A Survey on 3D Skeleton Based Person Re-Identification: Taxonomy, Advances, Challenges, and Interdisciplinary Prospects

Haocong Rao[1,2,3] and **Chunyan Miao**[1,2,3]*
[1] College of Computing and Data Science, Nanyang Technological University (NTU), Singapore
[2] Joint NTU-UBC Research Centre of Excellence in Active Living for the Elderly (LILY), NTU, Singapore
[3] Alibaba-NTU Global e-Sustainability CorpLab (ANGEL), NTU, Singapore
{haocong.rao, ascymiao}@ntu.edu.sg

Abstract

Person re-identification via *3D skeletons* is an important emerging research area that attracts increasing attention within the pattern recognition community. With distinctive advantages across various application scenarios, numerous 3D skeleton based person re-identification (SRID) methods with diverse skeleton modeling and learning paradigms have been proposed in recent years. In this paper, we provide a comprehensive review and analysis of recent SRID advances. First of all, we define the SRID task and provide an overview of its origin and major advancements. Secondly, we formulate a systematic taxonomy that organizes existing methods into three categories centered on hand-crafted, sequence-based, and graph-based modeling. Then, we elaborate on the representative models along these three types with an illustration of foundational mechanisms. Meanwhile, we provide an overview of mainstream supervised, self-supervised, and unsupervised SRID learning paradigms and corresponding common methods. A thorough evaluation of state-of-the-art SRID methods is further conducted over various types of benchmarks and protocols to compare their effectiveness, efficiency, and key properties. Finally, we present the key challenges and prospects to advance future research, and highlight interdisciplinary applications of SRID with a case study. A curated collection of valuable resources is available at https://github.com/Kali-Hac/3D-SRID-Survey.

1 Introduction

Person re-identification (re-ID) is an essential pattern recognition task of matching and retrieving a person-of-interest across different views or scenes, which has been widely applied to security authentication, smart surveillance, activity monitoring, healthcare, and embodied AI [Nambiar *et al.*, 2019; Ye *et al.*, 2021; Bao *et al.*, 2026]. Recent economical and precise skeleton-tracking devices (*e.g.*, Kinect [Shotton *et al.*, 2011]) have simplified the acquisition of 3D skeleton data, enabling them to be a prevalent and versatile data modality for gait analysis and person re-ID [Liao *et al.*, 2020; Rao *et al.*, 2024]. Unlike conventional person re-ID methods that rely on appearance or facial characteristics [Ye *et al.*, 2021], 3D **S**keleton based person **Re**-**ID** (SRID) models typically exploit body-structure features and motion patterns (*e.g.*, gait [Murray *et al.*, 1964]) from 3D positions of key body joints to identify different persons. With unique merits such as small input data, light-weight models, privacy-preserving without using appearances, and robustness against view and background variations [Han *et al.*, 2017], SRID has attracted surging attention from both academia and industry [Rao *et al.*, 2022].

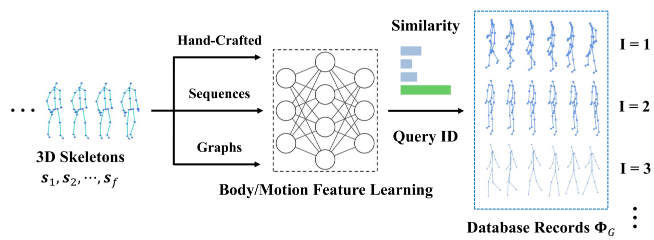
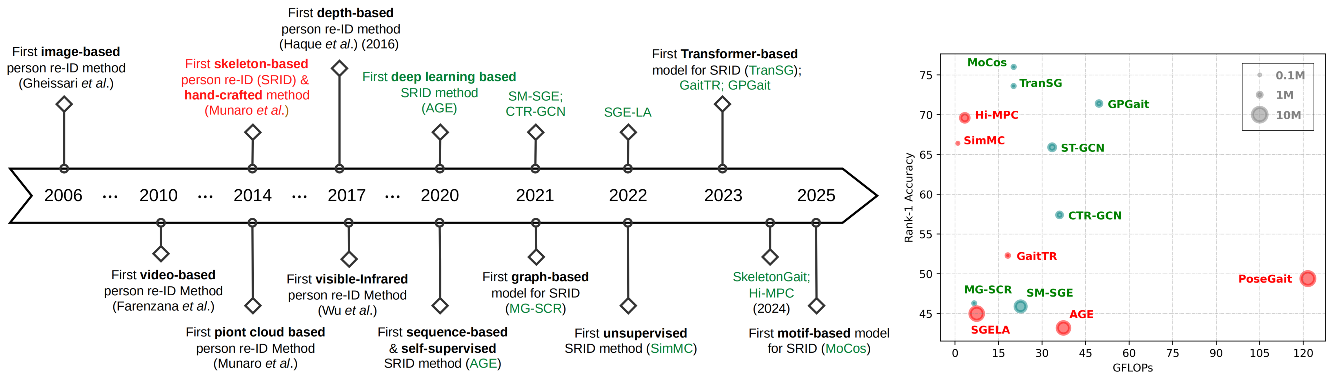


Figure 1: Overview of 3D skeleton based person re-ID (SRID) task with hand-crafted, sequence-based or graph-based modeling to learn effective body and motion features for identity recognition.

In recent years, research on SRID has gained significant momentum, leading to diverse skeleton modeling and learning paradigms. Early endeavors [Barbosa *et al.*, 2012; Munaro *et al.*, 2014a; Andersson and Araujo, 2015; Pala *et al.*, 2019] mainly extract hand-crafted features such as skeleton descriptors in terms of anthropometric, geometric and gait attributes of body. As these methods often require domain expertise such as anatomy and kinematics [Yoo *et al.*, 2002] for skeleton modeling, they lack the ability to fully mine latent high-level features beyond human cognition. To resolve this challenge, recent mainstream methods [Liao *et al.*, 2020; Huynh-The *et al.*, 2020; Rao *et al.*, 2022; Rashmi and Guddeti, 2022] leverage deep neural networks to automatically perform skeleton representation learning for SRID. One of exemplar methods (termed "*sequence-based modeling*") is to model sequential dynamics and motion semantics from raw or normalized skeletons (*e.g.*, joint trajectory) based on long

[1]Corresponding Author



(a) Origin and Advancements of SRID Research

(b) Model Accuracy & Efficiency Comparison

Figure 2: (a) Overview of research origin and technical advancements of SRID within the person re-ID community (Zoom in and follow the timeline for the best view). (b) Parameter sizes (Millions (M)), computational complexity (Giga Floating Point Operations (GFLOPs)), and KS20 Rank-1 accuracy of state-of-the-art deep learning methods for SRID (**Red**: Sequence-based models; **Green**: Graph-based models).

short-term memory (LSTM) and its variants [Wei *et al.*, 2020; Rao *et al.*, 2022]. However, they rarely investigate the intrinsic body relationships such as inter-joint motion correlations, thereby possibly overlooking some valuable skeleton patterns. Another paradigm (termed "*graph-based modeling*") mitigates this challenge by constructing skeleton graphs to model discriminative structural and actional features based on the interrelations of body parts [Rao and Miao, 2023]. This often requires multi-granularity body modeling and efficient relational reasoning mechanisms (*e.g.*, collaborative learning) based on skeleton graphs. Despite the great progress of SRID, this rapidly evolving technique still lacks a systematic review, making it difficult for researchers to gain a holistic view of this field and embark on new research endeavors.

In light of this, we present the first survey on SRID, elucidating recent advancements of skeleton modeling, learning paradigms, evaluation benchmarks, current challenges, and interdisciplinary applications. Firstly, we define the SRID task and provide a milestone overview to illustrate the origin and key advancements of SRID as shown in Fig. 2a. Secondly, we propose a systematic taxonomy of SRID methods to categorize them into hand-crafted, sequence-based, and graph-based modeling, and elaborate on their foundational mechanisms and representative approaches. We also illustrate the basic definitions and common methods within three mainstream SRID paradigms (supervised, self-supervised, unsupervised). Thirdly, we introduce existing public benchmarks, evaluation metrics, and protocols for SRID, while comprehensively evaluating state-of-the-art methods across different benchmarks to compare their performance and efficiency. Meanwhile, we conduct qualitative analysis of different methods to compare their key properties with a discussion of advantages and disadvantages. Finally, we discuss the current challenges in SRID and identify potential directions for future research. An overview of promising SRID applications in interdisciplinary areas, spanning healthcare, embodied AI, and security, is further provided and illustrated with a case study. The structure of this survey, including skeleton modeling (Sec. 3), learning paradigms (Sec. 4), benchmarks and evaluation (Sec. 5), challenges and prospects (Sec. 6) is shown in Fig. 3. We hope our survey can bring new insights to researchers and expedite future research in SRID.

2 Overview

2.1 Task Description

As illustrated in Fig. 1, the input of SRID task is a 3D skeleton sequence that belongs to a certain pedestrian, and the output is the predicted identity. Formally, we denote a 3D skeleton sequence as $\boldsymbol{S} = (\boldsymbol{s}_1, \cdots, \boldsymbol{s}_f) \in \mathbb{R}^{f \times j \times 3}$, where $\boldsymbol{s}_t \in \mathbb{R}^{j \times 3}$ denotes the t^{th} skeleton with 3D coordinates of j body joints. Each skeleton sequence $\boldsymbol{S}$ corresponds to a person identity I, where I $\in \{1, \cdots, \mathrm{C}\}$ and C is the number of different classes (*i.e.*, identities). In the SRID task, we generally have *training* set, *probe* set, and *gallery* set, respectively denoted as $\Phi_T = \left\{\boldsymbol{S}_i^T\right\}_{i=1}^{N_1}$, $\Phi_P = \left\{\boldsymbol{S}_i^P\right\}_{i=1}^{N_2}$, and $\Phi_G = \left\{\boldsymbol{S}_i^G\right\}_{i=1}^{N_3}$ that contain N_1, N_2, and N_3 skeleton sequences of different persons collected from different scenes or views. The task target is to learn a hand-crafted, sequence-based or graph-based model (detailed in Sec. 3) that maps 3D skeleton sequences into effective representations, so that we can query the correct identity of an encoded skeleton sequence representation in the probe set via matching it with the sequence representations in the database (*i.e.*, gallery set). SRID is essentially a retrieving and matching problem.

2.2 Origin and Advancements

As presented in Fig. 2a, the first SRID research [Munaro *et al.*, 2014a] using hand-crafted skeleton descriptors commenced in 2014, coming after the first RGB video based method [Farenzena *et al.*, 2010] and before the first depth-based approach [Haque *et al.*, 2016]. Then, Rao *et al.* proposed the first deep learning based SRID paradigm in 2020, followed by the first self-supervised, unsupervised, and supervised paradigms [Rao *et al.*, 2020; Rao and Miao, 2022a; Rao and Miao, 2023]. Over the past five years, an increasing number of innovative models have been devised specifically for SRID and related emerging tasks, including LSTM

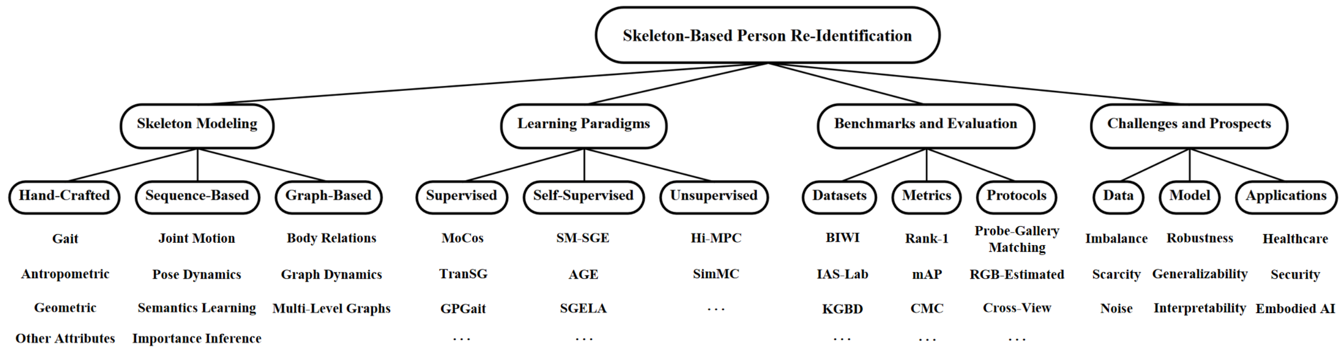


Figure 3: Structure of this survey with the taxonomy of SRID research. Representative branches and SRID methods are listed.

models ($e.g.$, AGE [Rao *et al.*, 2020]), CNN models ($e.g.$, SkeletonGait [Fan *et al.*, 2024]), Transformer models ($e.g.$, TranSG [Rao and Miao, 2023]), GCN models ($e.g.$, CTR-GCN [Chen *et al.*, 2021]), MLP models ($e.g.$, SimMC [Rao and Miao, 2022a]), and hybrid/composite models ($e.g.$, SM-SGE [Rao *et al.*, 2021a]) (compared in Table 2).

2.3 Taxonomy of SRID Methods

As shown in Fig. 3, we propose a systematic taxonomy for SRID approaches based on **skeleton modeling** (see Sec. 3) and **learning paradigms** (see Sec. 4). For skeleton modeling, we divide them into three categories, including (1) *hand-crafted modeling* using manually-extracted features ($e.g.$, skeleton descriptors), (2) *sequence-based modeling* that focuses on sequential features ($e.g.$, joint trajectory) of 3D skeletons, and (3) *graph-based modeling* that represents 3D skeletons as graphs, and further subcategorize them by different learning focuses such as body relations or graph dynamics. In terms of learning paradigms, we group them into (1) *supervised* SRID paradigms that require skeleton labels for feature learning, (2) *self-supervised* SRID paradigms that combine pretext tasks for skeleton representation learning with labeled fine-tuning, and (3) *unsupervised* SRID paradigms that learn skeleton features without using labels.

3 Skeleton Modeling

We elaborate on different skeleton modeling including their foundational mechanisms and representative approaches.

3.1 Hand-Crafted Modeling

Gait Attributes. Extracting gait features is a common way to characterize unique walking patterns of an individual [Cunado *et al.*, 2003], typically including (1) kinematic parameters ($e.g.$, angles of hips, knees, and feet), and (2) spatio-temporal parameters ($e.g.$, stride length, gait cycle time, velocity). They can be manually computed with:

$$f_{angles} = \{(\alpha_{ij}, \beta_{ij}) \mid (i,j) \in \Psi\}, \tag{1}$$

$$\alpha_{ij} = \arctan\frac{y_i - y_j}{x_i - x_j},\ \beta_{ij} = \arctan\frac{z_i - z_j}{\sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}}, \tag{2}$$

where x_i, y_i, z_i denote 3D coordinates of the i^{th} joint, the set Ψ defines *adjacent* joints constrained by the human skeleton model, and two kinds of angles α_{ij} and β_{ij} are calculated from these joint pairs. The velocity can be calculated by

$$Velocity = \frac{\sum_{i=1}^{n} \frac{strideLength}{n}}{cycleTime}, \tag{3}$$

where

$$strideLength = 2 * stepLength. \tag{4}$$

We calculate the step length by averaging the highest values of the difference between the right and left feet, and adopt the mean stride length over all n strides following [Andersson and Araujo, 2015].

Anthropometric Attributes. The Euclidean distance between two joints, such as bone lengths, limb dimensions, height, are usually computed as the anthropometric features (f_A) to differentiate individuals [Barbosa *et al.*, 2012]:

$$f_A = \{\|J_i - J_j\|_2 \mid (i,j) \in \Psi^*\}, \tag{5}$$

where J_i=(x_i, y_i, z_i), J_j=(x_j, y_j, z_j), Ψ^* not only contains adjacent joints but also can be extended to cover more anthropometric properties, such as 13 (D_{13}) or 16 (D_{16}) skeleton descriptors in [Munaro *et al.*, 2014b; Pala *et al.*, 2019].

Geometric Attributes. The incorporation of geometric skeletal measurements, including body-part ratios and inter-joint geodesic distances on the mesh surface, can enhance feature representations in person re-ID [Barbosa *et al.*, 2012].

The above hand-crafted features are often learned by different classifiers ($e.g.$, KNN) to perform person re-ID [Andersson and Araujo, 2015; Nambiar *et al.*, 2017]. They are also combined with different metric algorithms [Pala *et al.*, 2015] or other modalities such as 3D point clouds [Gharghabi *et al.*, 2015; Bondi *et al.*, 2016; Pala *et al.*, 2019; Munaro *et al.*, 2014a] to further boost person re-ID accuracy.

3.2 Sequence-Based Modeling

Joint Motion. The walking patterns are characterized by the motion of joints, which can be defined as the differences between body joint positions [Liao *et al.*, 2020]:

$$f_{motion} = \boldsymbol{s}_t - \boldsymbol{s}_{t-1}, \tag{6}$$

where $s_t = \{J_1^t, J_2^t, \cdots, J_j^t\}$, $J_i^t = (x_i^t, y_i^t, z_i^t)$ denotes the 3D coordinates of i^{th} joint in the t^{th} skeleton, and $i \in \{1, 2, \cdots, j\}$. The occluded or masked joints during walking can also help models learn useful motion patterns [Rashmi and Guddeti, 2022].

Pose Dynamics. The consecutive skeletons typically conveys dynamics of unique body poses, which can be encoded sequentially by temporal learning models ($e.g.$, LSTM):

$$\boldsymbol{h}_t = \begin{cases} \phi(\boldsymbol{s}_1) & \text{if } t = 1 \\ \phi(\boldsymbol{h}_{t-1}, \boldsymbol{s}_t) & \text{if } 1 < t \leq f \end{cases}, \tag{7}$$

where $\phi(\cdot)$ denotes the model function to encode the pose dynamics of skeletons, $\boldsymbol{h}_{t-1}$ represents the latent representation of previous $(t-1)$ poses, which provides the temporal context information to encode the long-term pose dynamics $\boldsymbol{h}_t$ for person re-ID [Rao *et al.*, 2022]. The joint distances, relative joint positions, and bones are combined in [Wei *et al.*, 2020; Zhang *et al.*, 2023] to enhance pose dynamics learning.

Semantics Learning. By encoding the skeleton sequence into latent high-dimensional representation, latent motion semantics such as motion continuity can be captured to enhance feature learning. Representative semantics learning tasks include skeleton sequence reconstruction and prediction, which can be simply formulated as follows:

$$\phi(\boldsymbol{s}_1, \boldsymbol{s}_2, \cdots, \boldsymbol{s}_f) = \begin{cases} \boldsymbol{s}_1, \boldsymbol{s}_2, \cdots, \boldsymbol{s}_f & \text{Plain Recon.} \\ \boldsymbol{s}_f, \boldsymbol{s}_{f-1}, \cdots, \boldsymbol{s}_1 & \text{Reverse Recon.} \\ \boldsymbol{s}_{f+1}, \boldsymbol{s}_{f+2}, \cdots, \boldsymbol{s}_{2f} & \text{Prediction} \end{cases} \tag{8}$$

The reverse reconstruction (Recon.) and prediction require the model function $\phi(\cdot)$ to capture high-level semantics ($e.g.$, order and correlations) to achieve the target output, which facilitates learning more meaningful gait representations for person re-ID [Rao *et al.*, 2022]. A few studies also explore the semantics learning of motion consistency [Rao and Miao, 2022a] and cross-scale body relations [Rao *et al.*, 2021a].

Importance Inference. Different skeletons and their feature representations typically possess different importance in characterizing poses and discriminative patterns of a person, which can be exploited to mine key skeletons or hard samples [Hermans *et al.*, 2017] for SRID learning. To aggregate key skeleton features for SRID, AGE [Rao *et al.*, 2020] explores a locality-aware attention mechanism to integrate features of important skeletons in the sequence, while SM-SGE [Rao *et al.*, 2021a] infers key correlations between different body-joint nodes. Rao *et al.* [2024] further propose a hard skeleton mining mechanism to adaptively infer importance of multi-level skeleton representations for SRID.

3.3 Graph-Based Modeling

Body Relations. Skeletons can be naturally modeled as graphs based on the physical connections of human body joints. For each skeleton $\boldsymbol{x}_t$, we represent it as the graph $\mathcal{G}^t(\mathcal{V}^t, \mathcal{E}^t)$, where $\mathcal{V}^t = \{v_1^t, v_2^t, \cdots, v_j^t\}, v_i^t \in \mathbb{R}^3, i \in \{1, \cdots, j\}$ and edges $\mathcal{E}^t = \{e_{i,j}^t \mid v_i^t, v_j^t \in \mathcal{V}^t\}, e_{i,j}^t \in \mathbb{R}$. $\mathcal{E}^t$ denotes the set of connections between different joints. Based on the skeleton graphs, we can model the inherent correlations, such as limb collaboration [Rao *et al.*, 2021b; Rao and Miao, 2022b] and channel-specific relations [Chen *et al.*, 2021] for SRID. The graph motifs can also be devised to enhance the structure and gait related relation learning [Rao and Miao, 2025].

Type	Dataset	Reference	Source	# ID	# Skeletons	# View
Sensor-Based	BIWI RGBD-ID	[Munaro *et al.*, 2014b]	Kinect V1	50	205.8K	S
	IAS-Lab RGBD-ID	[Munaro *et al.*, 2014c]	Kinect V1	11	89.0K	S
	KGBD	[Andersson and Araujo, 2015]	Kinect V1	164	188.7K	S
	KinectREID	[Pala *et al.*, 2015]	Kinect V1	71	4.8K	7
	UPCV1	[Kastaniotis *et al.*, 2015]	Kinect V1	30	13.1K	S
	UPCV2	[Kastaniotis *et al.*, 2016]	Kinect V2	30	26.3K	S
	Florence 3D Re-ID	[Bondi *et al.*, 2016]	Kinect V2	16	18.0K	S
	KS20	[Nambiar *et al.*, 2017]	Kinect V2	20	36.0K	5
Estimated	CAISA-B-3D	[Liao *et al.*, 2020]	Videos	124	706.5K	11
	3DGait*	[Wang *et al.*, 2023]	Videos	43	22.9K	S
	OUMVLP-Pose-2D	[Chen *et al.*, 2022]	Videos	10307	6667.0K	14
	PoseTrackReID-2D	[Chen *et al.*, 2022]	Videos	5350	53.6K	—

Table 1: Overview of SRID benchmark datasets. The number of skeletons in training is reported. "S" denotes single or egocentric view. *RGB-estimated* 3D and 2D skeleton datasets are listed. * denotes an interdisciplinary benchmark for gait and disease prediction.

Graph Dynamics. The dynamics of skeleton graphs can be leveraged to learn and capture the temporal evolution of joints' connection patterns and limbs' coordination for SRID. The process can be simplified as

$$\boldsymbol{g}_t = \begin{cases} \phi\left(f_G\left(\boldsymbol{s}_1\right)\right) & \text{if } t = 1 \\ \phi\left(\boldsymbol{g}_{t-1}, f_G\left(\boldsymbol{s}_t\right)\right) & \text{if } 1 < t \leq f \end{cases}, \tag{9}$$

where $f_G(\cdot)$ denotes the graph encoding model ($e.g.$, graph convolutional networks [Fu *et al.*, 2023]), $\phi(\cdot)$ is the temporal learning model ($e.g.$, LSTM) to encode the long-term dynamics of graph representations. To facilitate spatio-temporal graph learning, self-supervised pretext tasks, such as graph reconstruction and prediction, are incorporated to learn more effective skeleton features [Rao and Miao, 2023].

Multi-Level Graphs. Multi-level skeleton graph representations are devised to characterize coarse-to-fine body structure and motion [Li *et al.*, 2020], and various graph learning tasks such as sparse graph prediction and cross-scale graph inference [Rao *et al.*, 2021a] are proposed to help learn different-level graph semantics. Based on graph transformers, masked reconstruction [He *et al.*, 2022] and motifs [Rao and Miao, 2025] are further explored for high-level structural ($e.g.$, locality) and gait semantics learning.

4 Learning Paradigms

Supervised SRID paradigms leverage skeletal annotations or labels ($e.g.$, identity class) to guide the model to learn discriminative features, typically using cross-entropy (CE) loss as follows:

$$\mathcal{L}_{\text{CE}} = \frac{1}{N_1} \sum_{i=1}^{N_1} \sum_{j=1}^{C} -\mathrm{I}_{i,j} \log(\hat{\mathrm{I}}_{i,j}), \tag{10}$$

where $\mathrm{I}_{i,j}$ denotes the ground-truth identity label ($\mathrm{I}_{i,j} = 1$ iff the i^{th} sample belongs to the j^{th} identity otherwise 0), $\hat{\mathrm{I}}_{i,j}$ is the probability that the i^{th} sample is predicted as the j^{th} identity, and N_1 is the number of training samples. This loss has been widely applied to classic supervised models such as

	Method	Conference/ Journal	BIWI-S	BIWI-W	IAS-A	IAS-B	KS20	Algorithms/ Architectures	Network Parameters	Require Labels?	Pretext Task?	Multi-Scale Modeling?	Theory for Analysis?	Summary
Hand-Crafted	[Gharghabi *et al.*, 2015]	ICIEV	10.7	10.7	—	—	—	KNN	—	✓	✗	✗	✓	**Advantages:**
	[Munaro *et al.*, 2014a]	Person Re-Identification	32.1	39.3	—	—	—	NN	—	✓	✗	✗	✓	High theoretical explainability of features and models;
	D_{13} [Munaro *et al.*, 2014b]	ICRA	28.3	14.2	40.0	43.7	39.4	NN	—	✓	✗	✗	✓	Generally low complexity.
	D_{16} [Pala *et al.*, 2019]	Computers& Graphics	32.6	17.0	42.7	44.5	51.7	Adaboost	—	✓	✗	✗	✓	**Disadvantages:**
	[Elaoud *et al.*, 2017]	ACIVS	28.6	10.7	45.5	63.6	—	SVD	—	✓	✗	✗	✓	Require domain knowledge; Labor-expensive;
	PM [Elaoud *et al.*, 2021]	Multimedia Tools and Applications	39.3	39.3	36.4	81.8	—	RF	—	✓	✗	✗	✓	Low performance.
Sequence-Based	PoseGait [Liao *et al.*, 2020]	Pattern Recognition	14.0	8.8	28.4	28.9	49.4	CNN	8.93M	✓	✗	✗	✗	**Advantages:**
	AGE [Rao *et al.*, 2020]	IJCAI	25.1	11.7	31.1	31.1	43.2	LSTM	7.15M	✓	✓	✗	✗	Learn with raw sequences;
	SGELA [Rao *et al.*, 2022]	TPAMI	25.8	11.7	16.7	22.2	45.0	LSTM	8.47M	✓	✓	✗	✗	Capture temporal dynamics.
	GaitTR [Zhang *et al.*, 2023]	Expert Systems	43.2	16.3	43.7	47.8	52.3	Transformer	0.49M	✓	✗	✗	✗	**Disadvantages:**
	SimMC [Rao and Miao, 2022a]	IJCAI	41.7	24.5	44.8	46.3	66.4	MLP	0.15M	✗	✗	✗	✓	Lack body structure modeling;
	Hi-MPC [Rao *et al.*, 2024]	IJCV	47.5	27.3	45.6	48.2	69.6	MLP	3.32M	✗	✗	✓	✓	Ignore joint or limb relations.
Graph-Based	SkeletonGait [Fan *et al.*, 2024]	AAAI	15.1	10.8	31.4	31.5	22.2	CNN	11.11M	✓	✗	✗	✗	**Advantages:**
	MG-SCR [Rao *et al.*, 2021b]	IJCAI	20.1	10.8	36.4	32.4	46.3	GAT, LSTM	0.35M	✓	✓	✓	✗	Fully model body structure
	SM-SGE [Rao *et al.*, 2021a]	ACM MM	31.3	13.2	34.0	38.9	45.9	MGRN, LSTM	5.58M	✓	✓	✓	✗	and motion connections;
	CTR-GCN [Chen *et al.*, 2021]	CVPR	59.1	20.5	47.7	48.3	57.4	GCN	1.42M	✓	✗	✗	✗	Improved interpretability
	ST-GCN [Yan *et al.*, 2018]	AAAI	56.8	21.3	47.9	50.1	65.9	GCN	2.06M	✓	✗	✗	✗	with relations.
	GPGait [Fu *et al.*, 2023]	ICCV	54.1	29.0	50.9	60.1	71.4	GCN	1.30M	✓	✗	✗	✗	**Disadvantages:**
	TranSG [Rao and Miao, 2023]	CVPR	68.7	32.7	49.2	59.1	73.6	Transformer	0.40M	✓	✓	✗	✓	Graph topology pre-defining;
	MoCos [Rao and Miao, 2025]	AAAI	72.0	36.0	51.9	61.5	76.0	Transformer	0.40M	✓	✓	✗	✓	Complex relation modeling.

Table 2: Performance (R_1) and characteristics comparison of existing hand-crafted, sequence-based, and graph-based methods on different benchmark datasets (BIWI (S/W), IAS-Lab (A/B), KS20). Representative gait recognition methods using skeleton data are also compared following the same person re-ID evaluation protocol. We report the parameter sizes (**M**illion) and summarize their properties.

SVM [Munaro *et al.*, 2014b], KNN [Munaro *et al.*, 2014a; Gharghabi *et al.*, 2015; Nambiar *et al.*, 2017], MLP [Andersson and Araujo, 2015], Adaboost [Pala *et al.*, 2019], random forest [Elaoud *et al.*, 2021] to learn skeleton descriptors to classify different individuals. Recent studies explore supervised skeleton prototype learning paradigms by employing the contrastive loss

$$\mathcal{L}_{\text{Proto}} = \frac{1}{N_1}\sum_{k=1}^{C}\sum_{j=1}^{n_k} - \log \frac{\exp\left(\boldsymbol{S}_{k,j} \cdot \boldsymbol{p}_k / \tau\right)}{\sum_{i=1}^{C}\exp\left(\boldsymbol{S}_{k,j} \cdot \boldsymbol{p}_i / \tau\right)}, \quad (11)$$

where $\boldsymbol{S}_{k,j}$ denotes the j^{th} skeleton sequence belonging to the k^{th} identity, $\boldsymbol{p}_k$ represents the prototype generated by the feature centroid of k-class skeletons, and τ is the temperature for contrastive learning. Based on the losses in Eq. (10) and (11), diverse network architectures such as skeleton-based LSTM [Wei *et al.*, 2020; Rashmi and Guddeti, 2022], CNN [Liao *et al.*, 2020; Huynh-The *et al.*, 2020], and Transformer [Rao and Miao, 2023] are trained for SRID.

Self-Supervised SRID paradigms usually combine unlabeled learning of pretext tasks ($e.g.$, skeleton semantics learning) and labeled fine-tuning. A general form of pretext objective can be formulated as

$$\mathcal{L}_{\text{Pretext}} = \frac{1}{N_1}\sum_{i=1}^{N_1}\text{Dis}\left(\phi\left(T_1\left(\boldsymbol{S}_i\right)\right), T_2(\boldsymbol{S}_i)\right), \quad (12)$$

where $T_1(\cdot)$, $T_2(\cdot)$ denote the sequence transformation functions ($e.g.$, reverse and shift in Eq. (8)) to set the input and target of a pretext task, $\phi(\cdot)$ represent the model encoding function, and Dis$(\cdot)$ is the distance metric ($e.g.$, Euclidean distance). By employing sequence-based pretext tasks such as sparse sequential prediction [Rao *et al.*, 2021b], or graph-based tasks such as structure-trajectory prompted reconstruction [Rao and Miao, 2023], self-supervised learning encourages the model to capture high-level motion concepts and class-related spatio-temporal semantics for SRID.

Probe-Gallery	N-N		B-B		C-C		C-N		B-N	
Methods	mAP	R_1	mAP	R_1	mAP	R_1	mAP	R_1	mAP	R_1
AGE†	3.5	20.8	9.8	37.1	9.6	35.5	3.0	14.6	3.9	32.4
SM-SGE♣	6.6	50.2	9.3	26.6	9.7	27.2	3.0	10.6	3.5	16.6
SPC-MGR♣	9.1	71.2	11.4	44.3	11.8	48.3	4.3	22.4	4.6	28.9
SGELA†	9.8	71.8	16.5	48.1	7.1	51.2	4.7	15.9	6.7	36.4
SimMC†	10.8	84.8	16.5	69.1	15.7	68.0	5.4	25.6	7.1	42.0
TranSG♣	13.1	78.5	17.9	67.1	15.7	65.6	6.7	23.0	8.6	44.1
Hi-MPC†	11.2	85.5	17.0	71.2	14.1	70.2	4.9	27.2	7.5	50.1
MoCos♣	16.1	87.9	18.9	73.6	18.1	72.1	7.3	26.5	9.8	50.6

Table 3: Performance of state-of-the-art SRID methods on different conditions (**N**ormal, **C**lothes, **B**ags) of RGB-estimated CASIA-B dataset. "**C-N**" denotes "Clothes" probe set and "Normal" gallery set. † and ♣ denote sequence-based and graph-based methods.

Unsupervised SRID paradigms perform skeleton representation learning without using any labels. Existing methods mainly adopt unsupervised skeleton prototype learning frameworks, which use the feature centroids of clustering as prototypes (see Eq. (11)) [Rao and Miao, 2022a; Rao *et al.*, 2024]. Their performance could be affected by the robustness of clustering algorithms and contrastive learning mechanisms, while unlabeled pretext tasks are often combined to enhance general skeleton semantics learning.

5 Benchmarks and Evaluation

Benchmark Datasets. Table 1 summarizes the data source, identity number, skeleton amount, and viewpoint number of commonly-used datasets for SRID evaluation. They can be mainly categorized to two types: (1) **Sensor-based datasets**, where 3D skeleton data are captured from depth sensors such as Kinect, and (2) **RGB-estimated datasets**, in which skeleton data are estimated from RGB videos using 2D or 3D pose estimation models [Chen and Ramanan, 2017; Cao *et al.*, 2019]. Existing SRID datasets typically contain skeleton data collected from varying scenarios such as multiple views (KS20, KinectREID), appearance and clothing changes (BIWI RGBD-ID, IAS-Lab RGBD-ID), and different illumination conditions (KGBD), which enables a comprehensive evaluation of both short-term and long-term SRID performance. In addition to standard SRID datasets, the 3DGait dataset provides an **interdisciplinary benchmark** for evaluating the generalizability of SRID models on healthcare-related tasks, such as neurodegenerative disease prediction (see Sec. 6.2) [Rao *et al.*, 2025b].

Evaluation Metrics and Protocols. In SRID, the performance is typically evaluated based on several mertics, including Cumulative Matching Characteristics (CMC), Rank-1 accuracy (R_1), Rank-5 accuracy, Rank-10 accuracy, and Mean Average Precision (mAP) [Zheng *et al.*, 2015]. Multiple evaluation protocols are employed across varying datasets, including probe-gallery matching evaluation (main protocol evaluated in Table 2), RGB-estimated evaluation (representative protocol evaluated in Table 3), random view evaluation, cross-view evaluation [Rao *et al.*, 2024], zero-shot cross-dataset evaluation [Rao *et al.*, 2022], etc.

Comparison of Performance and Efficiency. As shown in Table 2, D_{16} and PM are two most competitive hand-crafted methods, performing well on both BIWI and IAS benchmarks. However, recent deep learning based models significantly surpass them: The latest sequence-based model (Hi-MPC) and graph-based model (MoCos) achieve superior performance to them across different datasets. Notably, the top three performers—GPGait, TranSG, and MoCos—all utilize skeleton graph representations, highlighting the efficacy of graph modeling and body relation learning for SRID task. In terms of efficiency ($cf.$ Fig. 2b), the sequence-based SimMC demonstrates the most compact design requiring only 0.15M parameters, followed by the graph-based MG-SCR (0.32M). Generally, methods utilizing Transformer architectures (GaitTR, TranSG, MoCos) exhibit significantly smaller parameter sizes than using other architectures such as LSTM. It is noteworthy that CNN-based approaches such as PoseGait incur substantially higher computational cost.

Comparison of Key Characteristics. As summarized in Table 2, hand-crafted methods design features based on domain knowledge, and employ well-established machine learning models, offering relatively high explainability and low model complexity. By contrast, sequence-based methods learn directly from raw data to capture sequential dynamics, but often overlook valuable features within body structure and joint relations. Graph-based approaches are designed to model these structural and motion connections via pre-defined topologies. By quantifying joint importance and collaboration, graph models provide a more interpretable framework. While both sequence-based and graph-based paradigms can integrate pretext tasks ($e.g.$, reconstruction) and multi-scale modeling ($e.g.$, multi-level body representations) to boost learning, only 4 out of 14 (28.6%) of these methods provide theoretical analyses of their effectiveness.

6 Challenges and Prospects

In this section, we elucidate key challenges in SRID data and models with a discussion of potential directions. We also present promising interdisciplinary applications of SRID.

Data Scarcity and Imbalance. Existing 3D skeletons are mainly collected from prevailing depth sensors such as Kinect [Shotton *et al.*, 2011], while diverse skeleton collection settings ($e.g.$, different devices in uncontrollable environments) have not be thoroughly explored. In contrast to existing RGB-based person re-ID data ($e.g.$, MSMT17 [Wei *et al.*, 2018]), the available skeleton data for person re-ID are relatively scarce and imbalanced. As shown in Table 1, existing Kinect-based SRID datasets contain 4.8 to 205.8 thousand skeletons with less than 200 identities, while the numbers of skeletons belonging to each identity often differ greatly ($i.e.$, imbalanced class distribution). Training on these datasets might incur limited generalization ability of data-driven SRID models using deep neural networks, with a risk of model over-fitting.

Data Noise. Kinect-based and RGB-estimated skeleton data possibly contain noise due to: (1) Internal factors: The key points are generated by device built-in algorithms (via depth images) or pose estimation models (via RGB images), thus the inherent limitation of algorithms ($e.g.$, precision, robustness) might result in inaccurate skeletons; (2) External factors: The quality of skeleton data may also be affected by device's tracking distance, illumination changes ($e.g.$, influence structured light in Kinect V1), and source data quality ($e.g.$, image resolution). Such inherent noise puts high demand on the model robustness against random perturbations.

To address these two challenges, higher-quality 3D skeleton data should be collected or generated. It is feasible to devise skeleton denoising models and augmentation strategies for skeleton generation. GAN-based pose generators [Yan *et al.*, 2017] and diffusion models [Ho *et al.*, 2020] could be transferred to generate and denoise 3D skeleton data. The future efforts include (a) collecting and opening new larger-scale SRID datasets, (b) transferring existing skeleton datasets from other areas to person re-ID datasets and formulating appropriate evaluation protocols, (c) estimating 2D/3D skeletons from large-scale public person re-ID datasets to construct new estimated SRID datasets, so as to advance SRID research and related person re-ID community.

Model Robustness. A few SRID models report unstable performance variations that are sensitive to different model parameter initialization, hyper-parameter settings ($e.g.$, clustering parameters) and varying data distributions [Rao and Miao, 2022a]. As these studies opt for the best-trained model with finely-tuned parameters, this practice often fails to reflect the architecture's true average performance and shows limited adaptability in real-world applications.

A key future direction is to systematically investigate how key factors ($e.g.$, model initialization, data quality) affect robustness, which is essential for developing more reliable models. Theoretical analyses of performance variations in terms of model-approximated functions and convergence conditions can also be provided for more robust model design. Moreover, multi-faceted evaluation metrics, such as performance average and standard deviation, should be reported

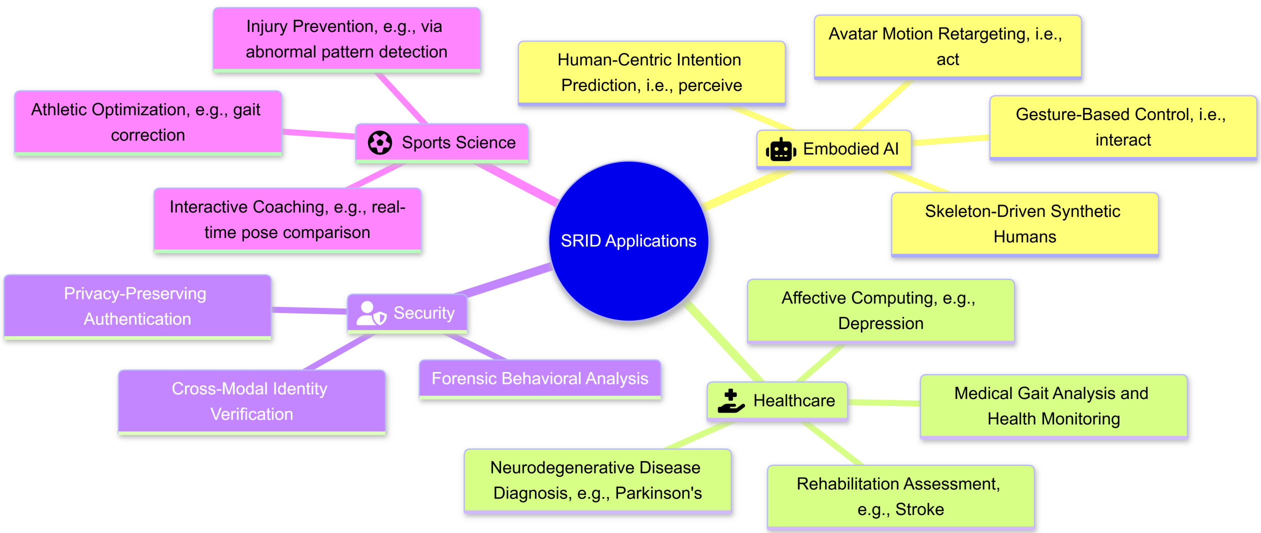


Figure 4: Interdisciplinary application landscape of SRID across three primary domains: healthcare (green box), embodied AI (yellow box), security (purple box), and sports science (pink box). Please zoom in for better view.

to better measure the overall robustness of models.

Model Generalizability. Most SRID models are trained on a single dataset with limited data sizes, views, scenes or conditions. As a result, they can hardly generalize to different data domains (*e.g.*, vulnerable to domain shifts) or real-world scenarios (*e.g.*, RGB-estimated skeleton data).

A potential solution is to exploit larger-scale SRID datasets to train the model across diverse scenarios, so as to learn more domain-general skeleton semantics. It is also feasible to explore domain adaptation or generalization techniques to co-train and transfer models. More benchmarks should be investigated for generality evaluation of SRID models.

Interpretability. Current SRID models typically lack intuitive explanations for the effectiveness of model architectures, skeleton features, and prediction results. This opacity not only poses risks of erroneous outcomes but also hinders their reliable applications at scale.

To this end, different human-friendly explanation including pose/feature visualization and corresponding language-based description can be considered. There also exist various architecture-specific explanation mechanisms, *e.g.*, class activation maps (CAM) [Zhou *et al.*, 2016] for CNN, knowledge graphs [Ji *et al.*, 2021] for graph neural networks, which could be applied to explainable skeleton learning. Moreover, large language models (LLMs) can serve as an agent to co-train the SRID model and interpret the learning process. A possible direction is to transform skeletons into time series text or pose images, and leverage LLMs or other agents to analyze the pose/feature importance.

6.1 Emerging and Open Directions

Multi-Modal Learning. Combining 3D skeletons and emerging modalities such as images [Bao *et al.*, 2025], radio frequency waves (*e.g.*, mmWave) is a promising direction, as they can provide pose or gait information from different dimensions (*e.g.*, silhouettes) to better perform recognition tasks. To this end, it is crucial to develop unified gait representations that generalize effectively across different data modalities.

Types	Methods	mAP	R_1
Hand-Crafted	D_{13}	35.0	21.9
	D_{16}	35.5	28.5
Sequence-Based	PoseGait	46.2	37.5
	SimMC	45.9	60.2
	Hi-MPC	47.5	64.1
Graph-Based	MoCos	47.4	57.8
	TransSG	48.2	60.2

Table 4: Generalization performance of representative SRID methods on the interdisciplinary benchmark 3DGait — A case study.

Cross-Modality Evaluation Protocol. To enable a fair comparison of person re-ID methods on multi-modal (*e.g.*, RGB-D) benchmarks, it is imperative to formulate a cross-modality evaluation protocol that standardizes re-ID settings (*e.g.*, probe/gallery settings, single/multi-shot recognition) of skeleton-based, depth-based, radar-based and other methods.

Skeleton Foundation Model. LLMs and pose generative models [Lucas *et al.*, 2022; Bao *et al.*, 2024] can be synergized to build a skeleton foundation model by training it on large-scale skeletons. It can hopefully be fine-tuned for diverse tasks such as skeleton visualization, augmentation, prediction, etc. Such foundation model will help investigate the limit of adaptability, generality, and interpretability of 3D skeletons, and can potentially serve as a healthcare copilot [Zhao *et al.*, 2025b] with expert rules and evidence-based reasoning [Zhao *et al.*, 2025a; Zhao *et al.*, 2021] to empower proactive health monitoring.

Privacy Protection. A unique advantage of SRID models is that they inherently protect privacy by avoiding the use of explicit appearance data. However, the illegal or irresponsible deployment of person re-ID technologies could jeopardize public security, making it crucial to establish SRID-related laws to protect social safety and personal privacy.

6.2 Interdisciplinary Prospects

As shown in Fig. 4, we list several promising applications of SRID in terms of **healthcare**, **embodied AI**, and **security**:

Healthcare. Since physiological and psychological states are intrinsically correlated with walking patterns (*e.g.*, Parkinsonian gait) [Lu *et al.*, 2025; Rao *et al.*, 2025a], the pre-trained SRID models can be transferred for neurodegenerative disease diagnosis and affective/psychological detection (*e.g.*, depression) [Lu *et al.*, 2023a; Lu *et al.*, 2023b]. Moreover, they can hopefully support medical gait analysis and rehabilitation assessment (*e.g.*, stroke) via non-intrusive monitoring in daily environments.

To demonstrate the above interdisciplinary applicability, we follow [Rao *et al.*, 2025b] to systematically evaluate various SRID models on a representative healthcare task: Neurodegenerative disease prediction. The case study in Table 4 shows that these models can capture generalizable discriminative gait patterns for this task, suggesting their promising potential for broader real-world healthcare applications.

Embodied AI. With smaller input data and lower resource requirement than RGB-based models, SRID can serve as a fundamental semantic link for intelligent agents and their interaction systems: Agents can act via skeleton-based avatar motion retargeting, perceive through human-centric intention prediction, interact using skeletal gesture based control, and evolve within skeleton-driven synthetic environments, which facilitate robust identity-aware collaboration in robotics.

Security. Leveraging the robustness of skeletal features against appearance variations [Han *et al.*, 2017], SRID advances public safety through privacy-preserving authentication and cross-modal (*e.g.*, face, fingerprint) identity verification. It also enables forensic behavioral analysis to track suspects and detect abnormal events across varying scenarios particularly where visual details are lacking.

Sports Science. By capturing fine-grained skeletal dynamics, SRID enables biomechanical analysis for athletic performance optimization (*e.g.*, running gait correction and swimming stroke refinement [Zhou *et al.*, 2026]) and sports injury prevention through abnormal movement pattern detection. Furthermore, it provides real-time skill acquisition feedback for interactive coaching (*e.g.*, pose comparison with expert templates), thereby supporting data-driven training in both competitive and recreational sports.

7 Conclusion

In this paper, we provide the first comprehensive review of 3D skeleton based person re-ID (SRID). We first define the SRID task and present a timeline to overview the major advancements of SRID. Then we formulate a taxonomy of SRID approaches in terms of different skeleton modeling and learning paradigms, systematically presenting foundational mechanisms and reviewing representative approaches. An empirical evaluation of state-of-the-art SRID methods is conducted across various benchmarks and evaluation protocols to compare their effectiveness, efficiency, and key properties. We further discuss critical challenges along with potential directions, and highlight interdisciplinary prospects of SRID.

Acknowledgments

This research is supported by the RIE2025 Industry Alignment Fund – Industry Collaboration Projects (IAF-ICP) (Award I2301E0026), administered by A*STAR, as well as supported by Alibaba Group and NTU Singapore through Alibaba-NTU Global e-Sustainability CorpLab (ANGEL). This research is also partially supported by the National Research Foundation, Singapore under its AI Singapore Programme (AISG Award No: AISG2-PhD/2022-01-034[T]).

References

[Andersson and Araujo, 2015] Virginia O Andersson and Ricardo M Araujo. Person identification using anthropometric and gait data from Kinect sensor. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, pages 425–431, 2015.

[Bao *et al.*, 2024] Peijun Bao, Zihao Shao, Wenhan Yang, Boon Poh Ng, Meng Hwa Er, and Alex C Kot. Omnipotent distillation with llms for weakly-supervised natural language video localization: When divergence meets consistency. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, 2024.

[Bao *et al.*, 2025] Peijun Bao, Chenqi Kong, Siyuan Yang, Zihao Shao, Xinghao Jiang, Boon Poh Ng, Meng Hwa Er, and Alex C. Kot. Vid-group: Temporal video grounding pretraining from unlabeled videos in the wild. In *2025 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2025.

[Bao *et al.*, 2026] Peijun Bao, Anwei Luo, Gang Pan, Alex C. Kot, and Xudong Jiang. Activityforensics: A comprehensive benchmark for localizing manipulated activity in videos. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2026.

[Barbosa *et al.*, 2012] Igor Barros Barbosa, Marco Cristani, Alessio Del Bue, Loris Bazzani, and Vittorio Murino. Re-identification with RGB-D sensors. In *ECCV Workshop*, pages 433–442. Springer, 2012.

[Bondi *et al.*, 2016] Enrico Bondi, Pietro Pala, Lorenzo Seidenari, Stefano Berretti, and Alberto Del Bimbo. Long term person re-identification from depth cameras using facial and skeleton data. In *International Conference on Pattern Recognition (ICPR) Workshop*, pages 29–41, 2016.

[Cao *et al.*, 2019] Zhe Cao, Gines Hidalgo, Tomas Simon, Shih-En Wei, and Yaser Sheikh. OpenPose: Realtime multi-person 2D pose estimation using part affinity fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(1):172–186, 2019.

[Chen and Ramanan, 2017] Ching-Hang Chen and Deva Ramanan. 3D human pose estimation= 2D pose estimation+ matching. In *CVPR*, pages 7035–7043, 2017.

[Chen *et al.*, 2021] Yuxin Chen, Ziqi Zhang, Chunfeng Yuan, Bing Li, Ying Deng, and Weiming Hu. Channel-wise topology refinement graph convolution for skeleton-based action recognition. In *ICCV*, pages 13359–13368, 2021.

[Chen *et al.*, 2022] Di Chen, Andreas Döring, Shanshan Zhang, Jian Yang, Juergen Gall, and Bernt Schiele. Keypoint message passing for video-based person re-identification. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, volume 36, pages 239–247, 2022.

[Cunado *et al.*, 2003] David Cunado, Mark S Nixon, and John N Carter. Automatic extraction and description of human gait models for recognition purposes. *Computer vision and image understanding*, 90(1):1–41, 2003.

[Elaoud *et al.*, 2017] Amani Elaoud, Walid Barhoumi, Hassen Drira, and Ezzeddine Zagrouba. Analysis of skeletal shape trajectories for person re-identification. In *International Conference on Advanced Concepts for Intelligent Vision Systems*, pages 138–149. Springer, 2017.

[Elaoud *et al.*, 2021] Amani Elaoud, Walid Barhoumi, Hassen Drira, and Ezzeddine Zagrouba. Person re-identification from different views based on dynamic linear combination of distances. *Multimedia Tools and Applications*, 80:17685–17704, 2021.

[Fan *et al.*, 2024] Chao Fan, Jingzhe Ma, Dongyang Jin, Chuanfu Shen, and Shiqi Yu. SkeletonGait: Gait recognition using skeleton maps. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, volume 38, pages 1662–1669, 2024.

[Farenzena *et al.*, 2010] Michela Farenzena, Loris Bazzani, Alessandro Perina, Vittorio Murino, and Marco Cristani. Person re-identification by symmetry-driven accumulation of local features. In *CVPR*, pages 2360–2367. IEEE, 2010.

[Fu *et al.*, 2023] Yang Fu, Shibei Meng, Saihui Hou, Xuecai Hu, and Yongzhen Huang. GPGait: Generalized pose-based gait recognition. In *ICCV*, pages 19595–19604, 2023.

[Gharghabi *et al.*, 2015] Shaghayegh Gharghabi, Faraz Shamshirdar, and Taher Abbas Shangari, *et al*. People re-identification using 3D descriptor with skeleton information. In *2015 International Conference on Informatics, Electronics & Vision (ICIEV)*, pages 1–5. IEEE, 2015.

[Han *et al.*, 2017] Fei Han, Brian Reily, William Hoff, and Hao Zhang. Space-time representation of people based on 3D skeletal data: A review. *Computer Vision and Image Understanding*, 158:85–105, 2017.

[Haque *et al.*, 2016] Albert Haque, Alexandre Alahi, and Li Fei-Fei. Recurrent attention models for depth-based person identification. In *CVPR*, pages 1229–1238, 2016.

[He *et al.*, 2022] Kaiming He, Xinlei Chen, Saining Xie, Yanghao Li, Piotr Dollár, and Ross Girshick. Masked autoencoders are scalable vision learners. In *CVPR*, pages 16000–16009, 2022.

[Hermans *et al.*, 2017] Alexander Hermans, Lucas Beyer, and Bastian Leibe. In defense of the triplet loss for person re-identification. *arXiv preprint arXiv:1703.07737*, 2017.

[Ho *et al.*, 2020] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems (NeurIPS)*, 33:6840–6851, 2020.

[Huynh-The *et al.*, 2020] Thien Huynh-The, Cam-Hao Hua, Nguyen Anh Tu, and Dong-Seong Kim. Learning 3D spatiotemporal gait feature by convolutional network for person identification. *Neurocomputing*, 397:192–202, 2020.

[Ji *et al.*, 2021] Shaoxiong Ji, Shirui Pan, Erik Cambria, Pekka Marttinen, and S Yu Philip. A survey on knowledge graphs: Representation, acquisition, and applications. *IEEE Transactions on Neural Networks and Learning Systems*, 33(2):494–514, 2021.

[Kastaniotis *et al.*, 2015] Dimitris Kastaniotis, Ilias Theodorakopoulos, Christos Theoharatos, George Economou, and Spiros Fotopoulos. A framework for gait-based recognition using Kinect. *Pattern Recognition Letters*, 68:327–335, 2015.

[Kastaniotis *et al.*, 2016] Dimitris Kastaniotis, Ilias Theodorakopoulos, George Economou, and Spiros Fotopoulos. Gait based recognition via fusing information from euclidean and riemannian manifolds. *Pattern Recognition Letters*, 84:245–251, 2016.

[Li *et al.*, 2020] Maosen Li, Siheng Chen, Yangheng Zhao, Ya Zhang, Yanfeng Wang, and Qi Tian. Dynamic multiscale graph neural networks for 3D skeleton based human motion prediction. In *CVPR*, pages 214–223, 2020.

[Liao *et al.*, 2020] Rijun Liao, Shiqi Yu, Weizhi An, and Yongzhen Huang. A model-based gait recognition method with body pose and human prior knowledge. *Pattern Recognition*, 98:107069, 2020.

[Lu *et al.*, 2023a] Haifeng Lu, Xiping Hu, and Bin Hu. See your emotion from gait using unlabeled skeleton data. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, volume 37, pages 1826–1834, 2023.

[Lu *et al.*, 2023b] Haifeng Lu, Shihao Xu, Shipeng Zhao, Xiping Hu, Rong Ma, and Bin Hu. Epic: Emotion perception by spatio-temporal interaction context of gait. *IEEE Journal of Biomedical and Health Informatics*, 28(5):2592–2601, 2023.

[Lu *et al.*, 2025] Haifeng Lu, Jiuyi Chen, Feng Liang, Mingkui Tan, Runhao Zeng, and Xiping Hu. Understanding emotional body expressions via large language models. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, volume 39, pages 1447–1455, 2025.

[Lucas *et al.*, 2022] Thomas Lucas, Fabien Baradel, Philippe Weinzaepfel, and Grégory Rogez. PoseGPT: Quantization-based 3D human motion generation and forecasting. In *ECCV*, pages 417–435. Springer, 2022.

[Munaro *et al.*, 2014a] Matteo Munaro, Alberto Basso, Andrea Fossati, Luc Van Gool, and Emanuele Menegatti. 3D reconstruction of freely moving persons for re-identification with a depth sensor. In *International Conference on Robotics and Automation (ICRA)*, pages 4512–4519. IEEE, 2014.

[Munaro *et al.*, 2014b] Matteo Munaro, Andrea Fossati, Alberto Basso, Emanuele Menegatti, and Luc Van Gool. One-shot person re-identification with a consumer depth camera. In *Person Re-Identification*, pages 161–181. Springer, 2014.

[Munaro *et al.*, 2014c] Matteo Munaro, Stefano Ghidoni, Deniz Tartaro Dizmen, and Emanuele Menegatti. A feature-based approach to people re-identification using skeleton keypoints. In *International Conference on Robotics and Automation (ICRA)*, pages 5644–5651. IEEE, 2014.

[Murray *et al.*, 1964] M Pat Murray, A Bernard Drought, and Ross C Kory. Walking patterns of normal men. *Journal of Bone and Joint Surgery*, 46(2):335–360, 1964.

[Nambiar *et al.*, 2017] Athira Nambiar, Alexandre Bernardino, Jacinto C Nascimento, and Ana Fred. Context-aware person re-identification in the wild via fusion of gait and anthropometric features. In *International Conference on Automatic Face & Gesture Recognition*, pages 973–980. IEEE, 2017.

[Nambiar *et al.*, 2019] Athira Nambiar, Alexandre Bernardino, and Jacinto C Nascimento. Gait-based person re-identification: A survey. *ACM Computing Surveys*, 52(2):33, 2019.

[Pala *et al.*, 2015] Federico Pala, Riccardo Satta, Giorgio Fumera, and Fabio Roli. Multimodal person reidentification using RGB-D cameras. *IEEE Transactions on Circuits and Systems for Video Technology*, 26(4):788–799, 2015.

[Pala *et al.*, 2019] Pietro Pala, Lorenzo Seidenari, Stefano Berretti, and Alberto Del Bimbo. Enhanced skeleton and face 3D data for person re-identification from depth cameras. *Computers & Graphics*, 79:69–80, 2019.

[Rao and Miao, 2022a] Haocong Rao and Chunyan Miao. SimMC: Simple masked contrastive learning of skeleton representations for unsupervised person re-identification. In *International Joint Conference on Artificial Intelligence (IJCAI)*, pages 1290–1297, 2022.

[Rao and Miao, 2022b] Haocong Rao and Chunyan Miao. Skeleton prototype contrastive learning with multi-level graph relation modeling for unsupervised person re-identification. *arXiv preprint arXiv:2208.11814*, 2022.

[Rao and Miao, 2023] Haocong Rao and Chunyan Miao. TranSG: Transformer-based skeleton graph prototype contrastive learning with structure-trajectory prompted reconstruction for person re-identification. In *CVPR*, 2023.

[Rao and Miao, 2025] Haocong Rao and Chunyan Miao. Motif guided graph transformer with combinatorial skeleton prototype learning for skeleton-based person re-identification. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, 2025.

[Rao *et al.*, 2020] Haocong Rao, Siqi Wang, Xiping Hu, Mingkui Tan, Huang Da, Jun Cheng, and Bin Hu. Self-supervised gait encoding with locality-aware attention for person re-identification. In *International Joint Conference on Artificial Intelligence (IJCAI)*, volume 1, pages 898–905, 2020.

[Rao *et al.*, 2021a] Haocong Rao, Xiping Hu, Jun Cheng, and Bin Hu. SM-SGE: A self-supervised multi-scale skeleton graph encoding framework for person re-identification. In *Proceedings of the 29th ACM International Conference on Multimedia*, pages 1812–1820, 2021.

[Rao *et al.*, 2021b] Haocong Rao, Shihao Xu, Xiping Hu, Jun Cheng, and Bin Hu. Multi-level graph encoding with structural-collaborative relation learning for skeleton-based person re-identification. In *International Joint Conference on Artificial Intelligence (IJCAI)*, pages 973–980, 2021.

[Rao *et al.*, 2022] Haocong Rao, Siqi Wang, Xiping Hu, Mingkui Tan, Yi Guo, Jun Cheng, Xinwang Liu, and Bin Hu. A self-supervised gait encoding approach with locality-awareness for 3D skeleton based person re-identification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(10):6649–6666, 2022.

[Rao *et al.*, 2024] Haocong Rao, Cyril Leung, and Chunyan Miao. Hierarchical skeleton meta-prototype contrastive learning with hard skeleton mining for unsupervised person re-identification. *International Journal of Computer Vision*, 132(1):238–260, 2024.

[Rao *et al.*, 2025a] Haocong Rao, Minlin Zeng, Xuejiao Zhao, and Chunyan Miao. A survey of artificial intelligence in gait-based neurodegenerative disease diagnosis. *Neurocomputing*, 2025.

[Rao *et al.*, 2025b] Haocong Rao, Jiachen Zhao, and Chunyan Miao. Llm-powered interpretable 3d gait visualization and analysis platform for interdisciplinary ai applications. In *27th International Conference on Human-Computer Interaction (HCII)*, 2025.

[Rashmi and Guddeti, 2022] M Rashmi and Ram Mohana Reddy Guddeti. Human identification system using 3D skeleton-based gait features and LSTM model. *Journal of Visual Communication and Image Representation (JVCIR)*, 82:103416, 2022.

[Shotton *et al.*, 2011] Jamie Shotton, Andrew Fitzgibbon, Mat Cook, Toby Sharp, Mark J Finocchio, Richard Moore, Alex Abenathar Kipman, and Andrew Blake. Real-time human pose recognition in parts from single depth images. In *CVPR*, pages 1297–1304, 2011.

[Wang *et al.*, 2023] Diwei Wang, Chaima Zouaoui, Jinhyeok Jang, Hassen Drira, and Hyewon Seo. Video-based gait analysis for assessing alzheimer's disease and dementia with lewy bodies. In *MICCAI Workshop on Applications of Medical AI*, pages 72–82. Springer, 2023.

[Wei *et al.*, 2018] Longhui Wei, Shiliang Zhang, Wen Gao, and Qi Tian. Person transfer gan to bridge domain gap for person re-identification. In *CVPR*, pages 79–88, 2018.

[Wei *et al.*, 2020] Chu-Chien Wei, Li-Huang Tsai, Hsin-Ping Chou, and Shih-Chieh Chang. Person identification by walking gesture using skeleton sequences. In *Advanced Concepts for Intelligent Vision Systems*, pages 205–214. Springer, 2020.

[Yan *et al.*, 2017] Yichao Yan, Jingwei Xu, Bingbing Ni, Wendong Zhang, and Xiaokang Yang. Skeleton-aided articulated motion generation. In *Proceedings of the 25th ACM international conference on Multimedia*, pages 199–207, 2017.

[Yan *et al.*, 2018] Sijie Yan, Yuanjun Xiong, and Dahua Lin. Spatial temporal graph convolutional networks for skeleton-based action recognition. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, pages 7444–7452, 2018.

[Ye *et al.*, 2021] Mang Ye, Jianbing Shen, Gaojie Lin, Tao Xiang, Ling Shao, and Steven CH Hoi. Deep learning for person re-identification: A survey and outlook. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(6):2872–2893, 2021.

[Yoo *et al.*, 2002] Jang-Hee Yoo, Mark S Nixon, and Chris J Harris. Extracting gait signatures based on anatomical knowledge. In *Proceedings of BMVA Symposium on Advancing Biometric Technologies*, pages 596–606. Citeseer, 2002.

[Zhang *et al.*, 2023] Cun Zhang, Xing-Peng Chen, Guo-Qiang Han, and Xiang-Jie Liu. Spatial transformer network on skeleton-based gait recognition. *Expert Systems*, 40(6):e13244, 2023.

[Zhao *et al.*, 2021] Xuejiao Zhao, Huanhuan Chen, Zhenchang Xing, and Chunyan Miao. Brain-inspired search engine assistant based on knowledge graph. *IEEE Transactions on Neural Networks and Learning Systems*, 34(8):4386–4400, 2021.

[Zhao *et al.*, 2025a] Xuejiao Zhao, Siyan Liu, Su-Yin Yang, and Chunyan Miao. Medrag: Enhancing retrieval-augmented generation with knowledge graph-elicited reasoning for healthcare copilot. In *Proceedings of the ACM on Web Conference 2025*, pages 4442–4457, 2025.

[Zhao *et al.*, 2025b] Xuejiao Zhao, Siyan Liu, Su-Yin Yang, and Chunyan Miao. A smart multimodal healthcare copilot with powerful LLM reasoning. *arXiv preprint arXiv:2506.02470*, 2025.

[Zheng *et al.*, 2015] Liang Zheng, Liyue Shen, Lu Tian, Shengjin Wang, Jingdong Wang, and Qi Tian. Scalable person re-identification: A benchmark. In *ICCV*, pages 1116–1124, 2015.

[Zhou *et al.*, 2016] Bolei Zhou, Aditya Khosla, Agata Lapedriza, Aude Oliva, and Antonio Torralba. Learning deep features for discriminative localization. In *CVPR*, pages 2921–2929, 2016.

[Zhou *et al.*, 2026] Kanglei Zhou, Ruizhi Cai, Liyuan Wang, Hubert P. H. Shum, and Xiaohui Liang. A comprehensive survey of action quality assessment: Method and benchmark. *Pattern Recognition*, page 113933, 2026. Review article.