

Keywords: motion capture, gap filling, BiLSTM, unscented Kalman filter

Kamil ŻELAZOWSKI ¹, Wojciech WOJCIECHEWICZ ¹,
Maria SKUBLEWSKA-PASZKOWSKA ¹, Paweł POWROZNIK ¹

¹ Lublin University of Technology, Poland, s97781@pollub.edu.pl, s97767@pollub.edu.pl,
maria.paszowska@pollub.pl, p.powroznik@pollub.pl

* Corresponding author: p.powroznik@pollub.pl

Automatic methods for 3D motion trajectories gap filling: Custom-based Kalman vs. BiLSTM

Abstract

Modern data analysis increasingly relies on time series data across diverse domains, such as biomechanics, medicine, sports, and computer animation. A special case of this data type is generated by optical motion capture systems, enabling precise spatiotemporal tracking of anatomical marker positions. However, during dynamic, fast-moving recordings, some markers may not be registered properly, mainly because they are obscured or because there are insufficient cameras. While continuous trajectories provide the basis for human movement analysis, their utility is heavily dependent on the accurate registration of all measurement points. The post-processing of motion capture data is highly time-consuming; thus, selecting an appropriate, highly reliable gap-filling method is extremely important for subsequent data analysis. In this study, we propose a novel, custom-designed Kalman filter for gap-filling dance motion-capture trajectories. This method integrates the three-dimensional positions, velocities, and accelerations of markers. The proposed method is compared with the basic Kalman filter and a bidirectional Long Short-Term Memory (BiLSTM) model, which is highly effective for imputing sequential data. Given these dual paradigms, this study aims to compare the classical mathematical model with the modern machine learning approach for reconstructing motion capture data, evaluating their respective effectiveness across varying spatial and temporal characteristics of missing data. Several examinations, based on the number of missing markers (from 2 to 14) in the given marker's neighbour group across various frame numbers (from 100 to 500), were performed. The BiLSTM model achieved lower RMSE values than the custom Kalman-based filter across almost all scenarios and was highly stable with a small number of missing markers. However, the accuracy of both methods deteriorated as the number of missing markers increased, indicating a significant impact of the spatial distribution of missing data on reconstruction quality.

1. INTRODUCTION

Modern data analysis increasingly relies on time-series data across diverse domains, such as biomechanics, medicine, sports, engineering, and computer animation (Dziuba-Kozieł et al., 2024, 2026; Millour et al., 2023; Powroznik et al., 2025; Skublewska-Paszowska, 2024; Skublewska-Paszowska et al., 2024; Zhou & Hu, 2008). A special case of this type of data is that generated by optical motion capture systems, which enable precise spatiotemporal tracking of anatomical marker positions. While these continuous trajectories provide the basis for human movement analysis, their utility is heavily dependent on the accurate registration of all measurement points. In practice, situations often arise where some markers are not recorded correctly, leading to missing marker trajectories (Van Der Kruk & Reijne, 2018). Marker self-occlusion, detection anomalies, or inherent hardware limitations typically cause these discontinuities. Such data gaps hinder further analysis and significantly degrade the quality of the obtained results, necessitating the use of robust reconstruction techniques. Over the years, numerous approaches to reconstructing trajectories of missing markers have been developed (Aristidou et al., 2008; Feng et al., 2014; Lai et al., 2011; Li et al., 2021; Mohaoui & Dmytryshyn, 2026). Classical methods primarily rely on deterministic dynamic models, most notably the Kalman filter. Conversely, rapid advances in machine learning have introduced data-driven models that are adept at capturing complex, nonlinear spatiotemporal dependencies. Among these, LSTM networks have emerged as highly effective for imputing sequential data due to their ability to model temporal dependencies in sequential data,

are particularly noteworthy (Yuhai et al., 2024). At the same time, classical estimation methods, such as the Kalman filter and its variants, remain widely used due to their interpretability, stability, and well-established theoretical foundations. Classical methods often rely on mathematical models that describe the process being studied, whereas machine learning algorithms enable the automatic extraction of relationships from data, even in complex and poorly understood systems. In practice, this translates into the possibility of achieving higher estimation or prediction accuracy while maintaining acceptable computational costs. Given these dual paradigms, this study aims to compare the classical mathematical model with the modern machine learning approach for (MoCap) data reconstruction, evaluating their effectiveness across varying spatial and temporal characteristics of missing data. A Kalman-based custom method was developed to improve estimation quality in the presence of noise, model uncertainty, or non-stationary data. A key study issue is to determine the extent to which the proposed solution can compete with modern machine learning methods, represented by LSTM architectures, and to identify the conditions under which each method achieves the best results.

Methods based on Kalman filters and machine learning algorithms are widely used in many areas of engineering and science (Bai et al., 2023), including navigation and positioning systems, mobile robotics, battery energy management, sensor networks, and power systems. In practical implementations, real-time requirements are particularly important. Kalman filters are characterised by low computational complexity and high interpretability, while LSTM models can detect hidden dependencies in historical data. Comparison of the author's Kalman filter-based method with solutions using LSTM networks will allow us to determine the trade-off among estimation accuracy, noise immunity, computational requirements, and the feasibility of implementation in real-time systems. Such analysis is particularly important for traditional folk dances, as they constitute cultural heritage that requires meticulous documentation and preservation for future generations. Motion capture data allow for precise representation of dancers' movements, but in practice, they often contain gaps due to markers obscured by elements of traditional costume, interactions between dancers, or the dynamic nature of the choreography. Effective reconstruction of missing trajectory fragments is therefore essential to maintaining the completeness and reliability of digital dance records.

The main authors' contributions are as follows:

- Curate a comprehensive dataset based on the Polish Polonaise dance in the form of motion capture recordings for LSTM training. The initial data consisted of 28 sequences performed by 2 subjects utilising a Vicon optical passive motion capture system. To ensure robust model training, the dataset was expanded to 5,920 sequences through a systematic augmentation protocol combined with simulated spatiotemporal occlusions;
- Prepare a custom mathematical model based on the Unscented Kalman Filter (UKF). This method is distinguished from a baseline Kalman filter by its hierarchical estimation logic, which utilizes the Kabsch algorithm and spatial sphere intersections to dynamically adjust predictions based on the real-time visibility of predefined neighbouring markers;
- Adapt and fine-tune the BiLSTM model using the biomechanical loss function. This loss function uniquely integrates a Mean Squared Error (MSE) objective with a structural penalty that strictly enforces constant rigid-body bone lengths, preventing physically impossible marker configurations and trajectory drift during prolonged occlusions;
- Compare the mathematical and machine learning methods for 3D data gap filling across 12 scenarios. These experiments systematically varied spatial sparsity (2 to 14 missing markers) and temporal duration (100 to 500 frames). When evaluated against Masked Root Mean Square Error and inference latency, the BiLSTM method proved superior, achieving significantly higher kinematic fidelity and requiring less than one-third of the computational time in 11 of the 12 tested configurations.

2. RELATED WORKS

In recent years, motion capture technology has experienced a rapid spike in popularity. It has been used in many complex domains, such as dance (Chen et al., 2023; Kico et al., 2020). By digitising movement data captured with MoCap, sophisticated virtual training systems enable users to practice at home without a real instructor (Esaki & Nagao, 2024). The addition of a gamified environment to virtual reality also demonstrated that users' dance skills can be further improved (Senecal et al., 2020). However, the efficiency of such systems is closely related to the quality of the underlying motion data. One challenge is marker self-occlusion. During rapid, complex dance moves, the dancer's limbs frequently obstruct the camera's view of other parts of the

body, leading to missing data for a marker or a group of markers in a given frame. Recovering missing marker positions has been a recurring problem. Existing solutions can be grouped into two major methodological categories: classical approaches that rely on deterministic, statistical, or model-based techniques, and AI-based methods that leverage deep neural networks to learn complex temporal dependencies in human motion (Table 1).

2.1. Classical methods

Classical approaches traditionally include deterministic, statistical, or model-based techniques such as interpolation, regression models, and Kalman filters. These methods remain important due to their robustness, interpretability, and low computational cost. For instance, Skurowski and Pawlyta (2024) evaluated tree-based regression methods against interpolation, low-rank matrix completion, and neural networks, finding that classical tree-based algorithms can be competitive for medium-length gaps (0.1–2 s) and offer a favourable accuracy-to-cost ratio when using high-quality Vicon data. Maggioni et al. (2025) conducted a comparative analysis of markerless (MediaPipe and Leap Motion Controller) and marker-based systems for detailed hand and finger kinematics, concluding that markerless systems still lag in precision, though fusing multiple data sources can significantly reduce errors. Hauenstein et al. (2024) focused on the reliability assessment of markerless MoCap for functional movement tasks using classical statistical metrics such as ICC (Intraclass Correlation Coefficient), SEM (Standard Error of Measurement), and minimal detectable change, showing that although many joint positions are reliably measured, some rotational components remain problematic. A broader review by Ji et al. (2023) on Human Pose Estimation (HPE) in marker-free MoCap noted that despite achieving high frame-based accuracy, modern HPE methods struggle with occlusion or self-penetration, highlighting the need to incorporate physical scene constraints traditionally associated with classical modelling. Gomes et al. (2021) proposed a Fully-Automatic Gap Filling approach based on the Unscented Kalman Filter (UKF), which estimates the occluded marker position by minimising the MSE of distance, movement, and geometric constraints relative to auxiliary markers. When tested against Vicon Nexus's spline, rigid body, and pattern fill methods, their approach demonstrated greater robustness, particularly for gaps at the start or end of a sequence, achieving an average RMSE of 3.9 cm in these cases. Lannan et al. (2020) presented a hybrid method in which a Tobit Kalman Filter (TKF) operates sequentially with a convolutional autoencoder (CAE). This approach provides an initial pose estimate, which the TKF then smooths and adds kinematic/dynamic constraints. This CAE, together with the TKF configuration, reduced the standard deviation by approximately 70% on simulated data and 30% on real-world data, outperforming configurations that use only CAE, only TKF, or CAE with Linear KF. Martini et al. (2024) introduced an Adaptive Kalman Filter (AKF) that replaces fixed noise covariances with runtime calculations considering marker velocity and uses a recurrent neural network (RNN) with two Gated Recurrent Units (GRU) as a learned non-linear transition model, followed by a Biomechanical Constraint Adjustment step and a Butterworth low-pass filter for final smoothing. This method achieved the best results in denoising and gap-filling tasks compared to Linear KF and SmoothNet, reducing the Mean Per Joint Position Error (MPJPE) by up to 31.9% in gap-filling.

2.2. AI-based methods

The second main category of solutions relies on machine learning and deep neural networks, which are particularly effective for modelling nonlinear motion patterns and recovering long gaps. Long short-term memory (LSTM) architectures are a common choice due to their ability to model nonlinear temporal dependencies. Yuhai et al. (2024) proposed a U-Bi-LSTM network to reconstruct missing marker trajectories, which demonstrated significantly lower reconstruction error than traditional methods, especially for long occlusions (up to 900 frames) or noisy input. Li et al. (2021) also used an LSTM-based method that first standardises bone length and transforms data to a local coordinate system to capture relative marker positions before reconstruction. Their approach showed strong adaptability for long gaps (up to 300 frames) and performed better as the number of missing markers increased, achieving a consistently low RMSE of around 2 cm where interpolation failed. Malekian and Lapeer (2024) addressed self-occlusion in monocular video MoCap using a Skinned Multi-Person Linear (SMPL) model, proposing two post-processing methods: an RF for frame-by-frame processing and an AE with one or two LSTMs for sequence-based processing. Kucherenko et al. (2024) examined the mismatch between standard numerical metrics (RMSE, MPJPE) and human perception of movement quality, introducing the Bone Distance Preservation (BDP) and Velocity

Distance (VD) metrics, which correlate better with human judgment, thereby opening a new direction in perceptual evaluation. He et al. (2021) addressed self-occlusion for challenging motions, such as yoga and rolling, using ChallenCap, a template-based approach. This method uses a Hybrid Motion Inference Stage with HybridNet, followed by a Robust Motion Optimisation Stage that uses multimodal references and 2D silhouette information. ChallenCap significantly reduced MPJPE by 60% and 44% compared to MonoPerfCap and VIBE, respectively. Skurowski and Pawlyta (2021) presented a novel approach in which neural networks are trained on data from an individual sequence to reconstruct its own gaps, comparing it with interpolation and low-rank matrix completion. Surprisingly, the results showed that for long gaps, the simpler FFNN outperformed the RNNs.

The literature reviewed indicates a research niche involving the comparison of classical methods for trajectory gap filling with modern artificial intelligence architectures. There is a need to assess the effectiveness of the developed proprietary filter at reconstructing missing trajectory fragments in motion-capture data. Comparison with the classical Kalman filter allows us to verify whether the introduced modifications improve motion reconstruction accuracy and increase robustness to interference and data loss. Comparing the results with the BiLSTM method allows us to assess the competitiveness of the proposed solution against modern deep learning architectures, which can model the complex temporal dependencies in motion sequences. The research will determine which method yields the best trajectory reconstruction quality while maintaining computational efficiency and applicability to real-time motion analysis systems.

Tab. 1. Summary of selected state-of-the-art methods for motion capture gap filling

Author (Year)	Method	Data / Gap Length	Results	Limitations
Skurowski & Pawlyta (2024)	Tree-based regression	Vicon data / 0.1–2 s	Favourable accuracy-to-cost ratio for medium gaps	Degrades for gaps exceeding 2 seconds
Maggioni et al. (2025)	Markerless (MediaPipe, Leap Motion)	Hand/finger kinematics	Reduced errors via data fusion	Lower precision compared to marker-based systems
Hauenstein et al. (2024)	Markerless MoCap	Functional movement tasks	Reliable joint positions	Rotational components remain problematic
Gomes et al. (2021)	Custom UKF	Vicon Nexus / Variable gaps	3.9 cm average RMSE	Primarily evaluated on start/end boundary gaps
Lannan et al. (2020)	Tobit KF + Convolutional AE	Simulated & real data	70% std dev reduction (sim)	Requires two-stage hybrid processing
Martini et al. (2024)	Adaptive KF + RNN (GRU)	Complex human motion	MPJPE reduced by up to 31.9%	Computationally demanding non-linear transition
Yuhai et al. (2024)	U-Bi-LSTM	Up to 900 frames	Lower error on noisy input	Requires full sequence (offline processing only)
Li et al. (2021)	Bone-constrained LSTM	Up to 300 frames / 2–14 markers	~2.0 cm RMSE	Accuracy drops with extreme spatial sparsity
He et al. (2021)	ChallenCap (HybridNet)	Yoga/rolling motions	44–60% MPJPE reduction	Requires multi-modal references & 2D silhouettes

3. MATERIAL AND METHODS

3.1. Dataset preparation

The dataset utilised in this study comprises motion capture recordings of the traditional Polish dance, the Polonaise. The kinematic data were acquired using an 8-camera Vicon motion capture system at 100 Hz, yielding 28 discrete sequences from one male and one female participant. Each registered sequence was supervised by a professional dance trainer. These sequences averaged approximately 3000 frames in duration and encompassed specific choreographic figures, including basic step (with and without bow), rotations and one- and two-measure bows. Each participant had 39 retro-reflective markers attached to their costumes in accordance with the Plug-in Gait model (Vicon Motion Systems, 2022). The captured data were post-processed utilising four steps. First, markers were labelled according to the Vicon Plug-in Gait model. Next, gaps in trajectories were interpolated, and noise was removed. Finally, the model was applied to reconstruct skeletal motion with full 3D joint kinematics.

To address the limited size of the initial dataset, an artificial data augmentation protocol was implemented. Each original sequence was subjected to four independent spatial transformations: random rotations along the X, Y, and Z axes, alongside a random spatial translation. Rotational angles were sampled from a uniform distribution over the interval -15° to 15° . Similarly, translational offsets for each spatial axis were drawn independently from -100 mm to 100 mm. To preserve the structural integrity and rigid-body kinematics of the subjects, these transformations were applied uniformly across all markers and all frames within a given sequence. To facilitate the training of the LSTM-based architecture, the augmented dataset was artificially corrupted to simulate data occlusion. A custom Python script was used to introduce gaps in the kinematic trajectories. The simulated occlusions were parameterised based on three primary variables:

- temporal duration: the duration of the induced gaps was set to discrete intervals of 100, 300, and 500 frames;
- temporal location: within each sequence, four non-overlapping temporal windows were randomly selected for occlusion. These gaps were permitted at any temporal index within the sequence, strictly excluding the initial starting frames to ensure sufficient data for initialisation;
- spatial sparsity: the number of concurrently occluded markers was explicitly set to 2, 6, 10, and 14. To ensure robust testing across varying anatomical locations, four distinct randomised marker dropout combinations were generated for each sparsity level.

The data augmentation culminated in a comprehensive corpus comprising 5,920 distinct CSV files. The initial 28 sequences were augmented using 4 spatial transformations, yielding 140 sequences (28 original + 112 transformed). Subsequently, combinations of simulated gaps (3 temporal durations, 4 spatial sparsity levels, and 4 randomised marker combinations) were systematically applied. Certain combinations were excluded if the original sequence was too short to accommodate the maximum gap length (e.g., 500 frames) or multiple non-overlapping windows. The valid combinations of these spatiotemporal occlusions across the augmented dataset resulted in exactly 5,920 files. To ensure full reproducibility of the data augmentation and simulated occlusions, all pseudo-random operations in the Python scripts were initialised with a fixed random seed (seed = 42).

3.2. Custom Kalman-based method

First, for each marker, neighbouring markers were manually defined based on their spatial distance. Manual definition of neighbouring markers was chosen over automatic clustering to ensure strict biomechanical relevance and rigid-body consistency. This approach guarantees that the selected neighbours belong to the same anatomical segment, which minimises interpolation errors during complex and dynamic choreographic movements. Each group of neighbours consists of three markers. A group of exactly three neighbouring markers was selected because it represents the mathematical minimum required to uniquely establish a local, three-dimensional coordinate system for rigid-body transformations. Using fewer markers would not provide sufficient spatial constraints to resolve the 3D position, while using more could introduce over-constraining issues and increase the probability of simultaneous occlusion within the reference group. During sequence processing, each marker is assigned an Unscented Kalman Filter (UKF) to track its frame-by-frame movement state. For UKF implementation, the FilterPy Python library was utilised (FilterPy, n.d.). The system's state

vector models 9 dimensions: 3D position, velocity, and acceleration (Eq. 1). The damped Singer acceleration model was employed as the transition model (Eq. 2)

$$s = [x, y, z, vx, vy, vz, ax, ay, az]^T \quad (1)$$

$$F = \begin{bmatrix} I_3 & dtI_3 & \frac{dt^2}{2}I_3 \\ 0_{3 \times 3} & \gamma I_3 & dtI_3 \\ 0_{3 \times 3} & 0_{3 \times 3} & \rho I_3 \end{bmatrix} \quad (2)$$

where: dt – time between frames ($dt = 0.01$ s, corresponding to the 100 Hz Vicon acquisition frequency),
 γ – velocity damping factor ($\gamma = 0.98$),
 ρ – acceleration damping factor ($\rho = 0.5$).

Because the Vicon system measures only 3D position, the observation model utilises a 3 x 9 mapping matrix to extract only the positional coordinates from the state vector (Eq. 3). The filter's uncertainty is controlled by a physically derived continuous white noise acceleration model for its process noise covariance matrix (Eq. 4) and observation noise covariance matrix scaled by an error factor (Eq. 5).

$$H = [I_3 \quad 0_{3 \times 3} \quad 0_{3 \times 3}] \quad (3)$$

$$Q = \begin{bmatrix} \frac{dt^5}{20}I_3 & \frac{dt^4}{8}I_3 & \frac{dt^3}{6}I_3 \\ \frac{dt^4}{8}I_3 & \frac{dt^3}{3}I_3 & \frac{dt^2}{2}I_3 \\ \frac{dt^3}{6}I_3 & \frac{dt^2}{2}I_3 & dtI_3 \end{bmatrix} q \quad (4)$$

$$R = I_3 \varepsilon^2 \quad (5)$$

where: q – process noise variance ($q = 11000$),
 ε – Vicon motion capture system error factor ($\varepsilon = 0.5$).

The filter's internal state vector was initialised with zeros. During the first frame in which a given marker is visible (cold start), its 3D position coordinates in the state vector are updated with the measured Vicon position, while the initial velocities and accelerations remain zero. The state covariance matrix was initialised as an identity matrix scaled by a factor of 10.0 to account for initial uncertainty.

Finally, for sigma-point generation, each UKF uses the following parameters: $\alpha = 0.1, \beta = 2.0, \kappa = 0.0$. All used parameters were based on the methodology proposed by Gomes et al. (2021). At each successive frame, each marker's UKF is updated with a new position. This measurement can be obtained from the marker data provided by the motion capture system if a marker is visible, or from the estimated position using rigid-body transformations. The algorithm performs different estimates based on the number of missing markers in the given marker's neighbourhood. Inspired by (Aristidou et al., 2008), the following logic was implemented:

- all markers in the group are visible: positions of neighbour markers from the current and previous frames are passed to the Kabsch algorithm to calculate rotation and translation matrices to transform the missing marker's position from the previous frame to the current frame;
- Two markers in the group are visible: first, it calculates an initial guess as the average of the distances between the neighbour marker positions in the current frame and the distance of this neighbour and the missing marker in the previous frame. Next, it calculates the intersection of 2 spheres centred and the neighbour markers' position in the current frame. Final estimation is the closest point to the initial guess on the intersection circle;
- one marker in the group is visible: in this situation, we assume that the distance between markers across frames is constant, and the estimation of the missing marker is the distance between the missing and neighbour marker in the previous frame, subtracted from the position of the neighbour marker in the current frame. The assumption of a constant distance, rather than a constant direction, stems from the biomechanical properties of the human skeleton. Bone lengths (distances between joints) remain fixed, while the direction (orientation) changes dynamically during dance movements. Assuming a constant

direction would lead to severe trajectory divergence during rotations, whereas maintaining distance preserves the structural integrity of the limb;

- all markers in the group are missing: this scenario is also based on the Kabsch algorithm, but for matrix calculations, it uses the markers' positions from frames $n-1$ and $n-2$. The final estimate is the position of the missing marker in frame $n-1$, translated to frame n .

Unlike previous UKF-based approaches that primarily rely on basic distance minimisation, the novelty of our custom UKF lies in the hierarchical incorporation of Kabsch-based rigid-body transformations and spatial sphere-intersection logic. This dynamic adaptation, based on real-time neighbour visibility, ensures strict satisfaction of geometric constraints and greater resistance to trajectory drift.

3.3. BiLSTM-based method

The second model was based on the LSTM architecture. Due to its capability to model long-term temporal dependencies in sequential data, the network utilises information from neighbouring frames to reconstruct missing marker trajectories directly in 3D space. The proposed model learns nonlinear kinematic relationships directly from motion capture data.

The motion capture recordings consisted of time-series 3D coordinates (x, y, z) for all tracked markers. To eliminate the influence of global body translation, all coordinates were transformed into a local coordinate system relative to the LASI marker according to (Li et al., 2021):

$$p_i^{local}(t) = p_i(t) - p_{LASI}(t) \quad (6)$$

where: $p_i(t)$ – original position of the i -th marker at time t ,
 $p_{LASI}(t)$ – LASI reference marker position.

To improve numerical stability during optimisation, all spatial coordinates were converted from millimetres to meters. Missing values represented as NaN were replaced with zeros. Additionally, a binary visibility mask was concatenated with the input features in order to inform the network which markers were missing explicitly:

$$m_i(t) = \begin{cases} 1, & \text{marker visible} \\ 0, & \text{marker missing} \end{cases} \quad (7)$$

The final input representation, therefore, consisted of marker coordinates together with corresponding visibility masks.

A bidirectional Long Short-Term Memory (BiLSTM) architecture was employed. The network consisted of two stacked bidirectional LSTM layers with 256 hidden units per direction, and a dropout rate of 0.1 for regularisation. The bidirectional structure enabled the model to exploit both past and future temporal dependencies simultaneously, which is particularly beneficial for gap reconstruction tasks. For an input sequence $X = (x_1, x_2, \dots, x_T)$, the hidden states were computed in forward and backward temporal directions (Li et al., 2021):

$$\vec{h}_t = \text{LSTM}(x_t, \vec{h}_{t-1}) \quad (8)$$

$$\overleftarrow{h}_t = \text{LSTM}(x_t, \overleftarrow{h}_{t+1}) \quad (9)$$

The final hidden representation was obtained through concatenation (Li et al., 2021):

$$h_t = [\vec{h}_t; \overleftarrow{h}_t] \quad (10)$$

The recurrent output was subsequently projected into 3D marker coordinates using a fully connected linear layer (Li et al., 2021):

$$\hat{y}_t = Wh_t + b \quad (11)$$

The complete model contained approximately 2.48 million trainable parameters.

3.3.1. Training

Training was performed using paired corrupted and ground-truth sequences generated from both the original motion capture recordings and the augmented dataset. The model was trained using the Adam optimiser with a learning rate of 0.001 and a batch size of 32. Artificial gaps were introduced at random temporal intervals according to the previously defined experimental scenarios (Section 3.1). To improve temporal learning efficiency and memory utilisation, the network processed the motion data in batches of 60 consecutive frames. Each training sample, therefore, consisted of a short temporal motion segment containing both visible and artificially occluded markers. This sliding-window strategy enabled the model to learn local kinematic dynamics while maintaining stable optimisation behaviour.

The model processes the motion data using temporal windows of 60 consecutive frames. Since the experimental gaps are longer than the input window (100, 300, and 500 frames), they are not processed as single sequences. Instead, the complete motion sequence is processed using consecutive 60-frame windows with a stride of 30 frames, resulting in a 50% overlap between adjacent windows. Predictions from overlapping windows are merged using a weighted overlap-add procedure. A triangular weighting function is applied to the predictions, assigning higher weights to frames near the centre of each window and lower weights near the window boundaries. This reduces potential discontinuities between predictions generated by neighbouring windows. If the final window contains fewer than 60 frames, it is zero-padded, while only the valid frames corresponding to the original sequence are retained. The bidirectional LSTM processes each 60-frame window in both temporal directions. Therefore, each prediction is based on the temporal context available within its corresponding 60-frame window, including both preceding and subsequent frames. For frames near the centre of a window, the available past and future context is approximately balanced, whereas frames near the window boundaries have less context on one side. Importantly, predictions are not generated recursively, but those from one window are not used as inputs to subsequent windows. All windows are processed using the original input sequence, including the missing-data mask. After inference, only coordinates identified as missing are replaced by the predicted values, while the originally observed coordinates are preserved. Following the weighted overlap-add reconstruction, a five-frame blending procedure is applied at the boundaries between observed and reconstructed trajectory segments to reduce potential discontinuities at these transition points. The complete dataset consisted of 5,920 generated motion capture files, which were split into training and validation subsets at 90%/10 %. Model optimisation was performed for 30 epochs using the Adam optimiser. It should be noted that the dataset contained recordings from two participants, and derived augmented versions of sequences from the same individuals were present in both the training and validation sets. To prevent the network from memorising participant-specific absolute coordinates, all spatial data were transformed into a local coordinate system relative to the LASI marker (Eq. 6) before training. This forced the model to learn generalised kinematic relationships rather than absolute positions, effectively mitigating the risk of participant-specific overfitting. The 90/10 training-validation split was performed after the augmentation process, but the assignment was made at the sequence level rather than at the level of individual generated files. Thus, all files derived from a given augmented sequence were assigned exclusively to either the training or validation subset. Consequently, different versions of the same source sequence were not present in both subsets.

3.3.2. Loss function

A composite objective function was utilised to simultaneously optimise positional accuracy, temporal smoothness, and biomechanical consistency (Yuhai et al., 2024):

$$L = L_{pos} + 0.6L_{smooth} + 0.6L_{bone} \quad (12)$$

The positional reconstruction loss was defined using Mean Squared Error (MSE):

$$L_{pos} = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2 \quad (13)$$

Temporal smoothness regularization minimized abrupt frame-to-frame trajectory changes, is given by (Yuhai et al., 2024):

$$L_{smooth} = \frac{1}{T-1} \sum_{t=2}^T \|\hat{y}_t - \widehat{y_{t-1}}\|^2 \quad (14)$$

To preserve realistic anatomical structure, an additional bone-length consistency constraint was introduced (Li et al., 2021):

$$L_{bone} = \frac{1}{B} \sum_{b=1}^B (\|\widehat{d}_b\| - \|d_b\|)^2 \quad (15)$$

where: d_b – ground truth bone vector,
 $\widehat{d}_b$ – reconstructed counterpart.

3.3.3. Learning behaviour

The learning curves show a consistent decrease in both training and validation loss, indicating stable convergence without significant overfitting (Figure 1). The reduction of both position and bone-related errors confirms that the model learns accurate reconstruction while maintaining realistic motion structure. Small oscillations visible in the learning curves can be attributed to the stochastic nature of the training process, the relatively limited number of original motion capture sequences, and the variability introduced by artificial gap generation. Such fluctuations are common in deep learning models trained on biomechanical time-series data and do not indicate instability of the optimisation process.

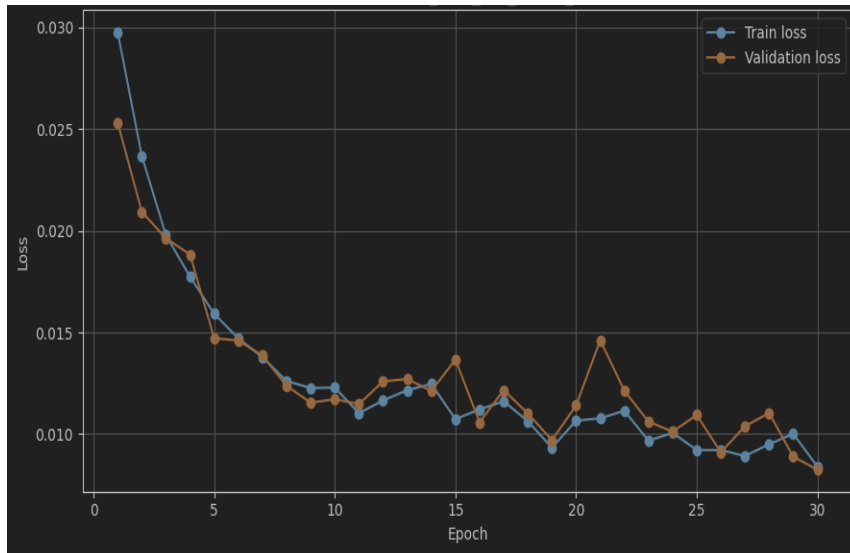


Fig. 1. Loss function plot during training

3.4. Metrics

To conduct a comparative analysis of both methods for reconstructing the positions of occluded markers, 12 evaluation scenarios were systematically defined using the naming convention [x]m_[y]f. In this notation:

- x denotes the number of concurrently missing markers, taking values from the set {2, 6, 10, 14};
- y represents the temporal duration of the missing data gap in frames, taking values from the set {100, 300, 500}.

These 12 sequences were specifically constructed to evaluate the model's robustness to previously unseen missing-data configurations. They were generated by introducing distinct, non-repeating combinations of temporal gaps and missing-marker patterns that were not used in the training or validation samples. Thus, while the underlying motion-capture recordings were shared, the missing-data configurations presented to the model during the final evaluation were unseen during model development.

The performance of evaluated methods was quantified using the following metrics:

- Mean Masked Root Mean Square Error (RMSE): to prevent artificial attenuation of the error metric, the RMSE was calculated only for frames with missing markers (Kucherenko et al., 2024). Incorporating intact ground-truth data into this calculation would inherently skew the overall error assessment. The masked RMSE is defined as:

$$\text{RMSE}_{\text{masked}} = \sqrt{\frac{\sum_{i=1}^N m_i (y_i - \hat{y}_i)^2}{\sum_{i=1}^N m_i}} \quad (16)$$

where: m – binary mask denoting whether data point is missing ($m_i = 1$) or visible ($m_i = 0$).

- Mean Estimation Time: the average computational time required to process each scenario.

4. RESULTS AND DISCUSSION

To provide a clear comparison of the applied modifications, the results obtained using a baseline, unconstrained Kalman filter implementation are gathered alongside the proposed methods in Figure 2. The results presented in the subsequent figures were calculated in the following manner: for each of the 12 occlusion scenarios, the evaluation was performed across a dedicated subset of 12 distinct sequences and averaged for a given configuration.

4.1. Reconstruction accuracy

Across all evaluated scenarios, the BiLSTM-based model consistently achieved lower RMSE values than the UKF-based (custom Kalman) approach and the baseline Kalman filter (Figure 2). The performance gap between the two methods increased with increasing occlusion complexity, becoming particularly pronounced in more demanding scenarios.

The UKF-based (custom Kalman) method exhibited high sensitivity to temporal scaling. For short gaps (100 frames), the UKF maintained moderate error bounds, but as the duration extended to 300 and 500 frames, it suffered severe trajectory drift. In contrast, the BiLSTM model demonstrated remarkable temporal robustness. For instance, in moderate sparsity conditions (6 missing markers), the LSTM's error increased by less than 20% when scaling from 100 to 500 frames, effectively neutralising the drift issues that caused the UKF's error to increase by over 200% under identical conditions. While the LSTM proved highly resistant to temporal scaling, spatial distribution proved its primary limiting factor. For configurations with low-to-moderate marker missingness (2m and 6m), the BiLSTM maintained errors strictly below 2.10 cm, outperforming the UKF by approximately 70–90%. However, as spatial sparsity increased to extreme levels (10m and 14m), the LSTM's accuracy degraded, yielding errors exceeding 10 cm in the most severe cases.

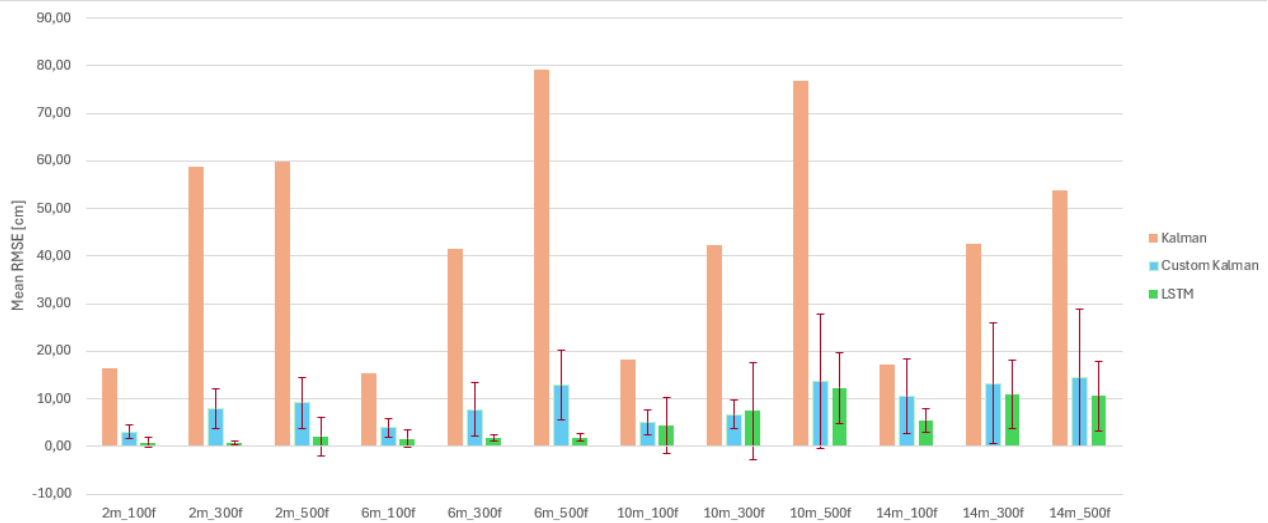


Fig. 2. Mean masked RMSE for every scenario. Error bars represent the standard deviation of the masked RMSE across the 12 evaluation sequences

Including bone-length and smoothness constraints in the BiLSTM loss function enabled the reconstruction of biomechanically plausible motion. The reconstructed trajectories preserved consistent inter-marker distances and exhibited smooth temporal transitions, even in highly corrupted sequences.

In comparison, the UKF-based method, despite incorporating rigid-body assumptions and local geometric constraints, exhibited reduced stability when multiple markers were missing simultaneously. This instability increased with spatial sparsity, leading to a significant escalation in reconstruction error, sometimes exceeding 100% relative to low-occlusion conditions. However, it must be emphasised that the proposed custom Kalman filter substantially outperformed the baseline unconstrained Kalman filter across all configurations. For example, in the 6-marker/500-frame scenario, the baseline Kalman drifted drastically, exceeding a 70 cm error. The custom UKF maintained an error of roughly 12 cm. This substantial improvement directly stems from the hierarchical estimation logic implemented in the custom UKF. By utilising the Kabsch algorithm and spatial sphere intersections based on visible neighbouring markers, the custom filter continuously corrects the trajectory with real-time structural anchors, whereas the baseline filter relies solely on linear extrapolation of the previous state's velocity, which inevitably diverges during non-linear human motion.

4.2. Estimation time

Evaluating execution time is a critical factor in modern motion capture analysis. When discussing computational efficiency, it is crucial to distinguish between total sequence throughput (offline processing) and per-frame latency (real-time processing). Applications such as virtual reality training systems and live biomechanical monitoring demand strict real-time execution; any delay in reconstructing missing markers disrupts the synchronisation between the user's physical movement and the system's digital response.

The processing time for the UKF-based method exhibited a relatively high baseline and considerable variance, with estimation times ranging from 47.97 s to 76.15 s per sequence (Figure 3). Notably, this computational burden did not correlate linearly with either spatial sparsity or the duration of temporal occlusion. Instead, execution times plateaued at approximately 60 seconds per sequence, reflecting the constant computational cost of internal-state update operations. Conversely, the BiLSTM-based architecture demonstrated significantly superior overall computational optimisation when processing large volumes of data. Inference latency for the neural network remained tightly constrained between 12.77 s and 21.80 s, with most evaluated configurations completing the entire trajectory reconstruction in 13 to 15 seconds. In parallel with the behaviour of the custom UKF, the neural model's execution time was largely independent of the severity of data degradation. Compared with the baseline Kalman filter, the custom UKF required approximately twice the processing time (the baseline averaged ~20-25 seconds). This increase is the expected trade-off for the advanced spatial logic; the baseline filter only performs matrix inversions for a single marker's state, whereas the custom UKF continuously queries the spatial coordinates of neighbouring markers and computes complex rigid-body transformations (via the Kabsch algorithm) at every frame.

While the BiLSTM model processes full sequences substantially faster in terms of throughput, its fundamental architecture prevents it from being a true real-time application. Because the network is bidirectional, it strictly requires access to both past and future frames to compute the temporal dependencies necessary for its high kinematic accuracy. Consequently, it cannot process live data streams without introducing substantial buffering latency. In contrast, the custom UKF processes kinematic data sequentially, computing the current frame's state utilising only past and present measurements. Although it requires more computational time to process a complete 3000-frame sequence than the BiLSTM, its frame-by-frame operational logic makes it fundamentally compatible with real-time tracking systems.

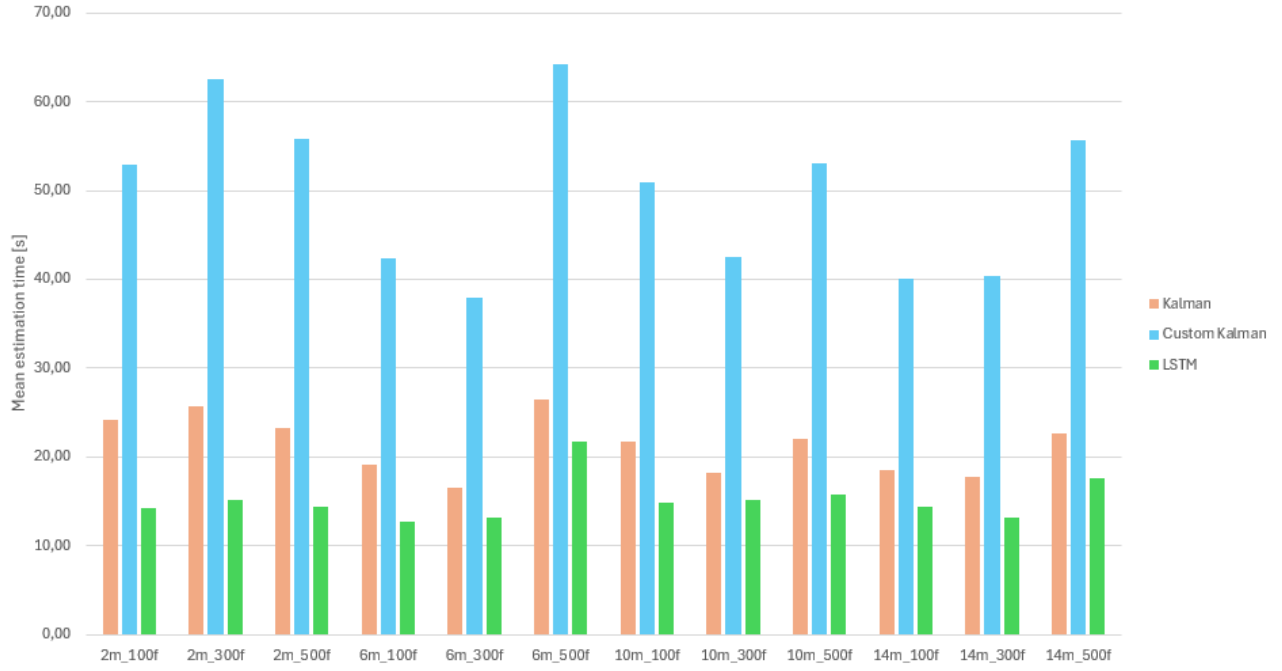


Fig. 3. Mean estimation time for every scenario

Overall, the empirical results and architectural paradigms dictate distinct use cases for both methods. The custom UKF method is well-suited to real-time environments and live tracking systems, where future frames are unavailable, offering a highly interpretable and mathematically transparent framework. On the other hand, the BiLSTM model is the optimal solution for offline post-processing of pre-recorded motion-capture datasets. In such scenarios, where complete sequences are available, and maximum trajectory reconstruction accuracy is prioritised over live execution constraints, the neural network effortlessly outperforms the classical filtering approach, particularly during severe, multi-marker occlusions.

4.3. Comparison with state-of-the-art methods

To provide a broader context for the proposed solutions, Table 2 compares our methods with existing gap-filling approaches documented in recent literature. Due to variations in datasets, marker configurations, and gap-generation protocols across studies, direct one-to-one comparisons are challenging. Nevertheless, the reported RMSE values provide a reliable benchmark for evaluating algorithmic robustness across varying levels of occlusion severity. It is crucial to emphasise that the RMSE values reported in Table 2 are derived from different datasets, marker sets, and disparate occlusion simulation protocols. Consequently, these metrics are not directly comparable and should be interpreted solely as contextual benchmarks for algorithmic behaviour under varying conditions.

Tab. 2. Performance comparison of the proposed methods against state-of-the-art

Method	Algorithm	Evaluated gap length	Average RMSE / Notes
Gomes et al. (2021)	Custom UKF	Variable (start/end gap)	~3.9 cm (highly effective for boundary gap)
Li et al. (2021)	Bone-constrained LSTM	Up to 300 frames	~2.0 cm (highly adaptable to spatial sparsity)
Skurowski & Pawlyta (2024)	Tree-based regression	10 – 200 frames	Competitive but degrades rapidly past 200 frames
Proposed Custom UKF	Modified UKF	100, 300, 500 frames	~3.0 – 14.5 cm
Proposed BiLSTM	BiLSTM	100, 300, 500 frames	~0.8 – 12cm

The proposed BiLSTM aligns closely with the performance of advanced neural models like Li et al. (2021), particularly in its ability to maintain sub-2 cm errors during extended temporal gaps when spatial sparsity is low. Meanwhile, our Custom UKF demonstrates that carefully designed geometric constraints can push the

boundaries of classical estimation, offering a robust alternative when machine learning deployment is not feasible.

5. CONCLUSION

The problem of reconstructing missing data in motion capture systems poses a significant challenge for motion analysis, particularly in ensuring the high quality and continuity of recorded sequences. Missing data resulting from hardware limitations or measurement interference can significantly affect subsequent processing and interpretation of results, so developing effective methods to reconstruct them is of great practical importance. The conducted research revealed clear differences between the analysed approaches. The BiLSTM-based model achieved lower reconstruction errors across the vast majority of analysed scenarios, confirming its high effectiveness in modelling complex temporal and spatial dependencies. In particular, its resistance to extended data gaps was noticeable, with the increase in error being relatively small compared to the classical method.

The Kalman filter performed correctly for short data gaps, but its performance decreased significantly as the gap length and the number of missing markers increased. This indicates its limitations in modelling more complex motion trajectories and its greater susceptibility to the accumulation of estimation errors. The obtained results partially confirm the research hypotheses. The LSTM model achieved lower RMSE values than the Kalman filter across almost all scenarios and remained highly stable even with a small number of missing markers. However, the accuracy of both methods deteriorated as the number of missing markers increased, indicating a significant impact of the spatial distribution of missing data on reconstruction quality. Computational efficiency was also a significant aspect of the comparison. In this respect, the BiLSTM model demonstrated a significant advantage, achieving estimation times several times faster while maintaining higher accuracy in most cases. This indicates its potential for use in systems requiring fast data processing.

Comparing the custom-based Kalman filter with the classic Kalman filter and the BiLSTM method allows us to determine which approach best reproduces the characteristic movement patterns found in folk dances. This is important not only for data quality but also for its subsequent use in ethnochoreological studies, biomechanical analysis, education, archiving cultural heritage, and the creation of digital reconstructions and animations of traditional dance forms.

In summary, machine learning-based methods, and particularly sequential models, offer an effective alternative to classical approaches for reconstructing missing data in motion capture systems. At the same time, the results indicate that selecting an appropriate method should take into account the nature of the data and the specific characteristics of the missing data. Importantly, while the BiLSTM model excels in offline reconstruction of severe occlusions, the proposed custom-based Kalman filter proved to be a highly effective and computationally viable alternative for real-time systems.

Limitations of this study should be acknowledged. First, the evaluation was conducted using a dataset representing a specific domain of human movement. Second, the missing data scenarios were artificially generated. While this allowed for systematic quantification, real-world marker occlusions may exhibit more complex and unpredictable patterns. Third, the position of the LASI marker was available at every frame, and the local coordinate transformation remained well-defined throughout the reconstruction process. Finally, the proposed custom UKF method relies on a manually defined neighbourhood of markers; therefore, its performance is highly dependent on the initial configuration. In the future, it is possible to expand the research to include other deep learning model architectures and to analyse the impact of additional biomechanical constraints and diverse datasets on reconstruction quality. The individual contributions of the bone-length and temporal-smoothness constraints to reconstruction consistency and computational efficiency should be systematically investigated.

Acknowledgements

All motion capture data from the current study are available in the repository at <https://polskietanceludowe.cs.pollub.pl/dane.html>.

The research programme titled “Analysis of folk dances” was approved by the Commission for Research Ethics, No. 8/2024. The study was carried out as a part of the project “Preserve from posterity” – 3D technology in protecting the traditions of national and regional dances (No. POPUL/SP/0414/2023/01).

The research program titled “Analysis of folk dances” was approved by the Commission for Research Ethics, No. 2/2016 dated 8.04.2016.

Conflicts of Interest

The authors declare no conflict of interest.

REFERENCES

- Aristidou, A., Cameron, J., & Lasenby, J. (2008). Real-time estimation of missing markers in human motion capture. In *2008 2nd International Conference on Bioinformatics and Biomedical Engineering* (pp. 1343–1346). IEEE. <https://doi.org/10.1109/ICBBE.2008.665>
- Bai, Y., Yan, B., Zhou, C., Su, T., & Jin, X. (2023). State of art on state estimation: Kalman filter driven by machine learning. *Annual Reviews in Control*, 56, Article 100909. <https://doi.org/10.1016/j.arcontrol.2023.100909>
- Chen, H.-Y., Cheng, Y.-H., & Lo, A. (2023). Improve dancing skills with motion capture systems: Case study of a Taiwanese high school dance class. *Research in Dance Education*, 24(4), 342–360. <https://doi.org/10.1080/14647893.2021.1980524>
- Dziuba-Kozieł, M., Kozieł, G., Harasim, D., Kisała, P., & Kochanowicz, M. (2024). Method of automatic calibration and measurement of the light polarisation plane rotation with tilted fibre Bragg grating and discrete wavelet transform usage. *Advances in Science and Technology Research Journal*, 19(1), 165–177. <https://doi.org/10.12913/22998624/194890>
- Dziuba-Kozieł, M., Kozieł, G., Markiewicz, J., Kochanowicz, M., Dorosz, D., Miluski, P., & Kisała, P. (2026). Machine learning-assisted design of inverted gradient multiring optical fibres for flat-top beam generation. *Optics Express*, 34(4), 7159–7171. <https://doi.org/10.1364/OE.587143>
- Esaki, K., & Nagao, K. (2024). An efficient immersive self-training system for hip-hop dance performance with automatic evaluation features. *Applied Sciences*, 14(14), Article 5981. <https://doi.org/10.3390/app14145981>
- Feng, Y., Xiao, J., Zhuang, Y., Yang, X., Zhang, J. J., & Song, R. (2014). Exploiting temporal stability and low-rank structure for motion capture data refinement. *Information Sciences*, 277, 777–793. <https://doi.org/10.1016/j.ins.2014.03.013>
- FilterPy. (n.d.). *FilterPy documentation*. Retrieved September 15, 2026, from <https://filterpy.readthedocs.io/en/latest/>
- Gomes, D., Guimarães, V., & Silva, J. (2021). A fully-automatic gap filling approach for motion capture trajectories. *Applied Sciences*, 11(21), Article 9847. <https://doi.org/10.3390/app11219847>
- Hauenstein, J. D., Huebner, A., Wagle, J. P., Cobian, E. R., Cummings, J., Hills, C., McGinty, M., Merritt, M., Rosengarten, S., Skinner, K., Szemborski, M., & Wojtkiewicz, L. (2024). Reliability of markerless motion capture systems for assessing movement screenings. *Orthopaedic Journal of Sports Medicine*, 12(3), Article 23259671241234339. <https://doi.org/10.1177/23259671241234339>
- He, Y., Pang, A., Chen, X., Liang, H., Wu, M., Ma, Y., & Xu, L. (2021). *ChallenCap: Monocular 3D capture of challenging human performances using multi-modal references*. *ArXiv*, *abs/2103.06747*. <https://doi.org/10.48550/arXiv.2103.06747>
- Ji, H., Wang, L., Zhang, Y., Li, Z., & Wei, C. (2023). A review of human pose estimation methods in markerless motion capture. *Computer-Aided Design and Applications*, 21(3), 392–423. <https://doi.org/10.14733/cadaps.2024.392-423>
- Kico, I., Zelniczek, D., & Liarokapis, F. (2020). Assessing the learning of folk dance movements using immersive virtual reality. In *2020 24th International Conference Information Visualisation (IV)* (pp. 587–592). IEEE. <https://doi.org/10.1109/IV51561.2020.00100>
- Kucherenko, T., Peristy, D., & Bütepage, J. (2024). *Evaluating the evaluators: Towards human-aligned metrics for missing markers reconstruction*. *ArXiv*, *abs/2410.14334*. <https://doi.org/10.48550/arXiv.2410.14334>
- Lai, R. Y. Q., Yuen, P. C., & Lee, K. K. W. (2011). Motion capture data completion and denoising by singular value thresholding. In *Eurographics 2011 — Short Papers* (pp. 45–48). The Eurographics Association. <https://doi.org/10.2312/EG2011/SHORT/045-048>
- Lannan, N., Zhou, L., Fan, G., & Hausselle, J. (2020). Human motion enhancement using nonlinear Kalman filter assisted convolutional autoencoders. In *2020 IEEE 20th International Conference on Bioinformatics and Bioengineering (BIBE)* (pp. 1008–1015). IEEE. <https://doi.org/10.1109/BIBE50027.2020.00171>
- Li, J., Xiao, D., Li, K., & Li, J. (2021). Graph matching for marker labeling and missing marker reconstruction with bone constraint by LSTM in optical motion capture. *IEEE Access*, 9, 34868–34881. <https://doi.org/10.1109/ACCESS.2021.3060385>
- Maggioni, V., Azevedo-Coste, C., Durand, S., & Bailly, F. (2025). Optimisation and comparison of markerless and marker-based motion capture methods for hand and finger movement analysis. *Sensors*, 25(4), Article 1079. <https://doi.org/10.3390/s25041079>
- Malekian, L., & Lapeer, R. (2024). Self-occluded human pose recovery in monocular video motion capture. In *2024 14th International Conference on Pattern Recognition Systems (ICPRS)* (pp. 1–6). IEEE. <https://doi.org/10.1109/ICPRS62101.2024.10677815>
- Martini, E., Boldo, M., & Bombieri, N. (2024). FLK: A filter with learned kinematics for real-time 3D human pose estimation. *Signal Processing*, 224, Article 109598. <https://doi.org/10.1016/j.sigpro.2024.109598>
- Millour, G., Velásquez, A. T., & Domingue, F. (2023). A literature overview of modern biomechanical-based technologies for bike-fitting professionals and coaches. *International Journal of Sports Science & Coaching*, 18(1), 292–303. <https://doi.org/10.1177/17479541221123960>
- Mohaoui, S., & Dmytryshyn, A. (2026). Low-rank completion for motion capture data recovery: Approaches, constraints, and algorithms. *Computer Science Review*, 60, Article 100878. <https://doi.org/10.1016/j.cosrev.2025.100878>
- Powroznik, P., Skublewska-Paszkowska, M., Dziedzic, K., & Barszcz, M. (2025). Feature fusion graph consecutive-attention network for skeleton-based tennis action recognition. *Applied Sciences*, 15(10), Article 5320. <https://doi.org/10.3390/app15105320>

- Senecal, S., Nijdam, N. A., Aristidou, A., & Magnenat-Thalmann, N. (2020). Salsa dance learning evaluation and motion analysis in gamified virtual reality environment. *Multimedia Tools and Applications*, 79(33–34), 24621–24643. <https://doi.org/10.1007/s11042-020-09192-y>
- Skublewska-Paszkowska, M. (2024). *Tennis motion recognition: Design of classification approaches and experimental studies*. Lublin University of Technology Publishing House.
- Skublewska-Paszkowska, M., Powroźnik, P., Barszcz, M., Dziedzic, K., & Aristodou, A. (2024). Identifying and animating movement of Zeibekiko sequences by spatial temporal graph convolutional network with multi attention modules. *Advances in Science and Technology Research Journal*, 18(8), 217–227. <https://doi.org/10.12913/22998624/193180>
- Skurowski, P., & Pawlyta, M. (2021). Gap reconstruction in optical motion capture sequences using neural networks. *Sensors*, 21(18), Article 6115. <https://doi.org/10.3390/s21186115>
- Skurowski, P., & Pawlyta, M. (2024). Tree based regression methods for gap reconstruction of motion capture sequences. *Biomedical Signal Processing and Control*, 88, Article 105641. <https://doi.org/10.1016/j.bspc.2023.105641>
- van der Kruk, E., & Reijne, M. M. (2018). Accuracy of human motion capture systems for sport applications; state-of-the-art review. *European Journal of Sport Science*, 18(6), 806–819. <https://doi.org/10.1080/17461391.2018.1463397>
- Vicon Motion Systems. (2022). *Plug-in gait reference guide* (Vicon Documentation).
- Zhou, H., & Hu, H. (2008). Human motion tracking for rehabilitation — A survey. *Biomedical Signal Processing and Control*, 3(1), 1–18. <https://doi.org/10.1016/j.bspc.2007.09.001>
- Yuhai, O., Choi, A., Cho, Y., Kim, H., & Mun, J. H. (2024). Deep-Learning-Based Recovery of Missing Optical Marker Trajectories in 3D Motion Capture Systems. *Bioengineering*, 11(6), 560. <https://doi.org/10.3390/bioengineering11060560>