Temporal Face Analysis for Intoxication Detection

Project Proposal

January 15, 2025

Contents

1 Introduction

This project aims to develop a system for detecting intoxication in individuals using temporal face analysis. The system will monitor individuals as they approach a camera, leveraging facial features and temporal patterns to identify signs of intoxication. This solution is designed for high-security environments, such as airports, where effective monitoring is crucial.

1.1 Objective

To design and implement a robust pipeline that uses temporal face analysis to classify individuals as intoxicated or non-intoxicated, based on video sequences.

1.2 Challenges

- Varying facial resolutions as individuals approach the camera.
- Subtle temporal changes in facial features indicative of intoxication.
- Differentiating intoxication from natural facial behavior variations.

2 Pipeline Overview

The solution consists of the following stages:

1. Face Detection and Tracking.
2. Preprocessing.
3. Feature Extraction.
4. Temporal Modeling.
5. Model Training and Evaluation.

3 Tasks

3.1 Task 1: Face Detection and Tracking

- Use YOLOv11 or RetinaFace for face detection.
- Align faces using facial landmarks to normalize for pose variations.
- Organize face crops as sequences based on video frame continuity.

3.2 Task 2: Preprocessing

- Normalize face resolution to a consistent size while preserving aspect ratio.
- Apply padding to avoid distortion.
- Group consecutive frames into sequences for temporal analysis (e.g., 30 frames per sequence).

3.3 Task 3: Feature Extraction

- Extract facial embeddings using pre-trained models (e.g., FaceNet or VGGFace).
- Detect facial landmarks for analyzing:
  - Eye openness (blink rate).
  - Mouth movements (smirking, yawning).
  - Head orientation (tilts or nods).
- Use emotion detection models (e.g., FER-2013, AffectNet) to generate emotion probabilities per frame.
- Combine embeddings, landmarks, and emotion features into a unified feature vector.

3.4 Task 4: Temporal Modeling

- Design a sequence model to process temporal data:
  - Use LSTMs or GRUs for short-term temporal dependencies.
  - Use Transformers for capturing long-term dependencies.
- Aggregate predictions across short sequences using:
  - Majority voting.
  - Weighted averaging of prediction confidence.

3.5 Task 5: Model Training and Evaluation

- Collect and preprocess a labeled dataset of intoxicated and non-intoxicated individuals.
- Split data into training, validation, and test sets.
- Train the temporal model using a sequence of extracted features.
- Evaluate using metrics:
  - Accuracy, precision, recall, and F1-score.
  - AUC-ROC for imbalanced datasets.

3.6 Task 6: Deployment

- Optimize the model for real-time processing using quantization.
- Deploy on edge devices like NVIDIA Jetson for real-time video analysis.
- Integrate with security monitoring systems to generate alerts for intoxicated individuals.

4 Optional Enhancements

- **Gaze Estimation**: Integrate gaze tracking to detect wandering or unfocused eyes.
- **Multi-Face Tracking**: Extend to monitor multiple individuals in a frame.
- **Multimodal Analysis**: Combine facial analysis with gait analysis for higher accuracy.

5 Tools and Frameworks

- **Face Detection**: YOLOv11, RetinaFace.
- **Landmark Detection**: OpenFace, Dlib.
- **Temporal Models**: TensorFlow, PyTorch.
- **Visualization**: OpenCV for overlaying results on video frames.

6 Conclusion

This project leverages temporal face analysis to identify intoxication through dynamic facial patterns. By integrating advanced computer vision techniques with temporal modeling, the system aims to provide robust and accurate results in real-world security contexts, such as airports.