Paper Summary and Project Mapping

# Paper Details

|  |  |
| --- | --- |
| Section | Details |
| Paper Title | A Review of Machine Learning and Deep Learning Methods for Person Detection, Tracking and Identification, and Face Recognition with Applications |
| Authors | Beibut Amirgaliyev et al. |
| Published In | Sensors, February 2025 |
| Purpose | Comprehensive survey of ML/DL methods for person detection, tracking, identification, and face recognition |
| Techniques Covered | HOG, DPM, Haar, YOLOv4/v8, Faster R-CNN, DeepSORT, FairMOT, ArcFace, FaceNet, UniPose, CRD, CCFA, Transformer-based Re-ID |
| Datasets Reviewed | COCO, CelebA, VGGFace, LFW, MOTChallenge, Market1501, CUHK03 |
| Evaluation Metrics | mAP, MOTA, IDF1, EER, F1-score |
| Applications | Smart cities, surveillance, transportation, crowd monitoring, and access control |
| Ethical Focus | Privacy, fairness, dataset bias, synthetic data use, federated learning |

# Pros and Cons of the Paper

|  |  |
| --- | --- |
| Pros | Cons |
| Covers 140+ studies with structured categorization | No novel algorithm proposed |
| Includes classical + deep learning methods | No experimental implementation or benchmarking |
| Details many relevant public datasets | Lacks edge deployment and compression strategies |
| Addresses real-world surveillance and smart city use cases | General overview of architectures, lacks training-specific insights |
| Discusses fairness and privacy in CV deployments | No sample code or model fine-tuning discussion |

# Project Relevance: CCTV-Based Person Attribute Extraction

|  |  |  |
| --- | --- | --- |
| Project Need | Supported By Paper | Suggested Tools / Models |
| Person Detection | YOLOv4/v8, Faster R-CNN, SSD | YOLOv8 |
| Person Tracking | DeepSORT, ByteTrack, FairMOT | DeepSORT |
| Attribute Extraction | FaceNet, ArcFace, UniPose, VGGFace | FaceNet (age/gender), UniPose (body pose), VGGFace (identity) |
| Re-identification (optional) | Transformer-based Re-ID, CRD, CCFA | CRD for cloth-changing re-ID |
| Dataset Recommendations | COCO, CelebA, LFW, MOTChallenge, VGGFace | CelebA (face attributes), COCO (detection), MOT (tracking) |
| Deployment Ethics | Privacy, fairness, bias, federated learning | Consider synthetic datasets and privacy-preserving techniques |

# Literature Survey (100 Words)

Amirgaliyev et al. (2025) present a systematic review of over 140 studies focused on machine learning and deep learning techniques for person detection, tracking, identification, and face recognition. They analyze classical approaches like HOG and Kalman filters alongside modern deep models such as YOLO, ArcFace, and DeepSORT. The paper highlights real-world applications in surveillance, transportation, and smart cities while addressing challenges such as occlusion, real-time constraints, and ethical concerns. This review serves as a valuable reference for developing intelligent video surveillance systems, especially for projects involving CCTV-based person attribute extraction using deep learning.

Paper Summary and Project Mapping

**Paper Details**

|  |  |
| --- | --- |
| Section | Details |
| Paper Title | Human Height Estimation Using AI-Assisted Computer Vision for Intelligent Video Surveillance System |
| Authors | K. Iyshwarya Ratthi, B. Yogameena, S. Saravana Perumaal |
| Published In | Measurement, Elsevier, 2024 |
| Purpose | To estimate human height from surveillance videos using AI with YOLOv7-OA and camera calibration. |
| Techniques Covered | YOLOv7-OA, Hybrid Attention Mechanism (HAM), Camera Calibration, Height Estimation Formula |
| Dataset Introduced | Sense-Height: Real-world dataset including children and adults under surveillance |
| Evaluation Metric | Mean Absolute Error (MAE), achieving as low as 0.02 cm |
| Applications | Missing child retrieval, surveillance analysis, intelligent monitoring |
| Ethical Considerations | Face blurring, consent collection, privacy-focused design |

**Pros and Cons of the Paper**

|  |  |
| --- | --- |
| Pros | Cons |
| High accuracy (0.02–0.8 cm error) | Focuses only on height, not other attributes |
| Real-world dataset with children | Relies on known camera parameters (tilt, focal length) |
| Occlusion-aware detection (YOLOv7-OA + HAM) | Not designed for real-time edge processing |
| Monocular camera usage (no depth sensor needed) | Can misestimate in seated or bent poses |
| Field of View (MDORI) zone modeling |  |
| Privacy-focused design with face blurring |  |

**Project Relevance: CCTV-Based Person Attribute Extraction**

|  |  |
| --- | --- |
| Component | How You Can Use It |
| Height Estimation | Use their mathematical model for attribute extraction |
| Occlusion Handling | Apply YOLOv7-OA + HAM in crowded scenes |
| Zoning Strategy | Incorporate MDORI zones for camera-based logic adjustment |
| Dataset Reference | Use Sense-Height to test your height estimation model |
| Camera Setup | Follow calibration approach to reduce measurement errors |
| Hybrid System Design | Combine height with gender, attire, etc. for robust identification |

**Literature Survey (100 Words)**

In their 2024 study, Iyshwarya Ratthi et al. introduced an AI-based human height estimation model for surveillance, leveraging monocular cameras and YOLOv7 with a hybrid attention mechanism (HAM). Designed to aid in missing child retrieval, the system uses camera calibration and a new dataset (“Sense-Height”) featuring adults and children. Unlike traditional models, this approach handles occlusion, diverse lighting, and motion conditions with high accuracy (error as low as 0.02 cm). The paper provides strong empirical validation and proposes a field-of-view (FOV) zoning strategy. This work is significant for integrating height as a soft biometric in intelligent video surveillance.

**📄 9.pdf One more paper**

**Title:** *Learning a Deep-Feature Clustering Model for Gait-Based Individual Identification*  
**Authors:** Kamal Taha, Paul D. Yoo, Yousof Al-Hammadi, Sami Muhaidat, Chan Yeob Yeun

**✅ Summary:**

This paper proposes a gait-based person identification model using **inertial measurement unit (IMU)** data from shoe-type sensors. The model extracts features through **stacked sparse autoencoders** and identifies attributes like **gender, age, height, and weight** using **clustering-based ensemble learning**. The system shows greater accuracy and robustness than traditional image-based gait recognition models, especially under real-world variations (e.g., clothing, lighting).

**✅ Pros:**

* Immune to lighting/appearance changes
* Extracts physical attributes (e.g., gender, height)
* High identification accuracy
* Uses deep learning and clustering synergy

**❌ Cons:**

* Not suitable for standard CCTV (needs wearables)
* No visual input—unsuited for traditional surveillance footage
* Complex and resource-intensive training

**📌 Project Relevance:**

While it doesn’t use CCTV, the **stacked autoencoder-based feature extraction** and **clustering logic** can inspire video-based gait recognition. You can adapt its learning strategy to improve soft biometric extraction from video gait patterns.

**📚 Literature Survey (100 words):**

Taha et al. (2024) proposed a gait recognition model using IMU data instead of conventional video. Their system collects gait features from shoe-embedded sensors and processes them using stacked sparse autoencoders. The high-level features are then clustered to identify physical characteristics like age, gender, and body size. The model shows greater robustness to occlusion and environmental variation than visual gait recognition systems. While it is not directly usable for CCTV-based projects, its deep learning approach and gait-based soft biometric extraction provide a strong conceptual base for designing attribute recognition models using motion cues in surveillance footage.

**📄 10.pdf last paper**

**Title:** *A Comprehensive Review of Face Recognition Techniques, Trends, and Challenges*  
**Authors:** H. L. Gururaj, B. C. Soundarya, S. Priya, J. Shreyas, Francesco Flammini

**✅ Summary:**

This is an extensive survey of **Face Recognition (FR)** methods, datasets, trends, and challenges. It covers techniques from traditional (PCA, LDA, ICA) to deep learning-based (CNN, hybrid models), evaluates video and image-based FR systems, and identifies challenges such as pose variation, occlusion, and aging. It categorizes FR techniques into **appearance-based, landmark-based, and hybrid** approaches and presents a taxonomy of systems and datasets used for facial recognition.

**✅ Pros:**

* Covers both 2D and 3D FR methods
* Surveys modern deep learning trends
* Explains FR phases: detection, feature extraction, classification
* Useful dataset and tool comparisons

**❌ Cons:**

* Review only, no experimental model
* Limited focus on real-time CCTV FR scenarios
* Doesn’t deeply explore attribute classification beyond facial identity

**📌 Project Relevance:**

This paper is an excellent **reference for face detection and recognition techniques**. You can leverage its insights on **hybrid models**, **feature extraction**, and **pose/illumination handling** for building robust face attribute extractors in your CCTV project.

**📚 Literature Survey (100 words):**

Gururaj et al. (2024) presented a detailed review of face recognition (FR) systems, covering traditional techniques like PCA and LDA, and advanced deep learning methods including CNN-based hybrid models. The paper explores FR challenges such as pose variation, occlusion, and aging, while classifying existing approaches into appearance-based, landmark-based, and hybrid methods. It also discusses video-based FR systems, dataset availability, and future directions. Although it does not introduce new models, this survey offers valuable insights into the selection of algorithms and datasets that can aid in developing accurate and real-time person attribute recognition from CCTV surveillance footage.