**Technical Documentation: Robust Multi-Video Person Tracking and Re-Identification**

**Overview**

This system tracks people across multiple shop surveillance videos, assigns consistent global IDs to the same individuals, and outputs annotated videos and a CSV report. It leverages YOLOv8 for detection, Deep SORT for short-term tracking, and a robust ReID pipeline combining body and face embeddings for global identity assignment.

**⚙️ Components**

**1. Person Detection**

* **Model:** YOLOv8m (yolov8m.pt)
* **Purpose:** Detects all person instances (class=0) in each video frame.

**2. Tracking**

* **Algorithm:** Deep SORT
* **Parameters Tuned For Shops:**
  + max\_age=20, n\_init=3
* **Purpose:** Associates detected bounding boxes frame-to-frame using Kalman filtering and appearance-based matching.

**3. Re-Identification (ReID)**

* **Model:** OSNet x1\_0 from TorchReID
* **Embedding Type:** Combined Body + Face embedding (1024-D)
* **Face Embedding Model:** InsightFace (buffalo\_l)
* **Logic:**
  + Crop the person’s body and detect the largest face within.
  + Extract:
    - **Body embedding** (512-D)
    - **Face embedding** (512-D)
  + Combine and normalize to form a single 1024-D embedding.
  + Compare cosine similarity with stored embeddings.
  + If similarity ≥ 0.75, assign existing global ID; otherwise, assign a new one.

**4. Persistent Global Identity Mapping**

* Maintains a mapping of DeepSORT track\_id ➝ Global ID
* Global IDs are reused across videos using consistent embedding matching logic.

**5. Optional Cross-Video Clustering (Commented)**

* Implements DBSCAN over all embeddings to refine and consolidate global IDs further.

**6. Output**

* **CSV File:** Contains global IDs, video name, frame index, and face bounding boxes.
* **Annotated Videos:** With face boxes and global ID overlays.

**Comparison: Current vs Previous Implementation**

| **Feature / Aspect** | **Previous Implementation** | **Current Implementation (Robust)** |
| --- | --- | --- |
| **Detection Model** | YOLOv8n (lightweight) | YOLOv8m (more accurate, robust to occlusions) |
| **Tracking** | Deep SORT (default params) | Deep SORT (tuned for longer life, delayed init) |
| **ReID Features** | Only body embedding (OSNet) | Concatenated Body + Face embedding (1024-D) |
| **Face Embedding** | Not used | InsightFace (buffalo\_l) integrated |
| **Embedding Similarity Threshold** | 0.8 (possibly stricter) | 0.75 (balanced for real-world variance) |
| **Persistent Global ID Mapping** | Per video | Cross-video memory using trackid\_to\_globalid |
| **Clustering / Cross-Video Merging** | Manual ID assignment or visual inspection | Optional DBSCAN-based cluster merging (more scalable) |
| **Bounding Box Annotation** | Person box only | Face box with ID (better visual clarity) |
| **Face Detection for Alignment** | None | YOLOv8n-face model for accurate facial region cropping |

**Key Improvements**

* **Multi-modal ReID:** Fusion of body and face cues makes ID assignment much more reliable under pose and angle changes.
* **Robust Similarity Logic:** Cosine-based matching over normalized vectors enhances robustness.
* **Cross-Video ID Tracking:** Ensures consistent identity even if a person appears in multiple videos with different appearances.
* **Face Overlay Instead of Full Body:** Easier human interpretability when validating results manually.
* **Scalability:** Structure is modular and scalable; DBSCAN-based clustering can be turned on for even better accuracy in large-scale setups.

**Recommendations**

* If runtime becomes a bottleneck, consider batching embedding inference for body crops.
* Enable the merge\_ids\_across\_videos() function when deploying for long video sets to improve cross-video identity accuracy.
* Consider saving per-frame cropped face + body images for further downstream training or analysis.