Coffee Mug Detection & Tracking – Solution Document

Problem Statement

To design and implement a solution that can detect and track coffee mugs in a given video using computer vision techniques. The system should identify mug objects, assign unique IDs, and track them through frames to create an annotated output video.

Approach and Tool Selection

1. Object Detection with YOLOv5 (Ultralytics):

We used a pre-trained YOLOv5s model for object detection.

This model is lightweight, fast, and accurate for real-time object detection tasks.

YOLOv5 identifies objects per frame and classifies them. We focused specifically on the "cup" label as a proxy for coffee mugs.

2. Tracking with OpenCV MultiTracker + CSRT:

OpenCV's MultiTracker API allows tracking multiple objects simultaneously.

We chose the CSRT (Discriminative Correlation Filter with Channel and Spatial Reliability) tracker due to its robustness in object tracking compared to KCF or MIL.

Tracking is initialized only in the first frame to keep it lightweight and suitable for real-time performance.

3. Video Handling and Output:

The video is read frame-by-frame using cv2.VideoCapture, and the annotated video is saved using cv2.VideoWriter.

A unique ID is assigned to each tracked mug for visual traceability.

4. Development Environment:

Entire project implemented in Google Colab for ease of access, GPU support, and cloud-based file management.

Dependencies are installed using pip (ultralytics, opencv-python-headless).

How We Know the Solution is Good

Correct Detection: The YOLOv5 model successfully detects "cup" objects in the video.

Tracking Consistency: Once initialized, the tracker follows the mugs through the video with stable bounding boxes and correct IDs.

Visual Output: The output video has clearly annotated bounding boxes and IDs for up to two coffee mugs, showing successful end-to-end processing.

Efficiency: The solution runs efficiently even on Colab CPU/GPU instances, processing the video within reasonable time.

Potential Improvements

Dynamic Detection: Currently, detection runs only on the first frame. Periodic detection (e.g., every 30 frames) would improve robustness in case mugs enter or leave the scene.

Use Deep SORT: Integrating a deep-learning-based tracking algorithm like Deep SORT could provide more accurate ID tracking and re-identification.

Fine-tune Detection: Train or fine-tune YOLOv5 on a dataset containing only coffee mugs for more specialized detection performance.

Add Tracking Loss Recovery: Implement logic to reinitialize tracking if confidence drops or if the tracker fails mid-way.

GUI/Stream Support: Create a simple UI or extend support for live webcam or RTSP stream inputs.