

EXERCISE TRACKING USING COMPUTER VISION

BY: ALONZO MAYS

OUTLINE OF PRESENTATION

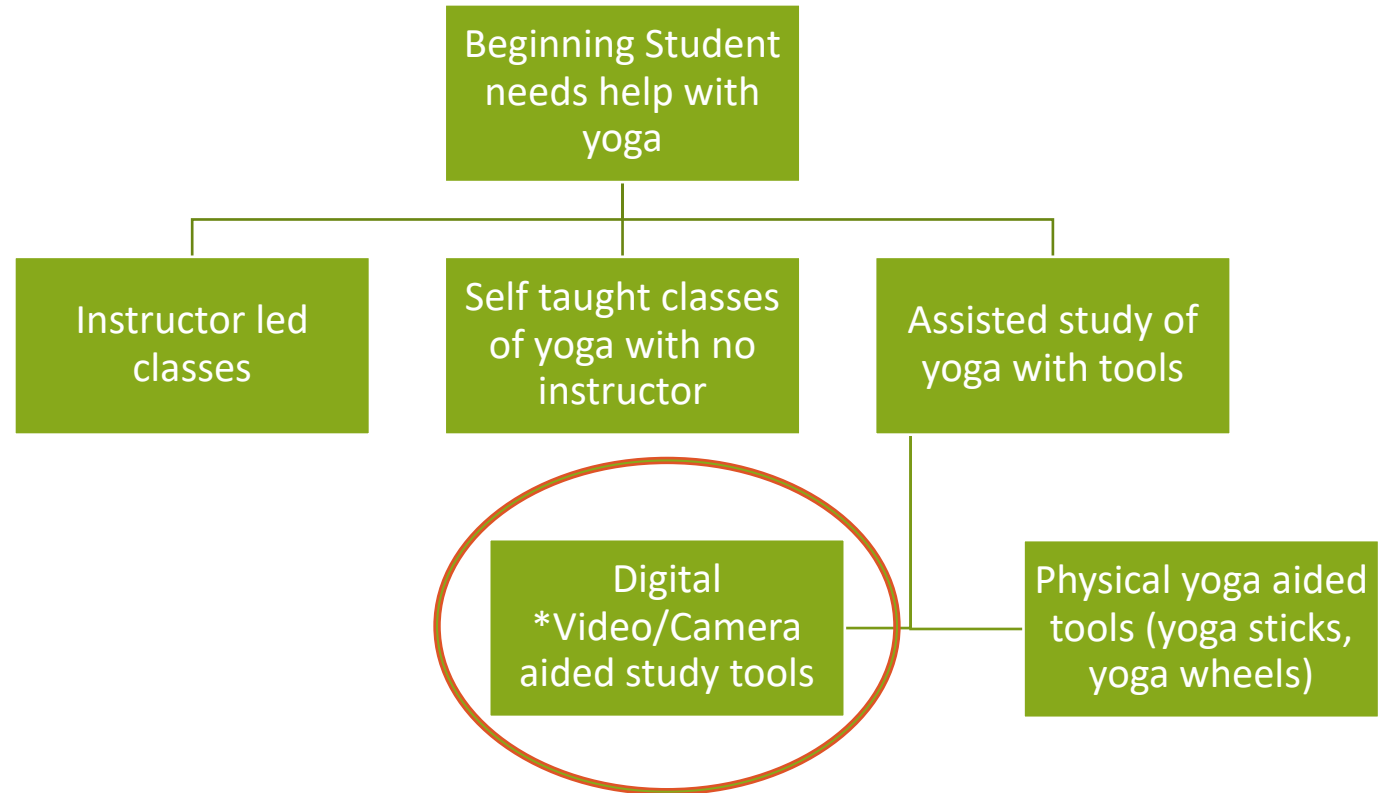
- I. Problem Statement
- II. What is Computer Vision?
- III. What is Media Pipe?
- IV. Exercise Example
- V. Future Considerations

PROBLEM STATEMENT

- I. PROBLEM STATEMENT
- II. WHAT IS COMPUTER VISION
- III. WHAT IS MEDIA PIPE?
- IV. EXERCISE EXAMPLE
- V. FUTURE CONSIDERATIONS

Learning yoga can be difficult when you start the practice. Beginning yoga students need help with doing basic poses (tree pose, mountain pose) in order to reduce injury and condition their bodies to become proficient practitioners.

PROBLEM STATEMENT, CONTINUED



I.	PROBLEM STATEMENT
II.	WHAT IS COMPUTER VISION
III.	WHAT IS MEDIA PIPE?
IV.	EXERCISE EXAMPLE
V.	FUTURE CONSIDERATIONS

WHAT IS COMPUTER VISION?

“Computer vision is a field of artificial intelligence (AI) that enables computers and systems to derive meaningful information from digital images, videos and other visual inputs — and take actions or make recommendations based on that information. If AI enables computers to think, computer vision enables them to see, observe and understand.”

-<https://www.ibm.com/topics/computer-vision>



SUBDOMAINS OF COMPUTER VISION


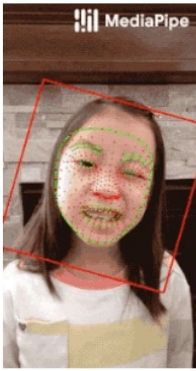

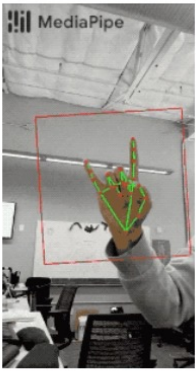




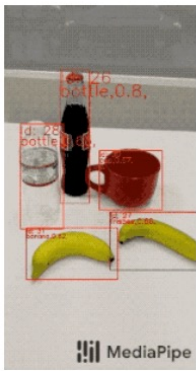

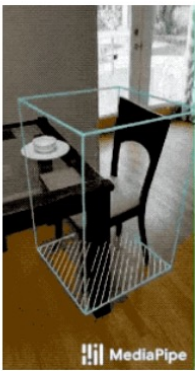

Scene Reconstruction	Object Detection	Event Detection	Video Tracking	Object Recognition	*Pose Estimation
<ul style="list-style-type: none">• 3D representation of objects	<ul style="list-style-type: none">• Finding objects in images or videos	<ul style="list-style-type: none">• Looking at videos and looking for anomalies to detect	<ul style="list-style-type: none">• Locating moving objects using a camera	<ul style="list-style-type: none">• Identifying objects in photos or videos	<ul style="list-style-type: none">• Prediction from user-defined reference pose.

-and more subdomains not listed..

WHAT IS MEDIA PIPE?

- I. PROBLEM STATEMENT
- II. WHAT IS COMPUTER VISION
- III. **WHAT IS MEDIA PIPE?**
- IV. EXERCISE EXAMPLE
- V. FUTURE CONSIDERATIONS

ML solutions in MediaPipe

Face Detection	Face Mesh	Iris	Hands	Pose	Holistic
					
Hair Segmentation	Object Detection	Box Tracking	Instant Motion Tracking	Objectron	KNIFT
					

MODEL SPECS

Model Type

Convolutional Neural Network

Model Architecture

Convolutional Neural Network: MobileNetV2-like with customized blocks for real-time performance.

Input(s)

Regions in the video frames where a person has been detected. Represented as a 256x256x3 array with aligned human full body part, centered by mid-hip in vertical body pose and rotation distortion of (-10, 10) .

Channels order: RGB with values in [0.0, 1.0].

Output(s)

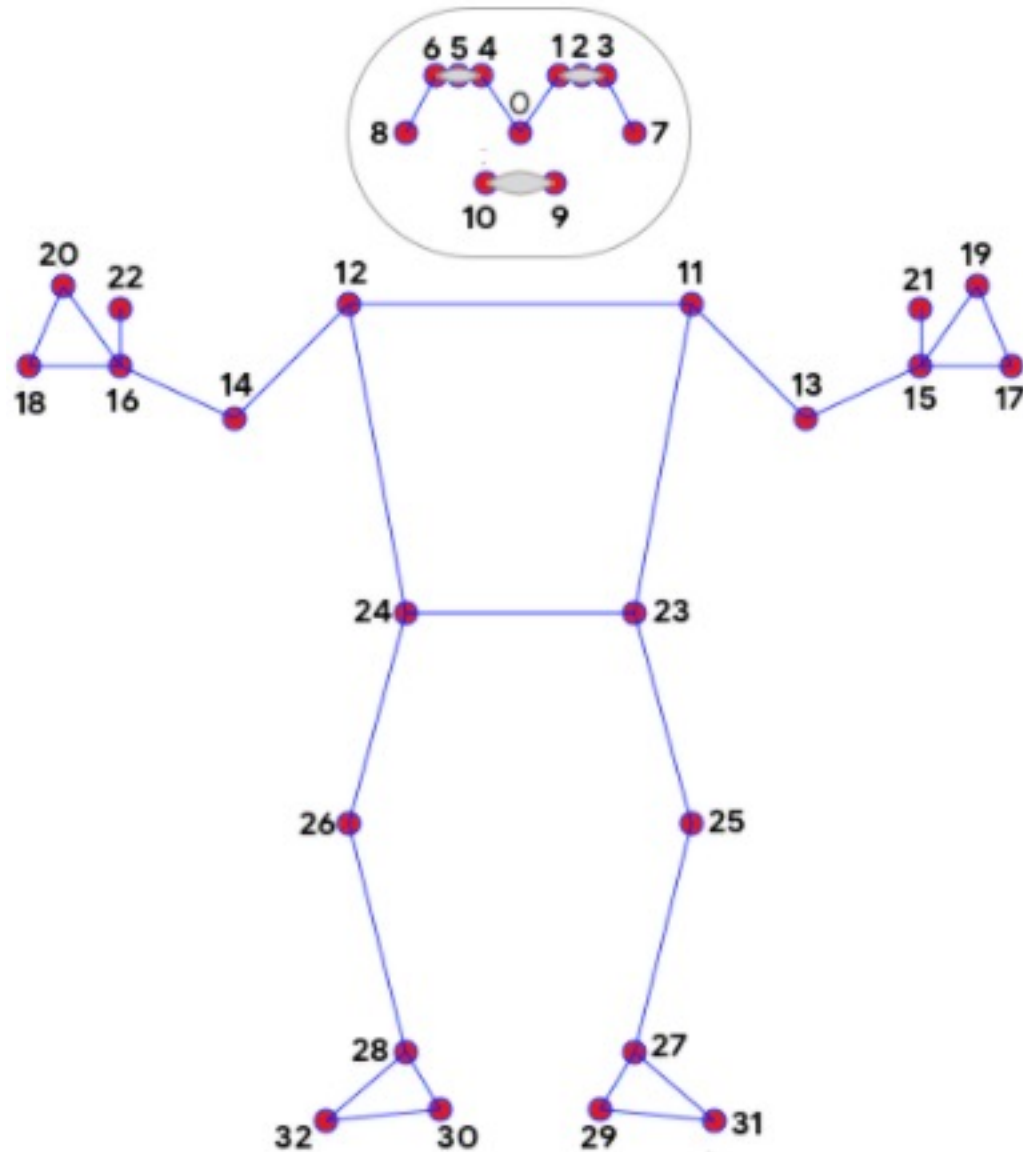
33x5 array corresponding to (x, y, z, visibility, presence).

- X, Y coordinates are local to the region of interest and range from [0.0, 255.0].
- Z coordinate is measured in "image pixels" like the X and Y coordinates and represents the distance relative to the plane of the subject's hips, which is the origin of the Z axis. Negative values are between the hips and the camera; positive values are behind the hips. Z coordinate scale is similar with X, Y scales but has different nature as obtained not via human annotation, by fitting synthetic data (GHUM model) to the 2D annotation. Note, that Z is not metric but up to scale.

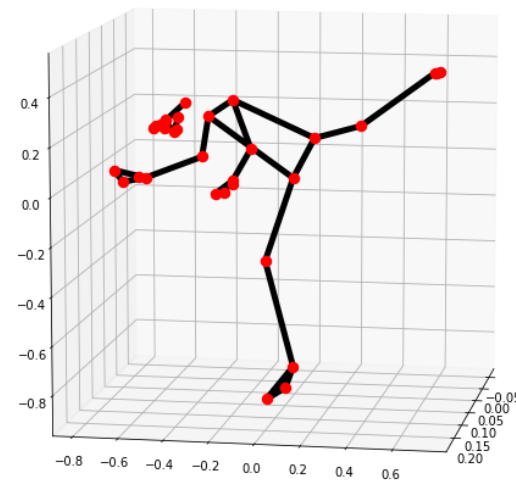
- Visibility is in the range of [min_float, max_float] and after user-applied sigmoid denotes the probability that a keypoint is located within the frame and not occluded by another bigger body part or another object.
- Presence is in the range of [min_float, max_float] and after user-applied sigmoid denotes the probability that a keypoint is located within

-<https://google.github.io/mediapipe/solutions/models#object-detection>

POSE OUTPUT FORMAT (MEDIAPIPE)



- | | |
|--------------------|----------------------|
| 0. nose | 17. left_pinky |
| 1. left_eye_inner | 18. right_pinky |
| 2. left_eye | 19. left_index |
| 3. left_eye_outer | 20. right_index |
| 4. right_eye_inner | 21. left_thumb |
| 5. right_eye | 22. right_thumb |
| 6. right_eye_outer | 23. left_hip |
| 7. left_ear | 24. right_hip |
| 8. right_ear | 25. left_knee |
| 9. mouth_left | 26. right_knee |
| 10. mouth_right | 27. left_ankle |
| 11. left_shoulder | 28. right_ankle |
| 12. right_shoulder | 29. left_heel |
| 13. left_elbow | 30. right_heel |
| 14. right_elbow | 31. left_foot_index |
| 15. left_wrist | 32. right_foot_index |
| 16. right_wrist | |



30.0 fps (Processing 0.9, latency 0.00ms, 1087.0 fps) 0 dropped



EDGE DETECTION

- Webcam

Fig 5. Example of MediaPipe Pose real-world 3D coordinates.



VIDEO EXAMPLE
WITH 3D
COORDINATE PLANE

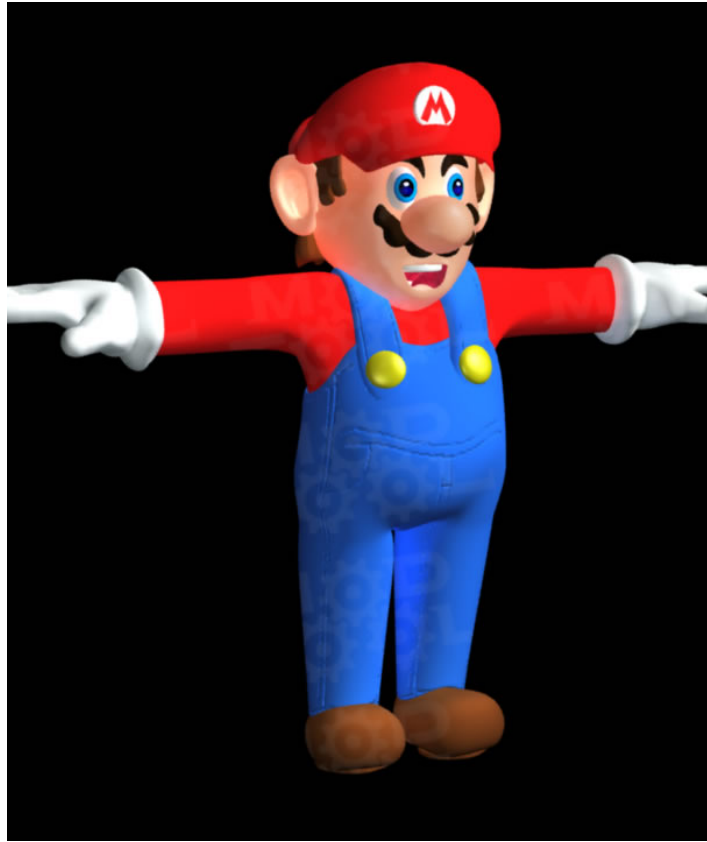
I.	PROBLEM STATEMENT
II.	WHAT IS COMPUTER VISION
III.	WHAT IS MEDIA PIPE?
IV.	EXERCISE EXAMPLE
V.	FUTURE CONSIDERATIONS

AIDED COMPUTER VISION YOGA EXAMPLE

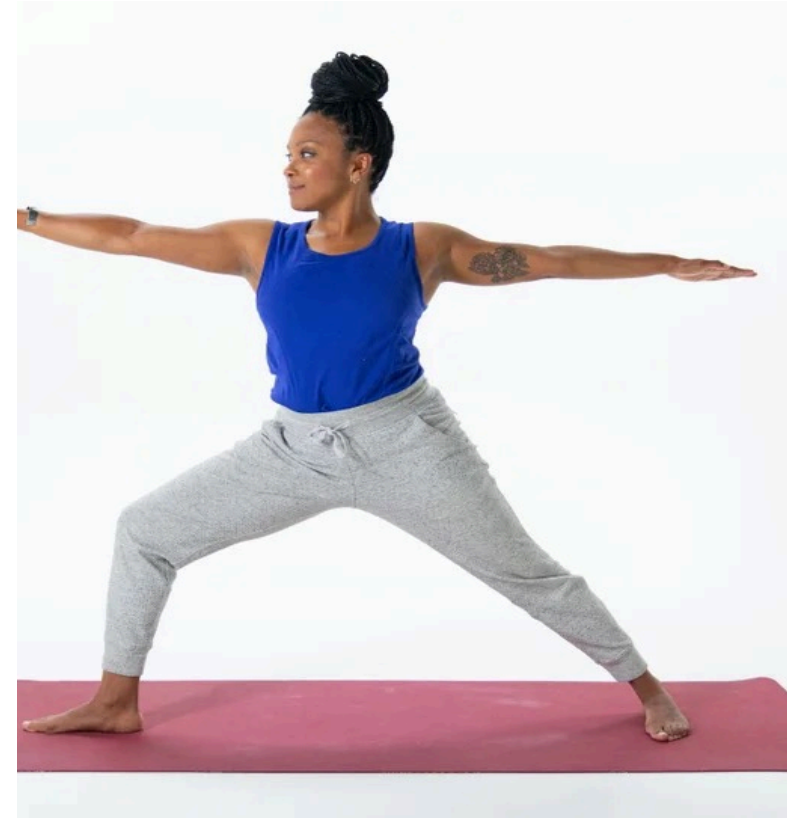
Tree Pose



T Pose (~yoga)

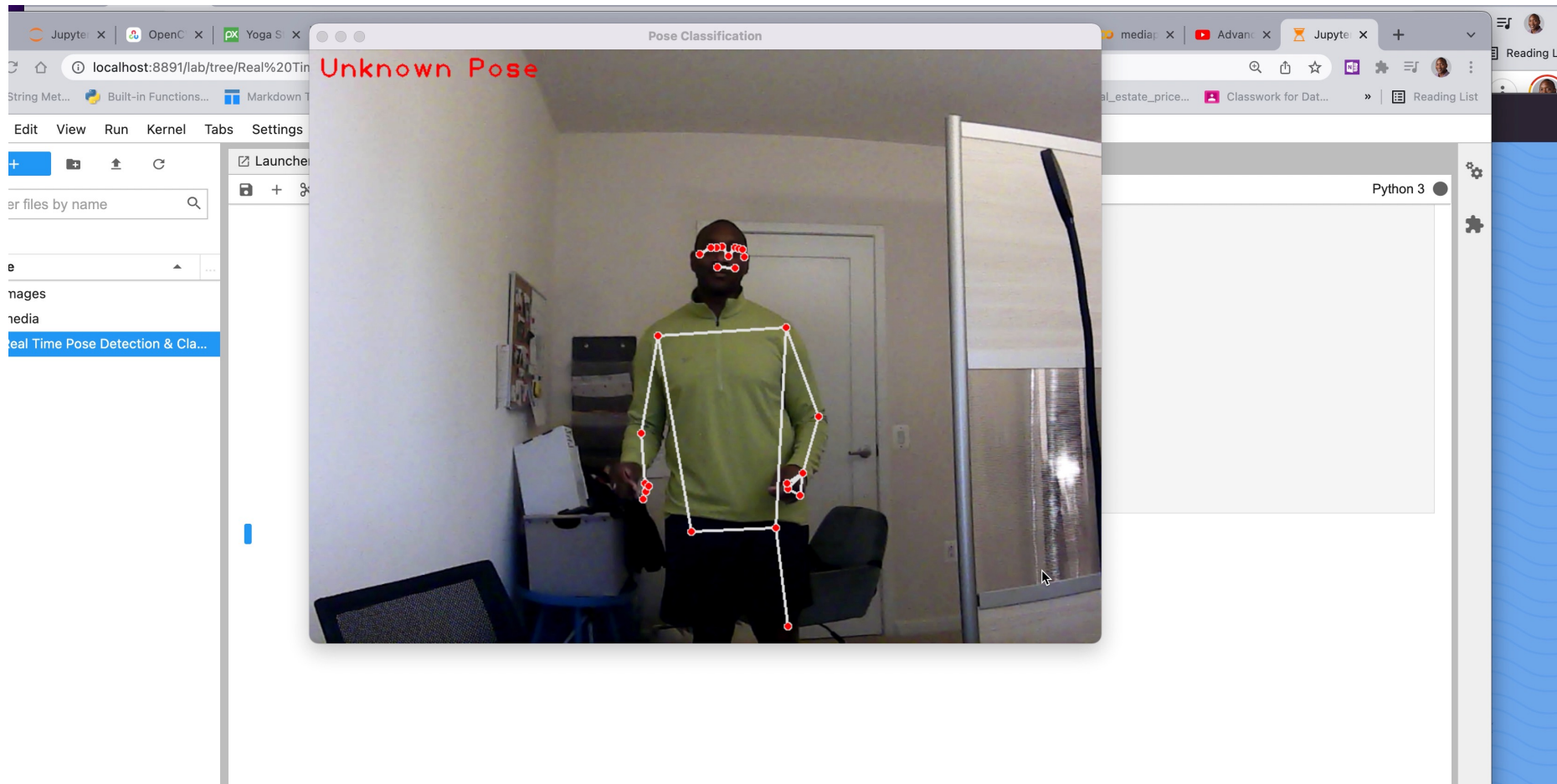


Warrior II Pose



ON TO THE NOTEBOOK!

IN CASE NOTEBOOK DOES NOT WORK, VIDEO OF POSES



- I. PROBLEM STATEMENT
- II. WHAT IS COMPUTER VISION
- III. WHAT IS MEDIA PIPE?
- IV. EXERCISE EXAMPLE
- V. **FUTURE CONSIDERATIONS**

CONCLUSION/RECOMMENDATION

IMPROVEMENTS

GAMIFICATION

- Uses verbal cues for correct or incorrect posture
- Keep a running score, add social component to

ADDITIONAL POSES

- Current heuristic approach is ok for 1-2 poses.
- Training 100+ images of several poses would be more efficient

VISUALIZER

[HTTPS://GOOGLE.GITHUB.IO/MEDIAPIPE/TOOLS/VISUALIZER.HTML](https://google.github.io/mediapipe/tools/visualizer.html)

THANK YOU!

ANY QUESTIONS?