Distributed Fall Detection via Temporal Signatures and Overlap-Aware Bidding in Multi-Camera Edge Systems

Jun Young Im[1], **Seon Woo Ok**[1] and **Moon Gi Seok**[1]
[1]Department of Computer Engineering, Dongguk University, Seoul, South Korea
{junyoung.im, seonwoo.ok, mgseok}@dgu.ac.kr

Abstract

The proposed system performs **distributed fall detection** using multiple RGB-D cameras operating on edge devices. Each camera independently tracks people within its own view and collaborates only when a person enters an overlapping area shared with other cameras. When overlap occurs, the camera sends lightweight metadata about the detected person to a coordinating **Auctioneer**, which determines which device should handle that person based on spatial and visual information. This architecture minimizes redundant computation and bandwidth usage, while maintaining **real-time coordination**, **scalability** by keeping all image data processed locally on each device.

Keywords: Fall detection, Multi-camera systems, Temporal signatures, Edge computing

1 Introduction

Multi-camera fall detection must maintain real-time response under occlusion and coverage gaps. Centralized fusion scales poorly, creating bandwidth and latency bottlenecks, while global re-identification (re-ID) is fragile to illumination and viewpoint changes and requires heavy cross-camera features. Recent distributed trackers such as DMMA [1], which performs decentralized multi-target association across cameras, and Uni-ID [2], which assigns globally consistent IDs for cross-camera re-identification, move computation to the edge but still rely on appearance cues or continuous consensus. Motion-centric methods like MAT [3], which associates targets via trajectory predictions, and UCMCTrack [4], which compensates for camera motion to maintain consistency, reduce appearance dependence yet incur high per-node cost or implicit centralization. Field of view (FoV) overlap is often estimated online through trajectory or homography cues [5], [6], which are sensitive to lighting and require continuous monitoring.

Privacy-aware systems restrict transmission to metadata [7], [8], but either operate independently—wasting computation—or keep always-on negotiation. In real facilities, lighting variation, moving objects, and varying scales frequently disrupt re-ID chains, while edge devices face strict power limits and cannot afford redundant detection. Existing paradigms solve parts of the problem but not the whole: centralized fusion causes bottlenecks, appearance-based methods degrade with viewpoint shifts, and motion-only schemes require dense communication.

We take a different route. The system precomputes 3D frustum intersections and activates collaboration **only when a person enters an overlapping region**. Each camera performs local tracking in a **2D-first manner**, using image-plane motion and confidence cues for stable association.

Within each overlap group, devices communicate through an **Auctioneer**, which coordinates identity matching and task assignment. Instead of sharing images or deep features, each node sends only lightweight metadata—such as center of mass (CoM), its velocity, and quality of view (QoV) score—minimizing bandwidth and preserving privacy while maintaining global consistency.

Contributions. (1) *Temporal Signatures* for robust local identity without global re-ID; (2) a *visibility-adaptive matching* scheme coupling 3D gating and normalized 2D fallback in a logistic–Hungarian objective with soft/sticky terms; (3) an *overlap-triggered, single-device execution* policy via lightweight bidding that preserves accuracy while eliminating redundant computation.

Organization. The remainder of the paper is organized as follows: Sec. II details the proposed

method, Sec. III presents experiments and ablations, and Sec. IV concludes the paper.

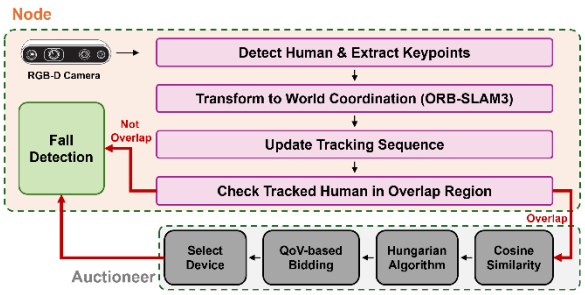

Figure 1. **Overall architecture of the proposed Auctioneer–Bidder system for multi-camera fall detection.**

2 Method

A. System Architecture

Overview. As illustrated in Fig. 1, each RGB-D node independently tracks humans within its FoV and estimates its own camera pose using ORB-SLAM3[10] in real time.

This allows every node to maintain a consistent local map and spatial alignment with other devices in the overlapping regions.

The estimated camera poses are used to transform keypoint trajectories into a shared world coordinate frame, enabling cross-view association under geometry consistency.

Upon detecting overlap, the node transmits lightweight metadata to the corresponding Auctioneer, which batches incoming messages by timestamp and performs two-stage processing: identity matching using cosine similarity and QoV–weighted assignment via the Hungarian algorithm [12]. This ensures one-node-per-person coordination with minimal bandwidth and full on-device privacy.

Definition of Temporal Signature. We define the Temporal Signature (TS) as a compact, time-evolving descriptor that characterizes a person's identity and motion across discrete timesteps:

$$TS_t = [\boldsymbol{r_t}, \boldsymbol{v_t}, \Phi_t^{IA}].$$

each TS is updated at every fixed timestep Δt. The term $\boldsymbol{r_t}$ denotes the center of mass (CoM) of the person in the world coordinate system, while its projected image center $\boldsymbol{c_t}$ (in pixels) is computed via the pinhole projection: $\boldsymbol{c_t} = \pi_{\mathbf{K_c},\mathbf{T_c}}(\boldsymbol{r_t})$, where $\pi_{\mathbf{K_c},\mathbf{T_c}}$ is defined by the camera intrinsics $\mathbf{K_c}$ and extrinsics $\mathbf{T_c}$. All image points are assumed undistorted after lens calibration. The projection Jacobian $\mathbf{J}_\pi = \partial\pi_{\mathbf{K_c},\mathbf{T_c}}(\mathbf{r_t})/\partial\mathbf{r_t}$ quantifies the local mapping between 3D position perturbations and their 2D image displacements. The velocity term $\boldsymbol{v_t}$ is computed as the finite difference of consecutive CoM estimates: $\boldsymbol{v_t} = \frac{\boldsymbol{r_t} - \boldsymbol{r_{t-1}}}{\Delta t}$. Finally, the depth-derived body-ratio anchor Φ_t^{IA} is obtained as a temporal average of normalized body-ratio descriptors over 3–5 recent timesteps, providing a long-term, view-invariant reference for person identity.

B. Local Tracking and Association (2D Projection Only)

a) Detector, pixel measurement, and 3D lifting: We use *YOLOv8n-pose* [9] (TensorRT-FP16) to obtain *KP* keypoints $\{(u_i, v_i, \text{conf}_i)\}_{i=1}^{KP}$. where (u_i, v_i) denote the 2D pixel coordinates of each joint and conf_i indicates its detection confidence.We follow the COCO keypoint convention ($KP = 17$). From these keypoints, we derive two complementary measurements: (A) a 2D pixel observation for tracking, and (B) a 3D body-ratio anchor for IA.

2D pixel measurement: the image-center $c_d \in \mathbb{R}^2$ is a confidence-trimmed average of torso-related keypoints with mask $[\,\text{conf}_i > 0.3\,]$; if fewer than three are valid, fall back to the detector box center. Pixel noise is

$$R_p = \sigma_{px}^2 I_2,$$

$$\sigma_{px}^2 = \text{EWM}\big(\sigma^2(\text{conf}_i)\big),$$

$$\sigma(\text{conf}) = a + b\,(1 - \text{conf}).$$

where EWM denotes an exponentially weighted median over recent frames to stabilize per-frame noise estimates. Defaults are $a = 0.5$ px and $b = 3.0$ px (refittable per site).

3D lifting for IA (only): with aligned depth d_i, the corresponding 3D camera-frame position is reconstructed as

$$\tilde{\mathbf{p}}_{\text{cam}}^{(i)} = d_i\,\mathbf{K_c}^{-1}\begin{bmatrix} u_i \\ v_i \\ 1 \end{bmatrix}.$$

From valid joints $\mathcal{J}$, compute height h and scale-invariant 3D ratios to form Φ^{IA}; standardized per device using Welford's online algorithm [11] for mean and variance estimation, which allows stable, incremental normalization of depth-derived features without storing past samples.

b) Kalman *State and prediction (CV with standard process noise):* The temporal components of the signature, $\mathbf{r_t}$ and $\mathbf{v_t}$ are modeled as a Kalman state $x_t = [r_t^\top, v_t^\top]^\top$. At each timestep Δt, the state is propagated using a constant-velocity (CV) motion model:

$$\mathbf{x}_{t|t-1} = \mathbf{F}\,\mathbf{x}_{t-1|t-1},$$

$$\mathbf{P}_{t|t-1} = \mathbf{F}\,\mathbf{P}_{t-1|t-1}\mathbf{F}^{\top} + \mathbf{Q}.$$

The state transition and process noise matrices are given by:

$$\mathbf{F} = \begin{bmatrix} \mathbf{I} & \Delta t\,\mathbf{I} \\ \mathbf{0} & \mathbf{I} \end{bmatrix}, \qquad \mathbf{Q} = \sigma_a^2 \begin{bmatrix} \frac{\Delta t^4}{4}\mathbf{I} & \frac{\Delta t^3}{2}\mathbf{I} \\ \frac{\Delta t^3}{2}\mathbf{I} & \Delta t^2\mathbf{I} \end{bmatrix},$$

The parameter $\sigma_a = 0.5\ \mathrm{m/s^2}$ denotes the standard deviation of process acceleration noise.

c) 2D measurement model and innovation: Let $\hat{\mathbf{f}}_{t|t-1}$ be the predicted CoM at time t. Its projection onto the image plane is obtained via the pinhole model: $\hat{\mathbf{c}}_{t|t-1} = \pi_{\mathbf{K}_c,\mathbf{T}_c}\left(\hat{\mathbf{f}}_{t|t-1}\right)$. The Jacobian of this projection, evaluated at the predicted CoM:

$$\mathbf{H}_{\pi} \equiv \mathbf{J}_{\pi}|_{\hat{\mathbf{f}}_{t|t-1}}.$$

maps infinitesimal 3D perturbations to the 2D image plane. The innovation covariance for the 2D pixel measurement is then given by $\mathbf{S}_{2D} = \mathbf{H}_{\pi}\,\mathbf{P}_{t|t-1}^{(r)}\,\mathbf{H}_{\pi}^{\top} + \mathbf{R}_{\mathrm{p}}$. Finally, the Mahalanobis innovation distance between the measured and predicted 2D centers is computed as:

$$z_m = \sqrt{(\mathbf{c}_d - \hat{\mathbf{c}}_{t|t-1})^{\top}\mathbf{S}_{2D}^{-1}(\mathbf{c}_d - \hat{\mathbf{c}}_{t|t-1})}.$$

d) Long-term 3D anchor residual and 2D-only consistency: To combine short-term 2D consistency($\boldsymbol{r_t}, \boldsymbol{v_t}$) with long-term 3D appearance(Φ_t^{IA}) stability, we define a composite residual between the detection anchor $\Phi_{t,d}^{\mathrm{IA}}$ and the track's running anchor $\Phi_{t,k}^{\mathrm{IA}}$.

The 3D anchor residual is measured as the cosine dissimilarity $z_s = 1 - \cos\langle \Phi_{t,d}^{\mathrm{IA}},\ \Phi_{t,k}^{\mathrm{IA}} \rangle \in [0,1]$. The overall 2D–3D consistency score is then formed as a visibility-weighted blend $z_{\mathrm{cons}} = \alpha(v)\,z_m + (1 - \alpha(v))\,z_s$, where z_m is the Mahalanobis distance of the 2D pixel residual, and the weight $\alpha(v)$ increases with visibility v: $\alpha(v) = \alpha_{\min} + (\alpha_{\max} - \alpha_{\min})\,v^{\kappa}$, defaults $\alpha_{\min} = 0.35$, $\alpha_{\max} = 0.85$, $\kappa = 1.5$.

Missing-IA guard: if fewer than three anchor joints in $\mathcal{J}$ are valid, hold Φ_k^{IA} and down-weight z_s via $\alpha'(v) = \min(1, \alpha(v) + \Delta_{\mathrm{hold}})$ (default $\Delta_{\mathrm{hold}} = 0.15$).

e) Gate (self-calibrating quantile): For robust association across varying visibility levels, we employ an adaptive gating threshold derived from the empirical distribution of consistency scores. Visibility v is discretized into three bands:

$$b \in \{[0,0.3), [0.3,0.7), [0.7,1]\}.$$

For each band, a target false-reject rate $\alpha_{\mathrm{FRR}}(b)$ defines the quantile threshold:

$$\delta_{\mathrm{gate}}(v) \coloneqq \mathrm{Quantile}_{\,1-\alpha_{\mathrm{FRR}}(b)}(z_{\mathrm{cons}}),$$

The quantiles are estimated online with t-digest or P^2 over 30–60 s window. During Cold-start, until at least $M_b \geq 200$ samples are collected per band, a pooled 95th percentile is. Used as a fallback threshold. A detection-track pair is accepted if $z_{\mathrm{cons}} < \delta_{\mathrm{gate}}(v)$, and otherwise rejected or deferred to probabilistic matching.

f) Kalman update (2D-only, Joseph form): The predicted state $x_{t|t-1}$ is corrected using the 2D pixel measurement cd. The Kalman gain is obtained as $\mathbf{K}_t = \mathbf{P}_{t|t-1}\mathbf{H}_{\pi}^{\top}\mathbf{S}_{2D}^{-1}$. The corrected state and covariance are updated as:

$$\mathbf{x}_{t|t} = \mathbf{x}_{t|t-1} + \mathbf{K}_t(\mathbf{c}_d - \hat{\mathbf{c}}_{t|t-1}),$$

$$\mathbf{P}_{t|t} = (\mathbf{I} - \mathbf{K}_t\mathbf{H}_{\pi})\mathbf{P}_{t|t-1}(\mathbf{I} - \mathbf{K}_t\mathbf{H}_{\pi})^{\top} + \mathbf{K}_t\mathbf{R}_{\mathrm{p}}\mathbf{K}_t^{\top}.$$

The covariance is updated in Joseph form.

g) From consistency to calibrated assignment cost: The combined features, $\mathbf{f} = [\,z_m,\ z_s,\ v,\ V,$ age] are mapped to a match probability $P(\mathrm{match}) = \sigma(\mathbf{w}^{\top}\psi(\mathbf{f}))$, where $\psi(\cdot)$ denotes feature normalization and interaction expansion. The corresponding cost is defined as $C = -\ln P(\mathrm{match})$ and all gated detection–track pairs are arranged into a global negative log-likelihood (NLL) cost matrix. The final data association is obtained by minimizing this matrix via the Hungarian algorithm, with default penalties are $C_{\mathrm{new}} = -\ln 0.1$ for new track creation and $C_{\mathrm{miss}} = -\ln 0.3$ for missed detections.

h) Spawning and state transitions: Duplicate births are avoided by deferring any unmatched detection within 0.35 m of an existing tentative track.

Each track follows a lightweight finite-state machine: Tentative → Confirmed after K=3 consecutive matches (≈0.10 s @ 30 fps); Confirmed → Deleted after M=10 consecutive misses (≈0.33 s); and Unmatched → Tentative if no nearby confirmed track exists. These rules ensure temporal stability while allowing rapid adaptation to new entries.

C. Coordination in Overlap Groups

When to coordinate. Each FoV $\mathcal{F}_{\mathrm{i}} = \{\chi : A_i\chi \leq b_i\}$ defines a 3D frustum in world space. For each camera i, the set of overlapping regions with other cameras is

$$\mathcal{O}_i = \{\mathcal{F}_i \cap \mathcal{F}_j \mid j \neq i,\ \mathcal{F}_i \cap \mathcal{F}_j \neq \emptyset\}.$$

All Overlap $\mathcal{O}_i$ are precomputed and cached. Coordination is activated only when a person's lifted

footprint enters any of the cached overlap regions (with entry/exit hysteresis).

a) Global matching uses using metadata: Each node device maintains its own tracking records and exports lightweight metadata (≈0.2 kB per person) whenever a person enters an overlap region. These compact messages are transmitted to the Auctioneer, which computes cosine-similarity costs between CoM and its velocity vectors and resolves associations via the Hungarian algorithm, ensuring consistent cross-view identity alignment with negligible bandwidth cost.

b) QoV-Based Bidding: For camera i and person p at time t, with keypoint confidence $conf_j^{(p)}(t)$ and validity mask $m_j^{(p)}(t) \in 0,1$, each keypoint is weighted by its anatomical importance w_j, defined as:

$$w_j = \begin{cases} 1.7, \{for\ shoulders\ and\ hips\}\ (j = 5,6,11,12) \\ 1.3, \{for\ head\ and\ eyes\}\ (j = 0,1,2) \\ 1.0, \{otherwise.\} \end{cases}.$$

$$KC_i^{(p)}(t) = \frac{\sum_{j=1}^{KP} w_j \,\mathrm{conf}_j^{(p)}(t)\, m_j^{(p)}(t)}{\sum_{j=1}^{KP} w_j\, m_j^{(p)}(t)},$$

$$KS_i^{(p)}(t) = \frac{\sum_{j=1}^{KP} w_j\, m_j^{(p)}(t)}{\sum_{j=1}^{KP} w_j}.$$

$KC_i^{(p)}(t)$ denotes the weighted mean confidence over valid keypoints, $KS_i^{(p)}(t)$ represents the weighted visible-keypoint ratio.

Then, the overall QoV score is defined using the *geometric mean:*

$$QoV_i^{(p)}(t) = \sqrt{KC_i^{(p)}(t)\, KS_i^{(p)}(t)}.$$

Finally, the winning device is selected as:

$$i^* = argmax_{i \in G} QoV_i^{(p)}(t).$$

The overall score $QoV_i^{(p)}$ represents a geometric mean of both confidence and providing a conservative yet stable estimate of view reliability. The runner-up camera is retained as a secondary bidder for temporal failover.

4 Experiments

Setup. 2× Intel RealSense D435i (30 fps), 2× Jetson Orin NX, Gigabit switch, elderly-care mockup (10 m×8 m). Camera intrinsics provided by the manufacturer and loaded from config (no re-calibration). NTP sync (<10 ms).

a) Local tracking:

Metrics. For each ground-truth person visible in a frame, tracking outcomes are classified as:

True Positive (TP): same track ID follows the same person as in the previous frame.

MISS: the track is lost or temporarily disconnected.

SWAP: two persons cross and exchange track IDs.

Formally, we compare mappings between gt_id and track_id from frame t-1 to t. Tracking accuracy is defined as

$$Accuracy = \frac{TP}{Number\ of\ GT}$$

In addition, we report two complementary continuity metrics:

ACRL (Average Correct Run Length): mean uninterrupted TP-segment length normalized by each person's visible duration—higher values indicate more stable tracking.

FRAG (Fragments per frame): number of TP→MISS or TP→SWAP transitions per frame—lower values indicate fewer interruptions.

These metrics emphasize temporal consistency—how long each person remains correctly tracked—rather than frame-level precision, which is crucial for fall detection tasks.

Dataset and Evaluation. Five short single-camera RGB–D videos (MV1–MV5) were recorded to evaluate tracking consistency under various motion and interaction conditions. MV1–MV3 contain single-person actions (entry, posture change), and MV4–MV5 include multi-person scenes with partial occlusion and crossings. Each sequence spans 53–89 frames, totaling 625 annotated person instances. Tracking performance was evaluated using Accuracy, ACRL, and FRAG, along with counts of SWAP and MISS events as continuity indicators.

Table 1. Single-camera tracking performance on five motion videos (MV1-MV5)

Dataset	Accuracy	SWAP	MISS	ACRL	FRAG
MV1	0.981	0	1	0.9811	0.000
MV2	0.986	0	1	0.9857	0.000
MV3	0.989	0	1	0.9889	0.000
MV4	0.910	6	3	0.3036	0.040
MV5	0.962	2	3	0.2790	0.0017
Mean	**0.966**	2	2	0.7077	0.014

Results. the proposed temporal-signature tracker maintains consistent person IDs even during crossings or occlusion, preserving continuous trajectories under challenging motion. Table 1 summarizes the quantitative results: the tracker achieves over 0.98 Accuracy and zero fragmentation in single-person videos (MV1–MV3), and a mean Accuracy of 0.966 across all sequences. Performance slightly decreases in multi-person videos (MV4–MV5) due to crossing-induced swaps (reduced ACRL, small nonzero FRAG). Overall, the results confirm strong temporal stability, providing a reliable basis for downstream fall detection relying on uninterrupted motion cues.

b) Identity Matching:

In the overlapping region between two cameras, 750 synchronized frames were collected, yielding **54 valid bidding instances** (i.e., persons simultaneously visible in both views). Each camera transmitted compact metadata to the coordinator, which computed cosine similarity between pairs and applied the Hungarian algorithm for final matching. As summarized in Table 2, the proposed bidding-based matching achieved **75.9 % accuracy**, **93.1 % precision**, **71.1 % recall**, and an **F1 score of 0.81**. These results confirm that the proposed logic ensures reliable identity alignment across overlapping cameras, maintaining high precision under partial occlusion and cross-view motion.

Table 2. Multi-Camera Identity Matching in Overlapping Region

Confusion Matrix	Predicted Same	Predicted Different
Actual Same	27	11
Actual Different	2	14

These results demonstrate that the proposed bidding logic ensures reliable identity matching across overlapping cameras, maintaining high precision even under partial occlusion and cross-view motion.

c) Fall Detection Accuracy by QoV:

To evaluate how QoV affects fall-detection performance, we employed the Spatial–Temporal Graph Convolutional Network (ST-GCN) [13], trained on the FUKinect-Fall dataset [14]. The dataset provides absolute 3D joint coordinates of humans performing both fall and non-fall activities, allowing robust joint-based learning. To further analyze model behavior under different QoV, we recorded additional fall sequences with a single camera while varying the subject's falling direction and posture. Using camera localization results, 3D keypoints were converted into a common world coordinate frame, and 20 consecutive frames (≈1.2 s) from each sequence were used as input. As shown in Fig. 2, the fall-detection accuracy exhibits a strong dependence on the QoV. When the **QoV score** is below 0.4, the mean accuracy remains low at around 0.42, indicating unreliable classification due to poor keypoint visibility. As **QoV** increases beyond 0.5, the accuracy sharply improves and stabilizes above 0.85 for $\mathbf{QoV > 0.6}$, demonstrating a clear positive correlation between QoV and detection reliability

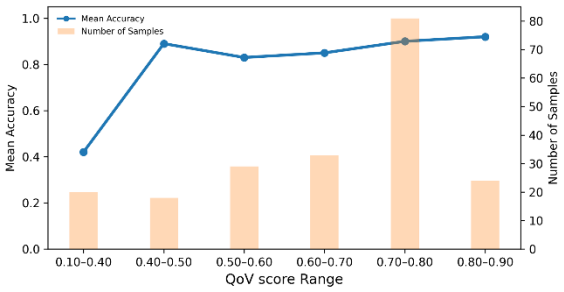

Figure 2. Mean fall-detection accuracy of the ST-GCN model across different QoV ranges. Each point indicates the averaged accuracy within a given QoV interval, while the bars represent the number of samples included in each bin.

5 Conclusion

We presented a distributed multi-camera fall-detection system combining temporal-consistency tracking, overlap-aware bidding, and QoV analysis. The tracker achieved 0.966 accuracy with stable continuity (high ACRL, near-zero FRAG), confirming its ability to maintain consistent identity tracking under occlusion and cross-view motion. The proposed bidding-based mechanism reliably associated identities across overlapping cameras, achieving an F1 score of 0.81 with high precision.

QoV analysis confirmed a positive correlation between observability and detection reliability. Overall, the system offers a scalable basis for reliable fall detection in distributed edge-camera systems.

Authors Contribution

All authors contributed equally to this work.

Acknowledgement

This research was supported by the "Regional Innovation System & Education (RISE)" program through the Seoul RISE Center, funded by the Ministry of Education (MOE) and the Seoul Metropolitan Government (2025-RISE-01-007-05), and by the Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Ministry of Science and ICT (MSIT), Republic of Korea, under the Artificial Intelligence Convergence Innovation Human Resources Development program (IITP-2025-RS-2023-00254592) and the IITP-ICAN (ICT Challenge and

Advanced Network of HRD) program (IITP-2025-RS-2023-00260248).

In addition, this work was supported by the Commercialization Promotion Agency for R&D Outcomes(COMPA) grant funded by the Korea government(Ministry of Science and ICT) (2710086167).

References

[1] S. Yang, F. Ding, P. Li, and S. Hu, “Distributed multi-camera multi-target association for real-time tracking,” Scientific Reports, vol. 12, no. 1, p. 11052, 2022.

[2] Y. Chen, L. Ma, S. Liu, M. Liu, C. Wu, and M. Li, “A real-time distributed multi-camera multi-object tracking system,” in 2022 2nd International Conference on Electrical Engineering and Mechatronics Technology (ICEEMT). IEEE, 2022, pp. 146–149.

[3] S. Han, P. Huang, H. Wang, E. Yu, D. Liu, and X. Pan, “Mat: Motion-aware multi-object tracking,” Neurocomputing, vol. 476, pp. 75–86, 2022.

[4] K. Yi, K. Luo, X. Luo, J. Huang, H. Wu, R. Hu, and W. Hao, “Ucmc-track: Multi-object tracking with uniform camera motion compensation,” in Proceedings of the AAAI Conference on Artificial Intelligence, vol. 38, no. 7, 2024, pp. 6702–6710.

[5] J. Jang, M. Seon, and J. Choi, “Lightweight indoor multi-object tracking in overlapping fov multi-camera environments,” Sensors, vol. 22, no. 14, p. 5267, 2022.

[6] K.-w. Park, Y.-J. Shim, M.-j. Lee, and H. Ahn, “Multi-frame based homography estimation for video stitching in static camera environments,” Sensors, vol. 20, no. 1, p. 92, 2019.

[7] M. Lupion, V. Gonzalez-Ruiz, J. F. Sanjuan, and P. M. Ortigosa, “Privacy-aware fall detection and alert management in smart environments using multimodal devices,” Internet of Things, vol. 30, p. 101526, 2025.

[8] Á. Carro-Lagoa, V. Barral, M. González-López, C. J. Escudero, and L. Castedo, “Multicamera edge-computing system for persons indoor location and tracking,” Internet of Things, vol. 24, p. 100940, 2023.

[9] Jocher, G., Chaurasia, A., & Qiu, J. (2023). Ultralytics YOLOv8 (Version 8.0.0) [Computer software]. https://github.com/ultralytics/ultralytics

[10] Campos, C., Elvira, R., Gómez, J. J., Montiel, J. M. M., & Tardós, J. D. (2021). ORB-SLAM3: An Accurate Open-Source Library for Visual, Visual-Inertial and Multi-Map SLAM. IEEE Transactions on Robotics, 37(6), 1874-1890.

[11] Welford, Barry Payne. "Note on a method for calculating corrected sums of squares and products." Technometrics 4.3 (1962): 419-420.

[12] Kuhn, H. W. (1955). The Hungarian method for the assignment problem. Naval Research Logistics Quarterly, 2(1-2), 83-97.

[13] S. Yan, Y. Xiong, and D. Lin, “Spatial temporal graph convolutional networks for skeleton-based action recognition,” *Proc. AAAI Conf. Artificial Intelligence*, vol. 32, no. 1, 2018.

[14] Aslan M., Akbulut Y., Sengor A., CevdetInce M. "Skeleton based efficient fall detection", J. Faculty Eng. Architecture Gazi Univ., 32 (4) (2017), pp. 1025-1034.