International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Cross-Camera Person Re-Identification in Urban Networks Using Machine Learning

Manish Borse [a], Tejas Barhate [b], Ujjwal Amodkar [c], Sumit Borse [d], Prof.N.L.Bhale [e]

[a] *Department of Information Technology, Matoshri College of Engineering & Research Centre Eklahar , Nashik-422105 , India*
[b] *Department of Information Technology, Matoshri College of Engineering & Research Centre Eklahare , Nashik-422105 , India*
[c] *Department of Information Technology, Matoshri College of Engineering & Research Centre Eklahare , Nashik-422105 , India*
[d] *Department of Information Technology, Matoshri College of Engineering & Research Centre Eklahare , Nashik-422105 , India*
[e] *Project Guide, Information Technology, Matoshri College of Engineering & Research Centre Eklahare, Nashik-422105 , India*

A B S T R A C T

Person re-identification (Re-ID) across multiple non-overlapping camera views is an emerging challenge in smart surveillance systems. Unlike traditional person detection within a single frame or camera feed, Re-ID requires consistent recognition of the same individual across different angles, backgrounds, lighting conditions, and time intervals. This paper proposes a robust machine learning-based system for cross-camera person re-identification in urban networks using a convolutional neural network (CNN) backbone and a feature embedding strategy based on triplet loss. We explore the use of state-of-the-art datasets like Market-1501 and DukeMTMC-reID to evaluate our model. The system demonstrates high accuracy and robustness across diverse conditions, showing promise for real-time smart city surveillance and public safety applications. Further, the paper discusses the technical challenges, recent advancements in Re-ID, and future enhancements to this approach, providing a comprehensive guide for researchers and developers working in intelligent surveillance.

Keywords: Person Re-Identification, Machine Learning, CNN, Smart Surveillance, Urban Networks, Cross-Camera Tracking, Triplet Loss.

1. Introduction

The advent of smart cities has encouraged the deployment of thousands of surveillance cameras throughout urban infrastructure. These cameras play a crucial role in monitoring public spaces, enhancing security, and supporting law enforcement agencies. However, tracking a person seamlessly across a network of non-overlapping cameras remains a complex task due to factors such as lighting variations, camera angles, occlusion, and temporal gaps between camera feeds.

Cross-camera person re-identification (Re-ID) refers to the task of associating images or video segments of the same individual captured by different camera views. This technology has potential applications in areas like criminal investigations, traffic analysis, crowd management, and even personalized navigation in public spaces. Unlike face recognition, e-ID considers whole-body appearance, clothing patterns, gait, and contextual cues.

This research addresses the need for intelligent, scalable, and real-time Re-ID systems. We propose an efficient architecture based on deep learning that utilizes a CNN backbone and a triplet loss function for embedding learning. The goal is to identify individuals reliably across multiple camera feeds despite visual variations.

2. Literature Survey

Person Re-ID has evolved significantly over the past decade. Early approaches relied on manually designed features such as color histograms, texture descriptors, and spatial orientation. These features, while easy to implement, were highly sensitive to environmental changes.

With the rise of deep learning, convolutional neural networks (CNNs) became the dominant approach for Re-ID. Notable works include DeepReID, which demonstrated the use of deep CNNs for person matching, and Part-based Convolutional Baseline (PCB), which improved performance by learning part-aligned features. Attention-based models such as HA-CNN and MGN (Multiple Granularity Network) further refined feature extraction by focusing on discriminative body parts.

Metric learning approaches such as triplet loss and contrastive loss have been instrumental in training Re-ID models. Triplet loss ensures that the embedding distance between samples of the same identity is minimized while maximizing the distance from samples of different identities.

Recent research has explored graph-based models and temporal reasoning for video-based Re-ID, improving robustness over time. However, most practical systems still depend on frame-based analysis and lightweight CNN architectures for real-time deployment.

3. Problem Statement

In urban networks, multiple non-overlapping CCTV cameras record individuals at different times and locations. Identifying a person who appears in one camera and reappears in another with different clothing, lighting, background, and pose is a non-trivial task.

Traditional video analytics tools are limited to activity detection within a single camera feed. Manual tracking is infeasible due to the volume of data generated. There is a pressing need for automated, accurate, and scalable systems that can continuously track individuals across different views and notify administrators in real time.

Challenges faced in Re-ID include:

- Pose and viewpoint variation
- Background clutter and occlusion
- Inconsistent illumination and resolution
- Appearance similarity between different people

4. Proposed Methodology

- ***4.1 Person Detection***

We utilize YOLOv5 for real-time person detection across frames from multiple camera feeds. YOLOv5 offers a balance between speed and accuracy, essential for urban environments.

- ***4.2 Feature Extraction***

Detected person images are passed through a CNN, specifically ResNet-50, pretrained on ImageNet and fine-tuned on person Re-ID datasets. The CNN captures rich visual representations by learning spatial hierarchies.

- ***4.3 Embedding Learning***

Using triplet loss, the model learns a feature space where samples from the same individual across cameras are closer than samples from different individuals. The embedding dimension is set to 256, and embeddings are normalized for consistency.

- ***4.4 Matching and Retrieval***

In the inference stage, embeddings are matched using cosine similarity or Euclidean distance. A threshold-based alert system notifies administrators when a match is found across camera views.

5. System Architecture

The architecture comprises the following:

- **Object Detection:** YOLOv5 identifies person regions.

Multi-Camera Input: Feeds from fixed surveillance cameras.

- **Feature Encoder:** ResNet-50 extracts person descriptors.
- **Feature Database**: Stores embedding’s of all detected individuals.
- **Re-ID Engine:** Calculates similarity and returns top matches.
- **Admin Interface**: Visual dashboard for monitoring matches.

This modular design supports horizontal scaling, making it suitable for large metropolitan areas. Integration with cloud-based storage and edge devices allows flexible deployment.

6. Experimental Setup & Results

- ***6.1 Market-1501:***
    - 32,668 images, 1,501 identities, 6 cameras
    - Our model achieved:
        - Rank-1 Accuracy: 87.2%
        - mAP (mean Average Precision): 74.6%
- ***6.2 DukeMTMC-reID:***
    - 36,411 images, 1,812 identities, 8 cameras
    - Our model achieved:
        - Rank-1 Accuracy: 82.3%
        - mAP: 71.4%

Inference was tested on an NVIDIA GTX 1660 GPU, achieving 12 FPS per stream, making the system suitable for real-time operation. Augmentation techniques like random erasing and horizontal flipping improved generalization.

7. Advantages Over Existing Systems:

1. Robust feature representation using deep CNNs
2. Higher accuracy through metric learning
3. Real-time inference with YOLOv5 + ResNet
4. Scalable architecture for city-wide networks
5. Visual interface for ease of use by security staff
6. Low false positives due to well-trained embedding's

8. Conclusions & Future Work:

This paper presents a real-time, accurate, and scalable approach to cross-camera person re-identification in urban networks using machine learning. The system effectively addresses key challenges using CNNs and metric learning, achieving high accuracy on benchmark datasets. With its modular design, the architecture can be scaled across city-level surveillance systems.

Future work includes:

- Incorporating temporal reasoning and trajectory prediction
- Gait recognition for identity verification
- Federated learning for decentralized data privacy
- Edge deployment on low-resource devices
- Integration with law enforcement databases
- Development of explainable Re-ID systems

By advancing cross-camera Re-ID, we take a step closer toward safer, smarter, and more connected urban environments.

References:

1. Zheng, L., Shen, L., Tian, L., Wang, S., Wang, J., & Tian, Q. (2015). Scalable Person Re-identification: A Benchmark. ICCV.
2. Sun, Y., Zheng, L., Yang, Y., Tian, Q., & Wang, S. (2018). Beyond Part Models: Person Retrieval with Refined Part Pooling. ECCV.
3. Hermans, A., Beyer, L., & Leibe, B. (2017). In Defense of the Triplet Loss for Person Re-Identification. arXiv preprint arXiv:1703.07737.
4. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. CVPR.

5. Yu, H. X., Wu, A., & Zheng, W. S. (2017). Cross-View Asymmetric Metric Learning for Unsupervised Person Re-identification. ICCV.
6. Zhou, K., Yang, Y., Cavallaro, A., & Xiang, T. (2021). Learning Generalisable Omni-Scale Representations for Person Re-Identification. TPAMI.
7. Ge, Y., Li, Z., Zhao, H., & Yin, Y. (2020). Structured Feature Learning for Person Re-Identification. CVPR.