MediaPipe Guide

By Behlul Vardal

Overview

MediaPipe is part of Google's AI Edge solution separate from google deepmind which can have edge solutions as well but its focused on AI Chats which is built by a different team. Looking through goggles Documentation there are 3 Sections: framework, Solutions, and API reference.

The framework section is a low-level component so you wont need to use it unless you're making your own version of the MediaPipe Solutions or want to understand how mediapipe solutions works on a deeper level. The framework has versions in C++, ios, android and python. Where the solutions have more variety of support python, web, ios, and android.

Mediapipe Studio which is part of MediaPipes Solutions, where you can see a demo of the solutions. However they don't show the code here directly, it's all front end. There is a way to find the example code, Mediapipe used codepen for these demos but not all demos are shown in the codepen, for example if you want to see the demo with the avatar at the time of writing this its not listed on all pens tab. You can still find them if you google for example Mediapipe Avatar demo codepen and the pen will appear even though it isn't listed in codepen. Google has also ended support for Legacy Solutions as of 2023.

Face Blendshapes and Landmarks

There are 478 face landmarks and 52 Face Blendshapes in Mediapipe Library:

0 - _neutral (ignore it)
1 - browDownLeft
2 - browDownRight
3 - browInnerUp
4 - browOuterUpLeft
5 - browOuterUpRight
6 - cheekPuff
7 - cheekSquintLeft
8 - cheekSquintRight
9 - eyeBlinkLeft
10 - eyeBlinkRight
11 - eyeLookDownLeftz
12 - eyeLookDownRight
13 - eyeLookInLeft
14 - eyeLookInRight
15 - eyeLookOutLeft
16 - eyeLookOutRight
17 - eyeLookUpLeft
18 - eyeLookUpRight
19 - eyeSquintLef
20 - eyeSquintRight
21 - eyeWideLeft
22 - eyeWideRight
23 - jawForward
24 - jawLeft
25 - jawOpen
26 - jawRight
27 - mouthClose
28 - mouthDimpleLeft
29 - mouthDimpleRight

30 - mouthFrownLeft
31 - mouthFrownRight
32 - mouthFunnel
33 - mouthLeft
34 - mouthLowerDownLeft
35 - mouthLowerDownRight
36 - mouthPressLeft
37 - mouthPressRight
38 - mouthPucker
39 - mouthRight
40 - mouthRollLower
41 - mouthRollUpper
42 - mouthShrugLower
43 - mouthShrugUpper
44 - mouthSmileLeft
45 - mouthSmileRight
46 - mouthStretchLeft
47 - mouthStretchRight
48 - mouthUpperUpLeft
49 - mouthUpperUpRight
50 - noseSneerLeft
51 - noseSneerRight

Limitations

If the camera's vision gets obstructed, for example you turn your head from the camera and parts of the face are not visible. Camera quality, resolution, and even the type of camera (like visible-light vs. infrared) can also impact how well a face is detected and tracked. This is why you either need multiple cameras at different angles or a single camera rigged to a person's face for acting or movements that would be more intricate.

Accuracy/Precision

When we are using webcams, we are also being limited to the precision of blendshapes. The reason why Apple is able to get more precise blendshape data with ARKit is because it's meant for using Apple's TrueDepth cameras that their products have.

- The TrueDepth camera creates a detailed 3D depth map by projecting over 30,000 invisible infrared dots onto your face.
- ARKit combines the depth map, infrared image, and the standard RGB image to build a comprehensive understanding of your face's geometry and expressions.

MediaPipe is using image analysis from the standard camera, this is what makes it more versatile. However you can still technically modify the MediaPipe framework for your own solutions. Although Google has created their own simple solution with MediaPipe, they've also documented their framework so that you can go in and create your own specific solution. They even have their own model maker so you can train your own ML model for real time images. This makes it so that if you had your own infrared cameras, and wanted to make a solution for yourself to get more accurate blendshapes, you could.