IJITCE

ISSN 2347- 3657

International Journal of Information Technology & Computer Engineering

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

International Journal of Information Technology & Computer Engineering

ISSN 2347–3657

Volume 12, Issue 3, 2024

HAND GESTURES RECOGNITION USING CONVOLUTION NEURAL NETWORKS

[1] K SUPARNA, [2]R. DEEPIKA

[1](Assistant Professor), MCA, **DANTULURI NARAYANA RAJU COLLEGE(A) PG COURSES, BHIMAVARAM ANDHRA PRADESH**

[2]MCA, scholar, **DANTULURI NARAYANA RAJU COLLEGE(A) PG COURSES, BHIMAVARAM ANDHRA PRADESH**

ABSTRACT

Hand Gesture Recognition (HGR) targets on interpreting the sign language into text or speech, so as to facilitate the communication between n deaf-mute people and ordinary people. This task has broad social impact, but is still very challenging due to the complexity and large variations in hand actions. Existing methods for HGR use hand-crafted features to describe sign language motion and build classification models based on those features. However, it is difficult to design reliable features to adapt to the large variations of hand gestures. To this problem, we propose a novel convolution neural network (CNN) which extracts discriminative spatial-temporal features from raw video stream automatically without any prior knowledge, avoiding designing features. To boost the performance, multi-channels of video streams, including color information, depth clue ,and body joint positions, are used as input to the CNN in order to integrate color, depth and trajectory information. We validate the proposed model on a real dataset collected with Microsoft Kinect and demonstrate its effectiveness over the traditional approaches based on hand-crafted features

1.INTRODUCTION

Sign language, as one of the most widely used communication means for hearing-impaired people, is expressed by variations of hand-shapes, body movement, and even facial expression. Since it is difficult to collaboratively exploit the information from hand-shapes and body

ISSN 2347–3657

Volume 12, Issue 3, 2024

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 to help 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-shape images from complex 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 as input 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 .

Kinect is a motion sensor which can 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 hands because CNNs have the capability to 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

ISSN 2347–3657

Volume 12, Issue 3, 2024

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 convolution and sub sampling on adjacent video frames. Experimental results demonstrate that 3D CNNs can significantly outperform Gaussian mixture model with Hidden Markov model (GMM-HMM) baselines on some sign words recorded by ourselves.

2. EXISTING SYSTEM

Hand gesture recognition using convolution neural networks (CNNs) is a well-researched area in computer vision and pattern recognition. The main components of such systems typically include data acquisition, pre-processing, feature extraction, classification, and post-processing.

1. Deep Learning-Based Systems

Advantages:

- **Accuracy and Precision:** CNNs provide high accuracy in recognizing complex patterns, making them ideal for hand gesture recognition.
- **Scalability:** CNNs can be scaled up with more data to improve performance.
- **Automation:** Automated feature extraction reduces the need for manual intervention in the recognition process.

Disadvantages:

- **Data Dependency:** Requires a large amount of labeled data for training, which can be difficult to obtain.
- **Computationally Intensive:** Training CNNs requires significant computational resources and time.
- **Complexity:** Designing and tuning deep learning models can be complex and requires expertise.

2.Traditional Machine Learning-Based Systems

Advantages:

ISSN 2347–3657

Volume 12, Issue 3, 2024

- **Less Data Required:** Compared to CNNs, traditional machine learning methods often require less data.
- **Faster Training:** Training time is generally shorter compared to deep learning models.
- **Simplicity:** Easier to implement and understand.

Disadvantages:

- **Lower Accuracy:** Often less accurate than CNNs, especially for complex gesture recognition tasks.
- **Feature Engineering:** Requires manual feature extraction and engineering, which can be time-consuming and less effective.
- **Limited Scalability:** Performance improvement with additional data is limited compared to deep learning methods.

3. Hybrid Systems (Combination of Traditional and Deep Learning Techniques)

Advantages:

- **Improved Performance:** Combines the strengths of both traditional and deep learning methods to improve accuracy and efficiency.
- **Flexibility:** Can adapt to various types of data and recognition tasks.

Disadvantages:

- **Complexity:** More complex to design and implement than pure traditional or deep learning systems.
- **Resource Intensive:** Requires resources for both traditional and deep learning components.

4. Real-Time Gesture Recognition Systems

Advantages:

- **Immediate Feedback:** Provides real-time recognition and feedback, useful for interactive applications.
- **User Experience:** Enhances user experience in applications like virtual reality and human-computer interaction.

Disadvantages:

- **Latency:** Achieving real-time performance can be challenging due to processing delays.
- **Resource Demands:** High computational requirements for real-time processing.

3. PROPOSED SYSTEM

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

End-to-End Deep Learning System

Components:

ISSN 2347–3657

Volume 12, Issue 3, 2024

- **Data Collection and Augmentation:** Use extensive datasets with diverse hand gestures. Augment data with rotations, flips, and lighting variations.
- **Preprocessing:** Normalize and resize images to a standard size.
- **CNN Architecture:**
- Input Layer: Standard size (e.g., 128x128x3 for RGB images).
- Convolution Layers: Multiple layers with varying filter sizes.
- Pooling Layers: Max pooling layers to reduce spatial dimensions.
- Fully Connected Layers: Dense layers for classification.
- Output Layer: Softmax layer for multi-class classification.
- **Training and Validation:** Split data into training, validation, and test sets. Use techniques like early stopping, dropout, and regularization.
- **Deployment:** Export the model and integrate it into an application for real-time gesture recognition.

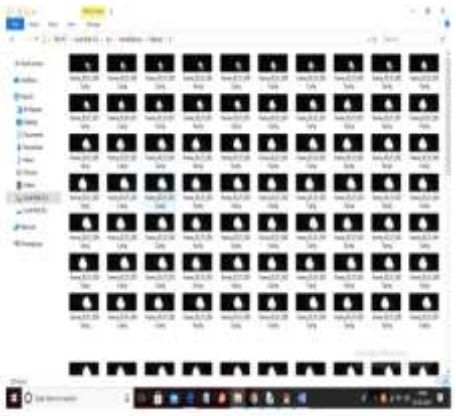
Advantages:

- **High Accuracy:** End-to-end learning optimizes the entire pipeline for gesture recognition.
- **Automated Feature Extraction:** Reduces the need for manual feature engineering.
- **Scalability:** Can be improved with more data and computational power.

Disadvantages:

- **Data Intensive:** Requires a large and diverse dataset for optimal performance.
- **Computationally Expensive:** Training and inference require significant computational resources.

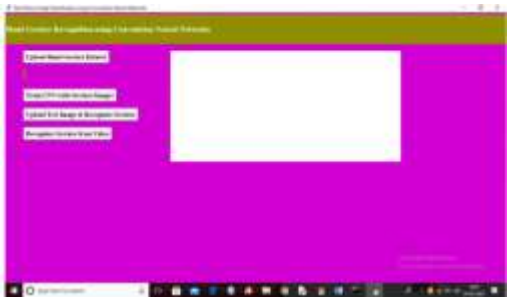
4.OUTPUTSCREENS

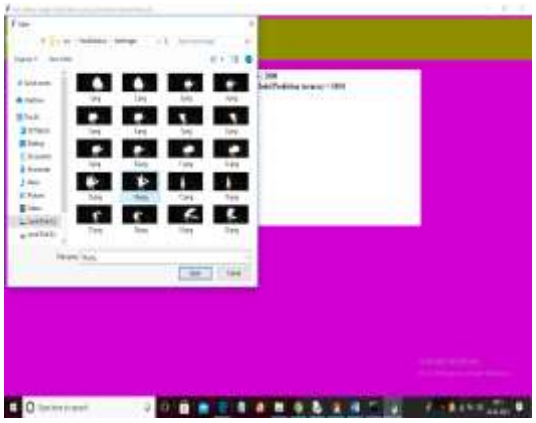


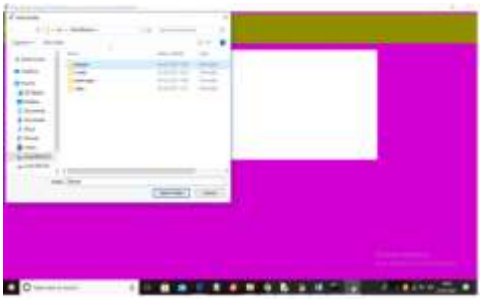
International Journal of Information Technology & Computer Engineering

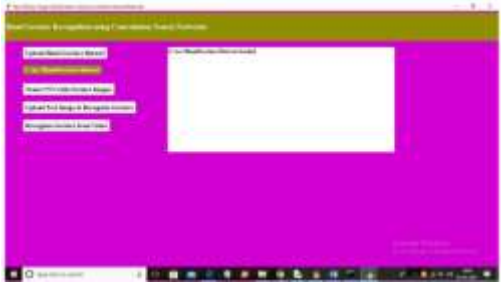
ISSN 2347–3657

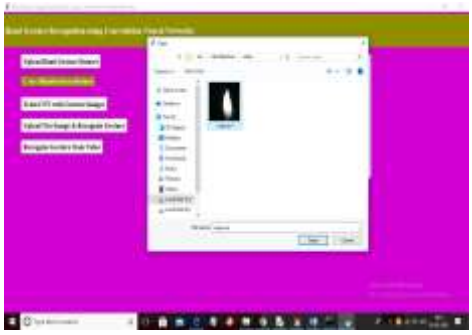
Volume 12, Issue 3, 2024

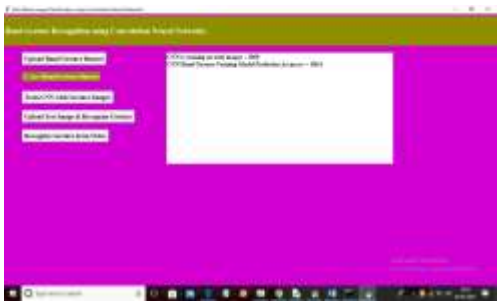






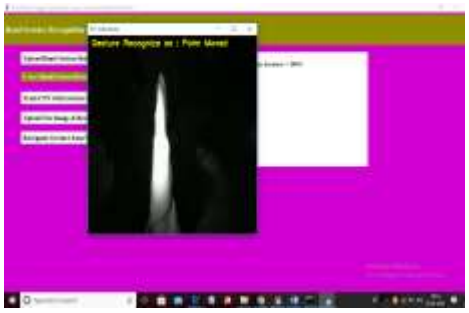


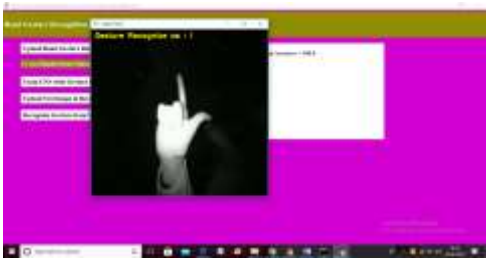




ISSN 2347–3657

Volume 12, Issue 3, 2024





5.CONCLUSION

We developed a CNN model for hand gesture recognition. Our model learns and extracts both spatial and temporal features by performing 3D convolutions. The developed deep architecture extracts multiple 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. For comparison, we evaluate both CNN and GMM-HMM on the same dataset. The experimental results demonstrate the effectiveness of the proposed method.

6. REFERENCES

[1] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton, "Imagenet classification with deep convolution neural networks," in Advances in neural 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," Proceedings of the 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,

ISSN 2347–3657

Volume 12, Issue 3, 2024

Ming Yang, and Kai Yu, "3D convolutional neural networks for human action recognition," IEEE TPAMI, 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 Compuster Vision, 1999. The Proceedings of the Seventh IEEE International Conference on. IEEE, 1999, vol. 1, pp. 116–122.

[9] Kouichi Murakami and Hitomi Taguchi, "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.