universidade de aveiro theoria poiesis praxis

Multimodal Interaction Fusion

IKEA Website Controlling

Roberto Rolão de Castro - 107133
Tiago Caridade Gomes - 108307

Course: Multimodal Interaction 2024/25

Contents

List of Figures

1. Introduction

This report presents the development of an **interactive application controlled through a combination of gestures and voice commands**, implemented using **Microsoft Kinect** and **Rasa**. The project builds upon previous work, combining voice recognition and gesture detection technologies to enhance the user's online shopping experience on the IKEA website. By interpreting voice commands and recognizing gestures, the system facilitates a wide range of shopping tasks, such as browsing products, managing the cart, and finalizing purchases, offering an intuitive, accessible, and efficient alternative to traditional input methods.

1.1 Objectives

- Utilize Microsoft Kinect to integrate gesture recognition.
- Leverage Rasa for voice command recognition.
- Enable the system to control a diverse set of functionalities within the application.
- Implement the Fusion Engine to aggregate information from Kinect (gestures) and Rasa (voice commands) for seamless communication with the **Multimodal Interaction Manager (MMI)**.
- Provide real-time, multimodal feedback via speech synthesis and visual cues for an enhanced user experience.

1.2 Application Context

The application merges two previous projects: one focused on voice-controlled interaction with an online shopping platform and another centered on gesture-based control. This iteration introduces a **Fusion Engine** that integrates input from Kinect and Rasa, enabling the **MMI** to process and coordinate multimodal commands. This combined approach demonstrates the potential of multimodal interaction to deliver a seamless and innovative user experience.

2. System Architecture

The system's architecture integrates several components to enable seamless gesture-based interaction:

- **Natural Language Understanding (NLU):** Rasa process voice commands, identifying intents and extracting relevant entities.
- **Microsoft Kinect:** The primary hardware for capturing user gestures. Kinect's depth-sensing cameras and motion tracking capabilities allow accurate recognition of predefined gestures and provide the foundation for gesture-based interaction.
- **Generic Gestures Modality:** The system implements a set of generic gestures, enabling users to interact through intuitive movements. These gestures are captured by the Kinect and mapped to specific application commands.

- **Fusion Engine:** The Fusion Engine is responsible for integrating and processing information from multiple modalities (gestures and voice commands). It combines incoming data to determine the user's intent and triggers the appropriate actions. This approach ensures the system responds cohesively and efficiently to multimodal inputs.
- **Multimodal Interaction Manager (MMI):** The MMI acts as the central hub for processing and managing inputs from multiple modalities (gestures and voice). It receives data from the Kinect and Rasa, aggregates it through the Fusion Engine, and determines the appropriate action to execute. The MMI ensures real-time synchronization of multimodal inputs, resolves conflicts between overlapping commands, and coordinates the flow of information between the system components.
- **Main.py:** This Python component acts as the central controller, managing the flow of data and coordinating the interaction between different components. It receives commands from the MMI via WebSockets and interacts with the IKEA website via Selenium WebDriver to execute actions such as scrolling and clicking.
- **Selenium WebDriver:**Used for automating interaction with the IKEA website. Selenium executes actions like scrolling, clicking buttons, and navigating product pages based on gesture inputs.
- **WebAssistantApp:**A web application that serves as the interface layer, enabling voice communication between the user and the system.
- **Speech Interface (TTS):**The speech interface complements the interaction by providing feedback through synthesized voice outputs, ensuring users are informed about system actions and statuses.

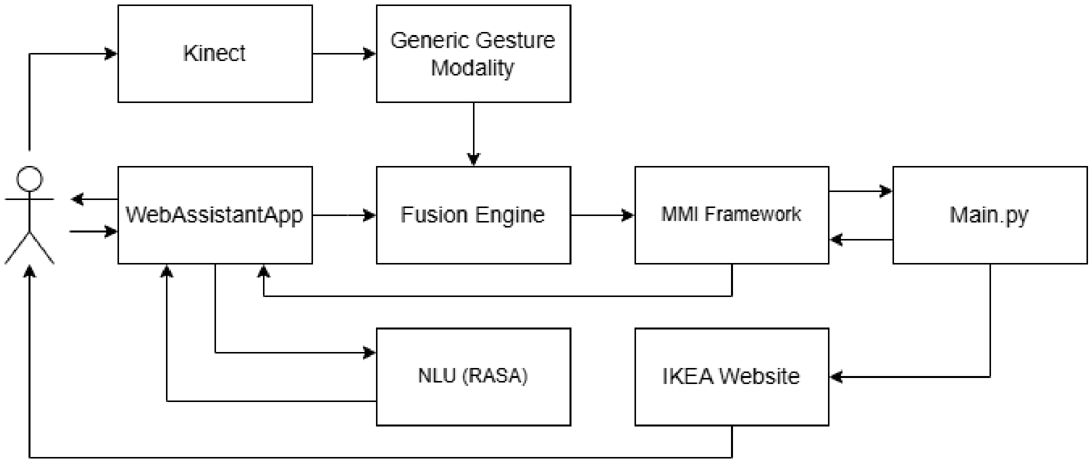

Figure 2.1: Architecture Diagram

This architecture ensures modularity and scalability, allowing components to operate independently while working together to deliver a cohesive user experience. Each module is optimized for its specific function, from gesture recognition to website interaction.

3. Overview of the Actions

The system implements the following gestures and voice to facilitate interaction with the IKEA website:

3.1 Single Actions

3.1.1 Voice

- **Open Site:** Say "Abrir o site" or "Vamos às compras" to open the IKEA Portugal website.
- **Search for a Category:** Say "Quero ver [cadeiras](category)" or "Procura por [sofás] category)" to search for a specific category of products.
- **Add to Cart/Favorites:**
  - Add a specific product to the cart by saying "Adiciona o produto ao carrinho" or "Gostaria de adicionar este item ao carrinho"
  - Add a specific product to the favorites list by saying "Adiciona-me este produto aos favoritos" or "Por favor, coloca isso nos favoritos"
- **Show Cart/Favorites:**
  - Say "Mostra-me o carrinho" or "Quero ver o que tenho no meu carrinho" to view the cart.
  - Say "Abre os favoritos" or "Gostaria de verificar os favoritos" to view the favorites.
- **Show More:** Say "Quero ver mais opções" or "Mostra mais opções" to display additional products on the page.
- **Finalize Purchase:** Say "Finaliza a compra" or "Por hoje, já escolhi tudo" to proceed to checkout.
- **Confirm Actions:** Say "Sim" or "Não" to confirm or cancel an action.
- **Filter Products:** Say "Mostra os produtos do [mais baixo ao mais elevado](criterio)" or "Ordena pelos produtos [Mais Recente](criterio), por favor" to organize the products displayed on the page according to the selected criteria, such as price, popularity, or dimensions.
- **Help:** Say "Preciso de ajuda!" or "Ajuda!" to open a webpage with all the voice and gestures commands that the user can do.

3.1.2 Gestures

- **Close Site:** Stand up and disappear of the image to close the website.
- **Navigate Products:** Use gestures to navigate:
  - Move the right open hand up to navigate upward.
  - Move the right open hand down to navigate downward.
  - Move the right close hand to the right to navigate rightward.
  - Move the left close hand to the left to navigate leftward.

3.2 Redundant Actions

3.2.1 Voice

- **Scroll the Page:** Say "Sobe a página" or "Desce a página" to navigate the page vertically.
- **Go Back:** Say "Quero voltar a trás" or "Volta para a última página" to navigate backward.
- **Main Page:** Say "Volta à página inicial" or "Quero voltar ao início" to navigate to the start.

3.2.2 Gestures

- **Scroll the Page:** Extend the arm to the side-up with the left hand open to scroll up the page and to the side-down with the left hand closed to scroll down the page.
- **Go Back:** Raise the right hand towards the shoulder to return to the previous page.
- **Main Page:** Raise the both arms towards the head like a house to return to the homepage.

3.3 Complementary Actions

- **Remove Product from Favorites/Cart:**
  - **Gesture:** Raise the both arms towards the head like a 'X' to remove the selected product.
  - **Voice:** Say "Remove o produto" or "Elimina o produto"
- **Select Product:**
  - **Gesture:** Point to the product with the right hand open to open the selected product.
  - **Voice:** Say "Selecionar" or "Quero este"

This gesture-voice-based interactions enables users to navigate and control the IKEA website efficiently, enhancing accessibility and providing a unique user experience.

4. Conclusion

The development of this multimodal interaction system demonstrates the significant potential of combining gesture and voice-based commands to enhance user experience in online shopping platforms. By integrating Microsoft Kinect for gesture recognition and Rasa for natural language understanding, the project successfully implements a seamless and intuitive interface for interacting with the IKEA website.

Through the implementation of the Fusion Engine, which aggregates inputs from both modalities, the system achieves high levels of accuracy and synchronization, allowing for advanced functionalities such as product selection, cart management, and purchase finalization. This multimodal approach bridges the gap between traditional input methods and more accessible, innovative solutions, making the application inclusive and user-friendly.