Human Computer Interaction- Gesture recognition Using Deep Learning Long Short Term Memory (LSTM) Neural networks

Dr. Ranga Swamy Sirisati[1], A. Eenaja[2], N.Sreeja[3],

T. Srajan Kumar[4], B. Naresh[5]

[1,5]Associate Professor, Department of CSE (AI&ML), Vignan's Institute of Management and Technology for Women, Kondapur, Ghatkesar, Telangana
[2]Assistant Professor, Department of Information Technology, Vignan's Institute of Management and Technology for Women, Kondapur, Ghatkesar, Telangana
[3]Assistant Professor, Department of CSE (DS), Vignan's Institute of Management and Technology for Women, Kondapur, Ghatkesar, Telangana
[4]Assistant Professor, Department of CSE, MallaReddy University Hyderabad.

Email id : sirisatiranga@gmail.com[1], eenajareddy979@gmail.com[2], sreeja@vmtw.in[3], teralasrajankumar@gmail.com[4], bhukyanaresh774@gmail.com[5]

Article Received: 06 Mar 2024 *Article Accepted: 30 Apr 2024* *Article Published: 30 May 2024*

Citation
Dr. Ranga Swamy Sirisati, A. Eenaja, N.Sreeja, T. Srajan Kumar, B. Naresh (2024). "Human Computer Interaction- Gesture recognition Using Deep Learning Long Short Term Memory (LSTM) Neural networks", Journal of Next Generation Technology (ISSN: 2583-021X), 4(2), pp. 9-18. May 2024.

Abstract

Computers that use gesture recognition can better comprehend human body language. This facilitates the development of a stronger connection between humans and machines than that which is possible with only text-based or graphical user interfaces (GUIs). The computer camera in this paper for gesture recognition reads the movements of the human body. In order to compare the results of hand identification, deep learning approaches including the Yolo model, Inception Net model+LSTM, 3-D CNN+LSTM, and Time Distributed CNN+LSTM have been explored for this paper. After then, the computer uses this data as input to run programs. The Yolo model performs better than the other three types. Twenty billion jester films, or 20% of the total, and Kaggle were used to train the algorithms. The following stage is to adjust the system loudness based on the direction of hand movement once the hand has been detected in the collected frames. Creating and finding the bounding box on the detected hand yields the direction of the hand movement.

Keywords: Human Computer Interaction, Machine Learning, Deep Learning, Long Short Term Memory (LSTM) Neural networks.

I. INTRODUCTION

Gesture-controlled interactive surfaces have proliferated in the past few years. Given that speech, gestures, and facial expressions make up the majority of real human communication, it seems sense for interface developers to attempt to mimic this behavior. With smartphones becoming more and more common, gesture-based interaction has become a part of daily life. Gestures provide users a new way to connect that feels natural and doesn't require interruptions or extra devices. It's like having a smartphone in the real world. Additionally, the user is not restricted to a single point of input; rather, a variety of interaction options are provided. [1].

Designing hand gestures and facial expression-based Human-Computer-Interaction (HCI) systems continues to be one of the most difficult tasks among many computer vision-based interactive systems. Finding a non-tangible method of computer interaction is our core goal. Computers that use gesture recognition can better comprehend human body language. This facilitates the development of a stronger connection between humans and machines than that which is possible with only text-based or graphical user interfaces (GUIs). Our paper seeks to create an audio player that can be operated by hand gestures made by people. [2][3].

Using mathematical algorithms, gesture recognition is a computational method that aims to identify and comprehend human gestures. An international conference devoted to gesture and facial recognition is held annually in the expanding subject of computer science known as gesture recognition. The movements would act as a clear instruction for actions like turning up or down the music. [4][5].

This paper aims to provide an interface that can both dynamically capture human hand gestures and adjust the level of loudness. Yolo model and other deep learning approaches are utilized for this. In order to improve human-computer interaction, gesture recognition computer processes can be implemented in a number of methods, including touch screens, cameras, and peripheral devices[6][7]. By contrast, the vision-based approach requires only a camera to recognize the real interaction between a person and a computer, eliminating the need for additional hardware. This system has persistent background issues, occasionally with people and lights as well. Along with the recognition methods, many procedures and algorithms employed in this system are explained here[8][9]. Segmentation is the process of looking for a connected area in an image where a pattern and algorithm may be adjusted, based on specified parameters like color or intensity. This procedure uses mathematical algorithms to try and identify and understand human gestures.[10][11].

An international conference devoted to gesture and facial recognition is held annually in the expanding subject of computer science known as gesture recognition[12]. The acronym for Open Source Computer Vision is OpenCV. Its name suggests that it is an open-source library for computer vision and machine learning[13]. In addition to many other things, it may be used for face identification, objection tracking, and landmark detection.

The goals of the paper are as follows:

1) To use a computer keyboard and mouse as little as possible.

2) To inexpensively incorporate gesture recognition capabilities into any computer.

3) To assist in the creation of an intangible method of system interaction[14].

4) The major goal of this endeavor is to determine the key benefits and drawbacks of each feature extraction approach in order to investigate the usefulness of two methods, namely hand contour and complex moments, to address the hand gesture identification problem[27].

II. RELATED WORKS

For many individuals these days, recognizing touch screen gestures comes naturally to them. The majority of people nowadays are aware that you may pinch-to-zoom on a touch screen to enlarge an image for a closer look, even though some computers and operating systems provide customizable gesture detection [15][16][17]. This particular motion works with almost every type of user interface, including those on PCs and smartphones. Touch screen interfaces make it relatively simple for people to interact with computers. For hand gesture recognition, the current technology makes use of a digital camera [18][19][20]. In order to record gestures, an item had to be in front of the camera, which served as an input device. It is quite expensive, restricts mobility, and prevents using a camera in dimly light areas [24][25][26]. GRS (Gesture Recognition System) qualities are heavily influenced by the camera specs [21][22][23].

III. METHODOLOGY

People must touch the screen in order to regulate the volume on a touch screen. Therefore, the suggested approach allows the user to adjust the volume using hand gestures without touching the screen, in place of using touch or buttons. The suggested system measures acceleration in the form of tilt, shock, and vibration based on the input signal from MEMS (Micro Electro Mechanical System) accelerometer sensors. It detects accurately and consistently. Additionally, a recognition system for six distinct static hand gestures—Open, Close, Cut, Paste, Maximize, and Minimize—was shown.

Whether it is for facial motions or whole body gestures, gesture recognition has been redesigned for various research applications (Dong, Yan, & Xie, 1998).

Although creating a recognition system that functions well in a variety of environments is difficult, it is more feasible since these obstacles actually exist in the real world. These requirements include distinct backgrounds for compounds and lighting, in addition to minimal translation, rotation, and scaling effects at specific angles. The cost of computation is another factor that has to be considered. A few feature extraction methods have the drawback of being erratic, which causes them to take longer. On the other hand, the majority of hand gesture recognition systems simply consider evaluation accuracy. In the last stage of result review, two factors must be taken into account: accuracy and computing cost. Accuracy is important since it helps identify results' strengths and weaknesses and pave the way for future applications. Touch-free systems are essential in our modern times. The user may communicate with the system without touching any buttons or the screen by using gestures.

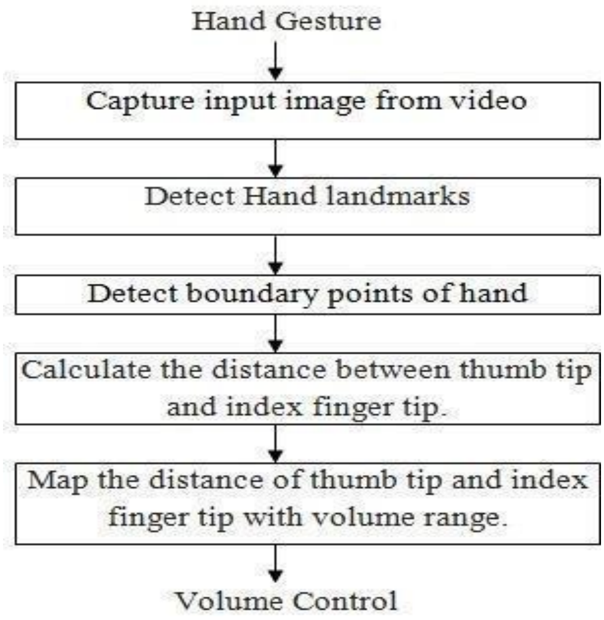

Fig 1. Flow chart

The flow chart is explained in Figure1. The development of technology based on the hand gesture system is a major business for companies such as Microsoft, Samsung, and Sony. Products based on this system include laptops, handheld devices, professional lighting, and LED lights. The field of hand gesture recognition is expanding rapidly for futuristic products and services. Adoption and utilization will get more affordable and cost-effective. It's an amazing feature that combines human wave with technology to transform data into features. The goal of the paper is to build a synchronous gesture classification system that can automatically identify movements under various illumination conditions. In order to do this, a real-time, synchronous gesture recognition system is created.

The goal of this paper is to create a comprehensive system that uses computer vision to recognize, locate, and interpret hand motions. This structure will function as one of the AI and computer vision concepts that allows for user involvement. It develops a function to recognize hand gestures using different parameters. The structure's primary goal is to make the system user-friendly, straightforward, and easy to manage without requiring any specialized gear.

This paper aims to provide an interface that can both dynamically capture human hand gestures and adjust the level of loudness. To compare the outcomes of hand detection, deep learning approaches such the Yolo model, Inception Net model+LSTM, 3-D CNN+LSTM, and Time Distributed CNN+LSTM have been researched. This paper uses OpenCV to change the current volume controller that is operated by human graphical motions.

IV. EXPERIMENTS AND RESULTS

The space consumption in the region between the convex hull and the hand's contour was used to compute the detection of hand gestures. By creating a description in which the edges are contained in the smallest convex set, a convex hull is applied to determine the edge of the user's hand. This paper uses OpenCV library files that may be utilized with any kind of image processing approach. Figure 2 depicts the suggested system's architecture. The files are in the (. png) format, together with their masked labels. The real-time video stream is used to collect the pictures, which are then processed further to increase the detection stage's effectiveness. During the detection phase, contours for the pictures are first determined, and the Convex Hull surrounding the hand region is then created using these contours. The computer system is then given control orders to carry out the related activity based on these visual outputs.

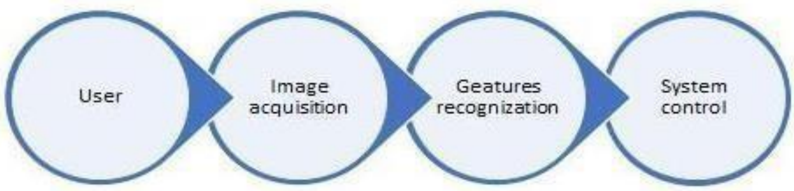

Fig. 2. Implementation

Fig. 3 depicts the suggested system's architecture. The files are in the (. png) format, together with their masked labels. The real-time video stream is used to collect the pictures, which are then processed further to increase the detection stage's effectiveness. During the detection phase, contours for the pictures are first determined, and the Convex Hull surrounding the hand region is then created using these contours. The computer system is then given control orders to carry out the related activity based on these visual outputs.

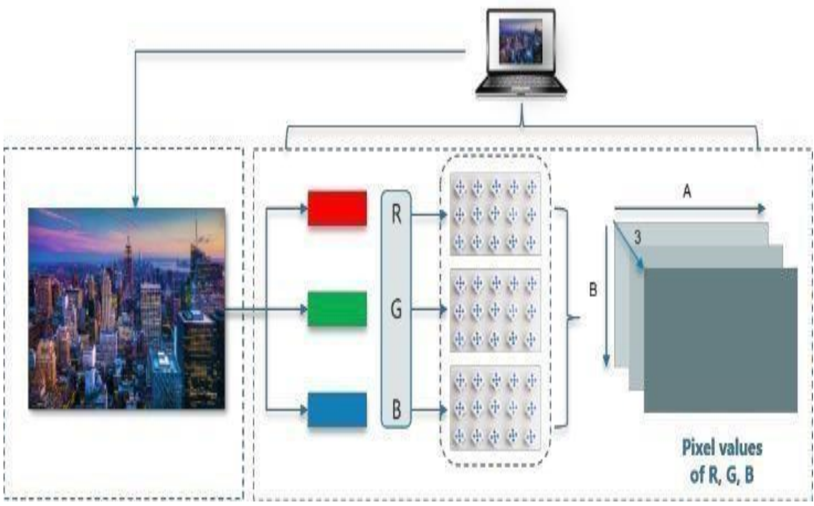

Fig. 3. Computer Understanding Image

It is evident from Figure that in order to differentiate between these photos, the three dimensions matrix pixel values (RGB values) for each image in are essential.

The basic goal of the semantic segmentation approach is to recognize several conceivable movements and only segment the hand area from the live video stream in order to locate distinct regions in an image and tag the labels that correspond to them.

The hand's area is covered by the peaks that make up the convex hull cluster. Here, it has to pass the convex set principle, which states that any lines connecting any two locations on the hull are included wholly within it (see Figure 4).

The specified functionality is carried out after the gesture has been identified. It is a dynamic procedure to identify the movement. Proceed to the initial stage to accept further photographs for processing, and so on, after executing the designated instruction from the gesture.

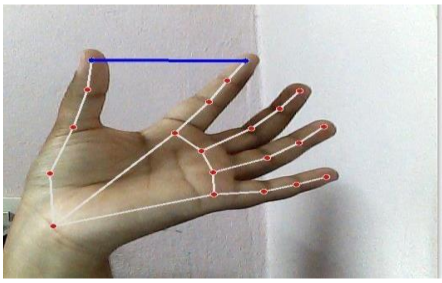

Figure 4: Hand Gesture

In order to measure the separation between the tips of the thumb and index fingers, the camera recognizes hands that have points in them. As seen in Figure 5, the distance between points 4 and 8 is exactly proportional to the device's volume.

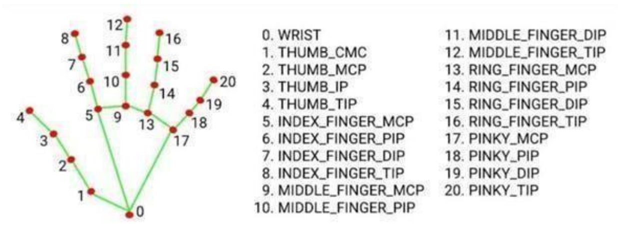

Figure 5: Coordinates of the key points of the hand

- Look for hand landmarks
- Measure the separation between the tips of your thumb and index finger.

• Chart the thumb and index finger tips' respective distances from the volume range. In this instance, the volume range was between -63. 5 and 0. 0, while the distance between the tips of the thumb and index finger fell between 30 and 350.

To escape, use the spacebar.

The movements are recognized and categorized into distinct gestures using OpenCV algorithms, and the output includes visual threshold and color representations are represented in Figure6, figure7 and figure8 respectively. In order to carry out the corresponding action command linked to volume, the extracted, processed, and converted streaming frames are combined with region segmentation during the live video. These picture outputs are then delivered as input into the system volume.

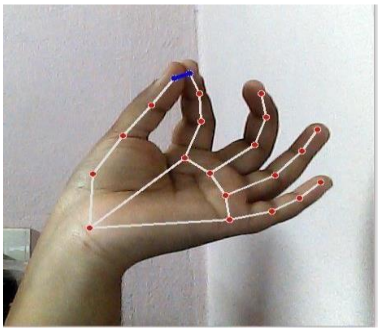

Figure 6: During the live video, streaming frame-1

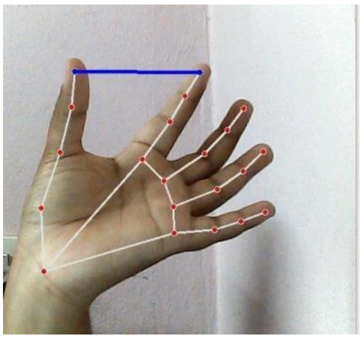

Figure 7: During the live video, streaming frame-2

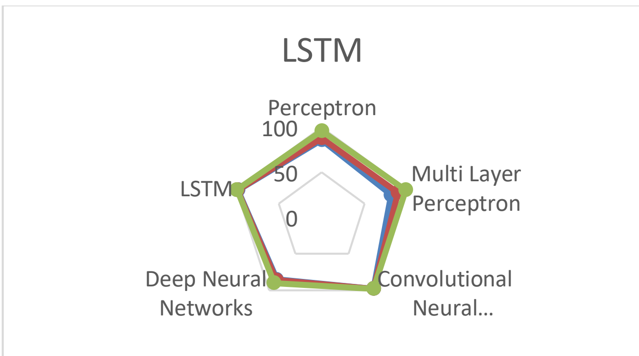

Figure 8: Differentiating LSTM with other Algorithms

V. CONCLUSIONS AND FUTURE SCOPE OF WORK

In conclusion, the paper's results demonstrate that the theories of hand segmentation and hand detection system can be used to the development of hand gesture recognition utilizing Python and OpenCV. In summary, the system has achieved many goals in this paper: Using Python and computer vision, create a comprehensive system for hand gesture detection, identification, and interpretation. able to adjust the system's loudness using a hand motion that corresponds to the paper's name on the system. Future work will focus on developing additional gestures and the corresponding activities needed to establish a computerized gesture detection system.. Since delay must be taken into account, a significant amount of study remains to be done in this area. The ultimate goal is to minimize the paper response time in real-world scenarios because even a single second of latency might result in distinct controls operating.

References

[1]. Ujitoko Y, Tokuhisa R, Sakurai S, Hirota K. Impact vibration source localization in two-dimensional space around hand. IEEE Trans Haptics. 2021; 14(4): 862-873.

[2]. Li B, Yang J, Yang Y, Li C, Zhang Y. Sign language/gesture recognition based on cumulative distribution density features using UWB radar. IEEE Trans Instrument Measur. 2021; 70: 1-13.

[3]. Aswini J, Yamini B, Jatothu R, Nayaki KS, Nalini M. An efficient cloud-based healthcare services paradigm for chronic kidney disease prediction application using boosted support vector machine. Concurr Comput Pract Exp. 2022; 34(10):e6722.

[4]. Sinha K, Kumari R, Priya A, Paul P. A computer vision-based gesture recognition using hidden Markov model. Innovations in Soft Computing and Information Technology. Springer; 2019: 55-67.

[5]. Bakheet S, Al-Hamadi A. Robust hand gesture recognition using multiple shape-oriented visual cues. EURASIP J Image Video Process. 2021; 2021(1): 1-18.

[6]. Suneetha M, Prasad MVD, Kishore PVV. Multi-view motion modelled deep attention networks (M2DA-net) for video based sign language recognition. J Vis Commun Image Represent. 2021; 78:103161.

[7]. Choi JW, Ryu SJ, Kim JH. Short-range radar based real-time hand gesture recognition using LSTM encoder. IEEE Access. 2019; 7: 33610-33618.

[8]. Noor S, Uddin V, Jabeen DE. Probabilistic hierarchical model using first person vision for scenario recognition. Wirel Pers Commun. 2019; 106(4): 2179-2193.

[9]. Swamy, S. Ranga, et al. "Dimensionality reduction using machine learning and big data technologies." Int. J. Innov. Technol. Explor. Eng.(IJITEE) 9.2 (2019): 1740-1745.

[10]. Swamy, R. S., S. C. Kumar, and G. A. Latha. "An efficient skin cancer prognosis strategy using deep learning techniques." Indian Journal of Computer Science and Engineering (IJCSE) 12.1 (2021).

[11]. Sirisati, Ranga Swamy, et al. "An Enhanced Multi Layer Neural Network to Detect Early Cardiac Arrests." 2021 5th International Conference on Electronics, Communication and Aerospace Technology (ICECA). IEEE, 2021.
[12]. Sirisati RS, Prasanthi KG, Latha AG. An aviation delay prediction and recommendation system using machine learning techniques. InProceedings of Integrated Intelligence Enable Networks and Computing: IIENC 2020 2021 (pp. 239-253). Springer Singapore.
[13]. Sirisati, Ranga Swamy, et al. "Cancer Sight: Illuminating the Hidden-Advancing Breast Cancer Detection with Machine Learning-Based Image Processing Techniques." 2023 International Conference on Sustainable Communication Networks and Application (ICSCNA). IEEE, 2023.
[14]. Sirisati, Ranga Swamy, et al. "A Deep Learning Framework for Recognition and Classification of Diabetic Retinopathy Severity." Telematique 23.01 (2024): 228-238.
[15]. Praveen, S. P., Vellela, S. S., & Balamanigandan, R. (2024). SmartIris ML: Harnessing Machine Learning for Enhanced Multi-Biometric Authentication. *Journal of Next Generation Technology (ISSN: 2583-021X)*, *4*(1).
[16]. adhuri, A., Jyothi, V. E., Praveen, S. P., Altaee, M., & Abdullah, I. N. (2023). Granulation-Based Data Fusion Approach for a Critical Thinking Worldview Information Processing. *Journal of Intelligent Systems and Internet of Things*, *9*(1), 49-68.
[17]. Dedeepya, P., Sowmya, P., Saketh, T. D., Sruthi, P., Abhijit, P., & Praveen, S. P. (2023, February). Detecting cyber bullying on twitter using support vector machine. In *2023 Third International Conference on Artificial Intelligence and Smart Energy (ICAIS)* (pp. 817-822). IEEE.
[18]. Praveen, S. P., Suntharam, V. S., Ravi, S., Harita, U., Thatha, V. N., & Swapna, D. (2023). A Novel Dual Confusion and Diffusion Approach for Grey Image Encryption using Multiple Chaotic Maps. *International Journal of Advanced Computer Science and Applications*, *14*(8).
[19]. Jyothi, V. E., Kumar, D. L. S., Thati, B., Tondepu, Y., Pratap, V. K., & Praveen, S. P. (2022, December). Secure data access management for cyber threats using artificial intelligence. In *2022 6th International Conference on Electronics, Communication and Aerospace Technology* (pp. 693-697). IEEE.
[20]. Praveen, S., Nguyen, H., Swapna, D., Rao, K., & Kumar, D. (2020). The efficient way to detect and stall fake articles in public media using the blockchain technique: Proof of trustworthiness. *International Journal on Emerging Technologies*, *11*(3), 158-163.
[21]. Swapna, D., & Praveen, S. P. (2020). An exploration of distributed access control mechanism using blockchain. In *Smart Intelligent Computing and Applications: Proceedings of the Third International Conference on Smart Computing and Informatics, Volume 2* (pp. 13-20). Springer Singapore.
[22]. Vemulakonda, R., Meka, S., Ketha, V., Surapaneni, P. P., & Kondapalli, S. V. (2016). An algorithm for basic IoT Architectures.

[23]. Sindhura, S., Praveen, S. P., Rao, N., & ArunaSafali, M. (2021, September). An energy efficient novel routing protocol in wireless sensor networks (WSN). In *2021 Third International Conference on Inventive Research in Computing Applications (ICIRCA)* (pp. 319-324). IEEE.

[24]. Marrapu, B. V., Raju, K. Y. N., Chowdary, M. J., Vempati, H., & Praveen, S. P. (2022, January). Automating the creation of machine learning algorithms using basic math. In *2022 4th International Conference on Smart Systems and Inventive Technology (ICSSIT)* (pp. 866-871). IEEE.

[25]. S. Phani Praveen , Thulasi Bikku, P. Muthukumar, K. Sandeep, Jampani Chandra Sekhar, V. Krishna Pratap. (2024). Enhanced Intrusion Detection Using Stacked FT-Transformer Architecture. Journal of , 13 (2), 19-29 (**Doi :** https://doi.org/10.54216/JCIM.130202)

[26]. Bikku, T., Chandolu, S. B., Praveen, S. P., Tirumalasetti, N. R., Swathi, K., & Sirisha, U. (2024). Enhancing Real-Time Malware Analysis with Quantum Neural Networks. *Journal of Intelligent Systems and Internet of Things*, *12*(1), 57-7.

[27]. Aruna, R., Kushwah, V. S., Praveen, S. P., Pradhan, R., Chinchawade, A. J., Asaad, R. R., & Kumar, R. L. (2024). Coalescing novel QoS routing with fault tolerance for improving QoS parameters in wireless Ad-Hoc network using craft protocol. *Wireless Networks*, *30*(2), 711-735.