General Observation

Performance Benchmarks

Tests conducted on a typical system (MacBook M2 Pro, 16GB RAM, macOS 24.3.0):

- **CPU Usage**: 15-25% on average
- **Memory Usage**: ~600MB
- **Average FPS**: 20-25 FPS at 720p resolution
- **Startup Time**: ~2-3 seconds including initial calibration

Optimal Setup

- Position camera at eye level, facing directly forward
- Ensure uniform, diffuse lighting on your face
- Avoid backgrounds with moving elements=
- Maintain 0.5-1.0 meter distance from camera

Calibration

The system requires an initial calibration period (5 seconds):

- Maintain a neutral expression
- Look directly at the camera
- Keep shoulders level and posture upright
- Remain still during the calibration process

Choice of Library/ Model

1. Mediapipe was chosen due to its simplicity and integration ease when using it for both face and body key-points. Other methods tried :
    1. YOLO : better accuracy, hard to integrate due to different landmark format
    2. RMTw : whole body-point by one module, hard to integrate, need optimisation, heavy overload
2. DISOpticalFlow (Fast) : due to its ease of integration and fast inference in FAST Mode
    1. Farnback : less accuracy, similar overhead
    2. LucasKanad : Less accuracy, similar overhead
    3. Deep learning : Very heavy computation

Thought Process

3. For emotion detection, I tried collecting data and training model but training on images have limitation like generalisation and also overhead in processing due to the sequential nature. In this regard i found that simple rules approach is faster in inference but gets limited due to some bottlenecks, which, if targeted will make it useful and later can be paired with deep learning methods. Now it's accuracy is limited by two bottlenecks :
    1. Less accurate feature detection : improving frown and smile detection by increasing complexity in rules
    2. very small micro-expression : hard to capture but can be mitigated by fusion of conditions like smile with frown with head condition
4. Hunch : Hunch detection using pose is hampered by accuracy of pose models because minor changes in shoulder is not picked up well by pose models and creating a baseline is very difficult because of one condition :
    1. HUMANS KEEP Moving, so if i keep a baseline height of the shoulder and then compare based on that, it will not be logical
    2. SOLUTION : keep the baseline as something that involves shoulder position in upright position with something on the body that moves relative to the shoulder or doesn't move during hunching.....(CHEST)
    3. so during calibration : we take the mid shoulder point and find a point roughly 1/4 distance below this point and take the distance between them as baseline and track this point using optical flow (SOLVED)
5. Face detection : we can mitigate the distance from the screen by always normalising the face keypoints with the face size to keep the ratios the same across different user distance from screen
6. LIMITATIONS :
    1. Hunch : user can move out of frame where optical flow fails
    2. Emotion : very limited rules basis
    3. Blinks : blinks have been consistent with various conditions, only fails when the user looks to the ceiling or to the ground
    4. Other pipeline : integrating other models for pose will become difficult due to their varied output format
    5. Lighting : work fairly well in all lights but depend on the model as it produces the keypoints