Sign Language to Speech Converter

Ushadevi M.B[1], Rohan G. Navale[2], Bhavana V. Dhongade[3], Nanditha H.G.[4], Nidhi R. Patil[5]
[1,2,3,4,5] *JNNCE, Shivamogga*

***ABSTRACT* - In the dynamic realm of technology-driven interactions, the demand for swift responses and streamlined operations has intensified. As digital experiences advance, focus shifts to innovative avenues of human-computer interaction beyond conventional methods. This project explores gesture recognition, leveraging the innate and instinctive nature of human gestures in day-to-day interactions. Through computer vision and deep learning, it transforms real-time hand movements into a sophisticated control mechanism for media players. Gesture-based communication establishes a new standard of user interaction, freeing individuals from the constraints of physical input devices. By defining seven distinct gestures, the project empowers users to seamlessly leverage their local device cameras, eliminating the need for additional hardware. This innovative approach enhances operational efficiency and enables users to exert control over laptops or desktops from a distance. The amalgamation of cutting-edge technology and intuitive human gestures addresses the evolving demands of interaction, serving as a testament to the transformative potential of gesture recognition in reshaping our digital interfaces. In this project, we focus on revolutionizing human-computer interaction through gesture recognition. The system, utilizing computer vision and deep learning, translates real-time hand movements into a sophisticated control mechanism for media players. With seven defined gestures, users can effortlessly leverage their local device cameras, eliminating the need for additional hardware. This approach augments operational efficiency and introduces a paradigm shift, allowing users to exert control over laptops or desktops from a distance. Adaptive learning mechanisms within the gesture recognition system ensure a personalized experience, continuously refining its understanding of individual gestures.**

1. INTRODUCTION

Particularly in the realm of computing. Human-Computer Interaction (HCI) stands as a crucial field, seamlessly integrating human abilities with the technical understanding of hardware and software technologies. This evolution has reshaped tasks, transformed complex calculations and wrote processes into effortless endeavors through the use of computers. As HCI rapidly evolves, emerging concepts like Gesture Recognition take center stage. Gestures, inherent and natural to all, serve as a non-cognitive computing interface allowing devices to capture and interpret human gestures as commands. This paper explores the profound applications of gesture recognition technology, ranging from virtual reality environment control to drowsiness detection in drivers. Despite the significant developments, challenges persist, such as the cost and inconvenience associated with glove-based systems. Neural networks, employing color segmentation and morphological operations, have proven efficient in addressing issues like noise in the region of interest and image background. Gestures have found applications across diverse fields, including entertainment, healthcare, and disaster relief. This paper introduces a groundbreaking approach to HCI by focusing on controlling media players with hand gestures, utilizing Convolutional Neural Networks (CNN). The proposed system offers touch-free and remote-free control, enhancing user experience and providing a productive solution for scenarios like watching movies or tutorials.

2. RELATED WORK

In Gyutae Park, Jinhwan Koh, celebrating its one-year anniversary, delves into the enhancement of accuracy in hand gesture recognition algorithms for human-machine interactions. The innovative approach presented in the paper combines 2D-FFT and convolutional neural networks (CNN) to analyze image data obtained through Ultra-Wide Bandwidth (UWB) radar. Notably, the classification results of this proposed method demonstrated comparable accuracy to well-established models, all while requiring less time for training. The paper underscores the significance of preprocessing techniques, particularly 2D-FFT, in augmenting accuracy, although not all preprocessing methods proved equally effective. Furthermore, the study reveals that prominent models boasting a substantial number of layers incurred longer execution times when compared to the efficiency of the proposed model. The experimental setup involved the utilization of IR-UWB radar equipped with two Vivaldi antennas and a bandwidth spanning from 6.0 to 8.5 GHz. In addition to presenting classification results for prominent CNN models, the research paper includes a confusion matrix comparing the Double Parallel CNN model with ResNet- 50, which exhibited the highest accuracy among the models examined. As the paper commemorates its one-year milestone, it stands as a testament to the ongoing pursuit of advancements in gesture recognition technologies.

3. METHODOLOGY

1. Data Acquisition : Then it captures each frame which gets converted from continuous videos to individual ones. The ability to acquire real-time data is vital as this forms the foundation of gesture analysis and interpretation that will follow.

2. Pre-processing : The black and white conversion of the frames is done through using OpenCV as a tool for this process. It makes the visual data simpler by emphasizing on important aspects but also eliminating clutter and distraction.

3. Feature Extraction : Accurate gesture recognition requires extracting appropriate features from the processed frames. The third stage involves extracting vital features or characteristics from the generated black and white images, which will eventually be used as input into the classifier model.

4. Classification : A model made with the Keras library is crucial for identifying these extracted features with certain hand gestures. The evaluation of the test accuracy confirms the model’s ability to give accurate predictions with the processed data.

5. Integration with Control Functions : PyAutoGUI enables the incorporation of the identified and classified gestures with the respective control functions. In this part, the identified gestures are transmitted into instructions that direct the multimedia player thus providing a smooth interaction experience.

6. Web App Integration using Streamlit: Set up Streamlit by installing it using pip and create the application with three pages: About, Project Demo Video, and Gesture Recognition. Develop the "About" page to describe the system, the "Project Demo Video" page to showcase a demonstration video, and the "Gesture Recognition" page to allow users to start the webcam and use the system. The main script navigates between pages and integrates the webcam feed and gesture recognition model. The application runs with a command to launch Streamlit, enabling an interactive web application for controlling a media player through hand gestures.

4. PROPOSED SYSTEM

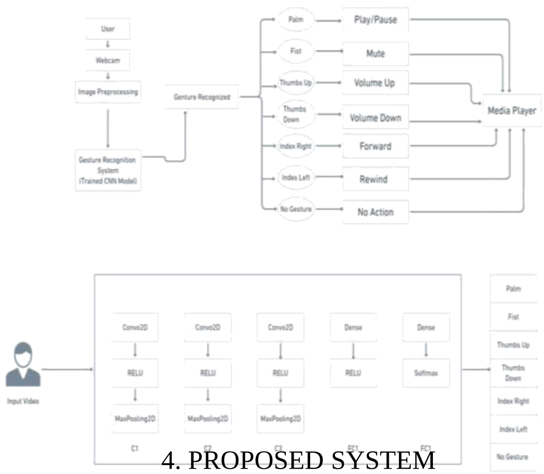

Fig 3.1[2] System design workflow

The purposed system for sign language to speech converter obtaining images when a user makes hand motions in front of the camera. OpenCV is used to gather the picture frames from the live video. To increase the accuracy of gesture prediction, these images are transformed into black and white, as illustrated in Figure 5, and are subsequently saved in the appropriate directories. Three distinct people's gestures are gathered for this project. For every gesture, there are 150 images in the dataset. To store images, a directory structure is made. For user convenience, there are two modes available: train and test. The option to select any one of the modes is provided to the user. The model is trained using the images acquired in train mode, and its accuracy is tested using the images obtained in test mode. The user-inputted directory determines where the images are kept. When the camera is turned on, two frames are shown on the screen, and the user can use the read function to take picture by picture while simulating the mirror image. The user must make the gestures while placing their hand inside the bounding box, or Region of Interest (ROI). After being taken out of ROI, the frames are resized to 120x120x1.

5. TECHNOLOGY USED

The technologies used here are Convolutional Neural Networks (CNNs) for image analysis, Python programming language for model development, and machine learning libraries like Open CV: Used for image capture and preprocessing. Keras : Utilized for building and training the Convolutional Neural Network (CNN). PyAuto GUI : Employed for integrating keyboard controls with predicted gestures. Streamlit : Used for creating the web application interface. Data Augmentation is the technique of artificially creating new data from preexisting data, known as data augmentation, is mostly used to train new machine learning (ML) models. Large and diverse datasets are necessary for the initial training of machine learning models, however finding sufficiently diverse real-world datasets can be difficult due to data silos, legal restrictions, and other issues. Data augmentation involves making minor adjustments to the original data in order to artificially enlarge the dataset. Many sectors are increasingly using generative artificial intelligence (AI) technologies for quick and high-quality data augmentation.

6. OUTPUT

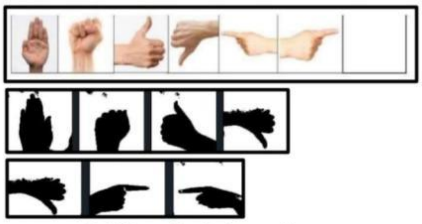

Fig 3.7 Gesture data after Pre-processing

FIG 3.28 Homepage of the webapp

Fig 3.29 Gesture recognition page

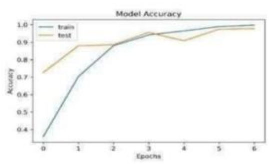

Fig 5.5 Graph of variations in accuracy with Epoch in CNN model.

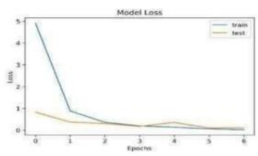

Fig 5.6 Graph of variations in loss with Epoch in CNN model

7. CONCLUSION

Based on computer vision and deep learning techniques, the suggested method of interacting with media players through hand gestures via an online application has demonstrated encouraging outcomes in terms of effectiveness and user-friendliness. It has been discovered that Convolutional Neural Networks (CNN) perform better for hand gesture recognition than earlier algorithms, offering superior evaluation metrics. Neural network-based gesture recognition has been investigated through a variety of approaches, such as image processing methods like color segmentation, contour extraction, and morphological filters. CNN models with fully connected SoftMax layers have been used, and the results show that they can recognize gestures with up to 96.83% accuracy. Although issues with background noise, illumination and processing speed persist, technological developments and methods such as data augmentation can help get around these problems and enhance the effectiveness of gesture recognition even more.

8. FUTURE SCOPE

To make gesture recognition systems more accurate and resilient in difficult situations, like cluttered areas and situations with people in the background, more research can be done. Investigating data augmentation methods can improve gesture recognition performance by expanding the variety and volume of training data. Other sensors, like infrared cameras or depth sensors, can be integrated to provide more data and increase the precision of gesture recognition systems. To make gesture-controlled systems more responsive and user-friendly, real-time gesture recognition algorithms that can handle dynamic gestures and continuous movements should be investigated. Increasing the number of gestures that are recognized and creating customizable gesture mapping options can provide users greater freedom and control over how they interact with media players and related apps. Investigating gesture recognition's potential in other fields, like robotics, augmented reality, and virtual reality, may create new opportunities for natural and engaging human-computer interaction.

9. REFERENCES

[1] Li, S., Chan, A. B., & Chung, A. C. S. (2018). "Hand Gesture Recognition Using Convolutional Neural Networks for Real-Time Applications." IEEE Transactions on

Multimedia, vol. 20, no. 8, pp. 2081-2090, August 2018.

[2] Wang, J., Chen, Y., Hao, S., & Pan, W. (2016). "Deep Learning for Hand Gesture Recognition on Skeletal Data." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 189-196, June 2016.

[3] Cao, Q., Pan, J., & Ngo, C. W. (2017). "Hybrid CNN and Dictionary-Based Models for Hand Gesture Recognition." IEEE Transactions on Image Processing, vol. 26, no. 5, pp. 2265-2278, May 2017.

[4] Wu, Y., Lin, Z., Wang, X., & Gao, W. (2018). "MV-CNNs for Real-Time Hand Gesture Recognition." IEEE Transactions on Cybernetics, vol. 48, no. 4, pp. 1166-1178, April 2018.

[5] Li, X., Yang, Y., Zhou, F., & Li, Y. (2019). "Real-Time Hand Gesture Recognition Using Depth and RGB Information." IEEE Access, vol. 7, pp. 135631-135641, September 2019.

[6] Yun, K., Honorio, J., Chattopadhyay, D., Berg, T. L., & Samaras, D. (2012). "Two-Handed Gesture Recognition in Ego-Centric View." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1919-1926, June 2012.

[7] Zhao L. Gesture Control Technology: An investigation on the potential use in Higher Education. University of Birmingham, IT Innovation Centre : Birmingham, UK. 2016 Mar.

[8] Harshada Naroliya, Tanvi Desai, Shreeya Acharya, Varsha Sakpal Enhanced Look Based Media Player with Hand Gesture Recognition. International Research Journal of Engineering and Technology, (ISSN: 2395-0072 (P), 2395-0056 (O)).2018;5(3): 2032-35.

[9] Wachs JP, Kölsch M, Stern H, Edan Y. Vision-based handgesture applications.

[10] Yashas J, Shivakumar G. Hand Gesture Recognition : A Survey. In2019 International Conference on Applied Machine Learning (ICAML) 2019 May 25 (pp. 3- 8). IEEE.

[11] Sharma P, Sharma N. Gesture Recognition System. In2019 4th International Conference on Internet of Things: Smart Innovation and Usages (IoT-SIU) 2019 Apr 18 (pp. 1-3). IEEE.

[12] Hakim NL, Shih TK, Kasthuri Arachchi SP, Aditya W, Chen YC, Lin CY. Dynamic hand gesture recognition using 3DCNN and LSTM with FSM context-aware model. Sensors. 2019 Jan;19(24):5429.

[13] Cardenas EJ, Chavez GC. Multimodal hand gesture recognition combining temporal and pose information based on CNN descriptors and histogram of cumulative magnitudes. Journal of Visual Communication and Image Representation. 2020 Aug 1; 71:102772.

[14] Adithya V, Rajesh R. A deep convolutional neural network approach for static hand gesture recognition. Procedia Computer Science. 2020 Jan 1;171:2353-61. [15] Oudah M, Al-Naji A, Chahl J. Hand gesture recognition based on computer vision: a review of techniques. journal of Imaging. 2020 Aug;6(8):73.

[15] Mohammed AA, Lv J, Islam MD. A deep learning-based Endto-End composite system for hand detection and gesture recognition. Sensors. 2019 Jan;19(23):5282