Real-Time Multimodal Hand Gesture-to-Text Conversion Using Advanced Machine Learning Architectures

A Research Proposal

Karishma M. Patel

October 9, 2025

Contents

1 Abstract

This research develops a software system that translates hand gestures into text captions in real-time, enhancing communication accessibility for individuals using sign language or gesture-based communication. The study bridges communication gaps for the deaf and hard-of-hearing communities through advanced machine learning and computer vision techniques. Objectives include designing a robust gesture recognition model, implementing a user-friendly interface, and evaluating accuracy and usability. Data will be collected via video recordings of hand gestures from diverse participants, analyzed using convolutional neural networks (CNNs). The expected outcome is a functional prototype with high translation accuracy, advancing inclusive communication technologies. This work empowers users with alternative communication methods and informs future assistive technology development.

2 Introduction

Effective communication is essential for social interaction, yet individuals relying on hand gestures, such as those in the deaf and hard-of-hearing communities, face barriers in text-based environments. Current gesture recognition systems are constrained by limitations in accuracy, real-time performance, and accessibility. This research develops a software solution that employs machine learning to translate hand gestures into text captions in real-time, overcoming these challenges.

The study addresses the need for inclusive communication tools that empower gesture-based communicators. The research objectives are: (1) to achieve high accuracy in translating hand gestures into text using machine learning; (2) to design an intuitive gesture-to-text interface; (3) to ensure the software accommodates diverse gesture vocabularies. This research enhances accessibility, promotes inclusion, and contributes significantly to the field of assistive technologies.

3 Literature Review

The literature on gesture recognition demonstrates significant advancements in computer vision and machine learning. Studies by **?** show convolutional neural networks (CNNs) achieving up to 95% accuracy in recognizing static hand gestures in controlled settings. Real-time translation, however, faces challenges due to variations in lighting, hand positioning, and gesture complexity (**?**). Wearable devices, as explored by **?**, provide high precision but are less accessible due to cost and hardware requirements.

Current literature lacks comprehensive datasets for diverse sign languages and user-friendly interfaces for non-technical users. There are also gaps in addressing cultural variations in gestures and ensuring real-time performance across devices. This research fills these gaps by developing a software-based solution that prioritizes accessibility, scalability, and adaptability to diverse gesture sets, offering a novel contribution to real-time gesture-to-text translation.

4 Research Design and Methods

This study employs a mixed-methods research paradigm, integrating quantitative performance metrics with qualitative user feedback. The methodology includes the following:

4.1 Data Collection

Video recordings of hand gestures will be collected from 50 participants, including native sign language users and non-signers, ensuring diversity. Gestures will be recorded under varied lighting and background conditions to enhance model robustness. A standardized gesture set, based on American Sign Language (ASL) letters and common phrases, will be supplemented by custom gestures to ensure adaptability.

4.2 Data Analysis

A convolutional neural network (CNN) will be trained on the dataset using TensorFlow and OpenCV for image processing. The model will be evaluated for accuracy, precision, and recall, targeting at least 90% accuracy in real-time translation. Statistical analysis will validate performance across diverse conditions, while qualitative feedback from user testing will refine the interface.

4.3 Ethical Considerations

The study ensures ethical integrity by obtaining informed consent from participants, safeguarding data privacy, and addressing potential biases in gesture recognition, such as variations in hand size or skin tone. Limitations, including model performance in low-light conditions, will be addressed through data augmentation and robust preprocessing techniques.

5 Timeline

The research will span 12 months, structured as follows:

- **Months 1-2**: Conduct literature review and design dataset.
- **Months 3-5**: Collect and preprocess data.
- **Months 6-8**: Develop and train the model.
- **Months 9-10**: Test prototype and collect user feedback.
- **Months 11-12**: Analyze data, write report, and disseminate findings.

6 Budget

The estimated budget is $2,000, covering:

- **Travel**: Participant recruitment and testing sessions ($1,500).
- **Miscellaneous**: Data storage and dissemination costs ($500).

These estimates ensure responsible use of funds while supporting project feasibility.

7 Expected Outcomes and Implications

The research will produce a functional software prototype that translates hand gestures into text with at least 90% accuracy in real-time. The software will feature an intuitive interface, accessible on standard devices such as laptops and smartphones. This work advances assistive technologies by delivering an affordable, scalable solution for gesture-based communication. It will benefit stakeholders, including the deaf community, educators, and policymakers, by fostering inclusive communication practices. Future implications include integration with augmented reality and support for broader sign language vocabularies.

8 References

A Data Collection Questionnaire

A questionnaire will document participants' gesture preferences and demographic information to ensure diverse data collection.

B Consent Form

A consent form detailing the study's purpose, data usage, and participant rights will be provided to all participants.