Final Project Report
Human Movement Detection

Computer Vision
Instructor: Dr. Duong Phung
Group: GTT-4
Tran Ngoc Anh - 200154
Nguyen Thu Huyen - 190033
Phan Nguyen Tuong Minh - 190048

Motivation

Human movement detection is a prominent and extensively researched area within the field of Computer Vision. Understanding the mechanisms by which computers can effectively identify and comprehend human movement, body language, and signals offers valuable insights for the integration of Computer Vision techniques into various aspects of our daily lives.

In line with our passion for Public Healthcare, our research group aims to delve into methods, algorithms, and technologies utilized for human movement detection. Our primary objective is to develop a warning system capable of promptly identifying instances of stroke happening in everyday life. By leveraging the advancements in Computer Vision, we seek to enhance the early detection and response to stroke cases, thus contributing to improved healthcare outcomes and potentially saving lives.

Given the absence of comprehensive technical and medical expertise, our group initiated our endeavor by constructing a rudimentary model focused on the identification of human movement through auditory cues. By doing so, we aim to establish a foundational understanding of movement detection techniques and pave the way for future advancements in our research. Our ultimate goal is to progress towards the development of a robust stroke movement detection system that incorporates sound-based alarms. This iterative approach allows us to build upon our initial findings

and expand the scope of our project as we acquire the requisite knowledge and resources, ultimately contributing to the improvement of stroke detection and intervention methods.

Project Overview

The primary objective of the project is to implement a system for 3D human pose estimation, focusing on the detection and movement tracking with sounds of specific key points, namely the right hand wrist, left hand wrist, and nose.

Specifically, the program we have developed encompasses two primary functionalities.

1. It detects and tracks the positions of the user's wrist joints and nose, thereby facilitating the monitoring of their corresponding movements. This capability enables accurate analysis of the user's gestures and kinematic behavior.
2. It incorporates an alarm system that triggers a sound notification upon detecting movement in at least one of the three aforementioned body parts. This alert mechanism serves as an effective means of drawing attention to significant movements or actions performed by the user.

By combining these features, the program offers a comprehensive solution for real-time monitoring of specific body points and provides an immediate auditory indication when relevant movements occur.

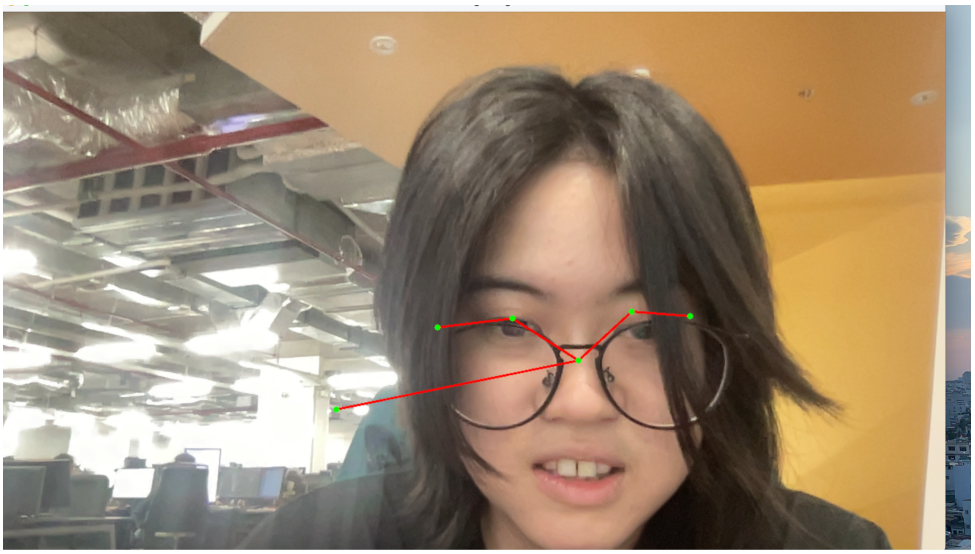
Technology

Initially, our approach was to start from scratch: we will collect a number of images of a specific human pose and conduct feature detection based on a predefined body frame. We would want to map the nose and two wrists of our images to the frame, thus retrieve the features. However, we realized that this is not a feasible approach to solve the problem as the accuracy fluctuates significantly among the test images. Therefore, we decided to switch to using a pre-built model.

Since pose estimation is a well-tackled problem with many pre-built models, we decide to choose one model and retrieve the results of joint detection from that model to generate our pose estimation and movement. We choose to use the Hand Pose detection model in Tensorflow for pose estimation. However, the model could not accurately detect the features in specific conditions for our demo. Therefore, we come up 2 approaches to improve this pre-built model:

- Apply imaging filtering for each input, which includes Gaussian blur, contour and histogram equalization to enhance some of the features that we want to retrieve from the images as well as minimize the lighting and unwanted details for our images.
- Retrain the model: along with applying the filters, we use an open-source human pose dataset to retrain the model. The dataset contains ~25 images that can be used for training. This will help enhance the accuracy of the model for the specific cases for our demo in class later on.

The below image shows our results after preprocessing the input images and retraining the model. The pose estimation can correctly detect the nose, but it also detects some mysterious features that do not belong to the pose. Furthermore, we still could not detect the two wrists correctly.



In the end, we ended up using a library called PoseNet by Tensorflow to retrieve the nose and two wrists. PoseNet is a well-trained library already, so we did not train the model again before conducting the real-time test. The features' location on the pose are returned as 2D points (x, y).

By successfully retrieving the joint features that we want to, we continue to detect the movement of a person. The idea is to collect the features' location of the two continuous frames, and calculate the distance change of the same feature. If the distance is greater than a certain threshold, the movement will be detected. Since the distance change in the nose is harder to recognize, we set the threshold for the nose to be smaller than the 2 wrists.

To denote a distance change of the feature, we integrate some sounds to notify the distance. Other than that, we implement a music rate change to define the speed of the movement. If the change of the distance is large, and small again in the next frame, we consider it as a quick movement, hence we increase the music tempo by 2 times, and vice versa for a slow movement. The sound implementation is created using the P5 library.

Result

Our software application demonstrates a good level of efficacy in detecting and tracking the position of the user's nose and both hand wrists. Additionally, it effectively triggers an alarm when it detects any movement of the nose. This program operates regardless of the proximity of the user to the camera, accommodating both close and far distances. Furthermore, the sound alarm mechanism functions reliably when detecting a single body part at a time. Notably, even in situations where the user conceals their nose, the program is capable of detecting the nose by relying on the remaining facial points.

Discussion

Limitation

The project at hand encounters some limitations that merit consideration.

Firstly, the current implementation of the project is limited to the detection and tracking of the movement of a single human individual at any given time. Consequently, the simultaneous detection of multiple individuals, which is often required in real-life scenarios, is not accommodated by the existing system.

Secondly, the accuracy of the detection algorithm is influenced by the distance between the user and the camera. Specifically, the proximity of the user to the camera amplifies the sensitivity of the algorithm, resulting in more acute movement detection. Conversely, greater distances may lead to reduced precision and accuracy in detecting movements.

Thirdly, the inclusion of a sound alarm for all three body parts simultaneously adversely impacts the program's overall accuracy and responsiveness. This is primarily attributed to inherent limitations in video frame processing and the time complexity associated with real-time analysis.

Further development

In order to fulfill our ambition of developing a stroke detection program, it is imperative to enhance the accuracy and capabilities of the existing system. This entails improving the detection accuracy for individuals situated at greater distances from the camera, enabling the simultaneous detection and tracking of multiple individuals, and precisely detecting and tracking all relevant body parts. These enhancements are essential to ensure the accurate recognition of stroke cases in order to facilitate timely medical intervention.

Source code

Please refer to this folder:
https://drive.google.com/drive/folders/1q9ySciYJU0xQcshxsQV2IMqs0ToYhe_X?usp=sharing

References

Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Rafal Jozefowicz, Yangqing Jia, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dan Mané, Mike Schuster, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. (2015). TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems.

George Papandreou, Tyler Zhu, Liang-Chieh Chen, Spyros Gidaris, Jonathan Tompson, & Kevin Murphy. (2018). PersonLab: Person Pose Estimation and Instance Segmentation with a Bottom-Up, Part-Based, Geometric Embedding Model.