AI Expression to Avatar System

Jorge Mejia Leon Do Ryan Hernandez Group 5

The Problem Space

High Barriers

Professional tools like Animaze require paid subscriptions, expensive motion capture gear, or complex technical rigging knowledge.

Lack of Expression

Free alternatives (e.g., Zoom avatars) are static and fail to capture genuine human emotion, limiting connection with audiences.

Privacy Needs

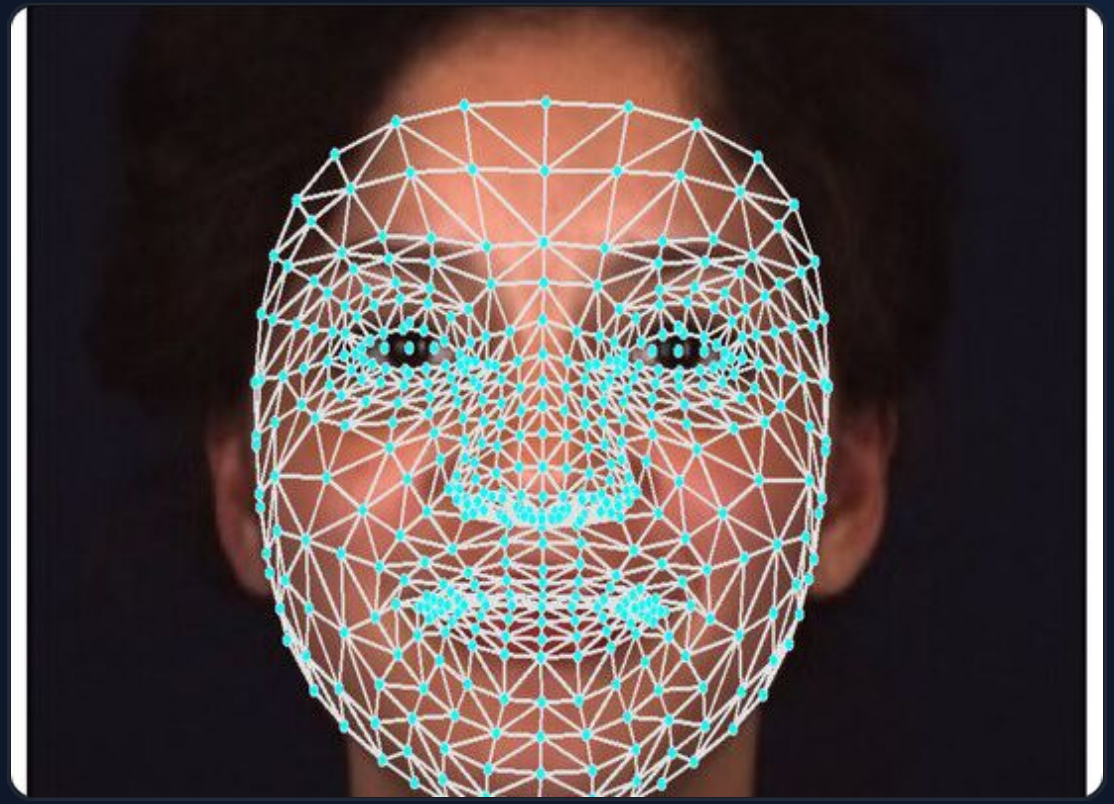
Creators want to express themselves authentically and engage audiences without revealing their physical identity or location.

Our Solution

Accessible & Expressive

We developed a lightweight, open-source desktop application that uses standard webcams to drive high-fidelity avatars in real-time.

- **Zero Cost:** Free and open-source alternative.
- **No Hardware:** Works with standard webcams.
- **Privacy First:** All processing is local; no video uploads.
- **Seamless:** Integration with OBS, Twitch, and Zoom.



Technical Stack

Computer Vision

MediaPipe Face Mesh: extracts 468 3D facial landmarks at 30 FPS to capture precise geometry (eyes, mouth, head pose).

Kalman Filters: Signal processing algorithm applied to smooth raw landmark data, eliminating "jitter" common in webcam tracking.

Deep Learning

DeepFace: A Convolutional Neural Network (CNN) that classifies 7 emotional states (Happy, Sad, Angry, etc.) at 15 Hz.

Neutral Calibration: Captures user's resting face baseline at startup to adapt the model to individual facial structures.

End-to-End Pipeline

Input

Webcam Video Feed

Processing

Fusion Layer
(Landmarks + Emotion)

Output

VTube Studio Avatar

Fusion & Calibration

1. Neutral Calibration

Users sit still for a configured amount of time, 2 seconds by default, upon clicking calibration button. The system calculates a baseline vector for their "neutral" face. This vector is subtracted from real-time frames to prevent resting faces from being misclassified as "Sad" or "Angry".

2. Parameter Fusion

We combine continuous signals (Eye Aspect Ratio for blinking, Mouth Aspect Ratio for speaking) with categorical emotion labels.

Data is transmitted via **WebSockets** to VTube Studio, controlling Live2D parameters like EyeOpenLeft and MouthOpen depending on the defined params in the config file.

Performance Metrics

Business Applicability

"The Colab of Avatars"

Free base tool for independent creators drives massive adoption and ecosystem growth.

Marketplace

Community store for developers to sell assets (props, outfits), creating a self-sustaining economy.

Enterprise Cloud

Premium tier for agencies (e.g., VShojo) offering cloud-hosted processing and dedicated support.

Future Roadmap

Phase 4 (Current)

Real-time landmark tracking, basic emotion mapping, VTube Studio integration.

Near Term

Dynamic background calibration (move while calibrating) and "Fear" expression support.

Long Term

Cloud offloading for low-end devices and full WebGL browser-based version.

"We have proven that high-quality digital expression does not require expensive hardware - just a webcam and smart AI."

- Group 5

DEMO

LIVE PROTOTYPE WALKTHROUGH