
Calibration-Free Passenger Re-ID in Mixed-Modality, Crowded East African Buses

1 Public transportation in sub-Saharan Africa (SSA) primarily relies on buses and matatus, often
2 operating above capacity. Accurate passenger flow data is essential for effective scheduling, route
3 optimization, crowd management, and transitioning to electric fleets. Reliable Origin–Destination
4 (OD) matrices are also crucial for equity analysis, identifying service gaps, and evaluating the
5 impact of improvements. However, the lack of such data is a significant challenge, as manual
6 surveys are costly and automated counting systems struggle in crowded environments. Consequently,
7 transportation agencies find it difficult to obtain reliable OD information for effective long-term
8 planning.

9 The recent transition from diesel to electric fleets has standardized onboard CCTV, enhancing the
10 feasibility of automated OD data collection. To explore this opportunity, we have partnered with an
11 electric bus operator in Nairobi and Kigali. The electric buses generally have two to three uncalibrated
12 interior cameras: **Cam-A**, located at the front and facing down the aisle (capturing boarding and initial
13 movements); **Cam-B** positioned above the exit door and facing the doorway (capturing alightings
14 and occasional rear-door boardings); and, in some instances, **Cam-C**, placed at the rear and facing
15 forward (capturing extended aisle trajectories). Our primary objective is to establish a connection
16 from each rider’s front-door tracklet to its continuations within the bus cabin, when available, and to
17 an exit-door tracklet, despite challenges such as heavy occlusion, background clutter, and low-quality,
18 mixed-modality CCTV that alternates between colour and monochrome.

19 Our baseline evaluation employed a pipeline that integrates YOLOv12 (fine-tuned on our dataset)
20 for detection, BoT-SORT for tracking, and OSNet (also fine-tuned) for cross-camera ReID, utilizing
21 CCTV footage collected from Nairobi and Kigali. This pipeline uses a region of interest (ROI) at
22 entry and exit doors to accurately count boardings and alightings at each stop, vital for creating OD
23 matrices. Under uncrowded conditions (Fig. 1, left), the performance is satisfactory due to full-body
24 visibility for reliable cross camera matching. In contrast, during crowded conditions (Fig. 1, right),
25 particularly during peak hours, the presence of standing passengers in the aisles and closely seated
26 individuals restricts visibility, providing only partial views, such as heads or shoulders. This heavy
27 occlusion affects ReID accuracy and, consequently, the reliability of OD estimation.

28 The onboard CCTV system alternates between low-resolution monochrome (such as IR night vision)
29 and color streams, which makes OD estimation more challenging. This variation removes key
30 features, such as color and full-body visibility, that conventional ReID models rely on, resulting in
31 reduced performance in crowded settings. Thus, there is a critical need for a solution optimized for
32 crowded, head-dominant perspectives in mixed-modality CCTV streams.



Figure 1: Baseline performance comparison. ReID succeeds when passengers are fully visible (left) but collapses under crowding (right).

33 To address these limitations, we are developing a calibration-free cascade approach that reinterprets
34 each cross-camera link as a tracklet-level assignment problem. This solution integrates motion-aware
35 gating (considering time windows, ROI overlap, and heading), dual-branch embeddings (combining
36 head and body crops) that are robust to mixed-modality CCTV, and height-rank cues. This is followed
37 by a Hungarian assignment with deferred Bayesian filtering. The design is intended to avoid metric
38 calibration, scale across varied bus layouts, and maintain robustness in the face of severe occlusion
39 and mixed-modality streams.