Grenze International Journal of Engineering and Technology, Jan Issue

Soundless Credential Validation based on Lip Synchronization in Image Processing

Dr. M. Senthil Kumar[1], Ms. M. Raghavi[2], A. Aafrin Nisha[3], R. Blesslinjaffy[4] and A. Dhakshayani[5]
[1-5]SRM Valliammai Engineering College/Department of Cyber Security, Chengalpattu, India
Email: msen1982@gmail.com, {raghavirgo, afraanmiza, blesslinjaffy, ayamini2003}@gmail.com

***Abstract*— In today's modern world, ensuring safe access to sensitive information and digital systems is essential. Traditional authentication techniques, such as passwords and PINs, are vulnerable to a wide variety of security concerns, including phishing attacks and password leaks. Silent Lip Password Recognition (SLPR) is an emerging biometric authentication method that relies on the visual analysis of an individual's lip movements during speech. This technique has garnered significant attention due to its non-intrusive nature and potential for enhancing security in various applications. Convolutional Neural Networks (CNNs) have proven to be highly effective in image-based tasks, making them a suitable choice for extracting meaningful features from lip motion data. This approach combines spatial and temporal information within the lip motion data, allowing the network to capture the nuances of individual lip patterns. The results show that CNNs can effectively distinguish and authenticate users with high precision, while also providing resistance to potential spoofing attempts.**

***Index Terms*— Biometrics, Neurons, Lip analysis, Authentication, Robustness, Facial Points.**

I. INTRODUCTION

Biometric authentication has become an integral part of modern security systems, offering reliable and convenient ways to verify the identity of individuals. Traditional biometric methods, such as fingerprint and iris recognition, have proven their effectiveness, but they often require direct interaction or physical contact, which may not always be practical or desirable. Silent Lip Password Recognition (SLPR) is an innovative and non-intrusive biometric authentication technique that has gained increasing attention in recent years [2]. This method leverages the unique visual patterns of lip movements during speech to verify a person's identity. The Convolutional Neural Networks (CNNs) have emerged as a powerful tool for automatically recognizing and authenticating users based on their silent lip movements. The human lip area is rich in information and exhibits a high degree of individuality, making it a suitable candidate for biometric authentication [11]. SLPR has several distinct advantages, including its non-intrusive nature, as it does not require physical contact or additional hardware, and it can be applied in various scenarios, such as secure access control, online identity verification, and password-protected applications. Moreover, this technique offers an additional layer of security since lip movements are not easily replicable, making it challenging for unauthorized individuals to gain access [1]. It is used for secure and user-friendly authentication methods, especially in a world where digital transactions and online activities are increasingly common. The research aims to provide insights into the development and application of a Silent Lip Password Recognition system that employs CNNs to improve user authentication accuracy and enhance the overall security of various applications [5]. This paper delves into the fascinating realm of Silent Lip Password Recognition,

Grenze ID: 01.GIJET.10.1.67_1

exploring its potential as a cutting-edge biometric authentication technique. In particular, the integration of Convolutional Neural Networks (CNNs) into the SLPR framework. CNNs have proven to be exceptionally proficient in image-based tasks, making them a natural choice for the visual analysis of lip movements [4]. By harnessing the power of deep learning and CNNs, we aim to enhance the accuracy and reliability of SLPR, thereby addressing the limitations of traditional password-based systems and contributing to the advancement of secure, convenient, and non-intrusive authentication methods.

II. Literature Survey

[1] According to X. Ai and B. Fang, traditional lip-reading models have two major components: an upstream visual module and a downstream encoder-decoder. A lip-motion representation is a pattern of representation created by combining all static frame characteristics along a time/position axis, and it is generated by a visual module composed of layered convolutions that consumes a video file or clips given to it frame by frame. Each frame's synchronized visual and phonetic features are combined to form a unified representation. The methodology for leveraging cross-modal language models to tackle the challenging task of lip-reading in silent videos is utilized. By effectively combining visual and textual information, synchronizing the modalities, and employing state-of-the-art deep learning techniques, this approach aims to significantly enhance the accuracy and robustness of lip-reading systems, with potential implications in accessibility, surveillance, and various human-computer interaction applications.

[2] G. Li et al. proposed an audio-visual multi-channel speech separation, dereverberation, and recognition system based on the visual modality's invariance to acoustic signal corruption. The method incorporates visual data into all stages of the speech processing pipeline. It develops a deep learning model for separating target speech from background noise and interference sources. This model can be based on architectures like Conv-TasNet or Deep Clustering. It utilizes the extracted audio features from different microphones as input to the speech separation model. They train the model using supervised learning, with ground-truth sources available in the dataset. The paper aims to develop a robust system that can extract clean speech from complex environments, reduce reverberation effects, and accurately transcribe the separated speech for various applications, including human-computer interaction and speech processing in real-world settings.

[3] S. S. Mahdi et al., narrated that the area of study leverages geometric metric learning techniques to associate facial attributes with 3D facial data. The process begins with the acquisition of 3D facial data. This can be obtained through 3D scanners, structured light sensors, or even from 3D reconstructions of 2D images using photogrammetry. The acquired 3D facial data often needs preprocessing to remove artifacts, align the face properly, and normalize the data. This step ensures that the data is suitable for analysis. Once the model is trained and evaluated, it can be used to predict demographic properties for new 3D facial data. The results can provide insights into the relationships between facial shape and demographic attributes.

III. Module Description

A. Module 1: Data Collection

In the silent lip password authentication system, the data collection involves the process of obtaining and recording video and audio information belonging to an individual’s silent lip movements while they utter a specific set of password or passphrase [6]. This data is obtained with the purpose of developing a biometric authentication system that can authenticate a person’s identity based upon the unique lip movement patterns correlated with silently pronounced a specific passphrase. The collected data consists of video clippings of lip movements, audio recordings and other appropriate facial features, which serves as credential points for authentication. Then, this biometric data is examined and processed to build a reliable and unique method for user authentication based on silent lip recognition [3]. The data can be collected in following ways:

Data Acquisition: It involves collecting video clippings of individuals uttering the password silently and record the corresponding audio, even though it is a silent lip analysis. This will serve as a reference point for syncing lip movement patterns with speech.

Data Annotation: The following data annotations is performed on each video recordings: Textual representation of the spoken passphrase, time stamps for specific sounds in the passphrase, analysis of lip movements [9] (key lip shapes, transition between shapes and lip trajectories) and some facial features such as facial landmarks, eye movements and some facial points.

Data Preprocessing: Lip regions from the video frames are extracted for preprocessing, normalizing these video frames and aligning those frames for compatible analysis and converting the audio frames to a specific format that will be used in synchronization with the video data [7].
Security and Data Storage: The collected and processed data for password analysis is stored securely to prevent data breaches or any misuse.

B. Module 2: Lip Segmentation and Dataset Training

Silent lip password recognition involves using lip movements to authenticate users without the need for audible speech. Lip segmentation is a crucial step in this process as it involves isolating and extracting the lip region from a given image or video frame. The dataset training involves the process of using a collection of data which trains the machine learning model [14]. It consists of images or video frames which has visual information of a user silently uttering or moving their lips. The dataset used for training is important as it serves as the foundation for the proposed model. This is a iterative process which continues until the model achieves a satisfactory level of performance on the training dataset. The success of the training process relies on the quality, diversity and representativeness of the dataset, as well as the appropriateness of the CNN architecture and training methodology [16]. In the Fig. 1, the maximum epoch of the trained dataset is identified with the help of the accuracy and loss ratio graph of the trained dataset. In training a neural network, the maximum epoch refers to a predefined limit on the number of complete iterations through the entire training dataset during the training process. i.e., an epoch is a single cycle where the neural network sees and learns from all the training data. The maximum epoch is used here to control the time taken for the training process and overfitting. The following steps provide a general guide for lip segmentation and dataset training for silent lip password recognition:

Preprocessing: Collect a dataset of images or video frames containing individuals speaking silently. Ensure diversity in terms of lip shapes, sizes, and lighting conditions. The input for lip detection and segmentation can be an image or video stream of the user's face. Preprocessing techniques may be applied to normalize the image, adjust for variations in lighting, and enhance contrast.
Lip Region Localization: Lip detection begins by identifying the approximate region where the lips are expected to be [6]. This can be based on prior knowledge of facial landmarks or by using facial detection algorithms like Haar Cascade classifiers or deep learning-based face detectors [10].
Lip Feature Extraction: Once the approximate lip region is located, specific features are extracted from that region to identify the lips more accurately. These features may include color information, texture, and edges that are characteristic of the lips.
Candidate Lip Region Selection: User lip regions are selected based on the extracted features. These regions are potential lip areas that need to be further refined [1].
Lip Model or Classifier Application: A trained model or classifier is used to verify and refine the candidate lip regions. In the context of a soundless password system, this model may be trained to recognize unique lip patterns associated with the authorized user.
Refinement and Post-processing: The detected lip region may undergo further refinement to precisely delineate the lip boundary. Techniques like contour tracing and edge smoothing can be applied for this purpose [19].
Lip Segmentation: The final step is to segment the lip region from the rest of the face or background. This segmentation isolates the lips for subsequent analysis and verification.
Train a Recognition Model: Prepare a dataset of silent lip movements associated with each user's password. Then train a deep learning model on the dataset, mapping the extracted features to specific passwords [18].
Model Evaluation: Evaluate the recognition model's performance on a separate test dataset. Adjust parameters or architecture as needed [13]. Finally, integrate the lip segmentation and silent lip password recognition models into your authentication system.

C. Module 3: Passphrase Detection

For the passphrase detection, the Convolution Neural Network (CNN) plays a major role in the process of detecting the passphrase being pronounced by the user [4]. In CNN, a classifier is a part of the network which is responsible for the process of making predictions or decisions based on the lip features extracted from the input data. The classifier follows the convolutional and pooling layers and precedes the output layer [8]. It is primarily performed to map the learned features to specific classes or labels, which makes the CNN well suited for image classification, object detection and various other pattern recognition processes. In the fully connected layers, each neuron is connected to every neuron in the layer before it, acting as a high-level feature extractor and learning the complex relationship within the collected data. Rectified Linear Unit (ReLU), an activation function that is applied after

each layer to introduce non-linearity, allowing the network to model complex patterns in data [5]. The network's prediction is generated in the classifier's final layer. The classifier is then trained to minimize a predefined loss function, which measures the difference between predicted and actual class probabilities. Once the system is trained, the CNN classifier is then used for the testing process. Here the video frames are fed into the network and the output of the softmax layer provides the predicted class probabilities [11]. The class which has the highest probability is the predicted passphrase. This process contributes to the robustness and effectiveness of soundless credential recognition systems.

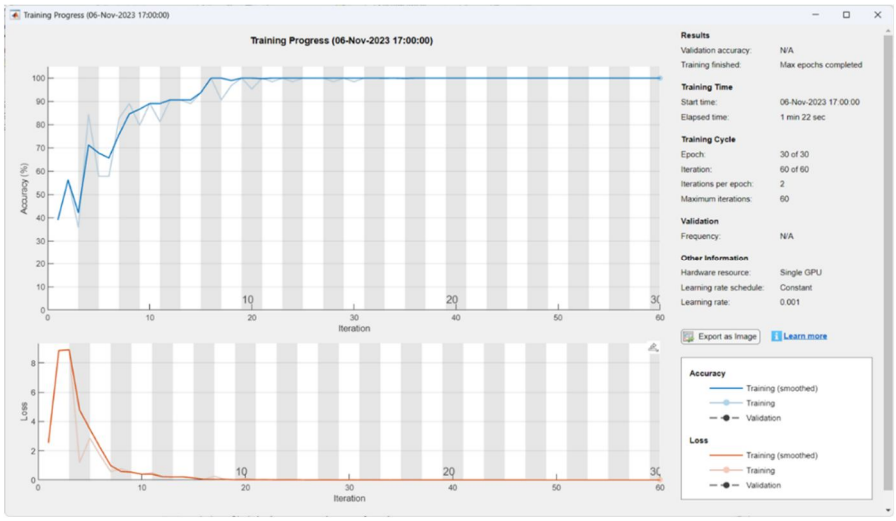

Figure 1. A graphical representation of training accuracy dataset progression and loss ratio of training dataset

IV. Proposed Methodology

A. Video Frame Extraction

In image processing, the process of extracting individual frames from a video is known as video frame extraction. This is a crucial step if you want to analyze or process each frame individually. Here's an overview of how this process is typically done by the input to the process is a video file in a format such as MP4, AVI, or any other commonly used video format [10]. Use a video processing library or tool (e.g., OpenCV in Python) to read the video file. This library allows you to access frames and information about the video, such as frame rate and resolution. Then iterate through the video frames, and at each iteration, extract a single frame. The frame rate determines how many frames are extracted per second. For example, at a frame rate of 30 frames per second (fps), 30 frames would be extracted in one second [14]. Once the frames are extracted, you can perform image processing operations on each individual frame. This might include tasks such as object detection, segmentation, feature extraction, or any other image processing task relevant to your application. The processed frames can be further analyzed individually, or you might apply algorithms that consider information from multiple frames, depending on the specific task.

B. Analogue-to-Digital Converters

ADCs are used to convert continuous analog signals (such as those from a sensor) into discrete digital values, which are then processed by digital systems [15]. In the context of image processing, the process starts with capturing an image using a sensor, such as a camera. The analog signals from the sensor are converted into digital values by an ADC. The resulting digital image consists of discrete pixels, each with a specific digital value representing its intensity or color. Lip synchronization often involves the processing of audio signals to match them with corresponding lip movements. An ADC could be used to convert analog audio signals, such as spoken words, into digital format for further analysis and synchronization. Lip synchronization typically occurs in the context of video processing. An ADC might be used to digitize analog audio signals captured alongside video footage. This digitized audio can then be synchronized with the visual information of lip movements. In some applications, lip synchronization might be enhanced by incorporating speech recognition. An ADC could be involved in converting analog speech signals into digital form, allowing for more detailed analysis and alignment with lip movements. Systems that aim for accurate lip synchronization might use multiple modalities, including

both visual (lip movements) and auditory (speech) information. ADCs could be part of the subsystem that handles the conversion of analog signals from microphones into digital signals for processing. However, apply an Analog-to-Digital Converter (ADC) to convert the analog lip movement signals (and optionally, the analog audio signals) into digital form. The ADC discretizes the continuous analog signals into a sequence of digital values, representing the intensity or color of pixels in the lip region of the captured images or video frames. Finally make decisions or generate output based on the results of the lip analysis. This could include lip-reading transcription, authentication results, or any other relevant information.

C. Convolutional Neural Networks

Gathering a dataset of individuals performing a set of predefined lip movements, gestures, or expressions. This dataset should include various lip patterns, including different lip shapes, movements, and combinations. Normalize and clean the lip movement data to account for variations in lighting conditions, camera angles, and facial expressions [7]. Remove noise and irrelevant data from the video or image feeds. Use computer vision techniques to identify and extract meaningful features from lip movements. These may include lip shape, lip contour, motion vectors, and other distinctive characteristics [17]. Extract the timing and duration of lip movements to create a temporal lip signature. During user enrolment, prompt the user to perform a set of specific lip movements or gestures. Record and extract the relevant lip movement features to create a unique lip analysis profile for the user. Store this profile securely in a database, associating it with the user's identity. When a user attempts to access a system or device, request them to perform a set of predefined lip movements further Capture and pre-process the user's lip movement data in real-time. Extract features from the live lip movements and create a temporal lip signature and compare the extracted lip movement features from the live data with the stored features in the database. Use a matching algorithm (e.g., dynamic time warping or machine learning classifiers) to determine the similarity between the live lip movements and the stored lip analysis profile [16]. Set a threshold for acceptable similarity scores to decide whether the user's lip movements match the stored profile. Adjust the threshold based on the desired security level and false acceptance/rejection rates. If the similarity score exceeds the threshold, grant the user access to the system [20]. If the similarity score falls below the threshold, deny access and, optionally, trigger additional security measures.

V. WORKING OF PROPOSED METHODOLOGY

Design a CNN architecture that can effectively learn and recognize the unique features in the lip movement. Convolutional layers are particularly useful for this purpose. Train the CNN on the lip movement data while using the user's silent password as the target label. Ensure that the dataset is balanced and includes enough samples for each user. Then users create their silent passwords by performing a specific lip movement or sequence of movements. The network maps these movements to a unique representation that can be used for verification. During the authentication process, the user is prompted to perform their silent password. The system captures their lip movement and processes it in real-time. The CNN analyzes the lip movement and compares it to the stored representations of the user's silent password. If the comparison is sufficiently close (based on a threshold or similarity measure), access is granted. To enhance security, additional factors or multi-factor authentication can be used in conjunction with lip synchronization, such as face recognition. In Fig. 2, the basic workflow of the validation using lip analysis was structured and demonstrated with the key references.

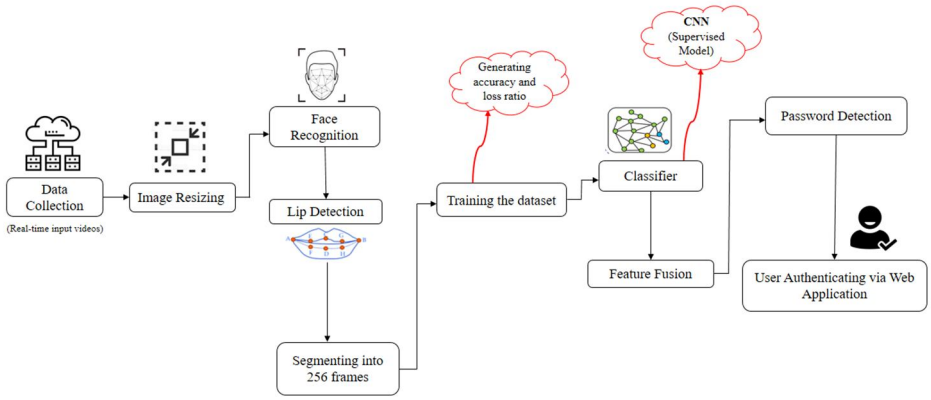

Figure 2. A diagrammatic representation of lip synchronization using CNN

VI. Related Works

The proposed related works aim to establish the feasibility, accuracy, and security of soundless credential validation through lip analysis. The key aspects include as follows:
Deep Lip Reading: Deep learning techniques, including CNNs and recurrent neural networks (RNNs), have been applied to lip reading tasks. These models learn to extract features from lip images and interpret them for speech recognition or authentication purposes.
Multimodal Biometrics: The research focuses on combining lip analysis with other biometric modalities for improved accuracy and security. This could involve integrating lip features with facial recognition or other behavioral biometrics.
Anti-Spoofing Techniques: The importance of security in biometric authentication, researchers are working on developing techniques to detect and prevent spoofing attempts. This includes studying how to distinguish between genuine lip movements and recorded or synthetic attempts to deceive the system.
End-to-End Lip-Reading Systems: The exploration of end-to-end lip reading systems where the entire process, from image input to final authentication, is learned by a single model. This approach can potentially simplify the system and improve overall performance.
Datasets and Benchmarks: The availability of larger and more diverse lip reading datasets, such as the Lip Reading in the Wild (LRW) dataset, has facilitated advancements in training deep learning models. Benchmarks help researchers compare the performance of different approaches and drive innovation.
Real-world Applications: Advancements include the testing and deployment of soundless credential validation systems in real-world scenarios. These applications may range from access control in secure environments to user authentication in consumer electronics.
Privacy Considerations: Researchers are increasingly focusing on addressing privacy concerns associated with biometric data. Advancements in privacy-preserving techniques, secure storage, and user consent mechanisms contribute to the ethical development and deployment of lip-based authentication systems.
Industry Adoption: The advancement of lip analysis for soundless credential validation is reflected in increased interest and potential adoption by industries. Companies working in biometrics, security, and technology are exploring and implementing these technologies for various applications.
In future it is expected that this authentication method will explore the integration of multiple biometric modalities, such as facial features, voice, and lip movements, for a more comprehensive and secure authentication system. Address challenges related to environmental factors, such as varying lighting conditions and camera angles, to ensure the system's reliability in real-world scenarios and Develop and implement privacy-preserving techniques to protect user data and ensure compliance with privacy regulations.

VII. Results

The primary result of a soundless credential validation system based on lip analysis would be its accuracy in correctly identifying individuals. This could be measured by the True Positive Rate (the rate at which legitimate users are correctly authenticated) and the False Positive Rate (the rate at which impostors are incorrectly authenticated). The system should ideally exhibit a high True Positive Rate and a low False Positive Rate. The system's robustness in the face of variations, such as different lighting conditions, angles, and lip shapes, would be a critical aspect of the results. Assessing how well the system performs under various conditions is essential. Measuring user acceptance is also crucial. It's important to gather feedback from users to understand how comfortable and convenient they find this method of authentication. Additionally, it is discussed that the balance between accuracy and user convenience is an important point of discussion. If the system is too strict, it may have a high False Rejection Rate, causing frustration for users. If it's too lenient, the False Acceptance Rate may rise, compromising security. A trade-off must be considered and a methodology of automated speed recognition system and blendshape working is proposed to overcome the problem and it discuss the importance of educating users about how the system works and potential limitations also lip analysis-based authentication can be less vulnerable to common attacks like key logging since there is no traditional password to be intercepted.

VIII. Discussion

The approach mainly discusses the capturing and analyzing the patterns to create a unique identifier for authentication purposes. Unlike some biometric methods, lip synchronization does not require physical contact. It can be captured through video or imaging, making it a non-intrusive form of authentication. One of the primary challenges is achieving a high level of accuracy. Lip movements can be affected by various factors such as lighting

conditions, facial hair, and changes in facial expressions, which can impact the reliability of the authentication system. As with any biometric system, there is a concern about spoofing or impersonation. Users may have concerns about the privacy implications of capturing and storing biometric data, even if it's based on lip movements. To overcome these problems, the proposed methodology, implement robust security measures to protect this sensitive information. Lip movements can vary across cultures and individuals. Ensuring that the system is inclusive and works well for a diverse user base is an important consideration and provides seamless integration of lip synchronization into the user experience. Finally, data protection and privacy laws has been taken into consideration to ensure that the collection, storage, and processing of biometric data comply with regulations in different regions. The design of Convolutional Neural Network (CNN) architecture for a soundless password system based on lip synchronization involves creating a model that can effectively extract features from lip movements and perform classification or authentication tasks that Crop and resize video frames, normalize pixel values, and pre-process the data to enhance the model's ability to learn and Depending on the task, the output layer for binary or multi-class classification is designed. Additionally, apply batch normalization to stabilize and accelerate training to introduce dropout layers to prevent overfitting. It is discussed that Training model works by splitting the dataset into training and testing sets then apply data augmentation techniques to increase the diversity of the training set and Train the model on the training set using the compiled model and monitor performance on the validation set. Furthermore, the approach evaluates the model on the testing set to assess its generalization performance and Use chosen evaluation metrics to measure performance.

IX Conclusions

The paper conclude that lip analysis offers a compelling alternative to traditional authentication methods, capitalizing on the uniqueness of everyone's lip movements. This innovation combines security and user convenience in a manner that is poised to redefine the future of digital access. Furthermore, it holds great potential in security while eliminating the inconvenience of traditional passwords such as PINs, passwords etc. This promising technology offers a glimpse into the future of secure and seamless user authentication. The integration of lip analysis into authentication systems represents a ground-breaking approach that not only enhances security but also streamlines the user experience. As technology continues to evolve, silent passwords based on lip analysis are well-positioned to play a central role in the field of digital security. Finally embracing this technology means not just securing our systems but also simplifying the user experience in our increasingly digital world. The accuracy enhancement through research and development efforts should focus on improving the accuracy of lip analysis for authentication. This includes refining algorithms, increasing the specificity of lip movement detection, and reducing false acceptance and rejection rates. Additionally implementing adaptive learning mechanisms that can continually update the user's lip analysis profile to account for changes over time. This ensures the system remains effective in recognizing users as their lip movements evolve and Integrate liveliness detection mechanisms to ensure that the lip analysis is performed on a live person rather than a static image or video. This can help throughout spoofing attempts. Also, explore the integration of lip analysis with other biometric or authentication methods, such as facial recognition, fingerprint scanning, or behavioural biometrics, to create more robust multi-factor authentication systems.in future this method conducts extensive real-world testing of lip analysis systems in various environments and conditions to assess their robustness and reliability and address privacy and ethical concerns associated with the collection and storage of biometric data, including compliance with data protection regulation. Finally, it focuses on user education to ensure users understand how to use and trust soundless password systems based on lip analysis. This includes providing clear instructions and addressing concerns.

References

[1] X. Ai and B. Fang, "Cross-Modal Language Modeling in Multi-Motion-Informed Context for Lip Reading," in IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 31, pp. 2220-2232,2023,doi:10.1109/TASLP.2023.3282109.

[2] G. Li et al., "Audio-Visual End-to-End Multi-Channel Speech Separation, Dereverberation and Recognition," in IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 31, pp. 2707-2723, 2023, doi: 10.1109/TASLP.2023.3294705.

[3] S. S. Mahdi et al., "Matching 3D Facial Shape to Demographic Properties by Geometric Metric Learning: A Part-Based Approach," in IEEE Transactions on Biometrics, Behavior, and Identity Science, vol. 4, no. 2, pp. 163-172, April 2022, doi: 10.1109/TBIOM.2021.3092564.

[4] H. X. Pham, Y. Wang and V. Pavlovic, "Learning Continuous Facial Actions From Speech for Real-Time Animation," in IEEE Transactions on Affective Computing, vol. 13, no. 3, pp. 1567-1580, 1 July-Sept. 2022, doi: 10.1109/TAFFC.2020.3022017.

[5] H. Liu, H. Cai, Q. Lin, X. Li and H. Xiao, "Adaptive Multilayer Perceptual Attention Network for Facial Expression Recognition," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 32, no. 9, pp. 6253-6266, Sept. 2022, doi: 10.1109/TCSVT.2022.3165321.
[6] H. Wang, G. Pu and T. Chen, "A Lip Reading Method Based on 3D Convolutional Vision Transformer," in IEEE Access, vol. 10, pp. 77205-77212, 2022, doi: 10.1109/ACCESS.2022.3193231.
[7] Z. Pan, R. Tao, C. Xu and H. Li, "Selective Listening by Synchronizing Speech with Lips," in IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 30, pp. 1650-1664, 2022, doi: 10.1109/TASLP.2022.3153258.
[8] D. Bigioi, H. Jordan, R. Jain, R. McDonnell and P. Corcoran, "Pose-Aware Speech Driven Facial Landmark Animation Pipeline for Automated Dubbing," in IEEE Access, vol. 10, pp. 133357-133369, 2022, doi: 10.1109/ACCESS.2022.3231137.
[9] S. Bu, Y. Zhao, T. Zhao, S. Wang and M. Han, "Modeling Speech Structure to Improve T-F Masks for Speech Enhancement and Recognition," in IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 30, pp. 2705-2715, 2022, doi: 10.1109/TASLP.2022.3196168.
[10] E. Villalobos, D. Mery and K. Bowyer, "Fair Face Verification by Using Non-Sensitive Soft-Biometric Attributes," in IEEE Access, vol. 10, pp. 30168-30179, 2022, doi: 10.1109/ACCESS.2022.3158967.
[11] L. Qin, F. Peng, S. Venkatesh, R. Ramachandra, M. Long and C. Busch, "Low Visual Distortion and Robust Morphing Attacks Based on Partial Face Image Manipulation," in IEEE Transactions on Biometrics, Behavior, and Identity Science, vol. 3, no. 1, pp. 72-88, Jan. 2021, doi: 10.1109/TBIOM.2020.3022007.
[12] Q. Li and X. Lin, "Efficient and Privacy-Preserving Speaker Recognition for Cybertwin-Driven 6G," in IEEE Internet of Things Journal, vol. 8, no. 22, pp. 16195-16206, 15 Nov.15, 2021, doi: 10.1109/JIOT.2021.3097266.
[13] C. -Z. Yang, J. Ma, S. Wang and A. W. -C. Liew, "Preventing DeepFake Attacks on Speaker Authentication by Dynamic Lip Movement Analysis," in IEEE Transactions on Information Forensics and Security, vol. 16, pp. 1841-1854, 2021, doi: 10.1109/TIFS.2020.3045937.
[14] B. -F. Wu, B. -R. Chen and C. -F. Hsu, "Design of a Facial Landmark Detection System Using a Dynamic Optical Flow Approach," in IEEE Access, vol. 9, pp. 68737-68745, 2021, doi: 10.1109/ACCESS.2021.3077479.
[15] M. Ezz, A. M. Mostafa and A. A. Nasr, "A Silent Password Recognition Framework Based on Lip Analysis," in IEEE Access, vol. 8, pp. 55354-55371, 2020, doi: 10.1109/ACCESS.2020.2982359.
[16] R. Tolosana, R. Vera-Rodriguez, J. Fierrez and J. Ortega-Garcia, "BioTouchPass2: Touchscreen Password Biometrics Using Time-Aligned Recurrent Neural Networks," in IEEE Transactions on Information Forensics and Security, vol. 15, pp. 2616-2628, 2020, doi: 10.1109/TIFS.2020.2973832.
[17] J. -W. Kim, H. Yoon and H. -Y. Jung, "Linguistic-Coupled Age-to-Age Voice Translation to Improve Speech Recognition Performance in Real Environments," in IEEE Access, vol. 9, pp. 136476-136486, 2021, doi: 10.1109/ACCESS.2021.3115608.
[18] W. Mu and B. Liu, "Voice Activity Detection Optimized by Adaptive Attention Span Transformer," in IEEE Access, vol. 11, pp. 31238-31243, 2023, doi: 10.1109/ACCESS.2023.3262518.
[19] W. Gao et al., "Digital Retina: A Way to Make the City Brain More Efficient by Visual Coding," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 31, no. 11, pp. 4147-4161, Nov. 2021, doi: 10.1109/TCSVT.2021.3104305.
[20] M. Szymkowski, E. Saeed, M. Omieljanowicz, A. Omieljanowicz, K. Saeed and Z. Mariak, "A Novelty Approach to Retina Diagnosing Using Biometric Techniques with SVM and Clustering Algorithms," in IEEE Access, vol. 8, pp. 125849-125862, 2020, doi: 10.1109/ACCESS.2020.3007656.