

HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG



BÁO CÁO HÀNG TUẦN
HỌC PHẦN: THỰC TẬP CƠ SỞ

ĐỀ TÀI:
NGHIÊN CỨU ƯỚC LƯỢNG KHOẢNG CÁCH
BẰNG CAMERA 2D

Giảng viên hướng dẫn: TS. Kim Ngọc Bách

Sinh viên thực hiện:

B22DCCN634

Trần Hữu Phúc

22-29/03/2025

A. BÁO CÁO TIẾN ĐỘ

1. Viết chương trình ước lượng tư thế, phát hiện điểm mốc trên bàn tay sử dụng Mediapipe

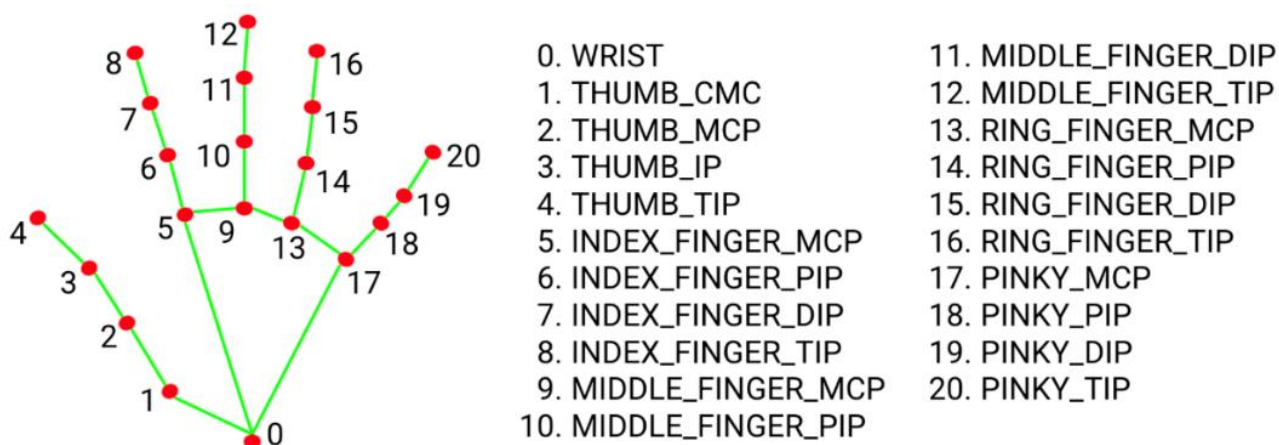
1.1. Phát hiện điểm mốc trên bàn tay bằng MediaPipe

Mediapipe là một framework mã nguồn mở do Google phát triển, dùng để xây dựng các ứng dụng xử lý video, âm thanh, và dữ liệu cảm biến thời gian thực. Framework này cung cấp các giải pháp mạnh mẽ dựa trên AI và DL giúp lập trình viên dễ dàng triển khai các chức năng xử lý dữ liệu đa phương tiện, đặc biệt trong lĩnh vực thị giác máy tính (Computer Vision) và nhận diện chuyển động.

Solution	Android	Web	Python	iOS	Customize model
LLM Inference API	●	●		●	●
Object detection	●	●	●	●	●
Image classification	●	●	●	●	●
Image segmentation	●	●	●		
Interactive segmentation	●	●	●		
Hand landmark detection	●	●	●	●	
Gesture recognition	●	●	●	●	●
Image embedding	●	●	●		
Face detection	●	●	●	●	
Face landmark detection	●	●	●		
Face stylization	●	●	●		●
Pose landmark detection	●	●	●		
Image generation	●				●
Text classification	●	●	●	●	●
Text embedding	●	●	●		
Language detector	●	●	●		
Audio classification	●	●	●		

MediaPipe cung cấp nhiều giải pháp tiên tiến, trong đó có MediaPipe Hands, một mô-đun chuyên dụng để phát hiện và nhận diện 21 điểm mốc (landmarks) trên bàn tay. Những điểm mốc này bao gồm các vị trí quan trọng như

đầu ngón tay, lòng bàn tay và các khớp nối, đóng vai trò thiết yếu trong việc phân tích cử động bàn tay. Mô hình này được xây dựng dựa trên tập dữ liệu huấn luyện lớn với hơn 30,000 hình ảnh thực tế, kết hợp với dữ liệu tay tổng hợp từ nhiều môi trường và cảnh vật khác nhau, đảm bảo khả năng nhận diện chính xác và ổn định trong nhiều điều kiện ánh sáng và góc nhìn.



Mỗi khung hình từ video sẽ được chuyển đổi sang định dạng RGB và gửi qua mô-đun MediaPipe Hands. Mô-đun này sẽ xác định và đánh dấu các điểm mốc trên bàn tay và trả về một landmarks list chứa thông tin về các điểm mốc này dưới dạng tọa độ 3D (x, y, z). Các tọa độ này sau đó được sử dụng để tính toán các khoảng cách giữa các điểm mốc.

MediaPipe Hands có khả năng xử lý dữ liệu video thời gian thực, mang lại độ chính xác cao và hiệu suất vượt trội. Điều này giúp nó trở thành một công cụ lý tưởng trong các ứng dụng như điều khiển bằng cử chỉ, theo dõi chuyển động tay hoặc tích hợp vào các giao diện tương tác người - máy (Human-Machine Interface - HMI)

1.2. Viết chương trình ước lượng tư thế tay

