Person Detection and Tracking in Crowded Scenes using Classical Computer Vision Techniques

Academic Year: 2024-2025

Realised by:

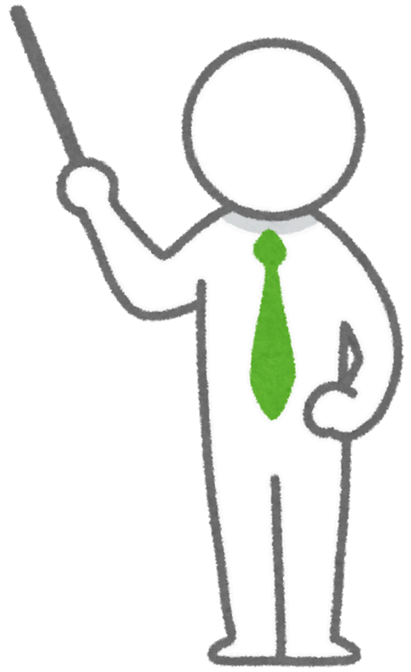
Manon LAGARDE
Maria KONTARATOU
Chaimae SADOUNE

Supervised by:

Céline Hudelot

Presentation outline

- *Problem Definition*
- *Methodology*
- *Evaluation Metrics*
- *Analysis of Results*
- *Conclusion*



Problem Definition

In dense urban environments and large-scale events, the ability to detect and track individuals is crucial for various applications such as public safety, crowd management, and behavioral studies.

Why is this important?

- Surveillance & security
- Crowd flow analysis
- Smart city applications

Why is this difficult?

- Scale Variations → People appear in different sizes depending on their distance from the camera.
- Occlusions → Individuals may be partially or fully blocked by others.
- Dynamic Backgrounds → Changing environmental conditions and moving cameras add complexity.

Objective →

1. Accurate individual detection without requiring large annotated datasets.
2. Robust tracking across frames, minimizing identity switches and false detections.
3. Computational efficiency, enabling real-time or near-real-time performance in practical applications.

Related Work

Over the years, several methodologies have been developed to address the challenges associated with this task.

- **Holistic Detection Approaches** : Early detection methods relied on scanning the entire image to detect pedestrians using feature-based techniques such as edge templates and Histogram of Oriented Gradients (HOG). However, these methods often struggle with background clutter and occlusions.

- **Part-Based Detection Approaches :** To address occlusion and pose variation, part- based models decompose a person into separate segments, detecting each part individually before assembling them into a whole. While robust to occlusions, their accuracy depends on the reliable detection of each segment.

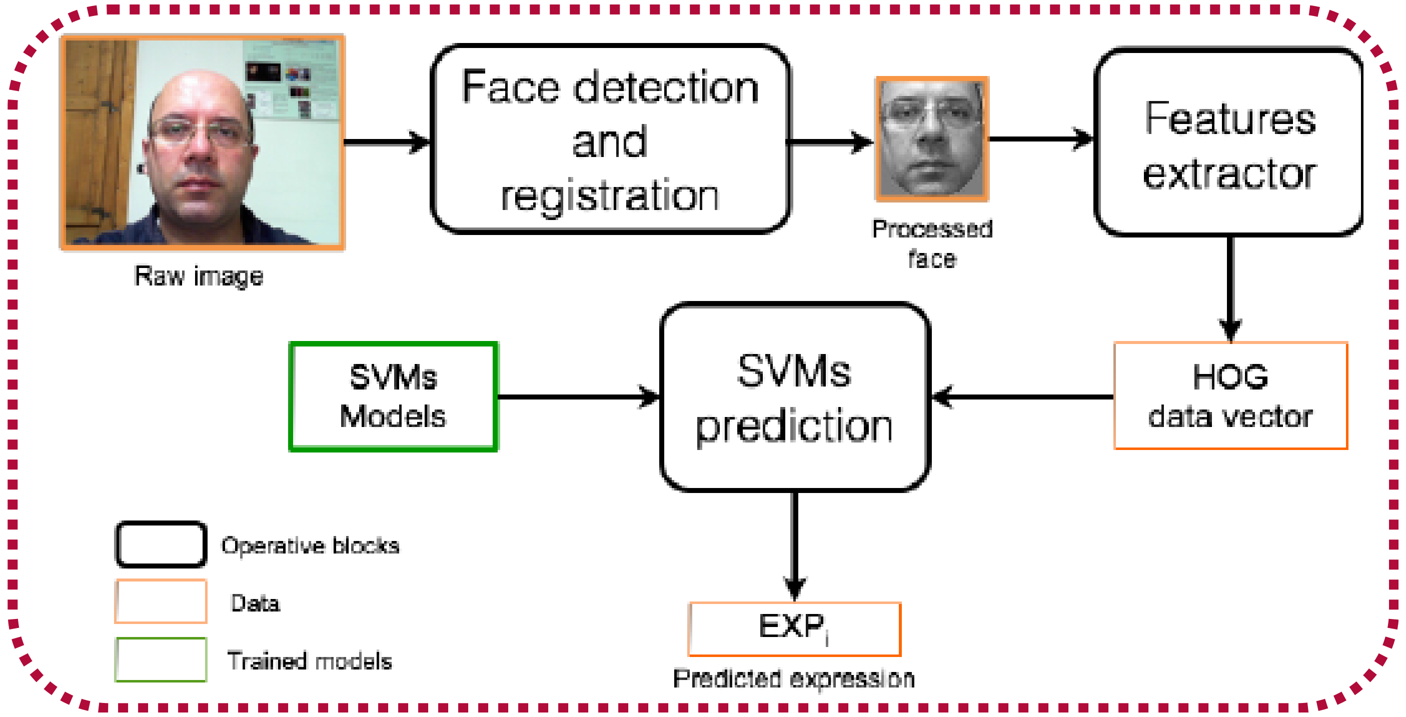
- **Motion-Based Detection Approaches :** These methods use background subtraction techniques to identify moving entities. Although effective in static environments, they are sensitive to lighting changes and dynamic backgrounds.

- **Deep Learning-Based Detection Approaches :** Recent advances leverage Convolutional Neural Networks (CNNs) such as You Only Look Once (YOLO) and Faster R-CNN for robust and efficient detection. However, these models require significant computational power and large-scale annotated datasets.

Methodology

The proposed methodology is based on classical computer vision techniques that balance accuracy and computational efficiency. The selected approaches are :

- **Pedestrian Detection using HOG and SVM**
- Pedestrian Detection and Tracking using HOG and Centroid Tracking
- Hybrid Pedestrian Tracking Pipeline



- HOG captures edge-based features of pedestrians effectively.
- SVM provides accurate classification with low computational overhead.

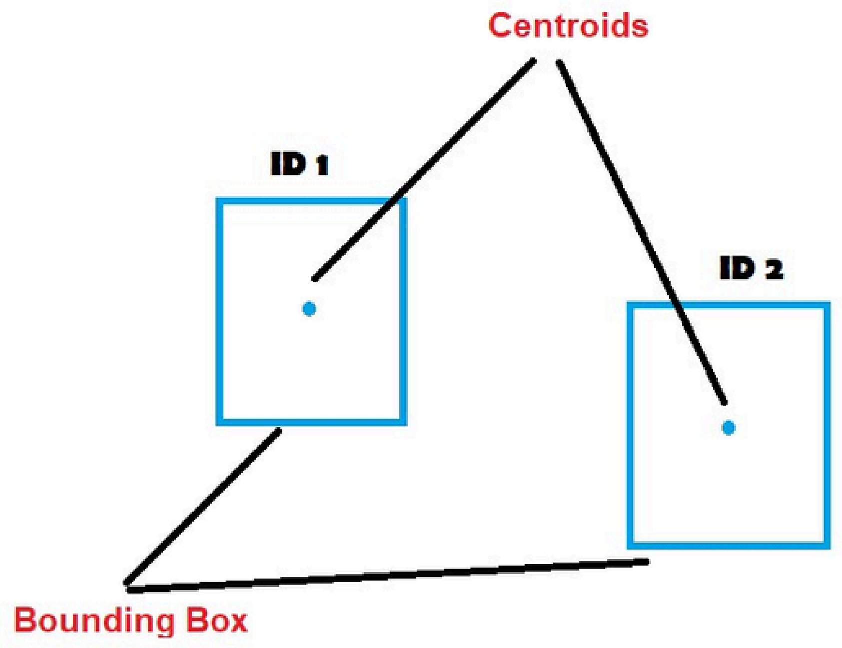
Methodology

The proposed methodology is based on classical computer vision techniques that balance accuracy and computational efficiency. The selected approaches are :

Pedestrian Detection using HOG and SVM

Pedestrian Detection and Tracking using HOG and Centroid Tracking

Hybrid Pedestrian Tracking Pipeline



- Detections are associated across frames based on centroid movement.
- Well-suited for moderate crowd densities with low computational cost.

Methodology

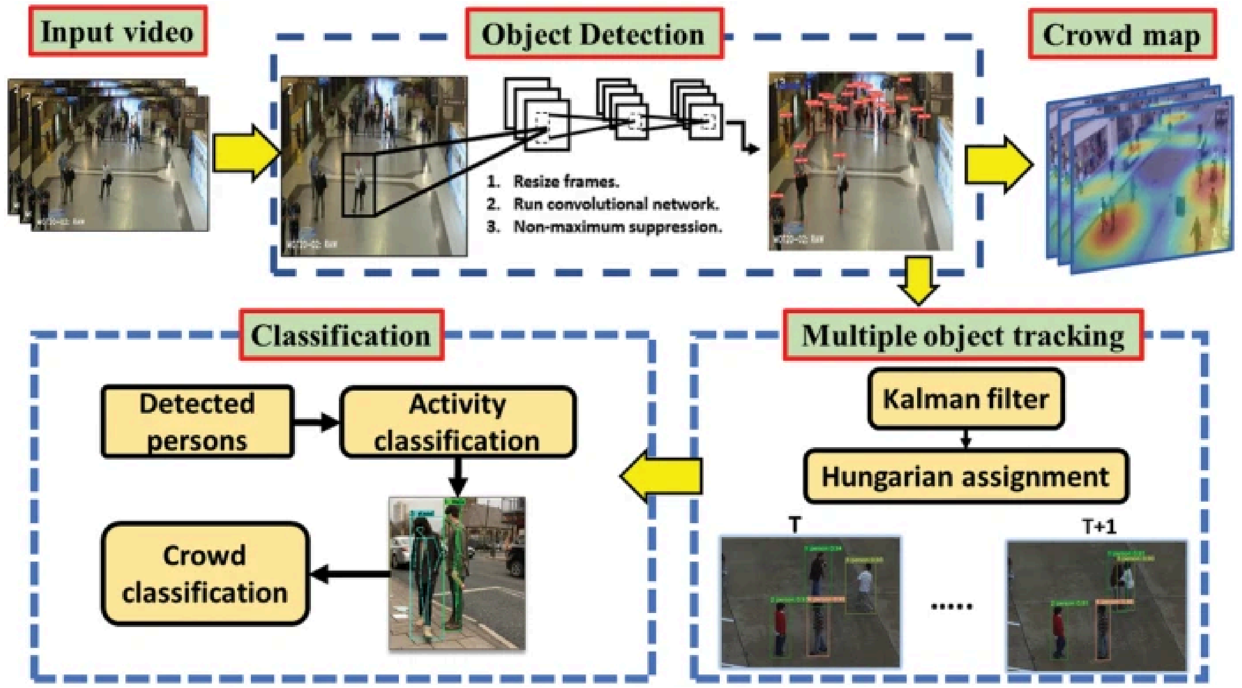
The proposed methodology is based on classical computer vision techniques that balance accuracy and computational efficiency. The selected approaches are :

Pedestrian Detection using HOG and SVM

Pedestrian Detection and Tracking using HOG and Centroid Tracking

Hybrid Pedestrian Tracking Pipeline

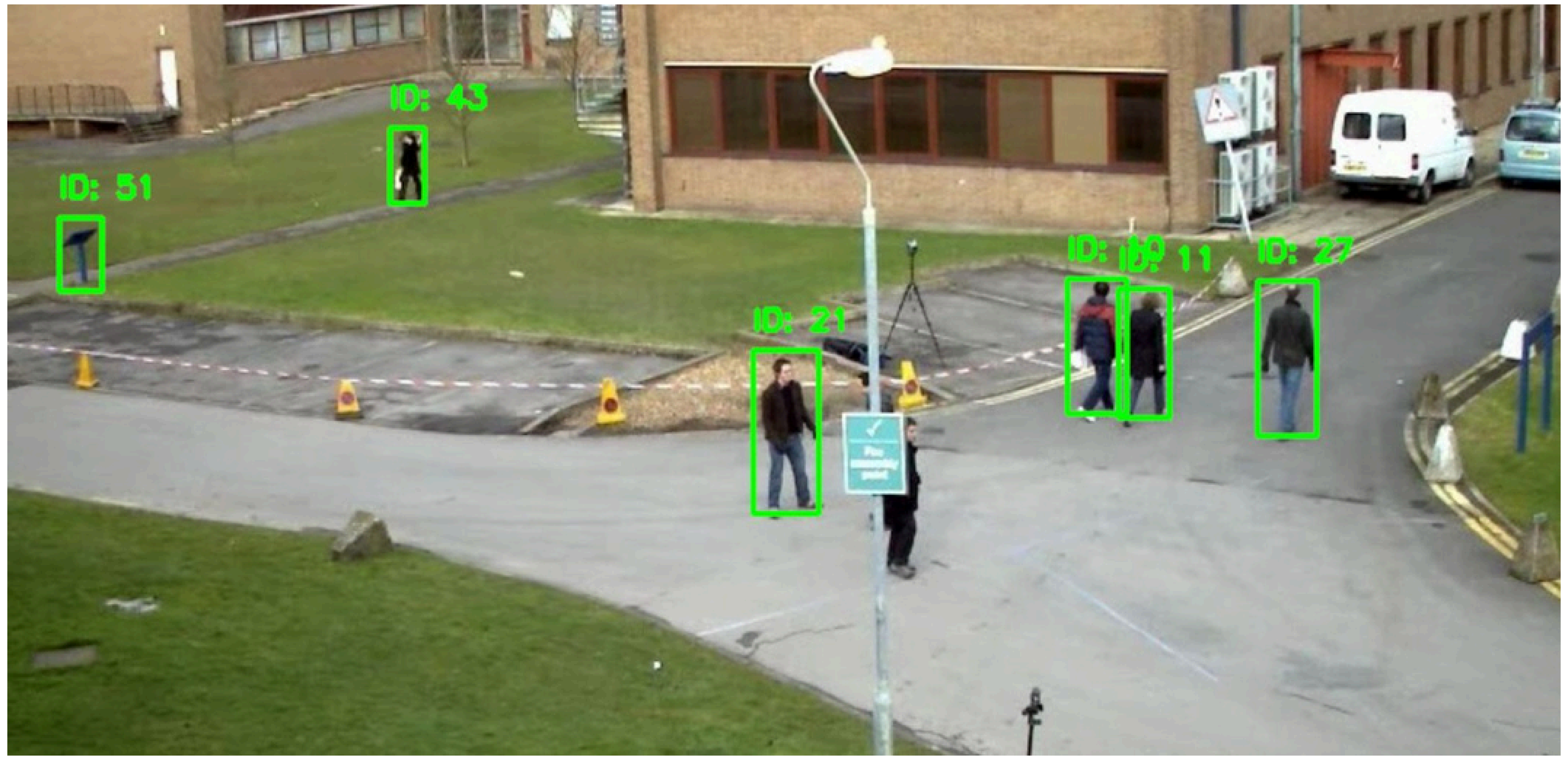
Background Subtraction, HOG + SVM Detection, and Kalman-Hungarian Tracking



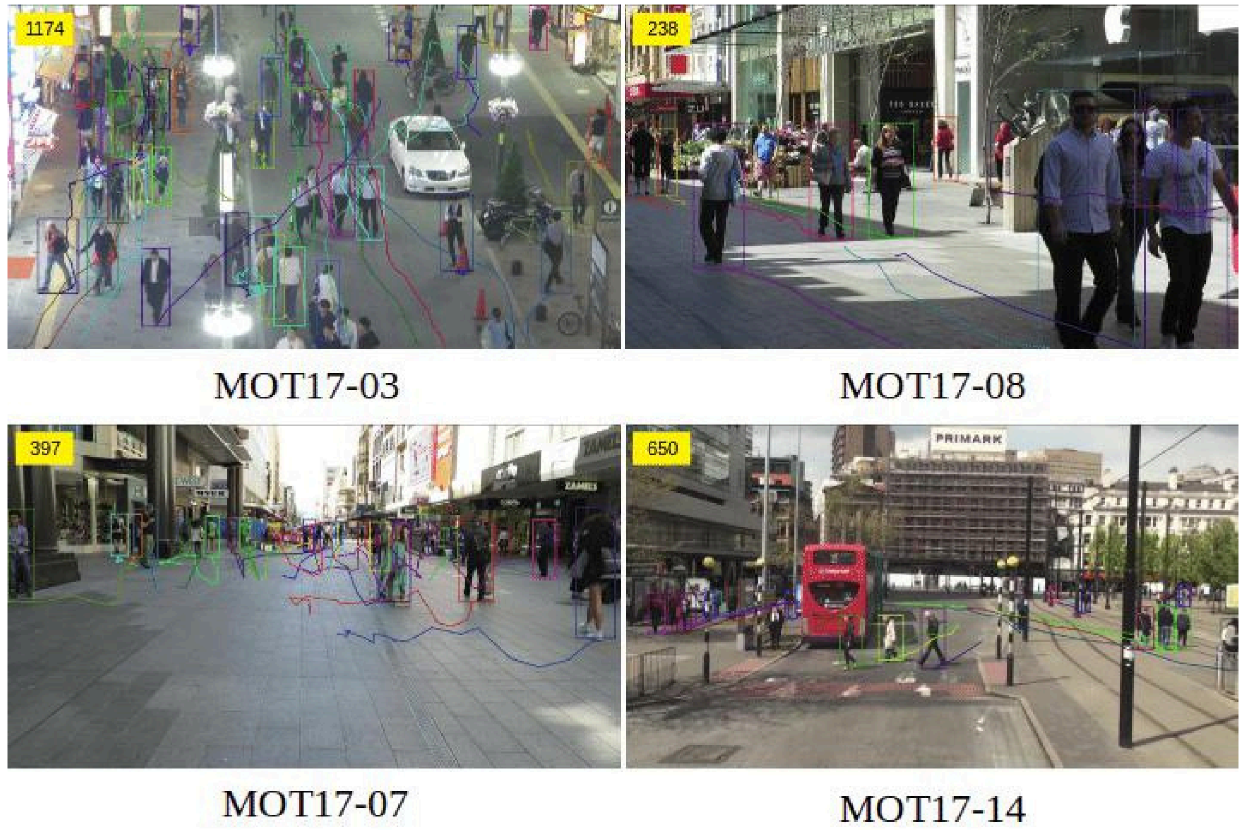
- **Background Subtraction (MOG2) →** Removes static elements and detects moving objects
- **Pedestrian Detection (HOG + SVM)**
- **IoU Filtering →** Merges overlapping detections from MOG2 and HOG to improve accuracy
- **Kalman Filter Prediction** → Estimates the next position of pedestrians based on motion history
- **Hungarian Algorithm Matching** → Matches new detections to existing tracks and updates Kalman if match is found

DataSet Used

To evaluate the performance of our pedestrian detection and tracking methods, we rely on well-known benchmark datasets that reflect real-world scenarios.



PETS2009



MOT Dataset

Before applying our detection models, we prepare the datasets through several preprocessing steps to enhance their suitability. After preprocessing, these datasets serve as the foundation for training and testing

Evaluation Metrics :

To evaluate the performance of our pedestrian detection and tracking methods, we used key evaluation metrics :

- **Precision & Recall** – Measure the model's ability to correctly detect pedestrians while minimizing false positives and false negatives.
- **F1 Score** – Balances precision and recall to provide an overall performance measure.
- **Intersection over Union (IoU)** – Evaluates how well predicted bounding boxes overlap with the ground truth.
- **Multi-Object Tracking Accuracy (MOTA)** – Penalizes false positives, false negatives, and identity switches.
- **Multi-Object Tracking Precision (MOTP)** – Measures how accurately detected objects align with ground truth over time.
- **Processing Time per Frame** – Determines real-time feasibility.

Evaluation Metrics :

To test the robustness of our models, experiments were conducted under diverse conditions :

- **Static & dynamic backgrounds** – Static scenes are easier for background subtraction, while dynamic scenes introduce motion-based detection challenges.

- **Different crowd densities** – Sparse pedestrian environments are easier to track, while dense crowds introduce occlusions and identity switches.

- **Scale variations** – People appear at different scales based on distance, challenging fixed-size feature extractors like HOG.

Dataset Used :

- **PETS09-S2L1** – A moderately crowded scene with occlusions.
- **PETS09-S2L2** – A highly dense crowd, making detection and tracking significantly harder.

Results Overview :

Model	Video	Precision	Recall	F1 Score	IoU	Dice Score	MOTA	MOTP
HOG + SVM	PETS09-S2L1	0.002	0.0013	0.0016	0.5427	0.0016	N/A	N/A
Hybrid Model	PETS09-S2L1	0.014	0.0118	0.0128	0.5639	0.0128	-0.8241	0.5697
HOG + SVM	PETS09-S2L2	0.0013	0.0004	0.0006	0.5226	0.0006	N/A	N/A
Hybrid Model	PETS09-S2L2	0.0079	0.003	0.0044	0.5516	0.0044	-0.3775	0.5631

Table - Evaluation Metrics Overview for Pedestrian Detection and Tracking

The **Hybrid Model (MOG2 + HOG + Kalman)** showed better precision and recall than the baseline model (HOG + SVM), but performance remains limited.

- Low Precision & Recall indicate frequent false positives and missed detections, limiting reliability.
- IoU (~0.55) suggests detections are somewhat aligned with ground truth but box placement is inconsistent.
- Negative MOTA scores highlight frequent identity switches and poor tracking stability, especially in high-density scenes.
- MOTP (~0.56) shows that when tracking succeeds, bounding box alignment is reasonable, but tracking failures dominate.
- PETS09-S2L2 (dense crowd) worsens performance, with recall dropping to 0.0030, making tracking unreliable.

While **classical methods provide a computationally efficient approach**, visual verification shows that person detection worked despite the weak tracking performance. Occlusion handling, feature extraction, and tracking consistency still require improvements for real-world applications.

Conclusion :

Advantages :

- Lightweight and interpretable approach using HOG + MOG2 + Kalman filtering, making it suitable for low-resource applications.
- Real-time pedestrian detection and tracking without requiring deep learning.
- Visual observation confirmed successful pedestrian detection, even though numerical scores were low.

Limitations :

- Struggles with occlusions and dense crowds, leading to identity switches and false detections.
- Tracking is unstable, as seen in negative MOTA scores, especially in high-density scenarios.
- Background subtraction introduces false positives, affecting detection accuracy.

Classical computer vision techniques offer a viable solution for pedestrian detection and tracking in low-resource environments. Despite challenges with occlusions and identity tracking, the approach provides a foundation for improvement. Enhancing feature extraction, tracking stability, and background modeling can make it a reliable alternative for real-time applications without deep learning.

References :

[1] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In CVPR, pages 886–893, 2005

[2] Z. Zivkovic. Improved adaptive Gaussian mixture model for background subtraction. In ICPR, pages 28–31, 2004

[3] P. Dollar, C. Wojek, B. Schiele, and P. Perona. Pedestrian detection: A benchmark. In CVPR, pages 304–311, 2009

[4] C. Wojek, S. Walk, and B. Schiele. Multi-cue onboard pedestrian detection. In CVPR, pages 794–801, 2009

[5] Performance evaluation of tracking and surveillance (PETS2009) dataset. University of Reading. Supplied as supplemental material pets2009.pdf, 2009

[6] MOT Challenge dataset. A benchmark for multi-object tracking. Supplied as supplemental material motchallenge.pdf. URL: https://motchallenge.net, 2015

[7] R. Kalmanman, A New Approach to Linear Filtering and Prediction Problems, Journal of Basic Engineering, vol. 82, no. 1, pp. 35–45, 1960.

[8] H. W. Kuhn, The Hungarian Method for the Assignment Problem, Naval Research Logistics Quarterly, vol. 2, no. 1-2, pp. 83–97, 1955.

[9] J. Munkres, Algorithms for the Assignment and Transportation Problems, Journal of the Society for Industrial and Applied Mathematics, vol. 5, no. 1, pp. 32–38, 1957.

Thank You