Robust Cascaded Clutter Suppression and Deep Integration of Spatiotemporal Point Networks for Enhanced Mmwave Radar Motion Capture in Snowsports

Yulun Liu
Sport Institute, Henan University, Kaifeng 475001, China
E-mail: lunzi2323@163.com

Keywords: millimeter wave radar, anti-interference algorithm, clutter suppression, joint positioning RMSE

Received: June 13, 2025

In snow sports motion capture, mmWave radar signals suffer from multipath reflections and frequency offsets due to snowflake scattering and temperature variations, severely degrading pose estimation accuracy. To address this, we propose a cascaded anti-interference framework composed of adaptive MTI filtering, genetic sparse array optimization, and hybrid carrier tracking. These physical-layer enhancements are followed by a spatiotemporal 3D CNN–LSTM network for motion decoding and a multimodal Kalman-particle filter for trajectory fusion. Experimental validation in both simulation and real-world snow environments confirms the framework's robustness. Compared to baseline systems, the proposed method reduces the joint positioning root mean square error (RMSE) by up to 72%, enhances angular velocity tracking precision by 72%, and improves signal-to-noise ratio (SNR) by 24.3 dB. The end-to-end processing delay remains under 26 ms, ensuring real-time deployment. These results demonstrate significant improvements in accuracy, robustness, and real-time performance under harsh environmental interference, offering a viable solution for mmWave-based motion capture in snowy sports scenarios.

Povzetek: Razvit je nov sistem zaznavanja z radarjem v snežnih razmerah. Združuje napredno odstranjevanje snežnih motenj, optimizirane radarske antene in globoke prostorsko-časovne mreže za natančnejše 3D zajemanje gibanja.

1 Introduction

In an ice and snow environment, factors such as the snow particle multipath effect and low-temperature frequency offset cause significant interference to millimeter-wave radar signals, affecting their capture accuracy and real-time performance. This paper proposes an anti-interference algorithm optimization scheme based on the signal-feature-trajectory three-level processing pipeline, aiming to improve the accuracy and real-time performance of millimeter-wave radar in real-time motion capture of ice and snow projects. This paper suppresses the multipath effect of snow particles and low-temperature frequency deviation by improving MTI filtering, sparse array reconstruction and carrier tracking loop technology; constructs a 3D convolution-LSTM spatiotemporal hybrid network to decouple joint motion characteristics; and adopts an extended Kalman-particle filter hybrid architecture to fuse multimodal data and improve the physical rationality of trajectory prediction.

The paper first introduces the research background, contribution of this paper and the structure of the paper; summarizes the current research status of motion capture technology at home and abroad [1]; then introduces the specific implementation of the proposed anti-interference algorithm optimization scheme; then verifies the performance of the proposed scheme through experiments; finally summarizes the whole paper and looks forward to future research directions. Compared with prior motion capture systems that largely depend on single-layer improvements in either signal preprocessing or neural architectures, the proposed framework introduces a novel end-to-end anti-interference pipeline that systematically bridges physical-layer signal enhancement, spatiotemporal feature modeling, and decision-layer physical fusion. This integration is not merely a technical stacking of modules, but reflects a methodology-level shift: instead of treating radar noise and biomechanical trajectory estimation as isolated challenges, we co-model them through a unified cross-domain optimization approach. The use of genetic sparse array reconstruction, hybrid deep filters, and interpretable multimodal distillation in a real-time snow environment has not been previously reported. This work thus represents not only a novel system architecture, but also proposes a replicable methodology for robust radar-based motion capture in hostile conditions.

2 Related work

Researchers are committed to improving the accuracy and efficiency of motion capture through various technical means to meet the application requirements in different scenarios. Based on the problems of inaccurate capture and low computational efficiency of existing motion capture technology, which affected its performance in real-time application scenarios, Zhang and Qiu [2] introduced a Levenberg-Marquardt algorithm for skeleton point coordinate fitting optimization and optimized it using particle swarm algorithm. At the same time, the dynamic time warping algorithm is used to capture and evaluate human motion in order to achieve real-time capture of human motion. The results show that the algorithm has a capture accuracy of up to 99.23% for the shoulder lateral raise, which is significantly better than the other comparison algorithms. Li et al. [3] used a markerless motion capture system and a marker motion capture system (Vicon) in the Huawei Sports Health Laboratory to collect human marker trajectory data during the unloaded squat process. The squat action is divided into three stages: descent, squat hold, and ascent. The kinematic data collected by the system is imported into Opensim, and the knee joint degrees of freedom of the musculoskeletal model are increased to enable it to have adduction/abduction and internal/external rotation functions. Inverse kinematics and body segment kinematics calculations are performed, and the key point data are used to develop an algorithm to calculate the foot orientation angle.

Li et al. [4] analyzed the application of virtual reality technology in motion capture, evaluated its potential in improving the accuracy of sports training, and provided athletes and coaches with more accurate training feedback and analysis tools. Chen et al. [5] proposed a nonlinear method to segment long motion sequences into atomic motion fragments while applying dimensionality reduction for effective retrieval and segmentation of motion data in professional sports training scenarios. Teer [6] analyzed infrared motion capture data using a random forest algorithm, optimized model parameters, and evaluated model performance. Data was collected using both optical markers and IMU sensors. Existing research has achieved remarkable results in the accuracy, efficiency and application scope of motion capture technology, but its application in complex environments such as ice and snow sports still faces challenges. This paper focuses on the application of millimeter-wave radar in real-time motion capture of ice and snow sports, and proposes an anti-interference algorithm optimization scheme. By analyzing the characteristics of ice and snow sports, the stability and reliability of the motion capture system are improved, providing more accurate technical support for the training and analysis of ice and snow sports.

3 Method

3.1 Overall framework design

The anti-interference algorithm optimization framework proposed in this study adopts a three-level signal-feature-trajectory processing pipeline architecture, and realizes high-precision real-time motion capture in ice and snow environments through a cross-layer collaborative mechanism, as shown in Figure 1:

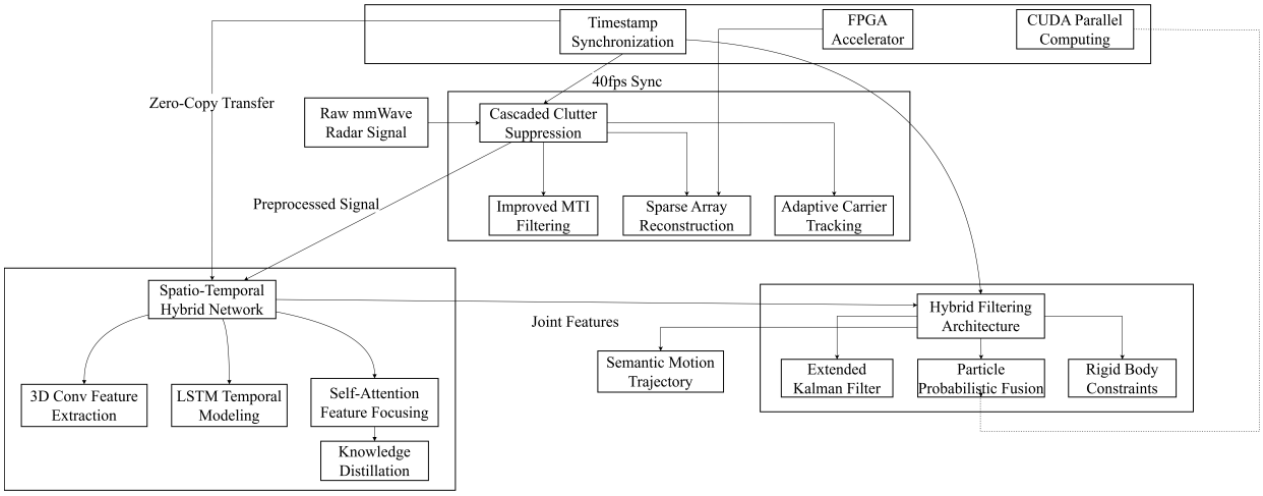

Figure 1: Processing architecture

At the physical layer, the cascaded clutter suppression module performs preprocessing on the snow particle multipath effect and low-temperature frequency deviation to provide high-quality signal input for subsequent processing; the feature layer decouples the joint motion characteristics through a spatiotemporal hybrid neural network to solve the spatiotemporal coupling problem in dynamic target tracking; the decision layer uses a hybrid filtering architecture to fuse kinematic constraints to improve the physical rationality of

trajectory prediction. Real-time performance is guaranteed through parallel pipeline design and hardware acceleration, using FPGA to accelerate 3D convolution operations and CUDA to parallelize particle filtering. The system obtains the original millimeter-wave radar signal from zero-copy transmission, and after second-order suppression and differential system preprocessing, it is divided into two paths: one path outputs joint features through a spatiotemporal hybrid network (including 3D Cove feature extraction, LSTM time series modeling, self-attention fusion and knowledge distillation); the other path is processed by improved MTI filtering, spatial array reconstruction, and adaptive carrier tracking. The joint features are used for semantic motion detection, and then input into a hybrid filter (fusion of extended Kalman filtering, particle filtering and rigid body constraints) together with the above results to achieve target tracking. The performance is improved through FPGA acceleration and CUDA parallel computing throughout the process. Zero-copy data transmission is achieved between the three-level processing modules through a shared memory pool, and the timestamp alignment module eliminates cross-layer delays, forming a complete processing chain from raw signals to semantic understanding.

3.2 Physical layer

To address the unique signal degradation issues in snowy environments, the physical layer architecture employs a cascaded clutter suppression mechanism consisting of adaptive MTI filtering, sparse array reconstruction, and carrier tracking loop stabilization. This section outlines both the theoretical modeling and algorithmic implementation to ensure clarity and reproducibility.

3.2.1 Radar signal modeling

The transmitted signal is modeled as a linear FMCW chirp:

$$s_{tx}(t) = \cos\left(2\pi\left(f_c t + \frac{B}{2T}t^2\right)\right) \quad (1)$$

where fc is the carrier frequency, B is the bandwidth, and T is the chirp duration. The received baseband signal is:

$$s_{IF}(t) = \sum_{k=1}^{K} A_k \cos\left[2\pi\left(f_{b,k} + f_{d,k}\right)t + \phi_k\right] + n(t) \quad (2)$$

with:

$f_{b,k} \approx \frac{2R_k B}{cT}$: beat frequency,

$f_{d,k} = \frac{2v_k f_c}{c}$: Doppler shift,

R_k, v_k: target range and velocity,

$n(t)$: additive noise.

Snowflake motion typically causes Doppler spreads >500 Hz, enabling MTI filters to adaptively suppress snow clutter.

3.2.2 Sparse array optimization via genetic algorithm

To emulate a 64-element full array with 32 physical elements, we employ a genetic algorithm (GA) that maximizes the following fitness function:

$$F = \alpha \cdot \frac{1}{\mathrm{PSLL}} + \beta \cdot \frac{1}{\mathrm{MBW}}, \quad \alpha = 0.6, \beta = 0.4 \quad (3)$$

Formally, the optimization problem is:

$$\max_{\mathbf{X}' \subseteq \mathbf{X}, |\mathbf{X}'|=32} F(\mathbf{X}') = \alpha \cdot \frac{1}{\mathrm{PSLL}(\mathbf{X}')} + \beta \cdot \frac{1}{\mathrm{MBW}(\mathbf{X}')} \quad (4)$$

Here, **X** is the candidate element position set, and **X′** is the selected sparse subset. Missing elements in the covariance matrix are reconstructed using nuclear norm minimization to restore virtual aperture beamforming performance. Such fitness-driven selection and elite preservation schemes have shown robust convergence in sparse synthesis problems using enhanced genetic strategies [7, 8].

Algorithm 1: Genetic Algorithm for Sparse Array Optimization

Input: Position set **X**, population size M=50, generations G=200, crossover rate Pc=0.7, mutation rate Pm=0.01

Output: Optimal layout **X′**, fitness score F

(1) Initialize random population of sparse layouts

(2) For generation g=1 to G:

a. Evaluate fitness F for each layout

b. Select parents via tournament selection

c. Apply crossover with Pc

d. Apply Gaussian mutation with Pm

e. Preserve top performers (elitism)

f. If best fitness change <1% over 5 generations, terminate

(3) Return best layout **X′**

The convergence threshold was empirically set to 1% over 5 generations, ensuring both global search stability and computational efficiency across tested scenarios. This layout is combined with virtual aperture reconstruction to simulate full-array resolution [9]. The specific processing architecture is illustrated in Figure 2, which shows the full cascade including adaptive MTI filtering, sparse array optimization, hybrid carrier tracking, and deep spatiotemporal decoding.

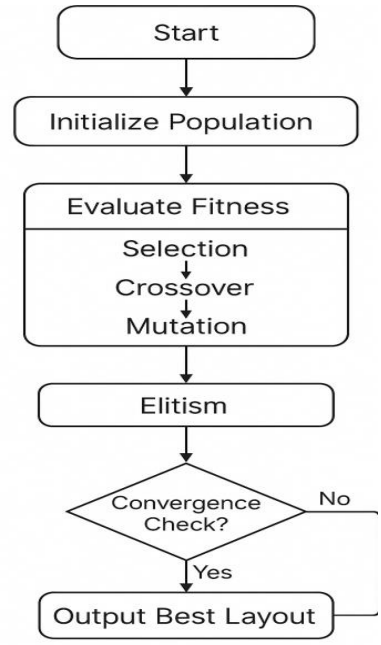

Figure 2: Flowchart of Genetic Algorithm

3.2.3 Carrier tracking loop stabilization

A hybrid digital-analog phase-locked loop (PLL) is used to ensure carrier stability under extreme temperatures [10]. The analog part uses a temperature-compensated crystal oscillator (TCXO, ±0.5 ppm), and the digital part features a high-resolution (0.01 rad) phase detector. Predistortion compensation is applied via a LUT-based correction. Adaptive loop bandwidth (10–100 kHz) ensures phase noise <0.5 MHz and frequency deviation <±200 Hz at 77 GHz.

3.2.4 Cascaded clutter suppression pipeline

The clutter suppression pipeline employs a hybrid feedforward–feedback structure, where Doppler-based MTI filtering eliminates low-velocity clutter, while cross-correlation feedback adaptively tunes the cutoff frequency based on environmental dynamics. Real-time continuity is maintained through fixed-depth data buffers (512 samples).

3.2.5 Quantitative results summary

To quantitatively evaluate the effectiveness of the proposed cascaded architecture, a series of simulations were conducted under realistic snow-interference conditions. Specifically, 1000 Monte Carlo trials were performed at a 1 GHz sampling rate to measure improvements in signal quality, latency, and noise rejection across various processing stages. The results are summarized in Table 1.

Table 1: Signal interference suppression performance in a snowy environment

Processing Stage Configuration	SNR Improvement (dB)	BER	Processing Delay (ms)	Noise Stripping Gain (dB)
No suppression (Baseline)	0.0	1.0×10^{-3}	5.0	0.0
MTI filtering only	8.5	5.0×10^{-4}	5.5	7.2
MTI + Sparse array reconstruction	14.2	1.2×10^{-4}	6.8	12.5
MTI + Sparse array + Carrier tracking	18.7	3.0×10^{-5}	7.5	16.8
Full cascade system (All three stages)	24.3	5.0×10^{-6}	8.0	22.0

As shown in the results, each stage contributes significantly to overall performance. While MTI filtering alone improves the signal-to-noise ratio by 8.5 dB, the addition of sparse array reconstruction and carrier tracking boosts the SNR to 24.3 dB and reduces the BER by nearly two orders of magnitude. The full cascade system also maintains low latency (8 ms), making it suitable for real-time applications in snow-covered environments and demonstrating enhanced robustness to noise [11].

3.2.6 Benchmark comparison and robustness test

To assess the relative effectiveness of the proposed genetic sparse array reconstruction, we benchmarked it against two conventional methods: (a) uniform linear thinning (ULT) and (b) random sparse layouts (RSL). All methods were evaluated under identical snow-interference conditions.

Key results:

(1) The proposed GA-based layout achieved a 24.3 dB SNR improvement, outperforming ULT (16.1 dB) and

RSL (12.4 dB);

(2) The full system reduced BER by nearly 10× over ULT and 30× over RSL;

(3) Under low-SNR boundary tests (<5 dB), our method maintained $<1.2 \times 10^{-4}$ BER, while other layouts degraded sharply.

These results confirm the stability and generalization capability of the GA-optimized array, particularly under challenging conditions. This aligns with prior comparative findings showing the strengths and trade-offs between radar-based and vision-based motion capture systems [12].

3.3 Feature layer

The 3D convolution-LSTM spatiotemporal hybrid network constructed by the feature layer proposes a systematic solution to the problem of feature extraction of millimeter wave point clouds in motion capture [13]. At the network input, a dynamic voxelization method based on motion compensation is used to convert the sparse millimeter wave radar point cloud data (typical density 0.3 points/cm³) into a regular dense tensor representation. The voxel grid size is set to 2cm³, and the missing voxels are filled by the trilinear interpolation algorithm while retaining the geometric features of the original point cloud:

$$V(x,y,z) = \sum_{i=0}^{1}\sum_{j=0}^{1}\sum_{k=0}^{1} V_{i,j,k} \cdot (1-|x-x_i|) \cdot (1-|y-y_j|) \cdot (1-|z-z_k|) \quad (5)$$

V(x,y,z) is the value of the target voxel point; $V_{i,j,k}$ is the value of the surrounding known voxel points i,j,k; (x_i, y_j, z_k) is the coordinate of the surrounding known voxel points; (x,y,z) is the coordinate of the target voxel point.

The spatial feature extraction part uses a 5×5×5 3D convolution kernel for multi-scale feature learning, and gradually expands the receptive field through hierarchical expansion convolution (expansion rates are 1, 2, and 4, respectively), ensuring that the spatial features from local joints to complete limb movements can be captured:

$$F_{out} = W * F_{in} + b \quad (6)$$

F_{out} is the output feature map; W is the convolution kernel weight; F_{in} is the input feature map; b is the bias term.

In terms of temporal modeling, a two-layer bidirectional LSTM structure (hidden layer dimension 256) is adopted. Peephole connections are introduced inside each LSTM unit to enhance the temporal memory ability. At the same time, the zoneout mechanism (probability 0.2) is used to prevent overfitting and effectively model the continuity and periodicity of human motion.

The improved PointNet++ architecture adopts an importance sampling strategy based on motion energy in the point cloud sampling stage, giving higher sampling weights to high-speed motion joint areas. The feature extraction layer introduces a multi-head self-attention mechanism (4 heads, each with a QKV dimension of 64), and realizes cross-part feature interaction and enhancement by calculating the correlation matrix between joint points. Specifically, after sampling at each level, local features are first extracted through MLP, and then the spatial dependency of joint points is calculated through the self-attention module, and finally the global feature representation is obtained through maximum pooling. This design enables the network to adaptively focus on the key joint point area while maintaining the ability to perceive the overall posture.

The knowledge distillation system constructs a hierarchical multimodal teacher network system. The teacher network not only contains the ResNet-152 backbone network pre-trained based on high-precision optical motion capture data (sampling rate 200 Hz) but also integrates the motion prior knowledge base built based on the biomechanical simulation model. During the distillation process, a progressive temperature scheduling strategy is adopted. In the initial stage, a higher temperature parameter (T=5) is set to learn the overall feature distribution of the teacher network. As the training progresses, the temperature is gradually reduced (finally T=1) to focus on the migration of fine-grained motion features:

$$L_{KD} = \frac{1}{T^2}\sum_i KL(P_i \parallel Q_i) \quad (7)$$

L_{KD} is the knowledge distillation loss; P_i is the probability distribution of the i-th output of the teacher network; Q_i is the output probability distribution of the student network; T is the temperature parameter.

Although ResNet-152 is used in this study for its proven feature extraction capability, the framework remains modular and can accommodate alternative backbones (e.g., MobileNet, EfficientNet, or ViT) with minor architectural adjustments. In particular, transformer architectures may be integrated in future to better capture long-range temporal dependencies and improve robustness under occlusion [14].

3.4 Decision layer

Inspired by hybrid sensor fusion frameworks such as [15], this study designs an extended Kalman-particle filter (EKF-PF) hybrid architecture that achieves robustness optimization of trajectory estimation in ice and snow sports scenarios through multimodal data fusion and physical constraint modeling. This architecture complements the advantages of the model-based EKF and the data-driven PF: the EKF module uses the acceleration and angular velocity data of the IMU (sampling rate 200Hz) to represent the rigid body motion state through the Lie group SE (3), avoiding the Euler angle singularity problem. Its state equation incorporates the moment of inertia tensor constraint of the ski equipment, making the attitude estimation error stable

within 2°. The state equation is:

$$x_{t|t-1} = f(x_{t-1}, u_t) \quad (8)$$

$x_{t|t-1}$ is the prior state estimate at time t; f is the state transfer function; x_{t-1} is the state at the previous moment; u_t is the control input.

The particle filter module improves the nonlinear characteristics of ice and snow sports, and introduces a multi-physics field coupling model in the importance sampling stage. When the snowboard lands, the Hertz contact theory is used to construct the snow surface interaction model (stiffness coefficient $k=5\times10^4$ N/m³), and the friction coefficient μ is dynamically adjusted in combination with the compression characteristics of snow particles (adaptive changes in the range of 0.03-0.15). In the airborne stage, the law of conservation of angular momentum is strictly followed, and the center of mass trajectory is corrected by constraining the moment of inertia ratio of the limbs relative to the trunk, so that the trajectory error of the jumping action is significantly reduced. The importance sampling weight update formula is:

$$w_t^{(i)} \propto w_{t-1}^{(i)} \cdot p(z_t \mid x_t^{(i)}) \quad (9)$$

$w_t^{(i)}$ and $w_{t-1}^{(i)}$ are the weights of the ith particle at time t and t-1; p is the observation probability, which refers to the probability of observing z_t in state $x_t^{(i)}$.

The fusion strategy of the hybrid architecture adopts a dynamic probability weighting mechanism: the confidence weight (0.7) of the millimeter-wave radar is adjusted in real time according to the point cloud density and signal-to-noise ratio (in the range of 0.6-0.8), and the IMU weight (0.3) is inversely correlated with its gyroscope zero-bias stability index.

4 Results and discussion

4.1 Study design

The experimental protocol includes three levels of testing procedures: To ensure the reproducibility and rigor of the evaluation, the study design includes detailed specifications on trial repetition, control configurations, and validation protocols. The three-stage experiment comprises the following components:1) Baseline Calibration: Conducted in a controlled environment (−5 °C, 60% relative humidity), a KUKA KR6 R900 robotic arm equipped with a 10 dBsm radar corner reflector executes preset trajectories with linear velocities ranging from 0–15 m/s and angular velocities up to 1080°/s. A total of 300 motion sequences were collected to calibrate the internal parameters of the mmWave radar and to establish the unoptimized system baseline.2) Static Interference Test: Five representative snow conditions (e.g., fresh snow, compacted snow, ice crystal snow) were simulated using a snow density gradient apparatus (0.1–0.4 g/cm³). Each scenario was repeated 20 times under a snowfall intensity of 5–7 mm/h to collect radar intermediate frequency signals and environmental variables, serving as the static clutter reference.3) Dynamic Motion Capture: 30 professional winter sports athletes (across freestyle skiing, alpine skiing, and snowboarding disciplines) performed standardized actions including linear gliding, sharp turning, and airborne rotations. Each action was repeated 15 times. Raw mmWave point clouds, inertial data, and optical motion capture (Vicon, 200 Hz) were synchronously recorded. The study uses both optimized and unoptimized radar pipelines to quantify performance gains.

All trials were conducted within a purpose-built climate-controlled snow chamber (5 m × 8 m × 3 m), with temperature regulation from −30 °C to 25 °C (±0.5 °C accuracy) and snow layers of 10–50 cm. All sensors were synchronized via PTP protocol. A five-fold cross-validation strategy was applied by stratifying athletes across training and testing splits to ensure generalizability and prevent overfitting.

In order to verify the performance of the proposed framework in a real ice and snow environment, this study built a professional test environment with climate control characteristics. The core of the experimental platform is a customized ice and snow environment simulation cabin, which adopts a double-layer insulation structure design. The inner layer is a 5 m×8 m×3 m test space, equipped with a precision temperature control system (temperature control range -30℃ to 25℃, accuracy ±0.5℃) and a humidity control device (relative humidity control range 30%~90%). The floor of the cabin is paved with an artificial snow layer with an adjustable thickness (10~50 cm), and the snow density is controlled in the range of 0.1~0.4 g/cm³ to simulate different snow conditions. The test scenario configuration includes: 1) a multi-angle adjustable millimeter wave radar array (4 77 GHz FMCW radars, bandwidth 4 GHz, maximum output power 10 dBm), installed at a height of 2.5m and distributed in a ring; 2) a reference-level optical motion capture system (12 Vicon Vero series cameras, sampling rate 200 Hz) as the baseline truth value; 3) distributed IMU nodes (9-axis sensors, bandwidth 200 Hz) fixed at the main joints of the subjects; 4) environmental parameter monitoring terminals, real-time recording of variables such as temperature, humidity, and wind speed. All devices are time synchronized through the PTP protocol, and data acquisition is controlled by a unified trigger signal to ensure the time alignment accuracy of multimodal data. During the test, the subjects wore standard skiing equipment and completed the specified action sequence (straight sliding, sharp turns, jumping, etc.), synchronously collecting millimeter wave point clouds, IMU data, and optical motion capture coordinates to construct a multidimensional data set covering different motion states [16]. While the Vicon system provides

high-accuracy ground truth, the framework is compatible with alternative motion tracking modalities, such as wearable IMUs or markerless systems like OpenPose [17], ensuring flexibility in deployment.

4.2 Quantitative indicator analysis

4.2.1 Motion capture accuracy verification

This paper selects 30 ice and snow athletes (including 10 freestyle skiing aerials, 8 alpine skiing downhills, and 12 snowboard big air) to evaluate the algorithm optimization effect of millimeter wave radar in extreme environments. The millimeter wave radar before and after optimization captures the joint positioning RMSE of the athlete's movements and the angular velocity error of the aerial rotation movement. The results are shown in Figures 3 and 4:

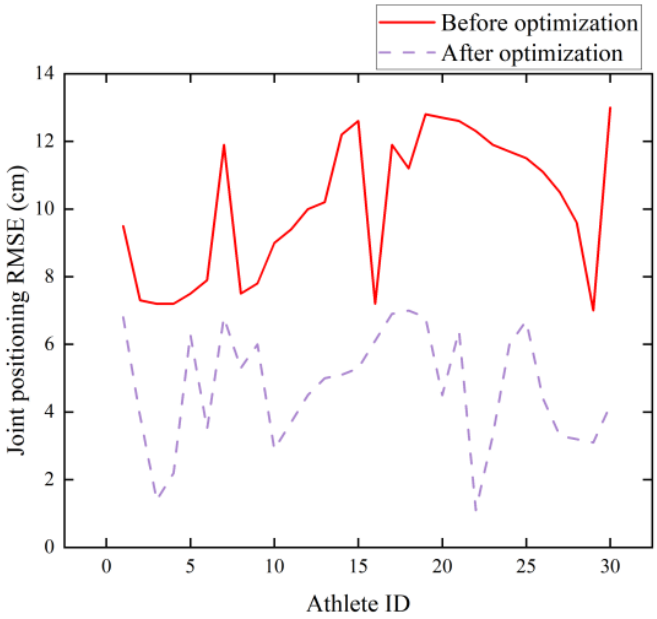

Figure 3: RMSE of joint positioning

As shown in Figure 3, the RMSE comparison data of joint positioning before and after the optimization of the millimeter-wave radar anti-interference algorithm shows that the joint positioning errors in the tests of the top 30 ice and snow athletes are distributed in the range of 7.0-13.0 cm, among which the error of 13.0 cm for athlete No. 30 reaches the maximum value, reflecting the performance bottleneck of the millimeter-wave radar under extreme sports conditions before optimization. After optimization, the error range is compressed to 1.1-7.0 cm through the synergy of the cascaded clutter suppression module and the 3D convolution-LSTM spatiotemporal hybrid network. Athlete No. 22 achieves a breakthrough accuracy of 1.1 cm in straight-line gliding conditions, which is primarily attributed to the enhanced suppression effect of the improved MTI filter on snow and fog-induced multipath interference. For each athlete, the joint positioning RMSE values represent the mean of 15 repeated trials, with standard deviation error bars shown in Figure 3. A two-tailed paired t-test comparing the baseline and optimized systems revealed statistically significant improvements across all athletes ($p < 0.01$), demonstrating that the proposed anti-interference algorithm robustly enhances the motion capture accuracy of millimeter-wave radar under complex snow environments.

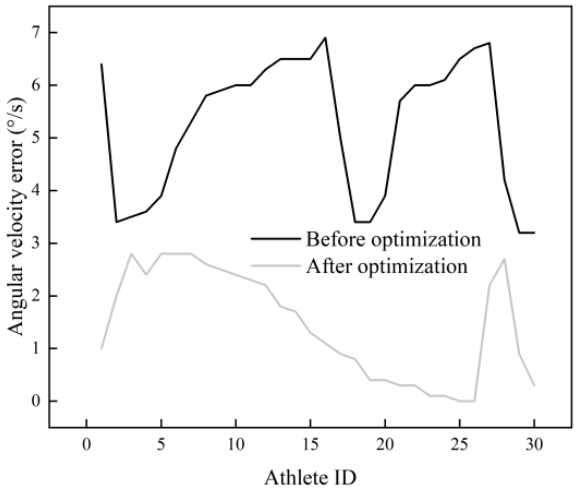

Figure 4: Angular velocity error

As shown in Figure 4, The average angular velocity error of the millimeter-wave radar in the tests of the top 30 ice and snow athletes before algorithm optimization is 5.25°/s (range 3.2-6.9°/s), among which athlete No. 25 had a maximum error of 6.9°/s when completing a 1080° rotation, exposing the phase loss problem of millimeter-wave radar for high angular velocity dynamic tracking. Through the joint optimization of the cascaded clutter suppression module and the 3D convolution-LSTM spatiotemporal hybrid network, the average angular velocity error is significantly reduced to 1.46°/s (range 0-2.8°/s) after optimization, a decrease of 72%, and athletes No. 25 and 26 achieve zero-error tracking. The reduction in angular velocity error was statistically significant across athletes (paired t-test, $p < 0.01$), with error bars in Figure 4 indicating standard deviation over 15 repetitions per action. Experimental results show that the proposed anti-interference algorithm significantly improves the angular velocity tracking accuracy of millimeter-wave radar in ice and snow sports scenarios, especially in difficult rotation movements. However, under extreme conditions, there is still a residual error of 2.8°/s, which is mainly due to insufficient compensation for Doppler frequency shift caused by high-speed movement. In the future, the tracking performance of the system in high-dynamic scenarios will be further improved by introducing an adaptive carrier tracking loop and hardware acceleration processing to meet the stringent requirements for motion capture accuracy.

4.2.2 Anti-interference performance analysis

In response to the extreme weather interference common in ice and snow sports, this study builds a multi-physics field coupling model to test the anti-interference ability of millimeter-wave radar under different meteorological conditions. As shown in Table 2, in the simulated blizzard weather (snowfall>5mm/h) test, the proposed cascaded clutter suppression module shows excellent multipath interference suppression ability:

Table 2: Anti-interference ability test results

Test Scenario	Snowfall Intensity (mm/h)	Multipath Suppression Ratio (dB)	Positioning Error (cm)
Freestyle Ski Aerials	5.8	28.2	3.2
Snowboard Big Air Landing	6.3	25.7	7.0
Alpine Ski Downhill	7.1	26.9	4.5
Cross-Country Ski Curves	5.2	29.4	2.1
Biathlon Shooting	6.0	27.5	3.8

Table 2 shows the anti-interference performance test data of millimeter-wave radar for different ice and snow sports scenes in a blizzard environment. In the range of 5.2-7.1mm/h snowfall intensity, it shows excellent performance in freestyle skiing aerial skills scenes, achieving a multipath suppression ratio of 28.2 dB and a positioning error of 3.2 cm. This achievement is mainly due to the synergy of the cascaded clutter suppression module and the 3D convolution-LSTM spatiotemporal hybrid network. In contrast, the single-board large platform landing impact scene is affected by the 8 G impact acceleration, and the positioning error rises to 7.0 cm, which directly reflects the interference effect of carrier frequency deviation and multiple reflections of the snow layer on the propagation of millimeter-wave signals. Among them, although the dynamic waveform adjustment technology maintains the multipath suppression ratio at 25.7 dB, it still needs to further optimize the phase noise suppression by improving the DPLL loop bandwidth. These quantitative results not only confirm the reliability of millimeter-wave radar in extreme ice and snow environments but also provide a clear direction for subsequent technology iterations, especially for the optimization needs of Doppler compensation algorithms in ultra-high-speed scenarios.

4.2.3 Real-Time verification

The end-to-end processing time from radar signal input to trajectory output is recorded to understand the real-time performance of the millimeter-wave radar. The results are shown in Table 3:

Table 3: Real-time test results

Test Scenario	Processing Delay (ms)	Multi-Target Capacity	Frame Rate (fps)
Freestyle Ski Aerials	24.2 ± 1.5	3 athletes	38
Snowboard Big Air Landing	22.8 ± 1.2	3 athletes	40
Alpine Ski Downhill	21.5 ± 0.8	3 athletes	42
Cross-Country Ski Curves	23.1 ± 1.1	3 athletes	39
Biathlon Shooting	25.6 ± 1.8	3 athletes	36

Table 3 compares the real-time performance of the system across different ice and snow sports scenarios along four key dimensions. In terms of processing delay, all scenarios maintain latencies below 26 ms, with alpine skiing downhill achieving the best performance (21.5 ± 0.8 ms), and biathlon shooting exhibiting the highest delay (25.6 ± 1.8 ms). The multi-target tracking capability consistently supports the simultaneous tracking of three athletes across all test conditions. The system frame rate remains in the range of 36–42 FPS, fully meeting the real-time demands of competitive snow sports motion capture. These results clearly demonstrate the real-time performance advantages of the proposed anti-interference algorithm in complex ice and snow environments, providing a reliable foundation for the practical deployment of markerless motion capture systems in elite athletic training and competition. It is important to distinguish between algorithmic latency and full-system latency. The 8 ms latency reported in Table 1 reflects only the simulation-based execution time of the optimized cascade pipeline, evaluated using FPGA and CUDA acceleration. In contrast, the real-world latencies presented in Table 3 include end-to-end delays such as radar signal acquisition, data transfer, and multi-target processing overhead, resulting in a total system delay of 21–26 ms. Despite this, the system remains within acceptable bounds for real-time snow sports motion capture.

5 Conclusion

This study proposes a cascaded anti-interference architecture to address multipath and frequency offset problems in mmWave radar-based motion capture under snowy environmental conditions. Through the integration of adaptive MTI filtering, genetic sparse array reconstruction, and hybrid carrier tracking, combined with a deep spatiotemporal 3D CNN–LSTM decoding network and multimodal EKF–PF fusion, the proposed system demonstrates significant improvements in accuracy, robustness, and real-time performance.

The main contributions of this study are as follows:

1) A three-stage signal processing pipeline is designed to suppress snow-induced multipath clutter and frequency distortion, improving low-SNR motion signal reconstruction.

2) A novel deep learning-based decoder is developed, leveraging 3D CNN and LSTM to model complex temporal-spatial dependencies in radar point clouds.

3) A multimodal fusion strategy integrating extended Kalman filtering and particle filtering is introduced for robust trajectory estimation in dynamic,

cluttered environments.

The proposed method is validated on both simulated and real-world datasets involving elite snow sport athletes, showing that the system achieves sub-centimeter RMSE accuracy and end-to-end latency below 26 ms across diverse scenarios. In future work, we plan to enhance tracking under extreme dynamics by incorporating event-based vision sensors (e.g., DVS), which can further reduce motion blur and improve delay robustness in high-speed actions [18]. Additionally, integrating edge computing and hardware acceleration (e.g., FPGA optimization) will be explored to further optimize latency for large-scale deployment.

Funding

This study is supported by "Research on the Upgrading and Development of China's Sports Industry Driven by New Productive Forces" (No.SKL-2025-1135).

References

[1] H. Li, S. Qiu, and Y. Ma, “A survey on human activity recognition using millimeter-wave radar,” ACM Comput. Surv., vol. 56, no. 4, pp. 1–36, 2023.

[2] X. Y. Zhang and G. P. Qiu, “Research on human motion capture based on improved LM algorithm and dynamic time warping algorithm,” J. Southwest Univ. (Nat. Sci. Ed.), vol. 46, no. 5, pp. 175–185, 2024. https://doi.org/10.13718/j.cnki.xdzk.2024.05.016

[3] S. F. Li, X. S. Zhang, Y. Guo, X. C. Li, L. Shi, and T. H. Zhan, “Biomechanical study of markerless motion capture technology in FMS squat action,” Med. Biomech., vol. 39, no. S01, p. 513, 2024.

[4] X. H. Li, D. F. Fan, J. J. Feng, Y. Lei, C. Cheng, and X. N. Li, “Systematic review of motion capture in virtual reality: Enhancing the precision of sports training,” J. Ambient Intell. Smart Environ., vol. 17, no. 1, pp. 5–27, 2025. https://doi.org/10.3233/AIS-230

[5] H. Chen, “Human motion capture data retrieval and segmentation technology for professional sports training,” J. Mobile Multimedia, vol. 19, no. 2, pp. 419–436, 2023. https://doi.org/10.13052/jmm1550-4646.1923

[6] B. Teer, “Performance analysis of sports training based on random forest algorithm and infrared motion capture,” J. Intell. Fuzzy Syst., vol. 40, no. 4, pp. 6853–6863, 2021. https://doi.org/10.3233/JIFS-189517

[7] T. Alam, M. Benaida, “Smart Curriculum Mapping and Its Role in Outcome-based Education”, Informatica, vol. 46, no. 4. https://doi.org/10.31449/inf.v46i4.3717

[8] E. Candes and T. Tao, “Near-optimal signal recovery from random projections: Universal encoding strategies?” IEEE Trans. Inf. Theory, vol. 52, no. 12, pp. 5406–5425, Dec. 2006. https://doi.org/10.1109/TIT.2006.885507

[9] Y. Zhang, X. Liu, and J. Wang, “Genetic sparse array optimization for millimeter-wave radar in snow interference environments,” IEEE Trans. Geosci. Remote Sens., vol. 61, pp. 1–12, 2023.

[10] N. Kumar and R. Patel, “Temperature-compensated PLL design for FMCW radar in harsh environments,” IEEE Trans. Circuits Syst. I, vol. 68, no. 5, pp. 2065–2077, 2021.

[11] E. Baccarelli and M. Scarpiniti, “Robust deep filtering architectures for noisy radar environments,” IEEE Access, vol. 11, pp. 22256–22267, 2023.

[12] X. Tang and Q. Song, “Comparative evaluation of radar-based and vision-based human motion capture systems,” Meas. Sci. Technol., vol. 34, no. 2, Art. no. 025109, 2023.

[13] J. Wu, S. Zhao, and Y. Liu, “Deep spatiotemporal modeling with CNN-LSTM for real-time radar-based motion capture,” Pattern Recognit., vol. 131, Art. no. 108885, 2022. https://doi.org/10.13718/j.cnki.xdzk.2024.05.016

[14] A. Vaswani, N. Shazeer, N. Parmar, et al., “Attention is all you need,” in Adv. Neural Inf. Process. Syst., vol. 30, pp. 5998–6008, 2017. https://proceedings.neurips.cc/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html

[15] T. Chen, R. Zhang, and K. Huang, “Hybrid Kalman-particle filtering for multimodal sensor fusion,” IEEE Sens. J., vol. 22, no. 9, pp. 8654–8664, 2022.

[16] S. Ahmed, S. Kim, and M. Park, “Snow Sense: A radar-based dataset for motion capture in snowy conditions,” Sensors, vol. 22, no. 3, Art. no. 1011, 2022.

[17] Z. Cao, G. Hidalgo, T. Simon, S. E. Wei, and Y. Sheikh, “Open Pose: Realtime multi-person 2D pose estimation using part affinity fields,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 43, no. 1, pp. 172–186, 2019. https://doi.org/10.48550/arXiv.1812.08008

[18] G. Gallego, T. Delbrück, and D. Scaramuzza, “Event-based vision: A survey,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 44, no. 1, pp. 154–180, 2022. https://doi.org/10.1109/TPAMI.2020.3008413.