A PERSONALIZED STATIC HAND GESTURE AND VOICE RECOGNITION SYSTEM FOR REAL-TIME SMART DEVICE CONTROL USING LIGHTWEIGHT DEEP LEARNING

Tharmathayalan Sankeethan*, Santhirasekaram Thanujan, Deluxshana Amalathas, S. H. D. Senanayake, M. Janotheepan

Department of Computer Science and Informatics, Faculty of Applied Sciences, Uva Wellassa University, Badulla, Sri Lanka

*Corresponding author: sankeethantharmathayalan@gmail.com

Abstract

Smart home automation systems frequently rely on manual switches or mobile applications, creating accessibility barriers for individuals with limited mobility. While gesture and voice control offer touchless alternatives, existing solutions often require continuous internet connectivity, lack user-level personalization, and depend on computationally heavy hardware. To address these limitations, this research proposes an offline, multimodal control system supporting 18 static hand gestures and 14 voice commands for real-time smart device operation. Personalization is achieved through a local on-device interface where users dynamically map their preferred inputs to specific appliance actions without cloud processing. The methodology utilizes lightweight deep learning models deployed directly on the AMB82 Mini microcontroller. For gesture recognition, a Convolutional Neural Network (CNN) trained on 96,558 augmented images classifies inputs, achieving 88.07% accuracy on an 8,095-sample test set with under 50 ms inference latency. For voice command processing, a Long Short-Term Memory (LSTM) network trained on 156,749 augmented clips extracts Mel Frequency Cepstral Coefficients (MFCCs), demonstrating 95.75% accuracy on a 4,305-clip test set while processing 1-second audio captures. Instead of feature-level fusion, inputs are integrated using a decision-level arbitration layer applying confidence thresholds, ensuring independent channel operation if one modality fails. Evaluated end-to-end under varied illumination and background noise, the deployed quantized system successfully actuated 85.1% of gesture commands and 90.6% of spoken commands. By executing these algorithms entirely on device within a unified framework, this study presents a responsive, cost-effective assistive solution that provides touchless device control for users with physical impairments.

Keywords: Smart home automation, Convolutional Neural Network, Long Short-Term Memory, Multimodal Control System, Assistive technology