

# Gesture based control app

Project conceptual documentation

**Vlad Bartolomei**

A thesis presented for part of the degree of  
Interacțiune Om-Calculator



MEMBRĂ A



Facultatea de Automatică și Calculatoare  
Universitatea Tehnică din Cluj-Napoca  
România  
20 Noiembrie 2024

# Contents

<b>1</b>	<b>Introducere</b>	<b>2</b>
<b>2</b>	<b>Functional requirements</b>	<b>2</b>
<b>3</b>	<b>Features</b>	<b>2</b>
3.1	Controlul Volumului . . . . .	3
3.2	Controlul cursorului (mouse) . . . . .	4
3.3	Oprirea programului de detecție a gesturilor . . . . .	4
3.4	Starea de pauză a unui video . . . . .	5
3.5	Captura de ecran - Screenshot . . . . .	5
3.6	Zoom in, zoom out . . . . .	6

# 1 Introducere

Scopul prezentului proiect este de a crea o aplicație care face mai ușor de vizionat conținut multimedia, propulsată de o tehnologie nouă, în speță MediaPipe. MediaPipe este o librărie dezvoltată pentru Python prin intermediul inteligenței artificiale care servește la detecția feței, mâinilor sau a posturii și amplasarea de *landmark*-uri.

## 2 Functional requirements

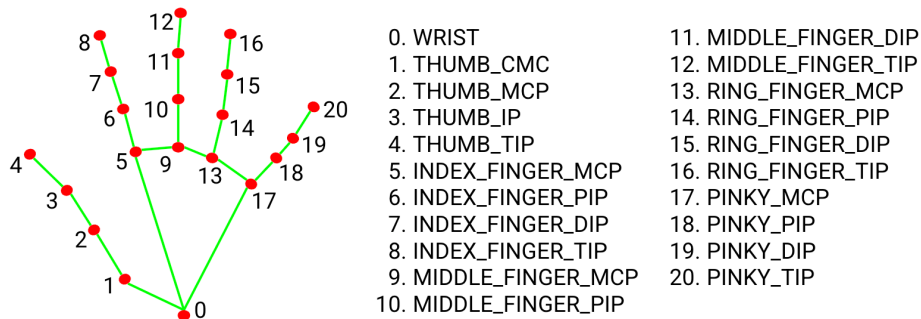
Aplicația își propune să elimine contactul fizic între utilizator și periferice și, astfel, să execute comenzile cerute de utilizator doar prin simpla recunoaștere a unor gesturi prestabilite. Pentru a simplifica problema, aplicația va aduce în fața utilizatorilor un set restrâns de instrucțiuni specifice controlului unor imagini statice sau video, precum și al controlului audio.

Plecând de la această idee fundamentală, aplicația va trebui să recunoască gesturi care să poată întreprinde următoarele:

- controlul volumului;
- mărirea sau micșorarea unei imagini/unui document înr focus;
- controlul cursorului prin mișcarea mâinii, precum și interacțiunea cu diverse elemente prin acțiunea de click;
- starea de redare/pauză (play/pause) a unui video;
- efectuarea unei capturi de imagini a întregului ecran - screenshot;
- oprirea programului.

## 3 Features

Esența rezolvărilor acestor funcționalități propuse constă în ceea ce MediaPipe oferă primordial, respectiv detecția mâinii printr-un set discret de 21 de puncte, după cum arată figura de mai jos:

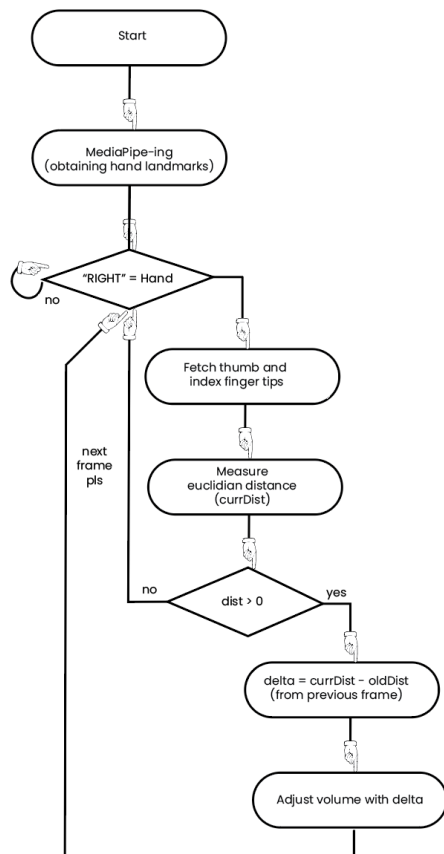


MediaPipe Hand Landmarks

În principiu, ambele mâini pot fi recunoscute, iar relații între aceste puncte (*landmarks*) pot fi determinate a.î. o acțiune să fie lansată în execuție.

### 3.1 Controlul Volumului

Pentru această problemă am decis să folosesc două vârfuri: al degetului arătător și al degetului mare (indecșii 4 și 8), reprezentând de altfel și un gest natural și uman pentru controlul volumului și al tonalității.



Organigrama modificării volumului

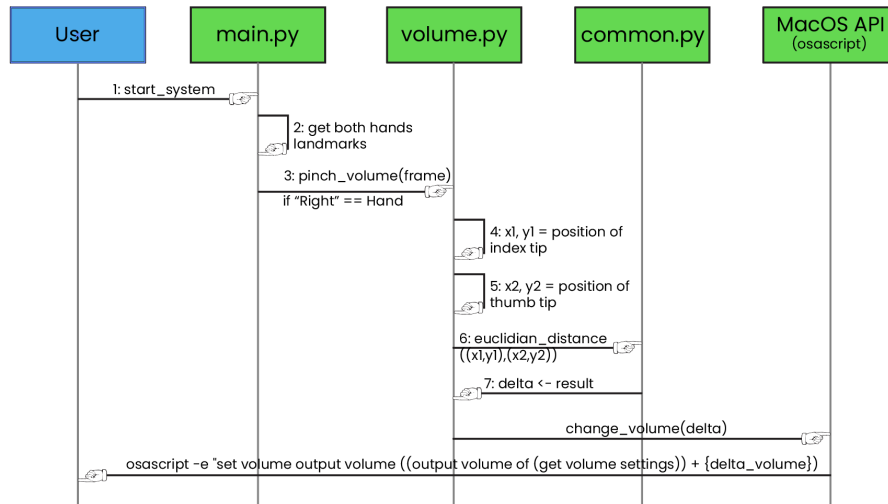


Diagrama de secvență aferentă modificării volumului

### 3.2 Controlul cursorului (mouse)

Pentru această capabilitate se va alege lucrul cu mâna dreaptă. O apropiere a degetului mare de arătător va genera un click, în timp ce mișcarea vârfului arătătorului va determina și mișcarea cursorului.

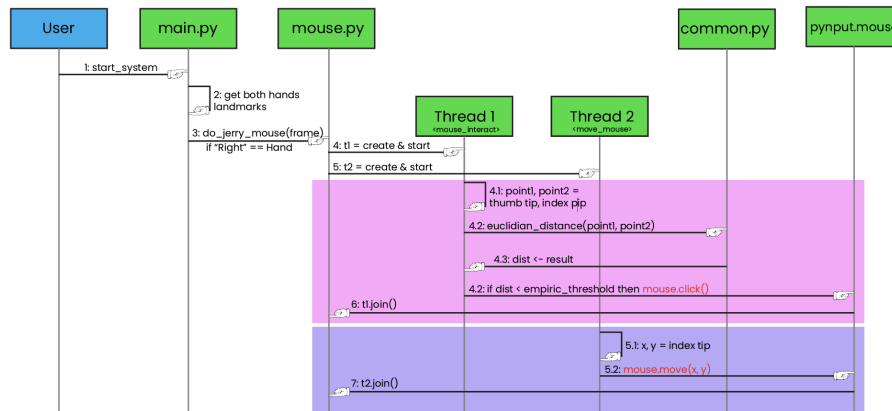


Diagrama de secvență pentru gestionarea mouse-ului

### 3.3 Oprirea programului de detecție a gesturilor

Ar fi destul de ironic ca un program care permite utilizatorului să facă un set de acțiuni prin intermediul gesturilor mâinii să trebuiască să fie oprit *dintr-un buton*. Așa că am implementat un gest pentru această capabilitate, anume ridicarea degetului mic.

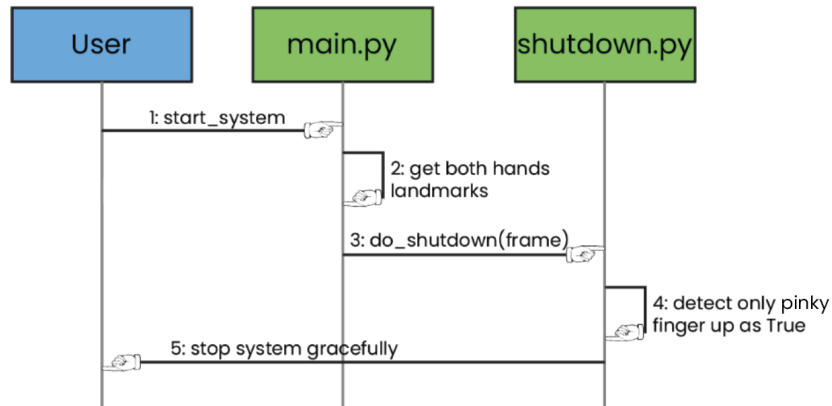


Diagrama de secvență pentru oprirea programului

### 3.4 Starea de pauză a unui video

Pentru acest feature, programul ar trebui să detecteze un **pumn închis**.

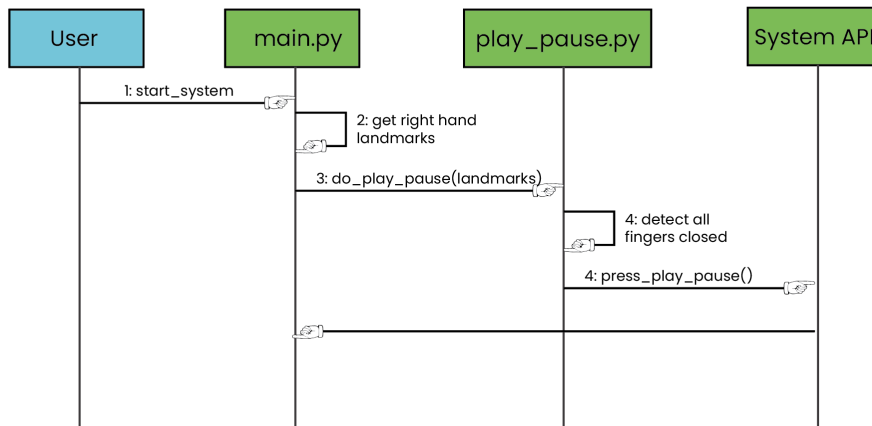


Diagrama de secvență pentru controlul redării unui video

### 3.5 Captura de ecran - Screenshot

Pentru acest feature, programul detectează vârfurile degetelor index și thumb de la ambele mâini, ale căror poziție trebuie să formeze un dreptunghi de laturi specific determinate. Distanța (euclidiană) de sus și jos este  $[0.17, 0.23]$  și cele laterale în intervalul  $[0.10, 0.15]$ .

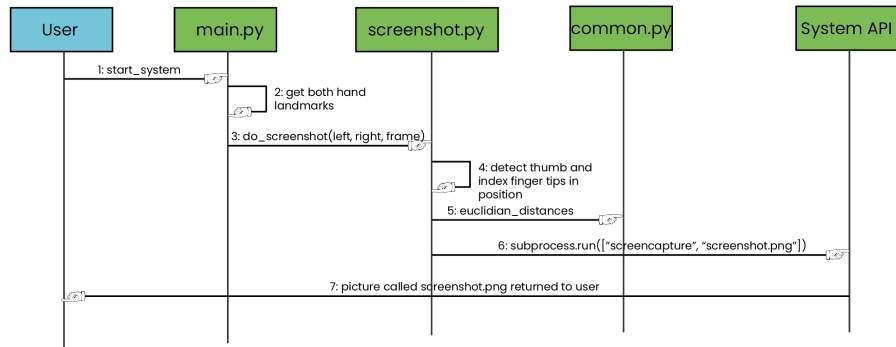


Diagrama de secvență pentru efectuarea unui screenshot

### 3.6 Zoom in, zoom out

Acest feature presupune detecția a două degete ridicate: index și middle:

- left fingers: zoom out
- right fingers: zoom in

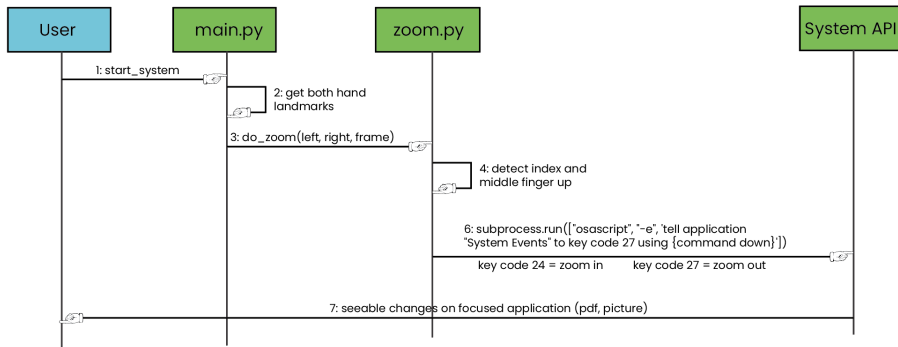


Diagrama de secvență pentru zoom in/out