IJARSCT

ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

Impact Factor: 7.301

Volume 3, Issue 1, June 2023

Voice Visual based Login Authentication using Machine Learning

Shruti Dhanak, Prasad Humbe, Jay Kakade, Paras Chavan, Prof. Santosh Kale
Department of Computer Engineering
NBN Sinhgad School of Engineering, Pune, India

Abstract: *Within the past few a long time, face and voice acknowledgment are drawing in much consideration. Combination of face and voice verification in a framework will guarantee secure human-machine intuitive instead of the conventional secret word. In this age of lockdown due to the Covid-19widespread, more exchanges take put online and it is critical to embrace these exchanges safely and securely. Conventional secret word frameworks and single confirmation frameworks can be amplified to the combination of confront acknowledgment and voice discovery. This paper presents quick login confirmation based on voice-visual combination.*

Keywords: Authentication System, Face recognition, Voice Recognition, Voice feature extraction, Voice feature, Identity authentication

I. INTRODUCTION

In today's period of information innovation, both sound and visual data plays an imperative part inexpanding the sum of information, coming about in an execution procedure that creates it simple to donate bits of knowledge on them. Confront acknowledgment and Voice acknowledgment are one of those two methodologies that point to confirm clients for secure and secure human-machine interaction.

Over the past decade, we've found that a traditional text password is not the best choice secure way to authenticate. Authentication has been done essential for many technical systems.

Mobile as people's demands grow Application-based biometric authentication systems attracted more attention not only because because of the high security control, but also because of that their high accuracy and speed.

With the improvement of Counterfeit Insightsand Mechanical technology, human errands are being supplanted bymachine errands. To secure this human-machineinteraction and to have distant better; a much better; a higher;a stronger;an improved">an improved client encounter, we arearranging to construct a system that employments both confront and

voice verification. Usually an elective toconventional confirmation frameworks. In later a long time,highlights of the human confront and voice are simple toget to, so we chosen to center our consideration onbuilding a bimodal biometric personality framework that employments voice-visual combination.

Driven by this, we designed the system for login for user authenticaton. Face and voice in this system unimodal biometric authentication system is merged and presented. The system meets real-time performance requirements and is built with an API that exposed using a REST endpoint that is Uses an Android app. The paper outlines the basics the principle of both single-mode systems their mixture was followed by our ongoing experiments organized on these topics.

II. RELATED WORK

Scientists have successfully developed various models and systems for face realization and voice authentication models either separately or a a fusion of both. Rapid identity exploration and deployment .The concept of OpenCV was put forward by Gary Bradski which had the capacity to perform on multi-level system. OpenCV has a number of critical capacities as well as utilities which appears from the beginning. The OpenCV makes a difference in recognizing the frontal confront of the individual additionally makes XML documents for a few regions such as the parts of the body. Deep learning evolved of late within the process of the recognition systems. Consequently profound learning

DOI: 10.48175/IJARSCT-11226

together with the confront recognition together work as the deep metric learning frameworks. In short deep learning in confront discovery and acknowledgment will broadly work on two regions the primary one being tolerating the solidary input picture or any other significant picture and the second being giving the finest yields or the comes about of the picture of the picture. We would be using dlib facial recognition framework that would be the simple way to organize the face evaluation. The two primary noteworthy libraries utilized in the system are dlib and face recognition. Python being a really capable programming dialects and one of the programming dialects that are being utilized all over the world has demonstrated to provide best comes about within the face recognition and discovery frameworks. Together confront recognition and discovery gets to be exceptionally simple and productive with the assistance of the python programming dialect and OpenCV.

Dlib could be a toolkit in C++ which contains algorithms for machine learning, which unravel issues within the genuine world. [9] Whereas composed for C , python ties are accessible to run in python. It has too the brilliant facial keypoint detector I utilized for a live time look following gadget in one of my prior posts. Dlib works to supply the frontal face finder with capacities taken from the Histogram of Arranged Slopes (Hoard), which are at that point transferred into an SVM. The dissemination of slope bearings is utilized as characteristics within the Hoard work descriptor. The concept behind Hoard is to extricate highlights into a vector and to bolster it into a classifying calculation such as, for instance, a vector supporting machine that will decide whether a confront is show or not in a field. The characteristics extricated are the conveyance of angle (arranged slope) headings of the picture. Gradients around edges and corners are more often than not wide which permits us to distinguish these areas. This show works with front and slightly non-front faces exceptionally well, compared to the other three, it may be a lightweight model additionally Little impediment doesn't influence it much. But it identifies small faces when it is ready with a minimum of 80 to 880 faces, which is the as it were drawback. You must at that point guarantee that the confront measure of the submission is bigger than that. Be that as it may, for littler faces, you ought to prepare your claim facial finder. Moreover, now and then portion of the front and indeed portion of the jaw are prohibited from the box and do not see down and up for side faces or serious non-front faces.

The voices of each person are effortlessly recognizable, even will recognize one another at the phone. In voice recognizing gathering the alternatives of the voice is imperative.

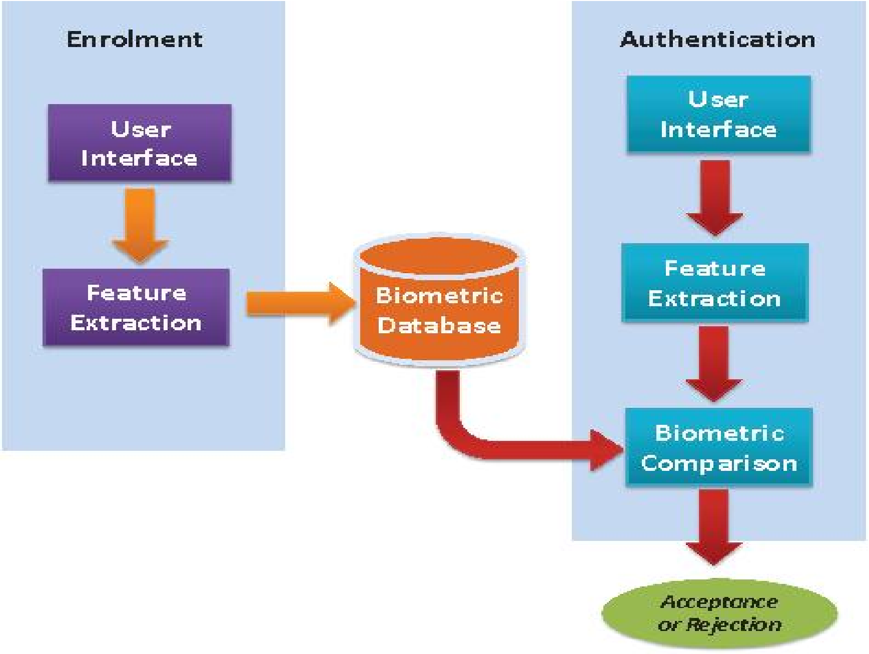
As per considered scrutiny and taken under consideration framework is designed with the set of rules of discourse, cryptography breakthrough inside the fashion of RGB (Ruddy, Green, and Blue) spectrograms. Display Methodology working like substitution addressing acknowledgment of voice, backed the précising the Cryptography almost framework and one's discourse characteristics. System is planned to arrange utilization of Convolutional Neural Networks, actualizing CNN is of bit a extreme errand since it has out appeared a astounding result and appeared threatening vibe to obscure frequency components which are not authorized to the system's database environment. Convolutional Neural Systems undergoes the vigorous training stage because it must be bolster with different voice modalities and each layer has its own allotted work to it.

III. PROPOSED SYSTEM

Our proposed voice-visual based verification framework, which offers a secure and effective implies of confirming client personalities based on their special voice designs. Our framework leverages the control of machine learning, profound learning, and voice recognition innovation to supply a vigorous verification arrangement.

3.1 System architecture

The proposed voice-based authentication system consists of several interconnected components that work together ensures accurate and secure user authentication. The system architecture is designed for collection, preprocessing, feature extraction, model training and real-time verification processes. Basic components of the system architecture described below:



Data Collection Component

This component is responsible for collecting a diverse and representative voice dataset along with visual data. It includes modules for capturing voice samples and image samples from individuals under controlled conditions. The collected voice samples are stored in a secure database for subsequent preprocessing and model training.

Preprocessing and Feature Extraction Component

The preprocessing and feature extraction component receives the voice data from various preprocessing techniques to remove noise, normalize volume levels, and eliminate artifacts from the voice signals. The preprocessed voice data and image data are then fed into feature extraction modules, which utilize techniques cepstral coefficients (MFCCs), Linear Predictive Coding (LPC), or Perceptual Linear Prediction (PLP) to extract relevant acoustic features. The extracted features serve as input for the subsequent model training phase. We use OpenCV’s built-in Haar Cascade XML files or even TensorFlow or using Keras. Over here especially , We apply a HOG (Histogram of Gradients) and Linear SVM (Support Vector Machines) object detector specifically for the task of face detection. We can also do it using Deep Learning-based algorithms which are built for face localization. Also, The algorithm will be used for the detection of the faces in the image. We obtain face bounding box through some method for which we use the (x, y) coordinates of the face in the image respectively.

Model Training Component

The model training component utilizes machine learning and deep learning algorithms to build voice recognition models. It takes the preprocessed voice data and corresponding user identity labels as input. The component employs popular algorithms such as Gaussian Mixture Models (GMMs), deep neural networks, or Long Short networks. The models are trained to map the extracted voice features to unique user identities, enabling accurate identification during the authentication process. The training process involves optimizing model parameters using appropriate optimization algorithms.

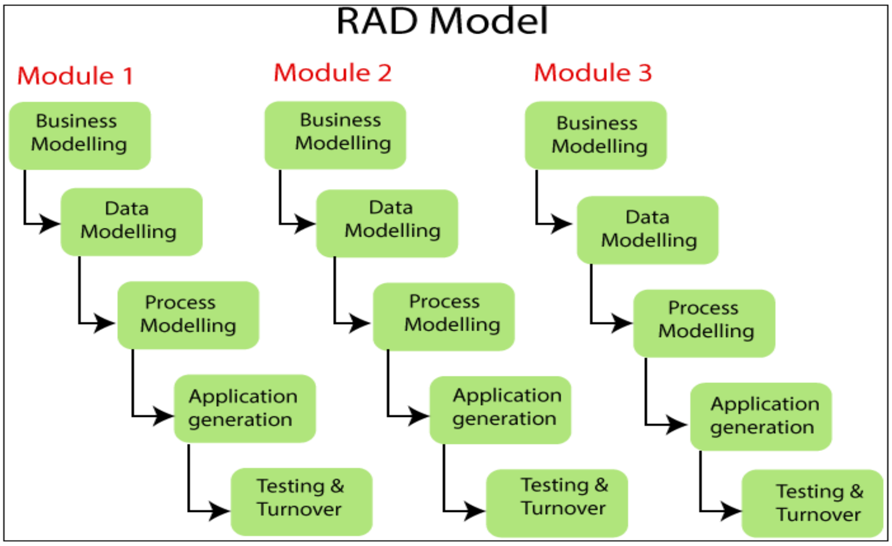
The model determines what you’ll use to locate faces in the input images. Valid model type choices are "hog" and "cnn", which refer to the respective algorithms used: HOG (histogram of oriented gradients) is a common technique for object detection. For this tutorial, you only need to remember that it works best with a CPU. CNN (convolutional neural network) is another technique for object detection. In contrast to a HOG, a CNN works better on a GPU, otherwise known as a video card.

Real-time Authentication Component

The real-time authentication component is responsible for the actual authentication process. It receives voice samples and image samples of users who seek authentication and applies the same preprocessing steps as in the data collection phase to match the format used during model training. The preprocessed voice and visual sample is then fed into the trained voice recognition model(s). The model(s) produce a similarity score or a probability distribution over known identities, indicating the degree of match between the user's voice and the stored voice patterns. If the similarity score surpasses a predefined threshold or the probability distribution exhibits a high confidence for a Otherwise, the authentication is rejected.

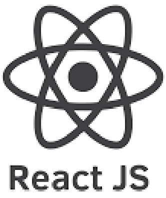
IV. TECHNOLOGIES USED

Rapid application development model (RAD)



The RAD model is a type of incremental process model in which there is extremely short development cycle. When the requirements are fully understood and the component-based construction approach is adopted then the RAD model is used. Various phases in RAD are Requirements Gathering, Analysis and Planning, Design, Build or Construction, and finally Deployment. The critical feature of this model is the use of powerful development tools and techniques. A software project can be implemented using this model if the project can be broken down into small modules wherein each module can be assigned independently to separate teams. These modules can finally be combined to form the final product. Development of each module involves the various basic steps as in the waterfall model i.e. analysing, designing, coding, and then testing, etc. as shown in the figure. Another striking feature of this model is a short time span i.e. the time frame for delivery(time-box) is generally 60-90 days.

React.js



React JS, commonly referred to as React is one frontend development framework, or to be more specific a library that has found a favorite with developers around the world. In the world of custom software development, it is very important that tech-based companies experiment with emerging frameworks to augment their digital capabilities. With React storming into the picture, it instantly fell into a race with its predecessors like Angular JS and Vue JS.

Flask(Python)

Flask is a micro-framework developed in Python that provides only the essential components - things like routing, request handling, sessions, and so on. It provides you with libraries, tools, and modules to develop web applications like a blog, wiki, or even a commercial website. It's considered "beginner friendly" because it doesn't have boilerplate code or dependencies which can distract from the primary function of an application.Flask is used for the backend, but it makes use of a templating language called Jinja2 which is used to create HTML, XML or other markup formats that are returned to the user via an HTTP request. More on that in a bit.

DLIB

The dlib library is arguably one of the most utilized packages for face recognition. A Python package appropriately named facerecognition wraps dlib's face recognition functions into a simple, easy to use API.The intricacies of face detection necessitate a wide range of face data. Having access to a diverse, well-curated dataset is invaluable in creating models that can handle variations in pose, expression, and lighting conditions.

OpenCV

OpenCV is a huge open-source library for computer vision, machine learning, and image processing. OpenCV supports a wide variety of programming languages like Python, C++, Java, etc. It can process images and videos to identify objects, faces, or even the handwriting of a human. When it is integrated with various libraries, such as Numpy which is a highly optimized library for numerical operations, then the number of weapons increases in your Arsenal i.e whatever operations one can do in Numpy can be combined with OpenCV.

CNN

Convolutional networks were the beginnings Hubel and Wiesel who found that a single network architecture could reduce complexity in the feedback neural network when studying neurons used for local sensitivity and orientation selection in the cerebral cortex of cats. CNN is often used with image processing that requires a two-dimensional matrix containing features and may be three-dimensional, the pixel values are in the horizontal and vertical coordinate indicators. CNN is a neural network model. Its architecture has three main ideas, Convolution layer, Pooling layer and Full connected layer . Each one of them has the susceptibility to improve speech recognition performance .

V. CONCLUSION

In conclusion, the development of a voice - visual based verification framework utilizing Python, Carafe, TensorFlow, and Keras offers a novel and strong approach to client recognizable proof and get to control. This extend leverages the control of machine learning and profound neural systems to analyze and confirm the one of a kind characteristics of an individual's voice and visual features.By utilizing the Jar framework, we have made a adaptable and versatile framework engineering that permits for consistent integration with different programming dialects and stages. The utilize of Python as the essential programming dialect offers straightforwardness and a wide run of libraries like OpenCV and dlib and instruments to bolster the improvement process.In rundown, this extend speaks to a noteworthy contribution to the field of verification frameworks by presenting a voice-based approach that combines the qualities of Python, Jar, TensorFlow, and Keras. The created framework offers an inventive and secure strategy for client identification, with potential applications in different spaces. Long-term holds promising conceivable outcomes for assist advancements and refinement of voice-visual based verification frameworks, clearing the way for improved security and client experiences within the advanced domain.

REFERENCES

[1] Tudor Barbu, Adrian Ciobanu, Mihaela Luca, “Multimodal Biometric Authentication based on Voice, Face and Iris”, IEEE, 2015.

[2] Girija Chetty and Michael Wagner ,“Audio-Visual Multimodal Fusion for Biometric Person Authentication and Liveness Verification”, ACM International Conference Proceeding,2006

[3] M. Acheroy et al., "Multi-modal person verification tools using speech and images", *Multimedia Appl. Services Techn. ECMAST*, 1996.

[4] Andrew Boles, Paul Rad, “Voice Biometrics: Deep Learning-based Voiceprint Authentication System”, International Journal of Advanced Research in Engineering and Technology, 2021.

[5] Z. Hachkar, A. Farchi, B. Mounir and J. El Abbadi, “A Comparison of DHMM and DTW for Isolated Digits Recognition System of Arabic Language,” International Journal on Computer Science and Engineering, vol.3, no.3, pp. 1002- 1008, 2011.

[6] Shoup, A., Tanya Talkar and J. Chen. “An Overview and Analysis of Voice Authentication Methods.” (2016).

[7]M. Trojahn, F. Ortmeier, “Biometric Authentication Through a Virtual Keyboard for Smartphones”, International Journal of Computer Science & Information Technology, vol.4, no.5, pp. 1- 12, Oct. 2012. .

[8] Singh, Nilu, R. A. Khan, and Raj Shree. "Applications of Speaker Recognition."Procedia Engineering 38 (2012): 3122- 3126.

[9] Q. Memon, Z. AlKassim, E. AlHassan, M. Omer, M. Alsiddig, “Audio-Visual Biometric Authentication for Secured Access into Personal Devices”, ICBBS '17: 6th International Conference on Bioinformatics and Biomedical Science

[10] Sarabjeet Singh, Yamini M, “Voice Based Login Authentication For Linux,” International Conference on Recent Trends in Information Technology (ICRTIT), 2013.

DOI: 10.48175/IJARSCT-11226