Email: editor@ijerst.com or editor.ijerst@gmail.com

ISSN2319-5991www.ijerst.com

Vol.21,Issue1,2025

HANDGESTUREBASEDHOMEAUTOMATION USING RASPBERRY PI

[1]MR.P.V.M.VIJAYBHASKAR,[2]N.V.KEERTHI,[3]S.INDIRA,[4]S. PUJITHA, [5]SD.NASIMA

[1](AssistantProfessor),ECE, RISE Krishna Sai Prakasam Group of Institutions::Ongole

[2345]B.TECH,scholar,ECE, RISE Krishna Sai Prakasam Group of Institutions::Ongole

ABSTRACT

Hand gesture-based home automation is a technology that allows users to control various electronic devices within theirhomes through simple hand movements. By utilizing the capabilities of a Raspberry Pi and sensors, this system detects and interprets hand gestures, enabling users to operate devices such as lights, fans, air conditioners, and security systems without the need for physical contact or voice commands. This automation system is designed to enhance convenience, accessibility, and energy efficiency in a smart home environment. Through the integration of computer vision, machine learning algorithms, and real-time processing, the Raspberry Pi system can identify specific gestures and trigger corresponding actions to control home appliances.Thisapproachoffersan innovative, intuitive, and user-friendly methodof interacting with smarthome systems.

KEYWORDS: Hand Gesture Recognition,HomeAutomation**,**Raspberry Pi**,** Computer Vision, Machine Learning, Smart Home Technology, Gesture Control, IoT (Internet of Things)**,** EmbeddedSystems**,** Smart Devices

1.INTRODUCTION

Sign language, as one of the most widely used communication means for hearing-impaired people, is expressed by variationsofhand-shapes,bodymovement, and even facial expression. Since it is difficult to collaboratively exploit the information from hand-shapes and body movement trajectory, sign language recognition is still a very challenging task. This paper proposes an effective recognition model to translate sign language into text or speech in order tohelp the hearing impaired communicate with normal people through sign language.

Technically speaking, the main challenge of sign language recognition lies in developing descriptors to express hand-shapes and motion trajectory. In particular, hand-shape description involves tracking hand regions in video stream, segmenting hand-shapeimagesfromcomplex

International Journal of Engineering Research and Science & Technology
ISSN2319-5991www.ijerst.com
Vol.21,Issue1,2025

background in each frame and gestures recognition problems. Motion trajectory is also related to tracking of the key points and curve matching. Although lots of research works have been conducted on these two issues for now, it is still hard to obtain satisfying result for SLR due to the variation and occlusion of hands and body joints. Besides, it is a nontrivial issue to integrate the hand-shape features and trajectory features together. To address these difficulties, we develop CNNs to naturally integrate hand-shapes, trajectory of action and facial expression. Instead of using commonly used color images asinput to networks like [1, 2], we take color images, depth images and body skeleton images simultaneously as input which are all provided by Microsoft Kinect .

Kinectisamotionsensorwhichcan provide color stream and depth stream. With the public Windows SDK, the body joint locations can be obtained in real-time as shown in Fig.1. Therefore, we choose Kinect as capture device to record sign words dataset. The change of color and depth in pixel level are useful information to discriminate different sign actions. And the variation of body joints in time dimension can depict the trajectory of sign actions. Using multiple types of visual sources as input leads CNNs paying attention to the change not only in color, but also in depth and trajectory. It is worth mentioning that we can avoid the difficulty of tracking hands, segmenting hands from background and designing descriptors for handsbecauseCNNshavethecapabilityto learn features automatically from raw data without any prior knowledge [3].

CNNs have been applied in video stream classification recently years. A potential concern of CNNs is time consuming. It costs several weeks or months to train a CNNs with million-scale in million videos. Fortunately, it is still possible to achieve real-time efficiency, with the help of CUDA for parallel processing. We propose to apply CNNs to extract spatial and temporal features from video stream for Sign Language Recognition (SLR). Existing methods for SLR use hand-crafted features to describe sign language motion and build classification model based on these features. In contrast, CNNs can capture motion information from raw video data automatically, avoiding designing features. We develop a CNNs taking multiple types of data as input.This architecture integrates color, depth and trajectory information by performing convolutionandsubsamplingonadjacent video frames. Experimental results demonstrate that 3D CNNs can significantly outperform Gaussianmixture model with Hidden Markov model (GMM-HMM) baselines on some sign words recorded by ourselves.

2. EXISTINGSYSTEM

Hand gesture recognition using convolution neuralnetworks(CNNs)isawell-researched area in computer vision and pattern recognition. The main components of such systemstypicallyincludedataacquisition,

ISSN2319-5991www.ijerst.com
Vol.21,Issue1,2025

pre-processing, feature extraction, classification, and post-processing.

1. DeepLearning-BasedSystems

Advantages:

- **Accuracy and Precision:** CNNs providehighaccuracyinrecognizing complexpatterns,makingthemideal for hand gesture recognition.
- **Scalability:**CNNscanbescaledup with more data to improve performance.
- **Automation:**Automatedfeature extraction reduces the need for manual intervention in the recognition process.

Disadvantages:

- **DataDependency:**Requiresalarge amount of labeled data for training, which can be difficult to obtain.
- **Computationally Intensive:** TrainingCNNsrequiressignificant computational resources and time.
- **Complexity:**Designingandtuning deep learning models can be complex and requires expertise.

2. TraditionalMachineLearning-Based Systems

Advantages:

- **LessDataRequired:**Comparedto CNNs,traditionalmachinelearning methods often require less data.
- **Faster Training:** Training time is generallyshortercomparedtodeep learning models.
- **Simplicity:**Easiertoimplementand understand.

Disadvantages:

- **LowerAccuracy:**Oftenlessaccuratethan CNNs, especially for complex gesture recognition tasks.
- **Feature Engineering:** Requires manual featureextractionandengineering,which canbetime-consumingandlesseffective.
- **Limited Scalability:** Performance improvementwithadditionaldataislimited compared to deep learning methods.

3. Hybrid Systems (Combination of TraditionalandDeepLearningTechniques)

Advantages:

- **Improved Performance:** Combines the strengths of both traditional and deep learningmethodstoimproveaccuracyand efficiency.
- **Flexibility:**Canadapttovarioustypesof data and recognition tasks.

Disadvantages:

- **Complexity:**Morecomplextodesignand implement than pure traditional or deep learning systems.
- **ResourceIntensive:**Requiresresourcesfor both traditional and deep learning components.

4. Real-TimeGestureRecognitionSystems

Advantages:

- **ImmediateFeedback:**Providesreal-time recognition and feedback, useful for interactive applications.
- **UserExperience:**Enhancesuserexperience in applications like virtual reality and human-computer interaction.

Disadvantages:

- **Latency:** Achieving real-time performance canbechallengingduetoprocessingdelays.
- **ResourceDemands:**Highcomputational requirements for real-time processing.

ISSN2319-5991www.ijerst.com

Vol.21,Issue1,2025

3. PROPOSEDSYSTEM

Recent advancement in deep learning have lead to CNNs. CNNs are deep neural networks designed for raw images and videos. This is usedtocommunicatethedefmutedpeople. Itis implemented to recognize the hand gestures easily from the muted people.

End-to-EndDeepLearningSystem

Components:

- **Data Collection and Augmentation:** Use extensive datasetswithdiversehandgestures. Augment data with rotations, flips, and lighting variations.
- **Preprocessing:** Normalize and resizeimagestoastandardsize.
- **CNNArchitecture:**
- InputLayer:Standardsize(e.g., 128x128x3 for RGB images).
- ConvolutionLayers:Multiplelayers with varying filter sizes.
- PoolingLayers:Maxpoolinglayers to reduce spatial dimensions.
- FullyConnectedLayers:Dense layers for classification.
- OutputLayer:Softmaxlayerfor multi-class classification.
- **Training and Validation:** Split data intotraining,validation,andtestsets. Use techniques like early stopping, dropout, and regularization.
- **Deployment:**Exportthemodeland integrate it into an application for real-time gesture recognition.

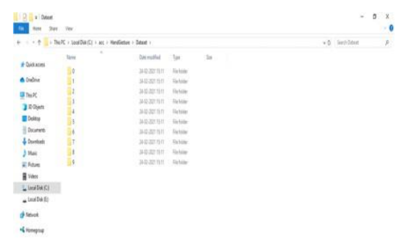
Advantages:

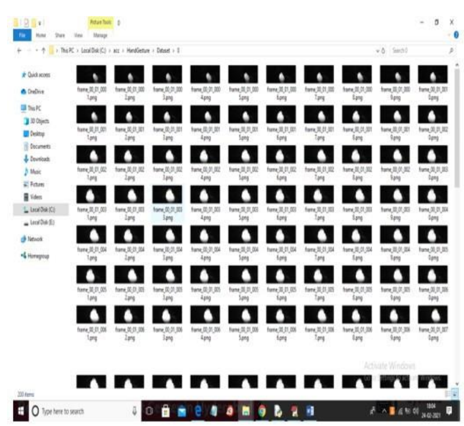
- **HighAccuracy:**End-to-endlearning optimizes the entire pipeline for gesture recognition.
- **Automated Feature Extraction:** Reducestheneedformanualfeature engineering.
- **Scalability:** Can be improved with moredataandcomputationalpower.

Disadvantages:

- **DataIntensive:**Requiresalargeand diverse dataset for optimal performance.
- **Computationally Expensive:** Training and inference require significantcomputationalresources.

4.OUTPUTSCREENS

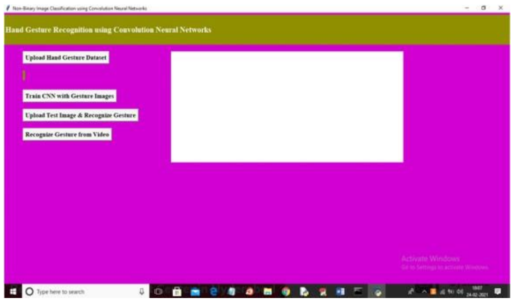


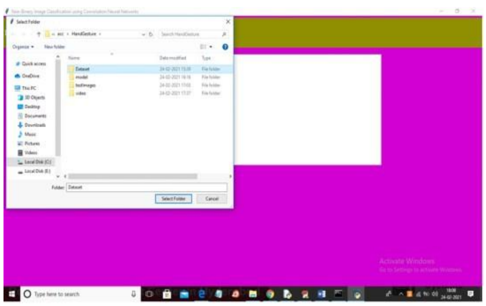


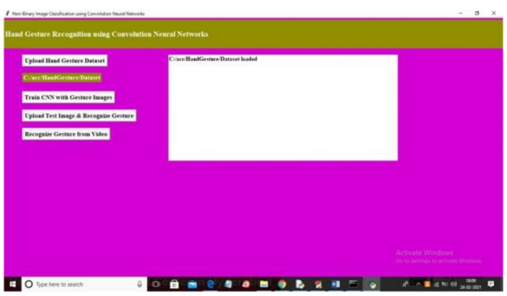
International Journal of Engineering Research and Science & Technology

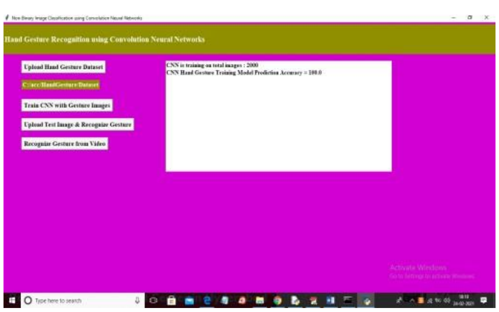
ISSN2319-5991www.ijerst.com

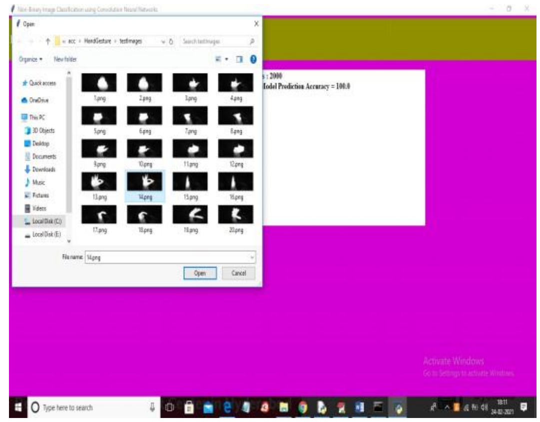
Vol.21,Issue1,2025

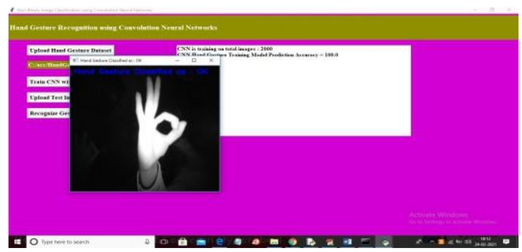


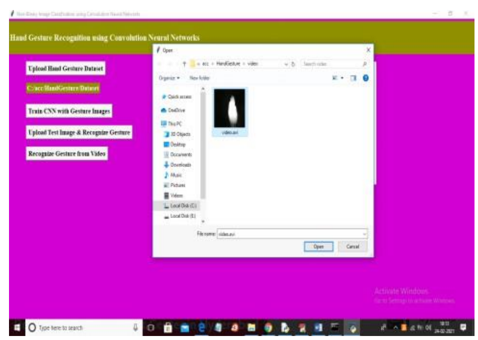


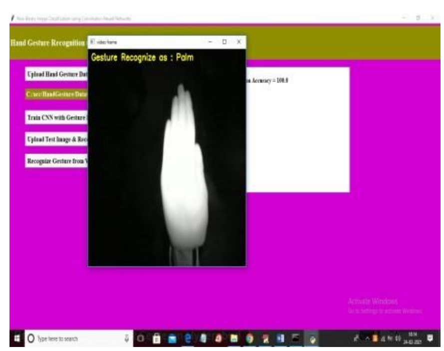








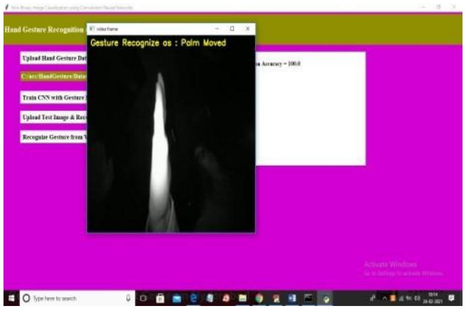


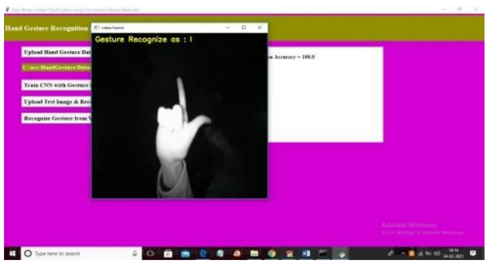


International Journal of Engineering Research and Science & Technology

ISSN2319-5991www.ijerst.com

Vol.21,Issue1,2025





5. CONCLUSION

We developed a CNN model for handgesture recognition. Our model learns and extracts both spatial and temporal featuresby performing 3D convolutions. The developed deep architecture extractsmultiple types of information from adjacent input frames and then performs convolution and sub sampling separately. The final feature representation combines information from all channels. We use multilayer perceptron classifier to classify these feature representations.Forcomparison,weevaluate both CNN and GMM-HMM on the same dataset. The experimental resultsdemonstrate the effectiveness of theproposed method.

6. REFERENCES

[1] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton, “Imagenet classification with deep convolution neural networks,”inAdvancesinneural information processing systems, 2012, pp. 1097–1105.

[2] Andrej Karpathy, George Toderici, Sanketh Shetty, Thomas Leung, Rahul Sukthankar, and Li Fei-Fei, “Large-scale video classification with convolutional neural networks,” in CVPR, 2014. [3] Yann LeCun, Leon Bottou, Yoshua Bengio, and Patrick ´ Haffner, “Gradient-based learning applied to document recognition,” Proceedingsofthe IEEE, vol.86,no.11,pp. 2278–2324, 1998.

[4] Hueihan Jhuang, Thomas Serre, Lior Wolf, and Tomaso Poggio, “A biologically inspired system for action recognition,” in Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on. Ieee,2007, pp. 1–8. [5] Shuiwang Ji, Wei Xu, Ming Yang, and Kai Yu, “3D convolutional neural networks for human action recognition,”IEEETPAMI,vol.35,no.1, pp. 221–231, 2013.

[6] Kirsti Grobel and Marcell Assan, “Isolated sign language recognition using hidden markov models,” in Systems, Man, and Cybernetics, 1997. Computational Cybernetics and Simulation., 1997 IEEE International Conference on. IEEE, 1997, vol. 1, pp. 162–167.

[7] Thad Starner, Joshua Weaver, and Alex Pentland, “Realtime american sign language recognition using desk and wearable computer based video,” IEEE TPAMI, vol. 20, no. 12, pp. 1371–1375, 1998.

[8] Christian Vogler and Dimitris Metaxas, “Parallel hidden markov models for american sign language recognition,” in Computer Vision, 1999. The ProceedingsoftheSeventhIEEEInternational

ISSN2319-5991www.ijerst.com

Vol.21,Issue1,2025

Conferenceon.IEEE,1999,vol.1,pp.116–122.

[9] Kouichi MurakamiandHitomiTaguchi, "Gesture recognition using recurrent neural networks," in Proceedings of the SIGCHI conference on Human factors in computing systems. ACM, 1991, pp. 237–242.

[10] Chung-Lin Huang and Wen-Yi Huang, "Sign language recognition using model-based tracking and a 3D hopfield neural network," Machine vision and applications, vol. 10, no. 5-6, pp. 292–307, 1998.