CAIRO UNIVERSITY
FACULTY OF ENGINEERING
CREDIT HOURS SYSTEM -SENIOR-1 LEVEL
IMAGE PROCESSING & COMPUTER VISION
CMPS446 -FALL 2024

Vision-Based Mouse Tracker with Hand Gesture Control

CMPS446 Project Proposal

Team Members:	Arwa Mohamed Mahmoud Mohamed Ibrahim	1210182
	Mohamed Ahmed Ibrahim Sobh	1210288
	Mohamed Nabil Elsayed	1210291
	Omar Ahmed Ibrahim	1210020
Submitted to:	Eng. Hannah Hatem	
Submission Date:	December 23, 2024	

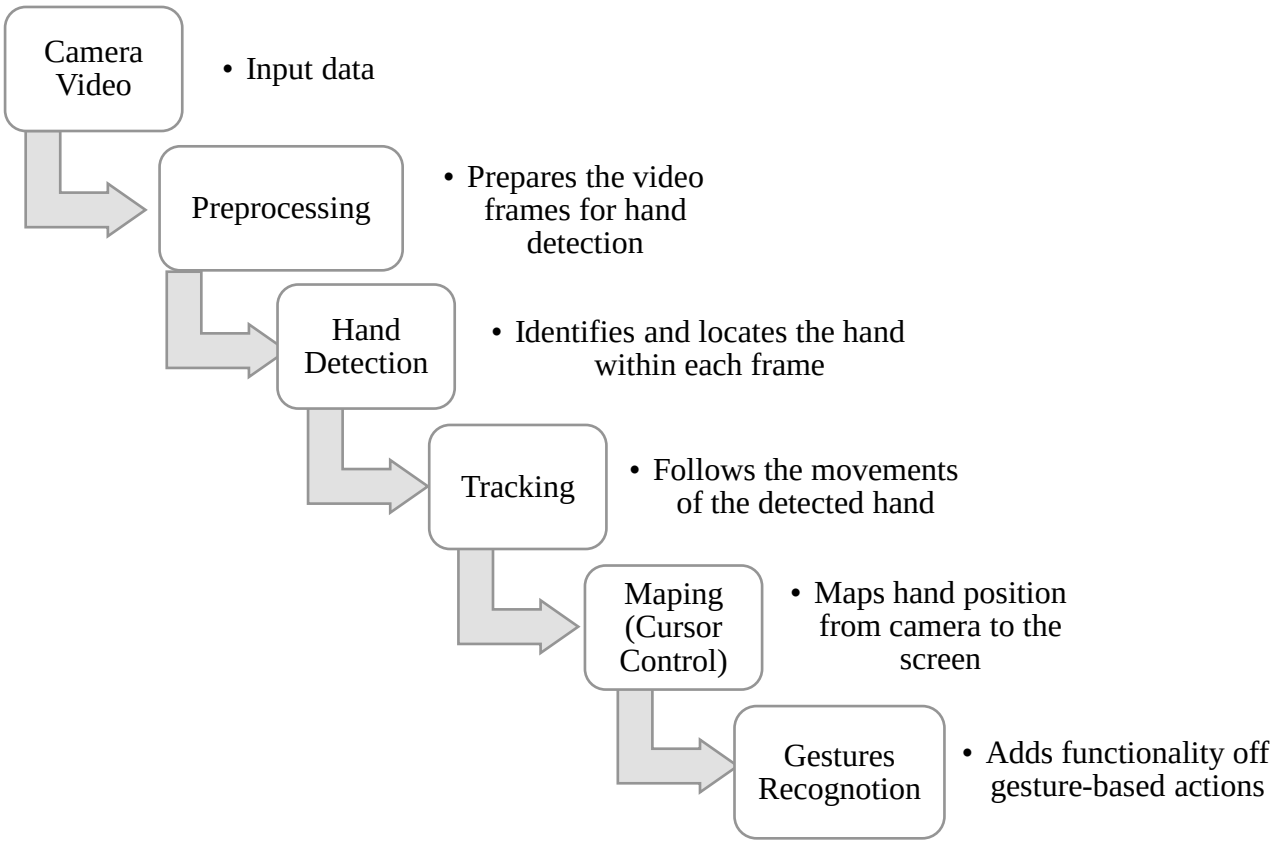
TABLE OF CONTENTS

THE PROJECT IDEA AND NEED:

The objective of the mouse tracker project is to create a hands-free control system that allows users to operate the computer's cursor through hand movements detected by a camera. The system will capture and interpret the user's hand position on the screen, providing a touchless alternative to input devices like a mouse or trackpad.

Such systems are highly demanded in environments where the traditional input may not be feasible, such as maintaining hygiene in healthcare sectors or providing a solution for users with limited mobility.

INFORMATIVE BLOCK DIAGRAM:



MODULES DETAILS:

The project is implemented twice:

- With classical applications
- With Neural Networks and deep learning libraries

Classical Approach:

Files and Scripts	Purpose
Util.py	Contains all general variables, constants, functions, and definitions for cleaner code.
Images_generator.py	Our first step is to generate images to train a ML classifier. This script uses a function defined in util.py called segment image using kmeans to segment the image, we use 2 clusters only for better training on HOG features.
Dataset_generator.py	The second step is to generate the dataset needed to train the model; we use HOG as the feature to train the model.
Model_generator.py	The third step is to generate the model, we use KNN, RandomForest, and Decision Tree for the classification algorithms, and then, we train the models and choose the model with the best accuracy.
Main	The final step is to run the main file, after the model is generated, we use this file to start the webcam, segment with kmeans, which is previously defined in util.py, and then predict using the model. It also gets X-Y of the hand from the K-means so we can move the mouse

USED ALGORITHMS:

Note: the algorithms mentioned below are implemented from scratch

Algorithm	Details
K-Mean Clustering	Segmentation of hand for k clusters.
Center of mass	Getting the X-Y of the hand.
K-Mean / Center of mass mixing	During the main application phase, the image will not be clustered into binary so that it can work with different environments, this leads to the disability of getting the center of mass, that's why we've decided to use the center of mass with a specific cluster that fits the hand color requirements.
Histogram of Gradients	Used for the classification of the image since it fits the most to detect gestures
RandomForest, KNN, Decision Tree	For the classification of the gesture

EXPERIMENTAL RESULTS AND ANALYSIS:

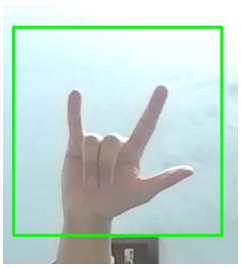
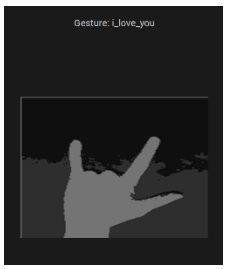
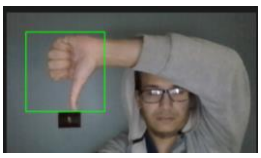
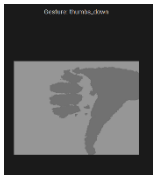
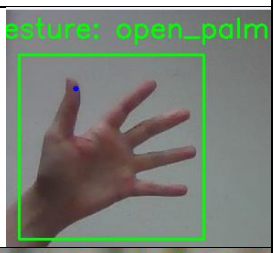
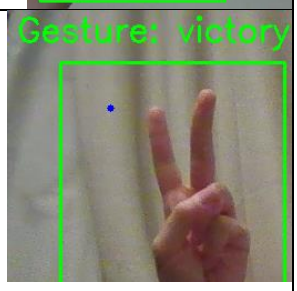
After experimenting and providing a great dataset with all positions available inside ROI, the classifier can easily classify the item in different environments with the help of 3 clusters, and for the center of mass, it is not 100% accurate since the segmentation is not binary, however, it is quite fine in terms of moving.

For test cases creation, we need to test it as a product, and as in implementations. For product-based testing, we need to test it simply as users, in conditions like plain background or busy background. For implementation testing, we need to train the model on a different environment than the one to be tested, or different clustering number than the one the model trained with.

When trained on different environment (different background environment and lighting environment), the model has proven itself quite a bit. It resulted success in 90% every case.

The failures came when we set our hands on position, while this position did not exist in the data training.

Test Cases

Test case	**Status**	**Illustration**
Same environment Good lighting Plain background Different cluster number	Success	 
Same environment Bad lighting Plain background Different cluster number		 
Different environment Plain background Same cluster number	Success	 
Different environment Poor lighting Busy background Same cluster number	Success	 

Same environment Good lighting Plain background Different cluster number Position that does not exist within the dataset	Failed	

Points of Strength and Weaknesses:

Strength	**Weakness**
• Can work with different environments • Independent of light • Independent of skin color	• The dataset must contain all types of positioning • Needs ROI

Comparison with NN Approach:

Classical Approach	**NN Approach**
Uses the same concept of gesture bclassification	Uses same concept of classification of gestures
Uses pure image processing techniques to get the hand position	Uses Mediapipe (a deep learning library) to get the hand position (No ROI)
	More accurate, works on a higher range, faster detection

WORK DIVISION BETWEEN TEAM MEMBERS:

Team Member	**Workload**
Arwa Mohamed Mahmoud	Classification pipeline, GUI
Mohamed Ahmed Ibrahim Sobh	
Mohamed Nabil	Segmentation
Omar Ahmed Ibrahim	Deep learning all alone + Center of mass

CONCLUSION:

To conclude, this project helped us understand the pipelining and helped us link the labs together.

RESEARCH PAPERS:

These are some research papers that might help implementing the project:

1. ***"MediaPipe Hands: On-Device Real-Time Hand Tracking"*** 2006.10214
2. ***"Gesture Recognition for Human-Computer Interaction"*** An Exploration into Human–Computer Interaction: Hand Gesture Recognition Management in a Challenging Environment | SN Computer Science

3. ***"Research studies on vision-based mouse control and touchless interactions"*** The effects of touchless interaction on usability and sense of presence in a virtual environment | Virtual Reality
4. ***"Robust Visual Tracking by Segmentation"*** 2203.11191
5. *Fast Online Object Tracking and Segmentation: A Unifying Approach*