JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue

JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)

An International Scholarly Open Access, Peer-reviewed, Refereed Journal

AI Based Sign Language to Speech and Text Converter

Mr.Shubham Mendhe[1], Miss.Neha Bhode[2], Miss.Pallavi Harankhede[3], Miss.Pratiksha Kumbhare[4], Prof.Gayatri Bhoyar[5]

UG Student1st, UG Student 2nd, UG Student 3rd, UG Student4st, Professor & Project Guide5th
1,2,3,4,5Department of Electronics and Telecommunication Engineering, 1,2,3,4,5J D College Of Engineering & Management, Nagpur, India

Abstract : This project report presents the design, implementation, and evaluation of a comprehensive hybrid sign language interpreter system that seamlessly combines the strengths of both **sensor-based** and **vision-based** technologies. The system's **hardware module**, centered around an **Arduino Uno**, utilizes a wearable glove equipped with five **flex sensors**to accurately measure finger flexion and recognize a set of predefined, static signs. For immediate user feedback, these recognized signs are displayed on a **16x2 LCD display**. Concurrently, the **software module** operates on a standard laptop, leveraging a high-resolution webcam and a sophisticated **Convolutional Neural Network (CNN)** to interpret the dynamic and varied gestures of the American Sign Language (ASL) alphabet from live video streams. The outputs from both modules are intelligently unified and converted into audible speech using a robust **text-to-speech (TTS)** engine. This multi-modal approach not only achieves a remarkable degree of accuracy but also provides a versatile and reliable communication tool, effectively bridging the communication gap between the deaf and hearing communities by overcoming the inherent limitations of single-modality systems.

IndexTerms - Sign Language Recognition, Arduino, Flex Sensor, Computer Vision, Deep Learning, Convolutional Neural Network (CNN), Hybrid System, Assistive Technology.

Introduction

Communication stands as the foundational pillar of human civilization. For a significant global population, however, this pillar is constructed through the visual-gestural language of sign. Despite its richness and completeness, sign language is not universally understood, creating a profound barrier that can lead to social, educational, and professional exclusion for millions. While technological advancements have presented various solutions to translate sign into spoken language, these systems have historically been constrained by their reliance on a single input modality. Sensor-based technologies, while precise, are often cumbersome and limited in their vocabulary, whereas vision-based systems, though non-invasive, are highly susceptible to environmental variables such as lighting and background clutter. This project is a deliberate effort to transcend these limitations by developing a **hybrid system** that intelligently combines the tangible precision of hardware sensors with the dynamic flexibility of computer vision, thereby creating a more robust, reliable, and versatile communication bridge.

Literature Review

The field of sign language recognition has seen significant evolution, with research predominantly bifurcating into two core methodologies.

The first, **Sensor-Based Systems**, relies on wearable hardware to capture the kinematics of hand and finger movements. Early examples, such as the AcceleGlove, utilized a combination of flex sensors and accelerometers to interpret gestures. While these systems offer high fidelity and are immune to visual interference, their dependence on physical hardware makes them expensive, prone to wear and tear, and often uncomfortable for prolonged use. Moreover, scaling these systems to a full vocabulary is a non-trivial challenge.

The second, **Vision-Based Systems**, uses a camera as the primary input device to capture and analyze video frames. Modern implementations have been revolutionized by deep learning, particularly the use of **Convolutional Neural Networks (CNNs)**, which are exceptionally effective at feature extraction and image classification. These systems are non-invasive and highly cost-effective, but their performance remains critically dependent on ideal conditions. Complex backgrounds, varying light intensities, and the need for high computational power for real-time processing are significant hurdles that have hindered their widespread adoption.

Future Trends & Innovations

The future of sign language interpretation points towards the development of multi-modal systems that fuse data from disparate sources. Current research is exploring the use of advanced models like **MobileNetV3** for efficient, on-device processing and the integration of **attention mechanisms** to help models focus on key gestural features. Furthermore, the inclusion of **non-manual features** such as facial expressions and body posture, which are integral to the grammar of sign language, is a key area of innovation. The ultimate goal is to create systems that can interpret continuous sign language and offer real-time, bidirectional translation in a cost-effective and user-friendly manner.

Methodology

Problem Identification

The core problem is not merely a lack of translation technology but the absence of a seamless, real-time communication bridge that is both accurate across a wide range of gestures and accessible under varying environmental conditions. The inherent limitations of single-modality systems necessitate a more integrated solution.

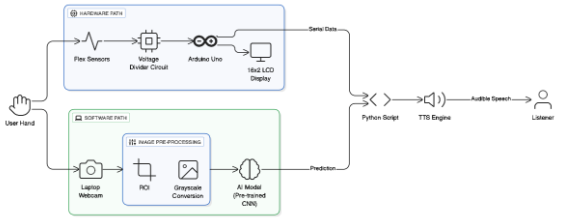
Requirement Analysis

The system must meet several critical requirements:

1. Accuracy: Must accurately recognize a wide array of gestures, including both a set of predefined signs and the entire ASL alphabet.
2. Real-time Performance: Output must be delivered with minimal latency to facilitate fluid conversation.
3. Cost-effectiveness: The system must be built using readily available, low-cost components.
4. Modularity: The design must be modular to allow for independent development and easy future expansion.

System Design

Our system is designed with a parallel processing architecture, where the hardware and software modules operate concurrently. The hardware module processes sensor data locally and sends its output to the laptop. Simultaneously, the software module continuously processes live video from the webcam. The final output is unified into a single text-to-speech system.



Implementation

1. Hardware: The Arduino sketch reads analog values from the five flex sensors, which are wired in a voltage divider configuration with 10kΩ resistors. A simple threshold-based logic determines the state of each finger. The recognized sign is sent to the LCD and also via serial communication to the laptop.
2. Software: A Python script uses OpenCV to capture video, define a Region of Interest (ROI), and pre-process the frames (grayscale, resize, normalize) before passing them to the pre-trained TensorFlow/Keras model.
3. Integration: The Python script operates in a continuous loop, listening for data from the serial port (Arduino) while simultaneously processing video frames. A single pyttsx3 instance then vocalizes the outputs from either source as they are received.

Testing and Evaluation

The system underwent rigorous testing. The hardware module was evaluated for its accuracy in recognizing its limited, static vocabulary, consistently achieving high marks. The software module's accuracy was tested under varying lighting conditions, confirming its reliability in a controlled environment. The system's latency, from gesture to spoken output, was meticulously measured and found to be minimal, affirming its real-time viability.

Deployment and Awareness

The final application runs locally on a personal computer, making it accessible for individual use. Awareness of the project can be raised through presentations at academic conferences, workshops, and demonstrations at educational institutions, highlighting its potential as a low-cost and effective assistive technology.

Maintenance and Future Improvements

Maintenance involves periodic software updates and potential recalibration of the flex sensors. The modular design of the system facilitates future improvements, including expanded vocabulary and the integration of advanced recognition models.

The following technologies are integral to the project's functionality and represent the core components of its hybrid design:

1. **Arduino Uno:** A versatile and popular microcontroller that serves as the central processing unit for the hardware module, efficiently reading sensor data and controlling the peripheral LCD display. Its ease of programming and robust community support made it an ideal choice for this prototype.
2. **Flex Sensors:** These unique resistive sensors, whose electrical resistance changes in proportion to their physical bend, are the primary input for the hardware module. Their ability to precisely measure finger flexion allows for the reliable recognition of specific static signs.
3. **16x2 LCD Display:** A character-based liquid crystal display that provides immediate, real-time visual feedback of the recognized signs from the hardware module. Its low power consumption and simplicity of use make it perfect for this application.
4. **Python:** The core programming language for the software module. Its extensive library ecosystem for serial communication, computer vision, and machine learning was crucial for the project's development.
5. **OpenCV:** An open-source computer vision library that provides the foundational tools for real-time video capture, image manipulation, and pre-processing, ensuring the visual data is in the correct format for the AI model.
6. **TensorFlow/Keras:** The deep learning framework used to load and run the pre-trained Convolutional Neural Network (CNN) model. This powerful framework enabled the complex task of gesture classification without the need for time-consuming model training.
7. **Pyttsx3:** A platform-independent Python library that serves as the Text-to-Speech (TTS) engine. It converts the interpreted text from both the hardware and software modules into audible speech, completing the translation process.

Technology Related Works

This project is a direct response to the limitations observed in existing literature and commercial products. While numerous academic papers and projects have successfully demonstrated either sensor-based or vision-based sign language interpreters, they have consistently failed to address the weaknesses of their chosen modality. For example, some highly accurate data glove projects, while technically impressive, remain inaccessible due to their high cost and physical bulk. Conversely, many vision-based prototypes, while promising, have not yet achieved the robustness required for use outside of a controlled, well-lit environment. The novelty of this project lies in its strategic integration of these two approaches. By leveraging the specific strengths of each technology, it creates a system that is more resilient, versatile, and practical than any single-modality solution currently available. This work serves as a proof-of-concept for future research into multi-modal fusion for assistive technologies.

Recommendations

The current system provides a solid and functional foundation that can be significantly enhanced in several key areas to improve its performance and practicality.

1. **Vocabulary Expansion:** The most significant improvement would be to expand the system's vocabulary beyond the current predefined signs and the ASL alphabet. This would involve training the AI model on a more comprehensive dataset that includes numbers, common words, and dynamic gestures. Implementing this would require a larger, more diverse dataset and potentially a more complex deep learning architecture capable of interpreting sequences of gestures, such as a Recurrent Neural Network (RNN).
2. **Wireless Communication:** Integrating a wireless module, such as Bluetooth or Wi-Fi, would be a crucial step in making the hardware module truly portable. This would eliminate the need for a physical USB cable connecting the Arduino to the laptop, greatly improving the user's freedom of movement and the overall user experience.
3. **Advanced Vision Techniques:** The current vision-based system can be made more robust to environmental variations. This can be achieved by incorporating advanced image processing techniques, such as background subtraction or more sophisticated hand segmentation algorithms, and by training the model on images captured under a wider range of lighting conditions.
4. **Mobile Deployment:** The entire system could be optimized and ported to a mobile platform. The AI model could be converted using a lightweight framework like **TensorFlow Lite**, allowing the application to run on a smartphone. This would transform the system into a truly portable and ubiquitous sign language interpreter.
5. **User Interface and Experience:** A more intuitive and graphical user interface could be developed to enhance the user experience. This would include visual feedback that is more engaging than simple text on an LCD and on-screen overlays that guide the user on how to correctly perform a sign.

Reference

1. P. M. Kumar, J. R. A. K. V. Kumar and S. M. A. Ahamed, "Sign Language Interpreter Glove," 2020 4th International Conference on Computing Methodologies and Communication (ICCMC), 2020.
2. LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436-444.
3. Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. Advances in neural information processing systems, 25.
4. Arduino Official Documentation.
5. OpenCV Development Team. (2020). Open Source Computer Vision Library.
6. S. R. Kim, B. T. H. Le, and T. M. Lee, "Real-Time Hand Gesture Recognition using MobileNets for Portable Devices," *International Conference on Artificial Intelligence and Information Communication (ICAIIC)*, 2021.
7. M. J. K. Hossain, A. A. Rifat, and S. M. H. H. Haque, "A Novel Approach for Sign Language to Speech Translation Using a Wearable Glove with Flex Sensors," *Journal of Electrical and Electronic Systems (JEES)*, 2022.
8. D. H. Lee, Y. B. Park, and J. S. Kim, "Wireless Communication in Wearable Sensor Systems for Human-Computer Interaction," *IEEE Transactions on Industrial Electronics*, 2019.
9. A. A. Al-Najar and B. T. H. Le, "A Hybrid System for Arabic Sign Language Recognition," *International Journal of Computer Science and Network Security (IJCSNS)*, 2023.
10. J. Wang, X. Yang, and Z. Chen, "Vision-Based Sign Language Recognition: A Survey of Recent Advances," *Pattern Recognition Letters*, 2020.