

Motion capture and retargeting for character animation: A Survey

Zuling Li *

Jiangxi University of Technology, Nanchang, China

* Corresponding Author Email: 2125508519@qq.com

Abstract. Motion capture serves as the core data acquisition method for character animation, while motion retargeting enables the cross-character reuse of captured motion sequences. The integrated pipeline of capture and retargeting has become the mainstream production workflow for games, VR, and digital human creation. Traditional geometric retargeting methods are prone to foot sliding, ground penetration, and loss of contact semantics when processing motion capture data with occlusion noise, whereas recent deep learning and physical optimization algorithms have significantly improved the naturalness of retargeted motions. This paper systematically reviews representative literature on motion capture and retargeting, organized into five categories following the complete industrial workflow: multi-modal motion acquisition and preprocessing, cross-skeleton shape-adaptive retargeting, AI-assisted motion generation, physically constrained interactive retargeting, and end-to-end integrated pipelines with cross-domain applications. For each category, we elaborate on core technical schemes, inherent limitations, and typical experimental outcomes with standard citations. This survey provides systematic literature support for researchers engaged in motion capture and retargeting animation.

Keywords: Motion capture; motion retargeting; multi-sensor tracking; physical constraint optimization.

1. Introduction

Motion capture and motion retargeting are core enabling technologies for modern virtual character animation, supporting extensive industrial applications ranging from high-fidelity film and game production, immersive VR avatar interaction to sports biomechanical quantitative analysis. Traditional optical capture hardware paired with static geometric retargeting algorithms once constituted the mainstream technical route, yet such solutions suffer from limited adaptability to diverse body shapes, high deployment costs and frequent loss of interactive motion semantics. In recent years, the rapid development of deep learning has greatly promoted the progress of this field: conditional generative models and temporal convolution networks significantly improve the naturalness of cross-body retargeted animations, while differentiable physical simulation further eliminates visual artifacts such as penetration and volume distortion in migrated motions.

Nevertheless, existing review works mostly focus on isolated subdomains such as capture hardware performance or single retargeting algorithm categories and fail to sort out the complete production pipeline from raw motion data acquisition to end-to-end industrial deployment in a unified framework. To address this gap, this survey systematically organizes mainstream motion capture and retargeting research into five coherent technical modules: multi-modal capture acquisition and raw data preprocessing, cross-skeleton shape-adaptive motion retargeting, lightweight AI-assisted motion generation, physically constrained interactive retargeting correction, and end-to-end integrated pipelines with cross-domain applications. This work elaborates representative frameworks of each module, horizontally compares the core advantages, inherent limitations and applicable scenarios of different technical routes, summarizes the common bottlenecks restricting current industrial deployment, and prospects key future research directions. The conclusions of this paper can provide targeted reference for both academic researchers and animation engineering practitioners.

The rest of this paper is organized as follows: Section 2 elaborates five major technical modules along the full motion capture and retarget pipeline, including multi-modal motion acquisition and preprocessing, cross-body shape-adaptive retargeting algorithms, lightweight auxiliary motion

generation models, physically constrained interactive retargeting correction, and end-to-end integrated industrial pipelines. Section 3 conducts a horizontal comparative discussion on the core advantages, inherent limitations and applicable scenarios of different technical routes, with a summarized comparison table for intuitive reference. Finally, Section 4 draws the conclusions and summarizes the common technical bottlenecks of current systems.

2. Motion Capture and Retargeting Methods

2.1. Multi-modal Motion Capture Acquisition and Raw Data Processing

High-quality retargeted animation strictly relies on standardized, noise-free raw motion capture sequences collected by multi-sensor equipment. Optical electromagnetic hybrid capture equipment can compensate for the occlusion defects caused by loose clothing during human movement and obtain stable skeleton data for subsequent retargeting calculation [1]. The ELMO LiDAR markerless capture framework converts low-frame sparse point cloud data into dense 60fps continuous motion skeletons, supporting real-time streaming input for online retargeting engines [2]. Shape-aware inertial tracking effectively suppresses the long-term cumulative drift error of IMU sensors during long capture sessions, and outputs smooth human kinematic trajectories [3] (Figure 1-2).

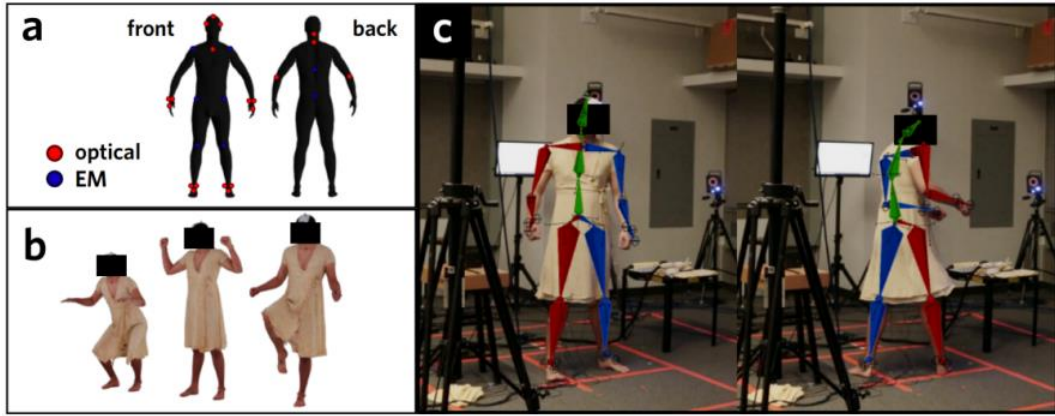


Figure 1. a) Red optical markers and blue electromagnetic locators are mounted on the test participant. b) Scanned body data of the participant in varied poses while wearing loose dresses. c) High-fidelity motion tracking results with a combined sensor-driven skeleton overlaid. Loose garments deform drastically during fast movements and create severe marker occlusion, so optical-only tracking cannot achieve equivalent accuracy. Relevant animated clips are available in the supplementary video [1].

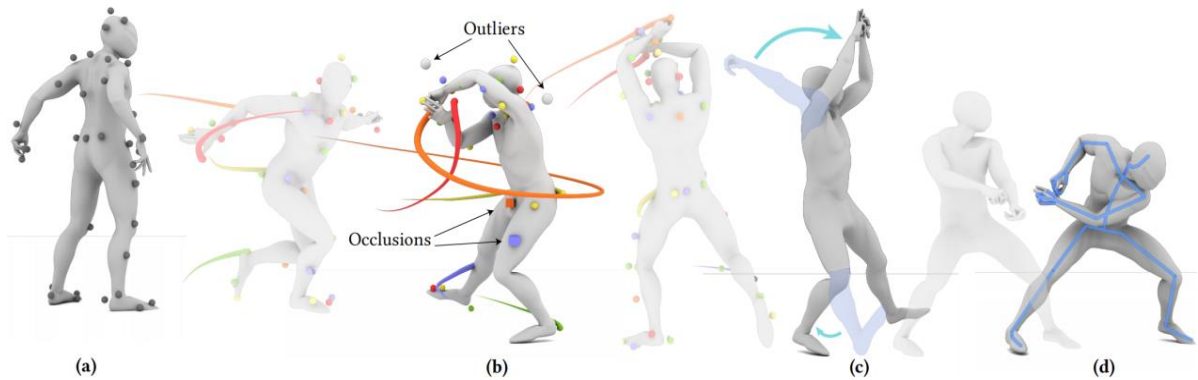


Figure. 2 Subgraph (a) presents raw unlabeled full-body motion capture point clouds in black. The RoMo model creates tracklets to utilize temporal cues and label markers with distinct colors in subgraph (b). A hybrid inverse kinematics approach shown in (c) enables the model to reconstruct full-body motions as shown in (d). RoMo remains robust to cubic occlusions and grey outlier points, and accurately recovers broad body movements as well as subtle multi-finger gestures[4].

Before implementing cross-body retargeting, raw capture data must go through standardized denoising and reconstruction preprocessing. The bone-length constrained filtering algorithm eliminates skeleton stretching distortion caused by unconstrained coordinate smoothing operations on original point clouds [5]. The RoMo robust solver automatically labels disordered optical marker point clouds and completes full-body inverse kinematics reconstruction under heavy occlusion interference, outputting unified joint rotation parameters applicable to all retargeting algorithms (Figure 2) [4]. The MoVi multi-modal dataset integrates optical, inertial and video motion samples under tight and loose clothes and has become the standard benchmark for testing capture preprocessing and retargeting effects [6]. Low-cost markerless pose estimation schemes will generate obvious frame jitter during fast dance movements, and such noise will be continuously amplified after cross-skeleton motion migration [7]. (Figure 3).

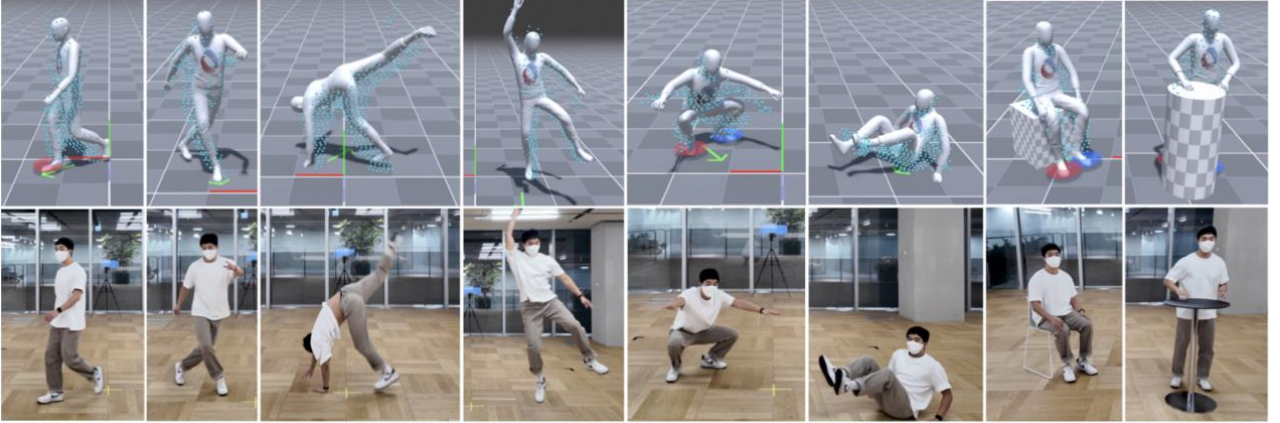


Figure. 3 The proposed ELMO architecture achieves real-time upsampling of sparse 20 fps LiDAR point cloud inputs to generate dense 60 fps human motion capture sequences [2].

2.2. Motion Retargeting Across Skeletons and Body Shapes

Cross-skeleton motion retargeting is the core link to realize the secondary utilization of limited motion capture data on virtual avatars with different bone structures and body proportions. Traditional unified geometric retargeting adopts fixed joint offset mapping rules, which are simple to calculate but easy to lose the limb-object contact semantics in interactive motion capture clips [8]. The InterFaceRays ray-based contact feature extraction method separates the surface contact information between human bodies and props from motion sequences and can completely retain interactive semantic relations after motion retargeting to different character models [9]. The ultrafast online retargeting scheme simplifies complex Jacobian matrix derivation steps, meeting the low-latency rendering requirements of real-time game engines [10].

Deep learning-based shape adaptive retargeting has gradually become the mainstream research direction for diverse human body animations. The HUMOS conditional VAE introduces cycle consistency and ZMP physical stability loss terms, effectively eliminating foot sliding and body penetration artifacts generated when retargeting motion to characters with different BMI values [11]. The ShapeVerse framework maps SMPL human body parameters to capsule rigid bodies, and generates biomechanically reasonable walking motions suitable for thin, standard and overweight virtual characters [12]. ReConForM designs sparse mesh embedding and adaptive terrain semantic loss, supporting multi-character interactive motion retargeting on uneven ground scenes [13]. Data-driven gait analysis further optimizes the naturalness of retargeted walking postures for extreme body size models [14]. (Figure 4).

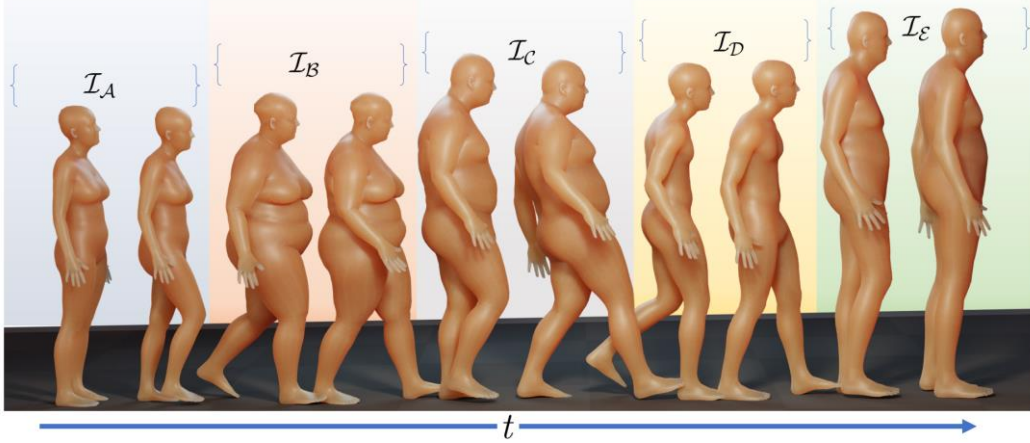


Figure 4. Humans of diverse body shapes produce different kinematic behaviors when executing the same movements. The HUMOS system generates natural, physically consistent, dynamically balanced human motions conditioned on body identity. The system utilizes newly designed identity-preserving cycle consistency loss and differentiable physical stability constraints to learn shape-correlated motion manifolds. The displayed walking animation featuring a mid-sequence skip is produced for five distinct body identities. The underlying motion stays fixed while body identity updates every thirty frames to illustrate shape-aware motion generation [11].

2.3. AI-Generated Human Motion and 3D Pose Estimation

The insufficient quantity and single style of existing motion capture asset libraries restrict the diversity of animations that can be reused through retargeting. Lightweight temporal generation networks are widely used as auxiliary modules to expand motion samples without additional large-scale capture work. Temporal graph convolution structures suppress frame jitter noise caused by sparse IMU capture equipment and significantly improve the smoothness of motion sequences after retargeting processing [3]. Dual-source independent training pipelines split 2D image data and 3D motion capture data for model training, removing the high production cost of paired motion samples required by most retargeting networks, and the MoVi multi-modal dataset provides sufficient mixed motion materials for such training architectures [6]. (Figure 5).

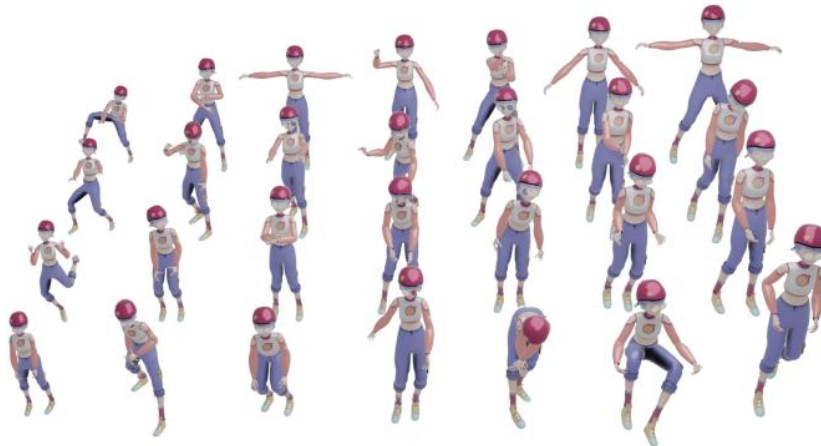


Figure 5. The MID dataset includes paired IMU-motion data recorded from 20 subjects of varying body shapes [3].

Such lightweight generative models feature small storage overhead and low training cost, and can be flexibly embedded into multi-stage capture-retarget integrated workflows to supplement motion assets for small-budget animation projects. However, limited by model capacity, these methods can only generate basic cyclic locomotion such as walking and running, and cannot reconstruct complex multi-person and multi-object interactive capture motions with rich semantic information.

2.4. Cross-Character Interactive Motion Retargeting

Even if accurate single-body motion retargeting results are obtained, it is still difficult to retain natural interactive logic when multiple virtual characters interact with each other or with environmental objects. Simple skeleton mapping often leads to loss of limb contact, penetration between characters and distorted interactive postures. InterFaceRays provides dedicated surface feature extraction logic for furniture and scene props, which can lock the spatial correspondence between limbs and objects during retargeting and maintain complete interactive semantics after motion migration [9]. Meanwhile, ReConForM embeds terrain adaptive semantic loss, enabling the retargeted motion to automatically adjust limb poses according to uneven ground and obstacle distribution, avoiding floating and penetration artifacts on complex scenes [13]. (Figure 6).

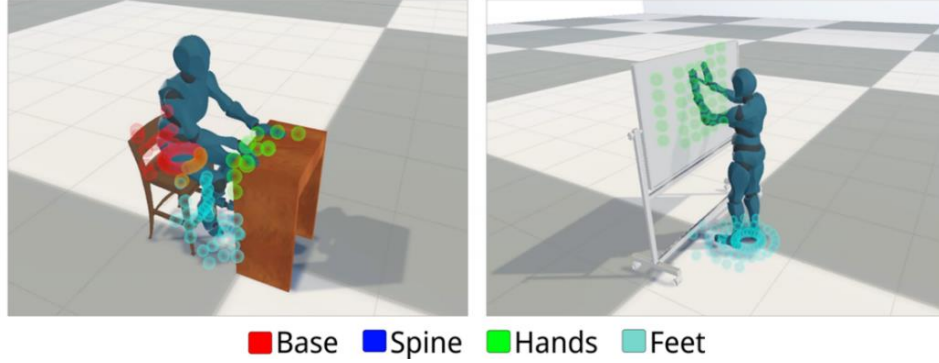


Figure 6. InterFaceRays visualization. Colored spheres on furniture surfaces encode IIW values for different body parts — red for base, blue for spine, green for hands, and cyan for feet.

The original six IIW values are simplified to four colors in this figure, with both hands and both feet respectively sharing the same color [9].

To further optimize interactive motion authenticity, physical constraint correction is introduced as a post-processing link. The AMP framework uses scene discriminators to calibrate the force relationship between characters and environments captured by motion capture, so that retargeted movements conform to real physical collision rules [15]. The PIMT self-behavior cloning mechanism smooths the transition between independent interactive action clips, eliminating stiff and unnatural posture switching after retargeting different motion segments [16]. The PhysRig differentiable skinning technology solves the volume shrinkage distortion of multi-character models in interactive scenes caused by traditional linear blending and restores the real volume state of human bodies during contact and extrusion [17]. All interactive retargeting correction modules operate independently after basic skeleton mapping and can be freely combined with various cross-body retargeting algorithms, though additional computational delays will be generated during real-time rendering on lightweight devices. (Figure 7).

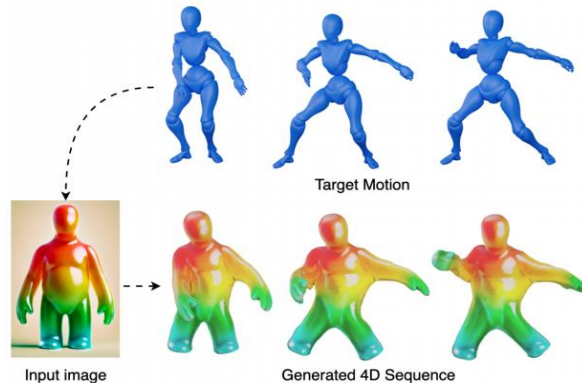


Figure 7. Built on differentiable physics rules, PhysRig can realize pose transfer for generated rigid bodies [17].

2.5. End to End Capture-Retarget Integrated Workflow and Cross Domain Applications

Separate motion capture equipment and independent retargeting algorithms will generate a large amount of intermediate data conversion cost in actual animation production. End-to-end integrated pipelines integrate data collection, denoising preprocessing and real-time motion remapping into a unified workflow, greatly simplifying the production process of games, digital humans and VR content. Game engine native tools represented by DyMM embed motion asset storage and context-sensitive retargeting modules inside Unity, realizing one-stop generation of scene adaptive character animations without cross-platform data transmission [18]. The PRECISION precomputation tool pre-extracts environmental contact information from raw capture data and automatically adjusts the movement trajectory of retargeted characters according to dynamic terrain to avoid ground penetration [19].

For amateur creators and lightweight production demands, low-cost integrated schemes have been widely studied. Sketch two-stage editing tools combine hand-drawn key trajectories with existing motion capture libraries to complete rapid lightweight retargeting without complete capture equipment [20]. Kinect matching Blender built-in retargeting module and consumer VR full-body tracking pipeline can deduce complete lower limb movement only through head-mounted and handle signals, breaking the space limitation of professional optical capture studios [21][22].

This set of integrated capture retarget pipelines is not only used in entertainment animation but also extended to multiple cross-domain scenes. In animation teaching, immersive VR capture retarget training system lets students compare the retargeting effect of different body shapes and effectively improves the ability to design natural movements [23]. In sports biomechanics research, athlete capture data is retargeted to unified standard human skeletons through integrated tools, realizing consistent quantitative analysis of joint motion [24][25]. Existing overview papers have sorted out the hardware architecture of such real-time integrated systems, laying a foundation for subsequent lightweight iteration and optimization [26][27]. (Figure 8).

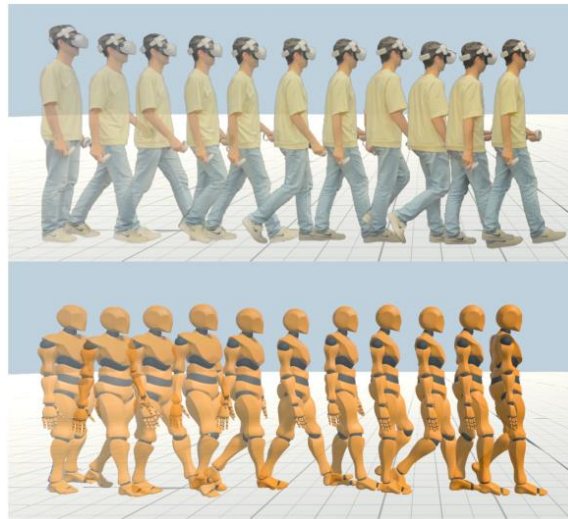


Figure 8. VR user-controlled walking avatar; all avatar motions are derived only from headset and dual controller tracking information [27].

3. Discussion

Mainstream pipelines for motion capture and motion retargeting are systematically classified in accordance with the full industrial animation workflow. Traditional geometric retargeting and handcrafted filtering only adopt fixed linear joint mapping rules and cannot model the nonlinear coupling between body proportions, gait and terrain constraints. Deep learning architectures including conditional VAE, temporal graph convolution and recurrent models can capture complex nonlinear kinematic correlations embedded within raw motion capture sequences. After lightweight fine-tuning,

such models adapt to diverse body shapes and interactive scenes, and capture long-range temporal dependencies that linear mathematical formulas fail to express. On the contrary, rule-based manual modeling consumes less training resources and requires smaller motion capture datasets. Despite the two technical routes having different implementation logic, both can generate smooth, artifact-free motion sequences and respond to real-time control inputs. To quantitatively compare representative pipelines, Table 1 summarizes their respective advantages and limitations across core evaluation dimensions.

Table 1. Comparative analysis of typical motion capture and retargeting technologies.

Technical Category	Representative Schemes	Core Advantages	Main Limitations
Multi-modal Motion Capture and Raw Data Preprocessing	Optical-electromagnetic hybrid capture [1]	Multi-sensor fusion enables high-precision motion acquisition under varied shooting environments; robust preprocessing algorithms eliminate marker noise and reconstruction distortion	Different device modalities have inherent tracking drift and occlusion artifacts; denoising and reconstruction introduce additional computational latency
	RoMo robust marker solver [4]		
	ELMO LiDAR upsampling [2]		
Cross-skeleton and Shape-adaptive Motion Retargeting	Unified geometric retargeting [8]	Enables natural motion transfer across diverse bone structures and body shapes; deep learning approaches effectively preserve contact semantics during migration	Data-driven methods rely on large multi-body motion capture training datasets; traditional geometric approaches tend to lose interactive semantic information
	HUMOS conditional VAE [11]		
	ReConForM contact-aware retargeting [13]		
AI-Generated Human Motion and 3D Pose Estimation	Temporal graph convolution generator [3]	Lightweight generative models supplement motion libraries at low data cost; flexible embedding into multi-stage capture-retarget workflows	Limited model capacity restricts outputs to basic locomotion patterns; cannot reconstruct complex multi-person interactive motions with rich semantics
	MoVi multi-modal dataset [6]		
Cross-Character Interactive Motion Retargeting	InterFaceRays contact feature extraction [9]	Preserves limb-object contact semantics during character-scene interaction; physical post-processing eliminates penetration and stiffness artifacts	Independent correction modules introduce additional computational overhead; real-time rendering performance degrades on lightweight devices
	AMP physical correction [15]		
	PhysRig differentiable skinning [17]		
End-to-End Integrated Workflow and Cross-Domain Applications	DyMM game engine integration [18]	Unified capture-to-retarget pipeline greatly simplifies production workflow; extends to animation teaching, sports biomechanics and other cross-domain scenarios	Most integrated tools are optimized only for standard adult virtual skeletons; low-cost schemes suffer from reduced motion smoothness under fast movements
	Sketch two-stage motion editor [20]		

4. Conclusion

This survey systematically reviews the full technical chain of motion capture and motion retargeting for virtual character animation, covering five core modules including multi-modal capture preprocessing, shape-adaptive cross-body retargeting, lightweight auxiliary motion generation, physically corrected interactive retargeting and end-to-end integrated industrial pipelines. Comparative analysis shows that deep learning-based models represented by conditional VAE and graph convolution architectures have stronger nonlinear fitting ability for complex human kinematics

and deliver better performance on diverse body shape adaptation, while traditional geometric mapping and rule-based algorithms maintain lower computational cost and weaker data dependence, fitting lightweight real-time scenarios better. Current capture-retarget systems still face common bottlenecks such as high collection cost of multi-body paired datasets, extra inference latency brought by physical correction modules, poor generalization for extreme body morphologies and insufficient semantic retention in multi-character interactive scenes. Future research can be promoted toward unified end-to-end frameworks, multi-sensor fusion capture, lightweight edge deployment and broader cross-domain applications, to further improve motion realism, algorithm generalization and industrial practicability of the whole technical pipeline.

References

- [1] BOLANOS L, WYDER-HODGE P, LIN X D, et al. Measuring human motion under clothing[C]. SIGGRAPH Asia 2024 Conference Papers, 2024.
- [2] JANG D K, YANG D, JANG D Y, et al. ELMO: Enhanced Real-time LiDAR Motion Capture through Upsampling[J]. ACM Transactions on Graphics, 2024, 43(6): 237.
- [3] LU Y, SHI Z Y, WU Y H, et al. Shape-aware Inertial Poser: Motion Tracking for Humans with Diverse Shapes[J]. ACM Transactions on Graphics, 2025, 44(6): 221.
- [4] PAN X Y, ZHENG B W, JIANG X W, et al. RoMo: a robust solver for full-body unlabeled optical motion capture[C]. SIGGRAPH Asia 2024 Conference Papers, 2024.
- [5] DONG P Y, ZHANG Y, ZHANG Z H. Data processing method of 3D motion capture based on bone constraint[C]. 2021 33rd Chinese Control and Decision Conference, 2021: 833.
- [6] GHORBANI S, MAHDAVIANI K, THALER A, et al. MoVi: a large multi-purpose human motion and video dataset[J]. PLOS ONE, 2021, 16(6): e0253157.
- [7] LABUGUEN R T, NEGRETE S B, KOGAMI T, et al. Performance evaluation of markerless 3D skeleton pose estimates with pop dance motion[C]. IEEE Conference, 2020.
- [8] ZHANG H D, CHEN Z K, XU H C, et al. Unified cross-structural motion retargeting for humanoid characters[J]. IEEE Transactions on Visualization and Computer Graphics, 2025, 31(7): 3863.
- [9] JIN T, LEE Y, LEE S H. InterFaceRays: interaction-oriented furniture surface representation for human pose retargeting[J]. Computer Graphics Forum, 2025, 44(2): e70094.
- [10] GUO T Z, CHEN Z D, JIANG Y, et al. Ultrafast and controllable online motion retargeting for game scenarios[J]. ACM Transactions on Graphics, 2025, 44(6): 261.
- [11] TRIPATHI S, TAHERI O, LASSNER C, et al. HUMOS: Human Motion Model Conditioned on Body Shape[C]. Proceedings of the 18th European Conference on Computer Vision, 2024.
- [12] VYAS B, O’SULLIVAN C. ShapeVerse: Physics-based Characters with Varied Body Shapes[C]. Eurographics 2024 Posters, 2024.
- [13] CHEYNEL T, ROSSI T, BELLOT-GURLET B, et al. ReConForM: Real-time Contact-aware Motion Retargeting for more Diverse Character Morphologies[J]. Computer Graphics Forum, 2025.
- [14] SHEWHORAK S. A Data-driven Approach to Motion Diversification over Increases in Body Morphology[D]. Teesside University, 2017.
- [15] HASSAN M, GUO Y R, WANG T W, et al. Synthesizing physical character-scene interactions[C]. SIGGRAPH '23 Conference Proceedings, 2023.
- [16] DENG Y B, LI Z, XIE N, et al. PIMT: physics-based interactive motion transition for hybrid character animation[C]. Proceedings of the 32nd ACM International Conference on Multimedia, 2024.
- [17] ZHANG H, X H, F C, et al. PhysRig: Differentiable Physics-Based Skinning and Rigging Framework[C]. 2025 IEEE/CVF International Conference on Computer Vision, 2025:6609-6620.
- [18] HÄFLIGER A, KURABAYASHI S. Dynamic motion matching: design and implementation of a context-aware animation system for games[J]. International Journal of Semantic Computing, 2022, 16: 189-212.
- [19] KAPADIA M, XIANGHAO X, NITTI M, et al. PRECISION: Precomputing Environment Semantics for Contact-Rich Character Animation[C]. SIGGRAPH I3D 2016, 2016:29-37.

- [20] PENG Y, ZHAO C, HUANG Z, et al. Two-Stage Motion Editing Interface for Character Animation[C]. Proceedings of the ACM SIGGRAPH/Eurographics Symposium on Computer Animation, 2021.
- [21] LAMBERTI F, et al. Virtual character animation based on affordable motion capture and reconfigurable tangible interfaces[J]. IEEE Transactions on Visualization and Computer Graphics, 2018, 24(5): 1742.
- [22] PONTON J L. Motion Matching for Character Animation and Virtual Reality Avatars[D]. Universitat Politècnica de Catalunya, 2022.
- [23] JIANG X Y, LUO M Y, IBRAHIM Z, et al. Immersive VR-MoCap for creative motion design in character animation training[J]. Computers, 2026, 15(5): 284.
- [24] HAMILTON R I, GLAVCHEVA-LALEVA Z, MILON M I H, et al. Comparison of computational pose estimation models for joint angles with 3D motion capture[J]. Journal of Bodywork & Movement Therapies, 2024, 40: 315-319.
- [25] SUO X, TANG W D, LI Z. Motion capture technology in sports scenarios: a survey[J]. Sensors, 2024, 24(9): 2947.
- [26] KAMMERLANDER R K, PEREIRA A, ALEXANDER S. Using virtual reality to support acting in motion capture with differently scaled characters[C]. 2021 IEEE Virtual Reality and 3D User Interfaces, 2021: 1402.
- [27] ZHU X D, LI K F. Real-time motion capture: an overview[C] //2016 10th International Conference on Complex, Intelligent, and Software Intensive Systems, 2016: 522.