Motion Capture and Gesture Recognitions/Controls

Nikal Prajapati   036 2096   22013 261   nikal22013261@iimscollege.edu.np
Sabu Dhungana   036 2741   22013 262   sabu22013262@iimscollege.edu.np
Shikshya Pokhrel   036 2493   22013 236   shikshya22013236@iimscollege.edu.np
Sujal Ratna Tuladhar   036 2483   22013 230   sujal22013230@iimscollege.edu.np

Artificial Intelligence
May 2024 – Semester 5
Section B Laboratory 3
Group 20

1) Introduction

Computer Vision (C.V.) is a rapidly growing field in Artificial Intelligence (A.I.) that allows computer and devices to understand and interpret real-world visual inputs. In this evolving digital age, applications relate to computer vision is an emerging technology that offers exciting potential, one of those abilities is allowing users to interact with multimedia applications through body movements as it translates actions to enhance user experiences and meaningful actions (Human-Computer Interaction) can also useful in controlling heavy-machineries and mobile-robots.

Although, traditional input devices like keyboards and mice are effective designs with huge audiences to cater to but can be troublesome and unintuitive for certain demographic, particularly while controlling in multimedia and entertainment applications. Our proposed project system specifically focuses on use and exploration of two powerful OpenCV and MediaPipe libraries to explore hand-based control of multimedia applications, volume adjustments, and artistic creation on a personal computer (PC), providing a platform to learn and experiment with core computer vision concepts

2) Problem Statement – Methods

Current methods of controlling multimedia applications rely heavily on traditional physical input devices (e.g., keyboards, mice, buttons and touchscreens). Which can limit user's natural interaction and accessibility mostly in inconvenient scenarios where users wish/need hands-free engaging interaction and such may not be as intuitive or efficient for tasks that involve dynamic and creative inputs, such as volume control, multimedia presentations, creative software's or digital art creation.

3) Aim

a) Learn about Computer Vision and Natural Language Processing through different algorithms, frameworks, tools

b) Using Python Language to develop specific idea of statistical programming and implement tangible machine learning solutions

c) Based on specific case study and example scenarios, find a way to improve the existing solution with research

d) Develop intuitive, accessible, and engaging ways to interact with multimedia applications control functions

e) Find replacement for conventionally challenging input devices for individuals with disabilities.

4) Objectives

a) Explore OpenCV and MediaPipe libraries and documentations to enable functionalities with different methodologies to develop basic recognition algorithms for various movements

b) Derive meaningful information from visual inputs of real time motion capture or hand gesture recognition as it is mapped to specific commands control for multimedia applications

c) Gaining practical knowledge and experience with OpenCV for image processing and video capture also utilization of Media Pipe's hand pose estimation module to understand hand landmark detection and tracking

d) Build a rudimentary system showcasing hand control of multimedia playback or simple on-screen interactions.

5) Preliminary Literature Review

OpenCV, by Intel, is a robust cross-platform-language open-source computer vision library, with capabilities of offering extensive tools for image and video processing, making it ideal for real-time feature tracking and recognition in Internet of Technology (IoT). MediaPipe, developed by Google, provides efficient and accurate pretrained pipeline-framework models for live and streaming media detecting and tracking hand landmarks in real-time. Their accuracy is measured in varying lighting conditions, visual odometry, differentiating between similar gestures, and ensuring real-time performance. robustness and ease of implementation effectiveness of hand landmark models provided for precise key point localization. for controlling a robot.

6) Methodology

We install and configure OpenCV and MediaPipe libraries in the python development environment. To capture and acquire the live feed for hand recognition we use an integrated webcam. Through tutorials and official documentations, we apply fun ctionalities for image manipulation, video capture, and basic outline provided by OpenCV. Media Pipe's pre-trained model allow tracking solution to access real-time landmark data (e.g., fingertip positions) from the video feed.

Control Logic & Algorithms, techniques like thresholding, classification, detection and estimation to recognize specific or pre-defined positions and movements can be created as set of hand to gestures and map them to corresponding multimedia applications control and creative software. The video feed can be preprocessed to improve the quality of the images and reduce noise. Communication is required between recognition module and target interface (media players, volume control, visual programs) and guides to different hand gestures associated with functionality.

7) Conclusion

While the project has an in-depth interactive element, our primary focus here lies on learning computer vision techniques. Implementing such learned concepts and functionalities will serve as a practical application allowing us to gain hands-on experience with popular libraries for computer vision tasks. Understand the potential of computer vision for various applications beyond such.

8) References

S, F. C., H, R., K, G., & S, J. (2023). Extraction of character from visuals and images using OpenCV. International Journal of Scientific Research in Computer Science, Engineering and Information Technology, 9(3), 194–200. https://doi.org/10.32628/CSEIT2390363