

Reidentificación de Personas basado en reconocimiento de marcha

Harold Canto

Juan Torres

Asesor: Edward Jorge Yuri Cayllahua Cahuina

Universidad de Ingeniería y Tecnología (UTEC)

2025

Contenido

1

Motivación

2

Problema Computacional

3

Justificación

4

Objetivos

5

Pipeline Metodológico

6

Dataset y Preprocesamiento

7

Resultados

8

Conclusiones y Recomendaciones

9

Trabajo Futuro

10

Referencias

Motivación

- Person Re-ID busca identificación entre cámaras sin depender solo del rostro.
- Los métodos supervisados alcanzan alto rendimiento pero requieren costosos conjuntos etiquetados.
- Métodos autosupervisados basados en consistencia temporal o pares frame/tracklet pueden aprender invariancias útiles sin etiquetas.
- Un enfoque híbrido puede aprovechar lo mejor de ambos mundos.

Fuente: Person Re-Identification - an overview | ScienceDirect Topics

Problema Computacional

¿Cómo lograr que el aprendizaje autosupervisado capture representaciones de identidad en secuencias de marcha no etiquetadas, que sean robustas, y que al combinarse con ajuste fino supervisado alcancen rendimiento competitivo con menores costos de etiquetado y recursos?

Justificación

Técnica:

- El supervisado necesita etiquetas y el SSL aprovecha grandes volúmenes no etiquetados.
- Un preentrenamiento SSL adecuado puede mejorar la generalización entre cámaras.

Práctica:

- Reduce costos operativos.
- Escalable para vigilancia real en entornos urbanos.
- Mejora robustez en condiciones de baja resolución.

Objetivo General

Desarrollar y evaluar un método híbrido para Re-ID basado en gait, que reduzca la dependencia de datos etiquetados y mejore la generalización en condiciones reales.

Objetivos Específicos

- Seleccionar datasets y protocolo.
- Diseñar un pretext-task SSL orientado a secuencias de marcha.
- Implementar fine-tuning con pérdidas métricas (Triplet) y clasificación (CrossEntropy).
- Evaluar con métricas según estado del arte.

Pipeline Metodológico

- Fase I: Preentrenamiento SSL (pares temporales / augmentaciones).
- Fase II: Transferencia y fine-tuning supervisado (PK Sampler, Triplet, CrossEntropy).
- Evaluación bajo Rank-1 y mAP, junto con pruebas de escalabilidad por data.

Motivación del Modelo Híbrido

- El aprendizaje autosupervisado captura patrones robustos sin necesidad de etiquetas.
- El aprendizaje supervisado refina las representaciones para la tarea específica de Re-ID.
- El modelo híbrido combina ambos enfoques para mejorar generalización y rendimiento.
- Esta estrategia reduciría la dependencia de datos etiquetados sin sacrificar exactitud.

CASIA-B ofrece una secuencia completa de imágenes para el estudio de *gait recognition*.

- 124 sujetos, 3 condiciones (normal, bolso y abrigo).
- Capturas desde 11 vistas angulares entre 0° y 180° .
- Adecuado para secuencias de marcha y SSL temporal.

Preprocesamiento

- Conversión a escala de grises; normalización de píxeles a $[0,1]$.
- Estandarización a size de 64×64 .
- Augmentaciones para SSL: recorte, rotación leve, flip horizontal, cambios leves de brillo/contraste.

Arquitectura General del Modelo Híbrido

Etapas del híbrido:

- ① Pre-entrenamiento SSL: backbone CNN aprende representaciones temporales.
- ② Transferencia: pesos de SSL como inicialización del backbone.
- ③ Fine-tuning supervisado: cabeza de clasificación + cabeza métrica (normalizada), optimizadas con CrossEntropy + Triplet Loss.

Componente autosupervisado (SSL)

- Basado en contraste temporal entre frames de la misma secuencia.
- Utiliza proyección SimCLR y pérdida InfoNCE para pares temporales.
- Augmentaciones aplicadas solo durante preentrenamiento SSL.

Componente supervisado

- Se añade una cabeza de clasificación sobre el embedding del SSL.
- Se usa Triplet Loss para maximizar separación entre identidades.
- Se emplea **P-K Sampler** para mejorar el aprendizaje métrico.
- Ajuste fino del backbone para adaptar las características al problema Re-ID.

Módulo de integración híbrida

El modelo híbrido combina:

- Representaciones autosupervisadas (*global invariances*).
- Ajuste supervisado para la identidad (*local discriminative features*).

Algunas ventajas del modelo híbrido:

- Mejora Rank-1 y mAP sin necesidad de más datos.
- Obtiene embeddings más consistentes entre cámaras distintas.

Resultados Cuantitativos

Modelo	Rank-1	mAP
Supervisado	62.1	55.3
SSL + Supervisado (Híbrido)	81.68	29.32

Conclusiones

- El modelo híbrido alcanzó un Rank-1 del 81.26% sobre 124 sujetos, mostrando que el preentrenamiento autosupervisado aporta información útil, aunque su rendimiento es bajo.
- El mAP de 29.23% evidencia que, si bien el modelo identifica correctamente a la persona en las primeras posiciones, presenta dificultad para ordenar todo el ranking, especialmente en vistas extremas y secuencias con pérdida de silueta.
- Las métricas adicionales demuestran:
 - Rank-5: 93.7%
 - Rank-10: 96.18%
 - Rank-20: 97.74%

Indicando que la identidad correcta aparece casi siempre dentro del top-20.

Conclusiones

- El valor $\text{minP} = 3.8$ confirma que el modelo suele recuperar rápidamente la identidad positiva, aunque con variabilidad según vista y calidad de la silueta.
- Los resultados confirman que el enfoque híbrido es funcional y competitivo dentro de métodos ligeros, pero requiere mejoras mecanismos view-aware, atención espacial o fusión temporal más profunda para acercarse a técnicas de referencia.

Trabajo Futuro y Recomendaciones

- Se recomienda implementar *Stochastic Weight Averaging* (SWA) para mitigar la sensibilidad a la inicialización y reducir la varianza durante el ajuste fino.
- Es necesario integrar mecanismos de atención espacial (View-Aware) para ponderar dinámicamente las regiones visibles y mejorar la recuperación en vistas frontales.
- Se sugiere transitar hacia arquitecturas *Vision Transformers* (ViT) para capturar dependencias espacio-temporales globales superiores a las CNNs tradicionales.
- Se propone investigar la fusión multimodal con datos RGB o esqueletos para compensar la falta de información de la silueta en condiciones de oclusión severa.

Referencias

- C. Joshua et al., “Using optical flow consistency for self-supervised person Re-ID”, 2025.
- X. Liu et al., “UCM-VeID V2: Multi-View End-to-End Re-ID”, CVPR 2025.
- M. Varenyk et al., “Self-supervised low-FPS tracking for Re-ID”, 2025.
- H. Rao et al., “Self-supervised gait encoding for Re-ID”, IJCAI 2020.
- Z. Dou et al., “Identity-seeking self-supervised Re-ID”, ICCV 2023.

Gracias