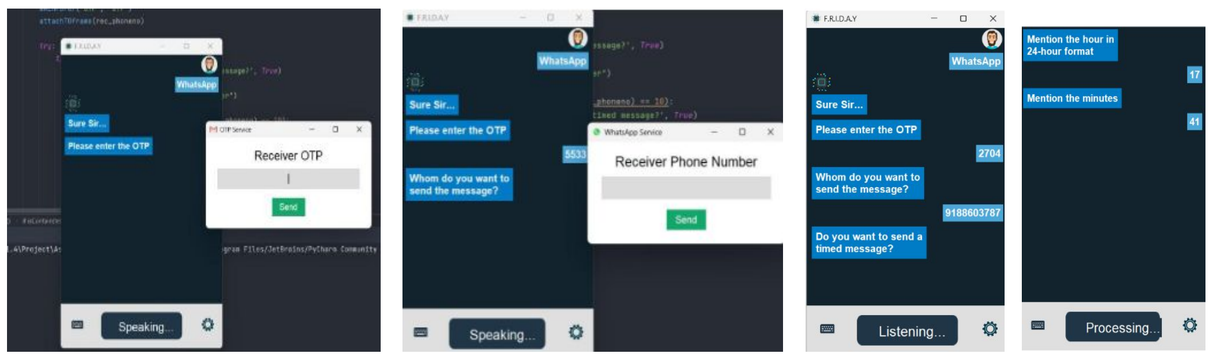
(iii)Volume control



(iv)Launching & remotely controlling web browsers



(v)Definition of words



(vi) Whatsapp Timed Message

Conclusion

Our project addresses security issues in popular voice assistants like Siri, Alexa, and Google, introducing innovative and convenient solutions unique to our system. These applications are increasingly integral to daily life for their hands-free, user-friendly nature. Integration into electronic devices enhances their utility. The Intelligent Voice Assistant efficiently executes user commands, ensuring project success.

References

[1] D. Anniappa and Y. Kim, "Security and Privacy Issues with Virtual Private Voice Assistants," *2021 IEEE 11th Annual Computing and Communication Workshop and Conference (CCWC)*, NV, USA, 2021, pp. 0702-0708, doi: 10.1109/CCWC51732.2021.9375964.

[2] C. M. H. Saibaba, S. F. Waris, S. H. Raju, V. Sarma, V. C. Jadala and C. Prasad, "Intelligent Voice Assistant by Using OpenCV Approach," 2021 Second International Conference on Electronics and Sustainable Communication Systems (ICESC), Coimbatore, India, 2021, pp. 1586-1593, doi: 10.1109/ICESC51422.2021.9532956. Hwang, Ing Shiou, Yi Ying Tsai, Bo Han Zeng, Chien Ming Lin, Huei Sheng Shiue, and Gwo Ching Chang: *Integration of eye tracking and lip motion for hands-free computer access*. Universal Access in the Information Society, 20:405–416, 2021.

[3] Ai, Haojun et al. “An Improvement of the Degradation of Speaker Recognition in Continuous Cold Speech for Home Assistant.” International Conference on Cryptography and Security Systems (2019).

[4] G. Iannizzotto, L. L. Bello, A. Nucita and G. M. Grasso, "A Vision and Speech Enabled, Customizable, Virtual Assistant for Smart Environments," 2018 11th International Conference on Human System Interaction (HSI), Gdansk, Poland, 2018, pp. 50-56, doi: 10.1109/HSI.2018.8431232.Al-Rahayfeh, Amer and Miad Faezipour: *Eye tracking and head move- ment detection: A state-of-art survey*. IEEE journal of translational engineering in health and medicine, 1:2100212–2100212, 2013.

[5] Vikramjit Mitra, , Zifang Huang, Colin Lea, Lauren Tooley, Sarah Wu, Darren Botten, Ashwini Palekar, Shrinath Thelapurath, Panayiotis Georgiou, Sachin Kajarekar, Jefferey Bigham. "Analysis and Tuning of a Voice Assistant System for Dysfluent Speech." (2021).