STM JOURNALS

Journal of
Computer Technology & Applications

ISSN: 2229-6964 (Online)
ISSN: 2347-7229 (Print)
Volume 15, Issue 1, 2024
January–April
DOI (Journal): 10.37591/JoCTA

http://computerjournals.stmjournals.in/index.php/JoCTA/index

Review JoCTA

Controlling Media Player Through Hand Gesture Recognition System Using CNN and RNN Models

Khallikkunaisa[1,*], Mayank N. Nandwani[2], Rohit Chauhan[2], Prabhant Kumar[2], Nitesh Kumar[2]

Abstract

Artificial intelligence markup language (AIML) project represents a pioneering endeavor in the realm of media player control through hand gesture recognition, merging advanced technologies like convolutional neural networks (CNN) and recurrent neural networks (RNN). By harnessing the image analysis capabilities of CNN, our system ensures accurate, real-time detection, and interpretation of intricate hand gestures, enabling users to interact with their media content naturally and seamlessly. What sets our project apart is the incorporation of RNN, which imbues the system with a temporal understanding of gestures, enhancing its ability to recognize complex sequences of gestures and commands, including actions like play, pause, or nuanced volume adjustments. This synergy of CNN and RNN not only exemplifies the transformative potential of AI and deep learning in human-computer interaction but also promises to redefine the user experience, offering an accessible, responsive, and immersive means of controlling digital entertainment. Our project addresses a critical need for more intuitive user interfaces, particularly for individuals with physical limitations, making media playback more inclusive and engaging.

Keywords: Hand gesture recognition, convolutional neural network (CNN), recurrent neural network (RNN), human-computer interaction, inclusive user interfaces

INTRODUCTION

In today's digital landscape, the fusion of cutting-edge technologies has paved the way for innovative interfaces that redefine human-computer interaction. Among these advancements, the integration of convolutional neural networks (CNNs) and recurrent neural networks (RNNs) has led to a pioneering endeavor known as the artificial intelligence markup language (AIML) project. This project marks a significant leap in the realm of media player control through hand gesture recognition [1–3].

***Author for Correspondence**
Khallikkunaisa
E-mail: khallikkunaisa.cs@hkbk.edu.in

[1]Associate Professor, Computer Science and Engineering, HKBK College of Engineering, Bengaluru, Karnataka, India
[2]Student, Computer Science and Engineering, HKBK College of Engineering, Bengaluru, Karnataka, India

Received Date: February 22, 2024
Accepted Date: February 28, 2024
Published Date: April 05, 2024

Citation: Khallikkunaisa, Mayank N. Nandwani, Rohit Chauhan, Prabhant Kumar, Nitesh Kumar. Controlling Media Player Through Hand Gesture Recognition System Using CNN and RNN Models. Journal of Computer Technology & Applications. 2024; 15(1): 29–34p.

The AIML project stands as a testament to the transformative potential of AI-driven systems, particularly in redefining how individuals interact with digital entertainment. At its core, this initiative harnesses the prowess of CNN, enabling accurate and real-time detection of intricate hand gestures. This functionality empowers users to seamlessly and naturally engage with their media content, transcending the constraints of traditional control mechanisms like remote controllers or touchscreens.

However, what sets the AIML project apart lies in its integration of RNNs, which imparts a temporal understanding of gestures to the system. This temporal awareness enhances the system's proficiency

in recognizing complex sequences of gestures and commands, accommodating actions ranging from fundamental tasks like play and pause to nuanced volume adjustments. This amalgamation of CNN and RNN not only showcases the synergistic potential of deep learning but also underscores its pivotal role in shaping the future of human-computer interaction.

Beyond its technological prowess, the AIML project is poised to make significant strides in addressing a critical need for more intuitive user interfaces. This innovation holds promise, particularly for individuals facing physical limitations, as it endeavors to make media playback more inclusive, engaging, and accessible to diverse user demographics.

This survey paper aims to delve into the intricacies of hand gesture recognition systems, examining the landscape of existing technologies, their methodologies, advantages, limitations, and the impact of incorporating CNN and RNN in reshaping the realm of media player control through hand gestures [4–9].

Proposed Method

The proposed device aims to interpret hand gestures for controlling media playback without requiring keyboard interaction. Utilizing OpenCV, the system accesses video feeds from the webcam, while PyAutoGUI integrates Python code to manage the media player. VLC media player compatibility is recommended for optimal functionality.

To initiate, the system captures video using the cv2 video capture feature, implementing cv2 Gaussian blur to reduce background noise and enhance clarity. Dilate and erode functions further refine the image, facilitating finger detection based on a 90-degree angle. PyAutoGUI then correlates recognized gestures with keyboard commands.

Data acquisition relies on the computer's built-in webcam. Segmentation involves diverse methods: a skin detection model for hand region identification and an approximate median technique for background subtraction. The recognition phase employs a decision tree as the classification tool. For Windows interaction, appropriate commands are triggered within the media player based on detected gestures as shown in Figure 1.

LITERATURE SURVEY

1. The research paper, "*Human-Computer Interaction using Hand Gesture Recognition*," authored by Selvarathi et al. [10], explores the broad applications of hand gesture recognition in connecting humans and computers. The abstract highlights the significance of gesture recognition in diverse areas, from therapeutic recovery to controlling consumer gadgets. The methodology section details a three-stage process for edge detection in an internet of things (IoT)-based system, utilizing discrete wavelet transform (DWT) and Fisher ratio (F-ratio) for feature extraction, showcasing its effectiveness in recognizing hand signals in an uncontrolled environment without relying on specific hand poses or equipment [10].
2. The research paper, "*Hand Gesture Recognition using Machine Learning Algorithms*," authored by Abhishek et al. [11], delves into the emerging field of gesture recognition, focusing on recognizing human gestures through mathematical algorithms for human-computer interaction (HCI). The paper highlights the limitations of traditional input devices like keyboards and mice and underscores the importance of gesture recognition for building user-friendly interfaces. The

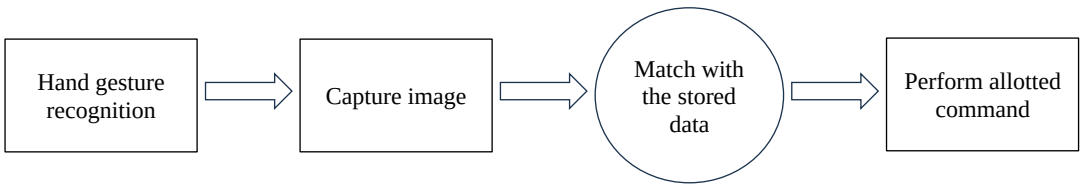

Figure 1. Modules.

methodology involves three key steps: learning, detection, and recognition, using a training dataset to train the system, capturing images through a webcam, and employing a 3D CNN for gesture recognition. The performance methodology evaluates the system's efficiency in real time HCI, considering metrics such as correct detection rates, responsiveness to user actions, and adaptability to various environments and lighting conditions. The advantages of the hand gesture recognition system include facilitating natural interaction, eliminating the need for physical input devices, operating in real-time for instant control, and demonstrating adaptability to diverse conditions.

3. The research paper, "*Static Hand Gesture Recognition Using Novel Convolutional Neural Network and Support Vector Machine*," authored by Veronica et al. [12], addresses the challenges of hand tracking and identification through visual means. The paper proposes a novel approach involving a CNN-based method for user-free hand gesture detection, emphasizing advantages such as higher accuracy and real-time recognition suitable for sign language and HCI [12]. The methodology introduces a custom model architecture for static hand gesture recognition, incorporating two sets of convolution layers, average pooling, fully connected layers, and a Softmax classifier. The implementation process involves image capture, pre-processing, feature extraction using the CNN, training, optimization, support vector machine (SVM) classification, and real-time implementation for computational efficiency. The performance methodology includes training, testing, and evaluation using standard metrics for accuracy, precision, recall, and F1 score, demonstrating the model's effectiveness in recognizing static hand gestures. The advantages of the proposed system include high accuracy, a customizable architecture for flexibility, efficient feature extraction using the novel CNN, and optimization for real-time applications.
4. The research paper, "*Hand Gesture Recognition Using EMG-IMU Signals and Deep Q-Networks*," authored by Vásconez et al. [13], introduces a novel approach to hand gesture recognition using reinforcement learning techniques. The developed system employs a deep Q-network (DQN) algorithm to classify and recognize both static and dynamic hand gestures, utilizing information from the Myo armband and G-force sensors [13]. The results demonstrate high accuracy, reaching up to 97.50%.
5. The research paper titled "*Hand Gesture Recognition System to Control Keyboard Functions*," explores hand gesture recognition for keyboard control [14]. The system utilizes convolutional and recurrent neural networks for accurate gesture interpretation. The methodology involves image capture, preprocessing, and real-time gesture mapping, aiming to enhance user-computer interaction. Performance evaluation includes real-time recognition metrics, user feedback, and comparative analysis against existing systems. Advantages include hands-free operation, improved accessibility, customization, and enhanced efficiency. Challenges include accurate gesture recognition, algorithm complexity, real-time processing, and lighting dependency. The paper presents an innovative approach to human-computer interaction, demonstrating the potential of gesture recognition in revolutionizing traditional keyboard input methods.
6. The paper, "*Gesture-Based Media Player Controller*" [15], explores the use of hand gesture recognition to control the VLC media player, providing an efficient and low-cost HCI solution. It emphasizes the increasing demand for quick responses in complex systems and the versatility of touch recognition. The system allows users to control their laptop/desktop remotely using hand gestures, offering a unique interface beyond traditional input devices. The application caters to individuals with physical disabilities by allowing them to define touches based on their abilities. The success of the system is demonstrated through actions like volume adjustment and video rewinding. Overall, the paper presents a practical and cost-effective solution for improving HCI with gesture-based media player control [15].
7. The study explores the application of deep learning techniques for image classification tasks. The study focuses on leveraging CNNs and transfer learning for accurate image classification. CNNs are known for their ability to automatically learn intricate features from images, while transfer learning allows the model to utilize pre-trained networks and adapt them to new classification tasks with limited data. The methodology involves training CNN models using a dataset of diverse images, fine-tuning pre-trained models, and evaluating their performance on image

classification tasks. The paper showcases the effectiveness of deep learning techniques in achieving high accuracy and efficiency in image classification, demonstrating the potential of CNNs and transfer learning for various real-world applications.

8. The paper authored by Toro-Ossaba et al. [16] introduces a novel approach using a RNN with long short-term memory (LSTM) units and dense layers for hand gesture classification based on electromyography (EMG) signals. The objective is to create a gesture classifier for hand prosthesis control, significantly reducing the number of EMG channels required and overall model complexity to enhance scalability for embedded systems. Their model, utilizing only four EMG channels, successfully recognizes five hands. This methodology establishes a pathway for reducing complexity in gesture recognition for human-machine interaction across different computational devices [16].
9. Haria et al.'s [17] paper introduces a marker-less hand gesture recognition system aiming to enhance HCI by minimizing reliance on traditional physical controllers like mice and keyboards [17]. The system efficiently tracks static and dynamic hand gestures, translating them into various actions such as opening websites, launching applications like PowerPoint, and facilitating slide navigation during presentations. The study demonstrates that an intuitive HCI can be achieved with minimal hardware requirements, fostering a more natural interface between users and computers. The methodology section emphasizes enriching HCI by integrating various sensory modes beyond conventional keyboard and mouse interactions.
10. The paper byKim and Rhee [18] addresses the utilization of human voices and body gestures in controlling intelligent devices, aiming to diminish reliance on conventional input methods like keyboards and mice. Focusing on hand gestures, often integral to non-verbal communication, the study employs fuzzy inference and recurrent neural networks with bidirectional LSTM architecture to recognize natural and continuous hand gestures for human-robot interaction. The approach involves selecting meaningful gestures using fuzzy theory, hand position interpolation, and Kalman filtering for occlusion handling. Despite challenges in selecting the correct meaningful motions, the system exhibits high hand gesture recognition accuracy once selected correctly. The methodology encompasses utilizing bidirectional LSTM architecture for recognizing meaningful hand gestures, emphasizing its effectiveness despite the difficulty in selecting the appropriate motion [18].
11. The paper by Lai and Yanushkevich [19] from the Biometric Technologies Laboratory, University of Calgary, Alberta, Canada, introduces a fusion-based approach using CNNs and RNNs to recognize dynamic hand gestures, leveraging depth and skeleton data. The proposed dual-network fusion aims to capture both spatial and temporal information for comprehensive hand gesture recognition. By combining CNN with LSTM networks for depth data and utilizing RNN for skeleton-based patterns, The methodology entails acquiring and preprocessing a dataset containing depth images and skeleton point sequences, establishing independent CNN + LSTM and RNN networks, and subsequently fusing predictions from both networks to create a robust model.
12. Hua et al.'s [20] paper introduces an online object tracking approach that addresses challenges posed by object transformations during tracking-by-detection scenarios. The methodology formulates object tracking as a proposal selection task and makes two main contributions. Firstly, it introduces novel proposals derived from geometric transformations undergone by the object, augmenting the candidate set for predicting the object's location. Secondly, it devises a selection strategy utilizing multiple cues, including detection scores, object edges, and motion boundaries, for robust proposal selection. Extensive evaluations on various benchmark datasets demonstrate superior performance compared to existing methods. The methodology involves a four-step framework. It begins with learning an initial detector model using a training set consisting of positive and automatically extracted negative samples [20].
13. The paper by Nunez and Cabido [21] introduces a deep learning-based approach for recognizing human activity and hand gestures from 3D skeleton data sequences. The methodology combines CNNs and LSTM networks, enabling effective temporal 3D pose recognition. The proposed methodology involves a fusion of CNN and LSTM networks, adapted to handle time series data of 3D skeleton key points, without requiring specific adjustments for different activity types or

the structure of 3D data. Additionally, the paper introduces a data augmentation strategy to counter overfitting and demonstrates the model's real-time processing capabilities. The performance methodology involves evaluating classification accuracy on benchmark datasets, analyzing the impact of CNN and LSTM fusion, assessing the benefits of data augmentation on small datasets, and measuring real-time processing efficiency on embedded platforms [21].

14. The paper by Agrawal and Gupta [22] from the Army Institute of Technology introduces a method for real-time hand gesture recognition in HCI using computer vision techniques. The proposed system, utilizing the Senz3D camera, analyzes 3D data in real time and employs classification rules to recognize hand gestures without the need for training data. The methodology involves advanced image processing and computer vision techniques optimized for efficiency, enabling precise hand gesture recognition for enhanced interaction with a personal computer. The performance methodology emphasizes real-time efficiency achieved through depth-based thresholding, color filtering, and mean-shift tracking for hand detection and tracking.
15. Carcow University of Technology presents a real-time hand gesture recognition system aiming to establish a user-independent interface with high recognition performance. The paper utilizes transfer learning and fine-tuning of pre-trained CNN models like AlexNet and VGG-16 for gesture recognition, addressing challenges in training deep CNNs from scratch due to limited labeled data [23]. The proposed method employs a score-level fusion technique to employ transfer learning by fine-tuning pre-trained CNN models on the target dataset. Score-level fusion enhances recognition accuracy by combining the strengths of two fine-tuned CNNs. Performance evaluation includes preprocessing, fine-tuning, normalization, fusion, and evaluation.

CONCLUSION

This technology of recognizing gestures to perform the tasks is growing very rapidly and this project is going to be helpful in the field of the technology which can perform tasks without any touch or physical contact.

Today when the world is growing too fast with advancement of technology, we are moving toward interfaces which do not require physical contact to handle and perform various tasks. Gestures serve as direct commands for these technologies and in future all the sectors like education, entertainment, automobile, etc. will adapt this technology to grow in their sectors.

Future Enhancements

In future the big companies are going to adapt these technologies into their techs. This will change the whole definition of using the electronic devices like laptops, television, mobile phones, etc. The continuous growth of this technology will create a gadgets sector which does not require and physical touch or support. Gadgets will work on the commands given by one's movement, and by this we can perform more with less effort.

REFERENCES

1. Nagalapuram GD, Roopashree S, Varshashree D, Dheeraj D, Nazareth DJ. Controlling media player with hand gestures using convolutional neural network. In: 2021 IEEE Mysore Sub Section International Conference (MysuruCon), Hassan, India, October 24–25, 2021. pp. 79–86.
2. Islam S, Matin A, Kibria HB. Hand gesture recognition based human computer interaction to control multiple applications. In: Vasant P, Zelinka I, Weber GW, editors. Intelligent Computing & Optimization. ICO 2021. Lecture Notes in Networks and Systems, volume 371. Cham, Switzerland: Springer; 2022. pp. 397–406.
3. Sen A, Mishra TK, Dash R. Deep learning-based hand gesture recognition system and design of a human machine interface. Neural Process Lett. 2023; 55: 12569–12596.
4. Paliwal M, Sharma G, Nath D, Rathore A, Mishra H, Mondal S. A dynamic hand gesture recognition system for controlling VLC media player. In: 2013 International Conference on Advances in Technology and Engineering (ICATE), Mumbai, India, January 23–25, 2013. pp. 1–4.

5. Liu X, Chen T. Video-based face recognition using adaptive hidden Markov models. In: 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Madison, WI, USA, June 18–20, 2003. pp. 1–1.
6. Nagalapuram GD, Roopashree S, Varshashree D, Dheeraj D, Nazareth DJ. Controlling media player with hand gestures using convolutional neural network. IEEE Mysore Sub Section International Conference (MysuruCon). 2021. pp. 79–86. doi: 10.1109/MysuruCon52639.2021.9641567.
7. Chaman S, Jani J, Fernandes H, Dhuka R, Mehta D. Real-time gesture to automotive control. In: 2018 International Conference on Current Trends towards Converging Technologies (ICCTCT), Coimbatore, India, March 1–3, 2018. pp. 1–6.
8. Jalab H, Omer HK. Human-computer interface using hand gesture recognition based on the neural network. In: 2015 5th National Symposium on Information Technology: Towards New Smart World (NSITNSW), Riyadh, Saudi Arabia, February 17–19, 2015. pp. 1–6.
9. Niranjani V, Keerthana R, Mohana Priya B, Nekalya K, Padmanabhan AK. System application control based on hand gesture using deep learning. In: 2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India, March 19–20, 2021. pp. 1644–1649.
10. Selvarathi C, Indu P, Kavyadarshini B, Logesh Kumar S, Mohamed Yasher R. Human computer interaction using hand gesture recognition. Int J Adv Trends Computer Sci Eng. 2020; 9 (2): 1600–1603. doi: 10.30534/ijatcse/2020/106922020.
11. Abhishek B, Krishi K, Meghana M, Daaniyaal M, Anupama HS. Hand gesture recognition using machine learning algorithms. Computer Sci Inform Technol. 2020; 1 (3): 116–120.
12. Veronica PG, Mokkapati RK, Jagupilla LP, Santhosh C. Static hand gesture recognition using novel convolutional neural network and support vector machine. Int J Online Biomed Eng. 2023; 19 (9): 131–141.
13. Vásconez JP, Barona López LI, Valdivieso Caraguay ÁL, Benalcázar ME. Hand gesture recognition using EMG-IMU signals and deep Q-networks. Sensors. 2022; 22 (24): 9613.
14. Shetty S, Joseph J, Thomas JS, Umesh MK, Vaishnav E. Hand gesture recognition system to control keyboard functions. Int J Creative Res Thoughts. 2023; 11 (5): b73–b78.
15. Shinde S, Mushrif S, Pardeshi A, Jagtap D, Vandana Rupnar P. Gesture based media player controller. Int J Res Publ Rev. 2022; 3 (5): 2289–2294.
16. Toro-Ossaba A, Jaramillo-Tigreros J, Tejada JC, Peña A, López-González A, Castanho RA. LSTM recurrent neural network for hand gesture recognition using EMG signals. Appl Sci. 2022; 12 (19): 9700.
17. Haria A, Subramanian A, Asokkumar N, Poddar S, Nayak JS. Hand gesture recognition for human computer interaction. Procedia Computer Sci. 2017; 115: 367–374.
18. Kim AR, Rhee SY. Intention estimation of a walker around pedestrian lights by using fuzzy rules. In: 2012 International Conference on Fuzzy Theory and Its Applications (iFUZZY2012), Taichung, Taiwan, November 16–18, 2012. pp. 233–237.
19. Lai K, Yanushkevich SN, Shmerko V. Reliability of decision support in cross-spectral biometric-enabled systems. In: 2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Toronto, Ontario, Canada, October 11–14, 2020. pp. 3401–3406.
20. Hua Y, Alahari K, Schmid C. Online object tracking with proposal selection. In: Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, December 7–13, 2015. pp. 3092–3100.
21. Nunez JC, Cabido R, Pantrigo JJ, Montemayor AS, Velez JF. Convolutional neural networks and long short-term memory for skeleton-based human activity and hand gesture recognition. Pattern Recogn. 2018; 76: 80–94.
22. Agrawal R, Gupta N. Real time hand gesture recognition for human computer interaction. In: 2016 IEEE 6th International Conference on Advanced Computing (IACC), Bhimavaram, India, February 27–28, 2016. pp. 470–475.
23. Sahoo JP, Prakash AJ, Pławiak P, Samantray S. Real-time hand gesture recognition using fine-tuned convolutional neural network. Sensors. 2022;22:706. doi: 10.3390/s22030706. PubMed: 35161453.

