Received 15 May 2025, accepted 22 May 2026, date of publication 30 September 2026
www.computingonline.net / computing@computingonline.net

Print ISSN 1727-6209
Online ISSN 2312-5381
DOI 10.47839/ijc.25.3.4850

High-Precision Markerless Human Motion Analysis in Real-Time

BOHDAN MYCHKOVSKYI, ANASTASIIA NIKOLAIENKO, EVGEN IVANOV

Department of Software Systems and Technologies, Taras Shevchenko National University of Kyiv,
Bohdan Hawrylyshyn str. 24, Kyiv, UA-04116, Ukraine

Corresponding author: Bohdan Mychkovskyi (e-mail: bogdanm@knu.ua).

ABSTRACT Markerless human motion analysis is important for a number of applications, in particular in the field of interactive computer systems. However, its widespread adoption is hindered by technical complexity. Existing systems either require specialized equipment, or lack robustness with respect to changes in lighting conditions, camera, and user appearance. This study presents a new technology for real-time, high-precision markerless human motion analysis that does not require specialized equipment. This technology is based on a robust and computationally efficient approach to comparison of video streams that represent reference and actual movements. The proposed technology was implemented in a modular software system. Experimental results confirmed the system's ability to identify and quantify movements under different conditions. During experiments the system achieved levels of accuracy and computational efficiency that make it potentially suitable for various applications. In the paper the main ideas of the proposed approach and experimental results are described and potential applications are discussed.

KEYWORDS computer vision, video analysis, MediaPipe, DTW, skeleton model, embedded systems.

I. INTRODUCTION

Human motion analysis plays a critical role in various fields, including interactive computer systems, sports, medical rehabilitation, ergonomics. Over the past decade, substantial progress has been made in motion capture technologies. However, achieving a truly universal solution that combines high accuracy and ease of use remains a challenge.

In accordance with [1], traditional marker-based motion capture systems are hindered by several fundamental limitations: high equipment cost, long setup time, reliance on trained personnel, and potential distortion of natural movement due to markers. While such systems provide high measurement accuracy, they often require artificial conditions that are not convenient for most applications.

Recent advances in computer vision have enabled markerless motion analysis systems, which offer a promising balance between accuracy and usability. For example, in [2] it was demonstrated that modern monocular, markerless systems have potential for the use in medicine. Nonetheless, these systems still face considerable challenges, especially under non-ideal conditions such as variable lighting, occlusions, and complex motion [3], [4].

Key issues include the ability of a system to adapt to different conditions and user appearance, and the need for automatic calibration without expert intervention [5]. Work [6] identified further issues such as complex scene dynamics that lower recognition accuracy.

Differences in user appearance exacerbate these challenges. Limb proportions vary significantly across ethnic groups and individuals [7]. Canonical anthropometric ratios represent average values and do not account for such individual differences [8], [9]. The ratio of leg length to total body height can vary significantly between individuals [8], [10].

From a computer vision perspective, difficulties also arise due to perspective distortions, camera angle variations, and partial occlusions. These factors can make it difficult to analyze movement based on its video recording [11], [12]. Objects closer to the camera appear disproportionately larger, making proportion-based methods sensitive to the camera configuration and location of the subject in the frame. With active dynamics, the apparent proportions relative to the real ones can significantly increase or decrease. These distortions are additionally complicated by the technical features of the video recording process [13].

In [14], it was noted that although convolutional neural networks have achieved great success in pose recognition, more in-depth research is needed in the areas of feature extraction and 3D training data generation. Work [15] confirmed that human pose estimation methods face serious challenges due to significant variations in user appearance, occlusions, noisy backgrounds, different angles, and other factors.

Despite rapid development of artificial intelligence and machine learning which underpin most modern markerless motion analysis solutions, many models still lack generalizability. As several studies show [16], [17], [18], models based on CNNs, LSTMs, or attention mechanisms can achieve high accuracy on specific datasets. However, their performance often becomes lower under unfamiliar conditions caused by changes in background, camera, movement speed, or movement type.

To address temporal variability in motion sequences, this study uses a variant of Dynamic Time Warping (DTW) algorithm which allows one to align time sequences with variations in speed and time [19], [20]. A detailed description of a modified version of this algorithm adapted for real-time human motion analysis tasks is presented in Section 1.5 "Adaptive Dynamic Time Warping Algorithm". This modification overcomes the limitations of traditional methods of time series comparison, providing flexible matching of points between sequences even in the presence of time shifts and variations in speed of movement.

In contrast to purely data-driven models, this work proposes a hybrid approach that integrates physically based motion modeling with computer vision. The proposed system incorporates anatomical constraints in order to minimize errors and enable physically plausible pose estimation important for applications. By taking into account biomechanics, the model used in the proposed system improves adaptability and robustness under real-world conditions. Moreover, the system self-calibrates to individual users without the need for supervision by an expert and maintains performance across diverse imaging conditions.

To incorporate anatomical constraints, the system proposed in this work uses BlazePose GHUM [21], [22], [23] – an architecture for real-time 3D pose estimation from single video input. Key features of BlazePose GHUM include high processing speed, a complete skeleton model, and robustness with respect to variable lighting, making it well-suited for deployment in uncontrolled environments.

The proposed solution implements comparison of human movement represented in video streams in real time and achieves high accuracy, adaptability, and convenience of deployment (in particular, the possibility of its use without specialized equipment). This allows it to overcome limitations of existing markerless systems.

The main tasks of the research presented in this work are as follows:

- select a human skeleton model;
- select an adequate method for determining the degree of similarity of the user's movements;
- develop a technology for efficient computation of the similarity degree of the movement sequence.

II. RESEARCH METHODOLOGY

A. SELECTION OF A HUMAN SKELETON MODEL

The proposed markerless human motion analysis technology is based on a new Adaptive Dynamic Time Warping (ADTW) algorithm which enables efficient computation of similarity measure for signals representing motion. Technical details of the ADTW algorithm are presented in Section 1.5 "Adaptive Dynamic Time Warping Algorithm". This section focuses on selection of a skeleton model that provides input data for ADTW.

To ensure accurate and efficient motion analysis, the proposed system employs MediaPipe BlazePose GHUM model. This model was selected for its favorable trade-off between computational efficiency and pose estimation accuracy in two-dimensional space. The selection process involved a comprehensive evaluation of available skeleton models based on their technical specifications and suitability for real-time motion analysis tasks.

Benchmarking was conducted using a standardized dataset of human movements with varying levels of complexity, captured under consistent lighting conditions and camera angles. Accuracy was quantified as the fraction of correctly detected keypoints, where a keypoint was considered correctly detected if the positional error did not exceed 5 % of the length of the corresponding body segment. The comparative results are summarized in Table 1.

Table 1. Comparison of keypoint detection models in terms of accuracy, performance, FPS, and latency.

Model	Platform	Keypoints	Accuracy (avg.)	FPS (avg.)	Latency
BlazePose	Browser (WebGL)	33	94.59 %	40-60 FPS	~15–25 ms
MoveNet	Browser (WebGL)	17 (Lightning), 21 (Thunder)	83.32 %	45-60 FPS	~15–23 ms
PoseNet	Browser (WebGL)	17	85.48 %	15-25 FPS	~40-66 ms
OpenPose (excluding face, fingers, feet)	Browser (WebAssembly)	25	95.71 %	2-5 FPS	~200-500 ms
OpenPose (body only)	Browser (WebAssembly)	18	96.48 %	4-7 FPS	~140-250 ms

Although OpenPose demonstrates the highest detection accuracy (95.71 % – 96.48 %), its extremely low frame rate (2 FPS – 7 FPS) and high latency (140 ms – 500 ms) make it unsuitable for real-time applications. Also, the substantial lag between actual movement and system response makes OpenPose impractical for the purpose interactive motion correction or feedback.

In contrast, BlazePose GHUM delivers an optimal balance: a competitive average accuracy of 94.59 %, frame rates of 40–60 FPS, and minimal latency (15–25 ms). These characteristics make BlazePose a suitable choice for practical, real-time motion analysis in many cases.

B. SELECTION OF A METHOD FOR DETERMINING SIMILARITY OF HUMAN MOVEMENTS

Historically, different methods of motion comparison were proposed. A comprehensive analysis of existing approaches [24] identifies five main directions for solving this task: absolute coordinates, Euclidean metrics, usage of body proportion ratios, neural network architectures and the angle-based approach to movement analysis. However, a significant part of these methods have shortcomings in practical application.

Traditional approaches based on absolute coordinates, Euclidean distance, and relative body proportions are inherently sensitive to extrinsic factors such as camera position, subject distance, and video scale. Variations in any of these parameters make these methods unreliable for motion comparison. Moreover, methods based on Euclidean distance struggle to adapt to nonlinear nature of human movement, particularly during complex actions [25], [26], [27]. In such cases these methods often lose their physical interpretability.

The angle-based approach to motion analysis is based on computing angles between body segments and offers higher robustness and interpretability. This approach relies on identifying three anatomical landmarks that form a joint angle, from which vectors are derived. Angular descriptors are inherently invariant to changes in image scale, anthropometric variability, and perspective distortion, making them particularly well-suited for markerless motion analysis [28], [29]. Scale invariance becomes especially important when analyzing video materials that have different resolutions or were recorded under different conditions.

The effectiveness of angular analysis is particularly significant for analysis of quick movements. The method can be implemented to correctly track and characterize movements with high angular velocities [30].

Despite its numerous advantages, the angle-based method has certain limitations that require critical consideration. One important challenge is insufficient accuracy of determining joint centers and angles, which limits the use of marker-free systems in clinical biomechanics and rehabilitation [31].

From the above mentioned discussion it can be concluded that angle-based methods are preferable for solving the problem considered in this work. So, a proposed implementation of an angle-based method will be described in the next sections.

C. COMPUTATION OF ANGULAR CHARACTERISTICS OF THE SELECTED SKELETON MODEL

Angle-based approach is based on calculation of angles between segments of a selected skeleton model (Fig. 1) [32].

Let us consider an example of calculating angular characteristics. Let point A be the right shoulder (11), point B be the right elbow (13), and point C be the right wrist (15). Let us create vectors BA (elbow – shoulder) and BC (elbow – wrist). The coordinates of the vectors are calculated as the difference of the corresponding coordinates of the points:

$$\overrightarrow{BA} = (A_x - B_x, A_y - B_y);\ \overrightarrow{BC} = (C_x - B_x, C_y - B_y). \quad (1)$$

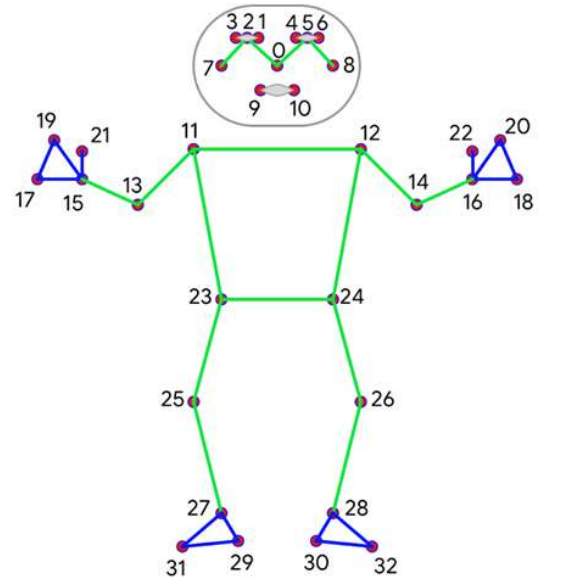


Figure 1. A skeleton model of 33 body points determined by the BlazePose GHUM algorithm is used for markerless real-time body pose determination. Source of the image – [32].

Vectors $\overrightarrow{BA}$ and $\overrightarrow{BC}$ form an angle θ (Fig. 2), which is a key parameter for assessing limb movement in two-dimensional space.

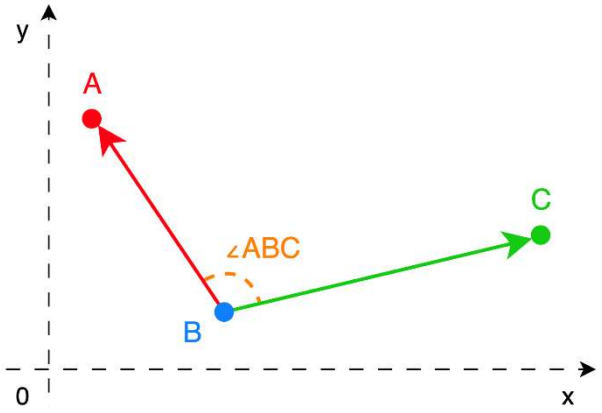


Figure 2. Visualization of the angle θ

The oriented angle between vectors on a plane can be determined using the *atan2* function as follows:

$$\angle ABC = \operatorname{atan2}\left(\overrightarrow{BA_x}\cdot\overrightarrow{BC_y} - \overrightarrow{BA_y}\cdot\overrightarrow{BC_x},\ \overrightarrow{BA_x}\cdot\overrightarrow{BC_x} - \overrightarrow{BA_y}\cdot\overrightarrow{BC_y}\right), \quad (2)$$

where $\overrightarrow{BA_x}$ and $\overrightarrow{BA_y}$ are the components of the vector $\overrightarrow{BA}$, $\overrightarrow{BC_x}$ and $\overrightarrow{BC_y}$ are the components of the vector $\overrightarrow{BC}$.

Using *atan2* allows one to obtain an oriented angle in the range [0°, 360°), eliminating the need for additional conditional logic to determine the direction of rotation.

Implementation of the angle-based motion analysis can be used to achieve high measurement accuracy and invariance to image scale. However, special attention should be paid to selection of the most informative angular characteristics. A skeleton model of 33 points allows for the calculation of a significant number of angles, but not all of them have the same importance. In addition to traditional anatomical joint angles, alternative configurations can be considered that include different combinations of skeleton model points. Optimization of the set of angular parameters can be based on the principle of minimal sufficiency: using the minimum set of parameters that provides maximum informativeness for a specific application domain and task. Such an approach can be used to reduce computational complexity of the resulting implementation and increase its robustness by eliminating redundant and potentially unstable parameters.

D. DETERMINATION OF THE OPTIMAL REFERENCE FRAME

The proposed iterative frame processing algorithm implements sequential accumulation of pose characteristics obtained from the skeleton model. The results are ranked in descending order of their significance and relevant data is extracted in order to determine reference proportions. Changes are calculated relative to reference values.

Characteristics used to determine whether a frame is suitable as a reference frame are described below.

The first characteristic $\theta_{parallel}$ reflects whether shoulder and hip lines are parallel. It is calculated as the angle between the vectors (S1, S2) and (H1, H2), where (S1, S2) are the left/right shoulder, (H1, H2) are the left/right hip.

The formula used is:

$$\theta_{parallel} = \arccos\left(\frac{\overrightarrow{S_1S_2}\cdot\overrightarrow{H_1H_2}}{|\overrightarrow{S_1S_2}|\cdot|\overrightarrow{H_1H_2}|}\right), \quad (3)$$

where the permissible deviation is no more than 5°.

The second characteristic is the angle of inclination of shoulders. It is computed in a similar way using the coordinates of the left (xS1; yS1) and right shoulder (xS2; yS2).

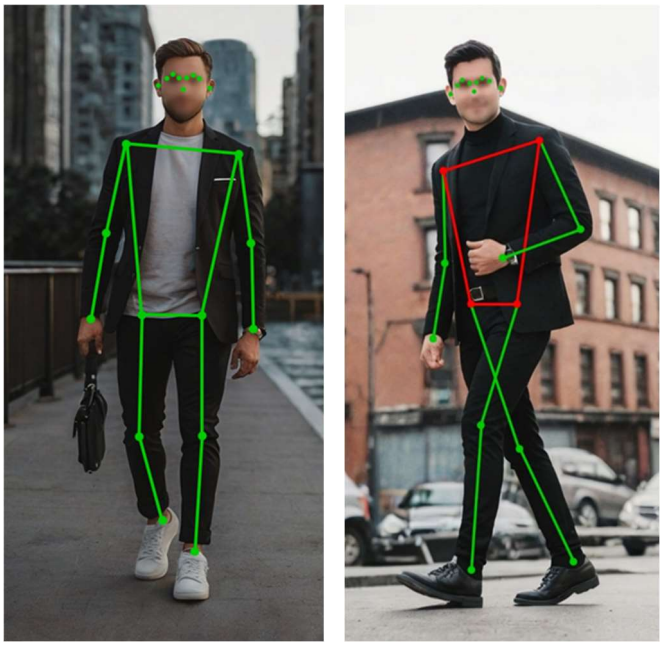
The next characteristic ΔL is defined as the relative difference between the lengths of the right (S2, H2) and left (S1, H1) sides:

$$\Delta L = \frac{|L_{left} - L_{right}|}{\max(L_{left},\ L_{right})}\cdot 100\ \%, \quad (4)$$

where $\Delta L \leq 5\ \%$.

The final characteristic concerns the position of the head. The optimal angle is considered to be 70°-90° between the segments formed by the pairs of points (0; 11) and (0; 12) in Fig. 1.

As an example, consider the frames shown in Fig. 3.



a *b*

Figure 3. Examples of frames determined by the algorithm as suitable for the use as a reference (a) and not suitable for the use as a reference frame (b).

The frame shown in Fig. 3*a* satisfies all the criteria: the shoulder and hip lines are almost parallel, the inclinations are minimal, and the body proportions are symmetrical. This makes it suitable for the use as a reference.

In contrast, Fig. 3*b* shows a frame that the algorithm considers unsuitable as a reference due to violations of the parallelism and inclination criteria. Visible violations include a significant inclination of the shoulders.

Using the above mentioned characteristics and criteria, an algorithm for selecting an optimal reference frame can be given. Such an algorithm is described below and is illustrated in Fig. 4.

Explanations for Fig. 4:

- Select the required number of frames for analysis;
- Process each frame sequentially;
- Calculate the values of 5 characteristics described above for the current frame;
- Compute the total quality score as a weighted sum of values of 5 characteristics described above;
- Sort frames in descending order by the total quality score;

- Return the best frame data (with the highest total quality score) or return the first frame if there are no other options.

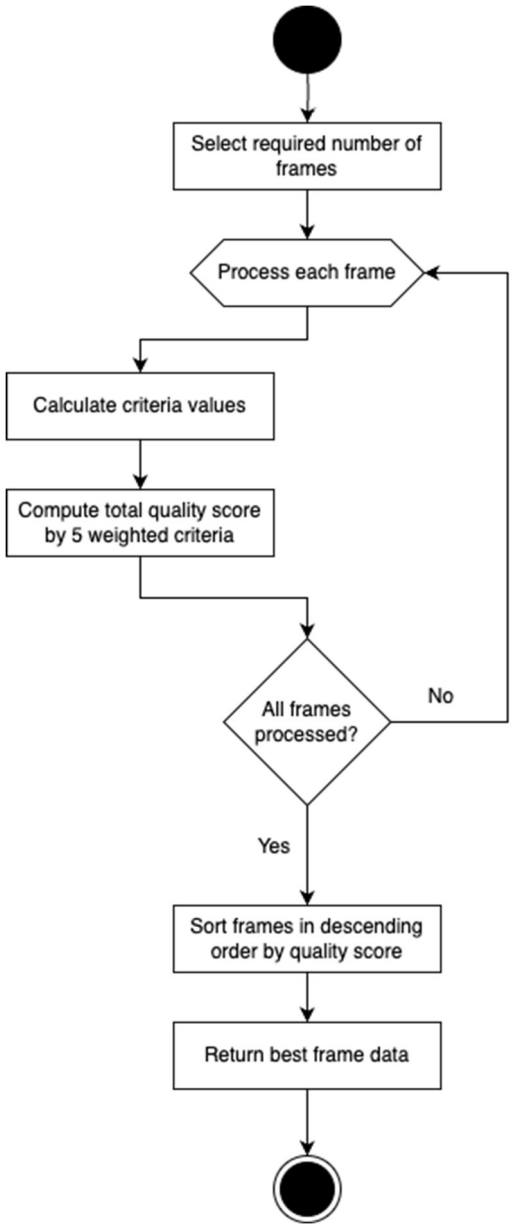


Figure 4. Reference frame selection algorithm based on 5 criteria.

As can be seen from the algorithm, if no frame meets the criteria, the algorithm automatically selects the first frame of the sequence. This decision is based on the fact that the initial frame often reflects the starting position which is the basis for most exercises.

E. ADAPTIVE DYNAMIC TIME WARPING ALGORITHM

Dynamic Time Warping (DTW) is a classical algorithm used to identify the optimal alignment between two temporal sequences [33]. Unlike Euclidean distance, which rigidly compares points with identical time indices, DTW allows for alignment of corresponding points, thereby enabling robust comparisons despite temporal distortions.

As shown in Fig. 5, DTW creates nonlinear mappings between sequences, allowing points to be associated with their best matches.

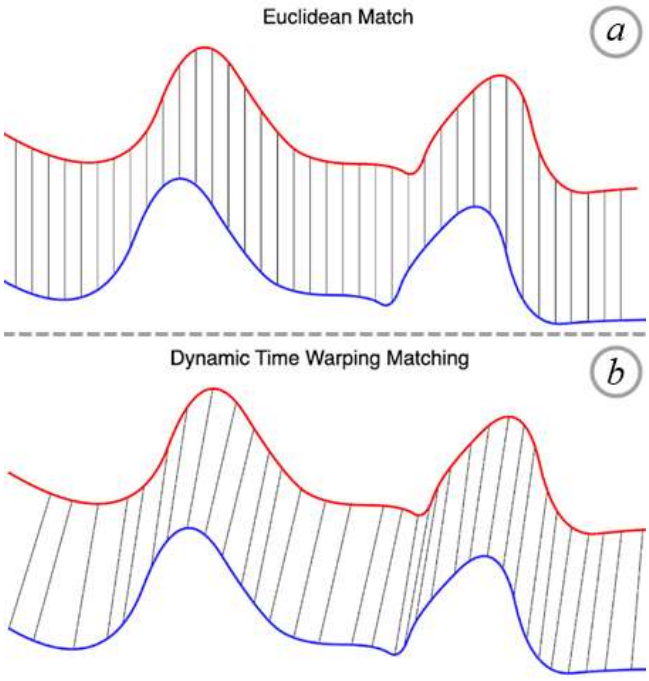


Figure 5. Examples of matching sequences (a), (b)

Despite this, direct application of both Euclidean comparison and classical DTW turns out to be inadequate in the context of the task of analysis of changes of angular characteristics of a human skeleton model in real time. Euclidean distance is based on a rigid comparison of points with the same time indices and does not take into account possible variations in speed of motion, which are inevitable in real human movements. This leads to a significant accumulation of errors even with minor deviations in time. The classical DTW algorithm works with two fully known sequences and seeks to minimize the global distance without taking into account local accuracy requirements. However, in the task considered in this work, only one sequence (the reference one) is fully formed, while the other is gradually formed in real time. Such asymmetry imposes new requirements on an adequate algorithm: it must be able to incrementally process new values, ensuring comparison not over the full window, but over the available part of the stream, which is constantly changing. This complicates the alignment task and requires modification of the classical approach in order to correctly handle such scenarios.

Let $R = (r_1, r_2 \dots, r_m)$ denote the fully known reference sequence and $Q = (q_1, q_2 \dots, q_t)$ be incrementally observed query stream. For each incoming observation q_t, ADTW restricts alignment candidates to the reference window $j \in [max(0, t-k), min(m-1, t+n)]$, where k and n are the backward lag and forward lead parameters. The local cost function $s(q_t, r_j)$ incorporates a per-segment validation weight w_j and is computed as a weighted sum of absolute differences across all d feature dimensions:

$$s(q_t, r_j) = \sum_{\text{dim}} w_{j,\text{dim}} \cdot |q_{t,\text{dim}} - r_{j,\text{dim}}|, \qquad (5)$$

where dim denotes each feature dimension of the comparison parameters representation, $w_{j,\text{dim}}$ is the per-segment weight for

dimension dim, $q_{t,\text{dim}}$ is the dim-th feature of the incoming frame, and $r_{j,\text{dim}}$ is the corresponding reference value.

The cumulative cost matrix $D(t,j)$ is updated according to:

$$D(t,j) = s\big(q_t, r_j\big) + \min\big(\mathcal{P}(t,j)\big), \qquad (6)$$

where $\mathcal{P}(t,j)$ is the set of finite-valued predecessors:

When $\mathcal{P}(t,j) = \emptyset$ (base case - no reachable predecessor exists), $D(t,j) = s\big(q_t, r_j\big)$. Only predecessors with finite values are admitted; cells outside the window are initialized to ∞ and excluded implicitly. The frame is classified as compliant $D(t,j^*) \le \tau$ if $j^* = argmin_{j\in[j_{min}, j_{max}]} D(t,j)$ and τ is a predefined tolerance threshold.

Pseudocode:

Input: R[0..m-1], D, q_t, t, k, n, W[0..m-1], τ
Output: COMPLIANT or NON-COMPLIANT

1. j_min ← max(0, t − k); j_max ← min(m − 1, t + n)
2. for j = j_min to j_max:
3. s ← Σ_dim W[j][dim] · |q_t[dim] − R[j][dim]|
4. predecessors ← ∅
   if t > 0 and j > 0 and D[t−1][j−1] < ∞ : add D[t−1][j−1]
   if t > 0 and D[t−1][j] < ∞ : add D[t−1][j]
   if j > j_min and D[t][j−1] < ∞ : add D[t][j−1]

   if predecessors = ∅:
   D[t][j] ← s
   else:
   D[t][j] ← s + min(predecessors)
5. j* ← argmin_{j ∈ [j_min, j_max]} D[t][j]
6. return (D[t][j*] ≤ τ). COMPLIANT : NON-COMPLIANT

The inner loop of ADTW_process_frame iterates over the alignment window $[j_{min}, j_{max}]$. The window width is $j_{max} - j_{min} + 1 = \min(m-1, t+n) - \max(0, t-k) + 1 \le k + n + 1$ and is therefore bounded by the constant $k + n + 1$, independent of both m and t.

Each iteration computes the local cost over d feature dimensions and performs a fixed number of predecessor lookups, giving $O(d)$ per iteration. The per-frame cost is thus $O\big((k+n)\cdot d\big)$. For a stream of T frames the total execution time is $O(T \cdot (k+n) \cdot d)$.

This is the key computational advantage of the windowed formulation: whereas a full DTW comparison of two sequences of length T and m requires $O(T \cdot m \cdot d)$ time, ADTW replaces the factor m with the fixed window size $k + n + 1$, enabling real-time per-frame processing regardless of reference sequence length.

A single distance metric does not allow one to differentiate the level of accuracy for individual data fragments or values, which is critical in the given task, where individual time intervals or parameters may require different validation coefficients. For example, in scenarios with sudden changes in speed or external obstacles, the standard DTW calculates distances between all points in a series using a rigidly defined trajectory [34], [35]. To overcome these limitations, a modification of the DTW algorithm, the Adaptive Dynamic Time Warping (ADTW) is proposed (Fig. 6).

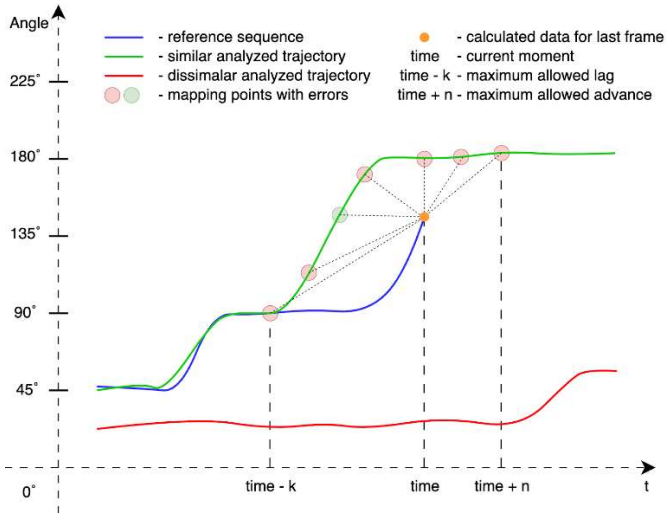


Figure 6. Visualization of the ADTW method for analyzing angular characteristics of skeleton model

The proposed trajectory analysis method (see Fig. 6) uses a time window to compare the input data with the reference sequence. Each new flow point (marked in orange at the time position) is compared with reference points within the allowable lag (time-k) and lead (time+n). The green line represents a similar trajectory, which, despite the deviation from the reference (blue line) in the range of 90°-180°, remains acceptable, while the red line represents an unacceptable trajectory. This fact cannot be established using Euclidean distance or standard DTW, but using ADTW it is possible to establish similarity in this case. The dotted connections and error circles visualize the process of assessing the compliance of the movement with the given parameters in real time.

In addition to increased robustness, ADTW reduces computational complexity by limiting the number of candidate comparisons and prioritizing points that are temporally close. This optimization yields significant performance gains, which are essential for real-time applications. However, such optimizations have trade-offs: the limited number of elements can lead to loss of relevant data outside the selected window, especially in conditions of sharp contextual changes, such as sudden stops or trajectory changes. In addition, introduction of additional parameters requires careful tuning to avoid suboptimal operation of the algorithm in unpredictable scenarios.

The described approach demonstrates better robustness compared to traditional DTW. Experiments have shown that dynamic selection of validation metrics allows one to reduce the average positioning error in the cases of unstable measurements, where the standard algorithm accumulates errors due to the rigidity of the single cost function. Further research should focus on automating parameter selection, possibly via machine learning to further improve performance and eliminate the need for manual calibration during deployment.

III. DESCRIPTION OF THE ADTW-BASED SOFTWARE SOLUTION FOR MOVEMENT EXECUTION CONTROL

The proposed software solution for real-time analysis of human movement is built around the Adaptive Dynamic Time

Warping (ADTW) algorithm and is composed of four interdependent modules:

- Video input and data processing module
- Pose validation module
- Analysis module
- Visualization module.

The video input module receives live video data from the webcam or device camera and transmits each frame directly to the BlazePose model which detects and returns the two-dimensional coordinates of 33 skeleton keypoints. These coordinates are then used to compute angular characteristics and inter-point distances required for further analysis.

The position validation module compares the current skeletal configuration with a predefined reference pose corresponding to the initial state of the exercise. The comparison is based on angular and spatial criteria. If the user's body position deviates from the required configuration, the system generates prompts in order to guide the user to get the correct starting position. Once all conditions are satisfied, the module emits a readiness signal allowing the movement analysis phase to begin.

The reference values for the video sequence are obtained using the method of selection of the optimal reference frame. The analysis module compares the calculated values obtained from the video stream with the scaled values from the video sequence using the ADTW method, taking into account time correspondences and deviations from the permissible range for each body segment.

The visualization module provides real-time feedback to the user during exercise execution. It displays hints and recommendations received from other modules, provides information about correctly positioned segments of the model and highlights those that are outside the permissible range (Fig. 7). Visual indication of the positions of body segments in real time allows one to improve exercise performance. The users can see whether their movements correspond to the reference movements and, if they do not, can try to correct them.

The interaction between software modules is based on the principle of sequential data processing with minimal delay between stages. The key feature is the absence of additional stages of video post-processing. This minimizes delay and ensures maximum performance.

IV. EXPERIMENTAL RESULTS

The proposed algorithm was evaluated using the Penn Action dataset [36], which contains 2,326 video recordings of physical exercises performed by individual subjects. Of these, 2,039 sequences were retained for evaluation, while the remaining 287 sequences were excluded because the video recordings, which showed multiple people in the frame, caused target identification errors in the single-subject skeleton detection model.

From each video sequence, motion parameters were extracted using the skeleton model. To evaluate algorithm performance, each original sequence was compared against a synthetic set of approximately 24,000 variations generated by applying pseudo-random keypoint displacement within defined ranges of 1–2%, 3–5%, and 6–10% of the corresponding keypoint value range, combined with temporal shifts simulating motion lag and lead relative to the original sequence.

Each variation was compared against the original sequence using the ADTW algorithm. A result was considered correct if the algorithm classified the variation consistently with its actual deviation: as compliant if it was within the permitted range and as non-compliant if it was outside it. For simple motion patterns, classification accuracy reached 97.8%. For complex multi-phase movements, accuracy was 92.3%, reflecting greater trajectory variability and higher sensitivity to simultaneous deviations across multiple angular dimensions.

In addition to dataset-based evaluation, the system was assessed under real operating conditions to verify runtime stability and robustness.

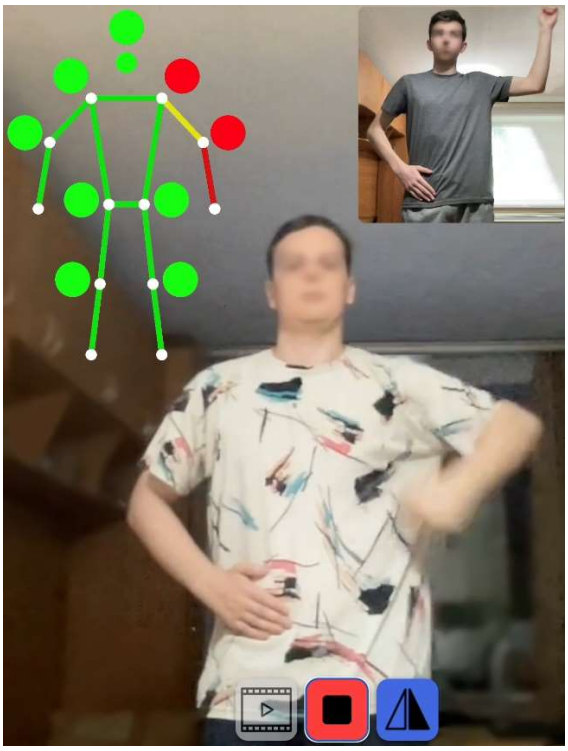


Figure 7. Visualization of real-time exercise feedback

The system uses a single camera and does not require synchronization of multiple devices. This reduces the computational load and makes the underlying technology suitable for the use on devices with limited resources, including smartphones.

The accuracy of results obtained by the system largely depends on the quality of the video stream. Critical factors that reduce efficiency include low resolution (less than 720p), poor or unstable lighting, shadows, atmospheric effects (fog, rain), and the presence of outsiders in the frame. High angular velocity of movements also leads to a decrease in accuracy. Optimal performance is achieved with a frontal view, camera location at body level, and mandatory head visibility.

Under favorable conditions, the system demonstrated stable operation without the need for recalibration or restart during extended sessions of up to 25 minutes. The ADTW-based analysis module maintained consistent performance even in dynamic scenarios involving partial occlusions, with tracking continuity ensured through coordinate buffering and spline interpolation.

A special calibration procedure can be implemented to partially compensate for negative conditions. To ensure efficiency and stable real-time performance, the system uses

asynchronous processing, adjustable frame rate (15–60 FPS), caching of reference data, and incremental change processing.

V. DISCUSSION

The developed markerless motion tracking technology can be potentially used in a wide range of application areas. The experimental validation results indicate a high level of accuracy and performance for the use in real-world conditions.

In the field of sports training, the described approach can be integrated as a component of analytical platforms for assessing exercise and movement techniques. In particular, the proposed method allows coaches to obtain quantitative data on athletes' movement patterns without the use of specialized equipment.

Another promising application of the developed technology is in rehabilitation medicine, where accurate tracking of patients' movements is crucial for assessing the progress of recovery. Implementation of marker-free tracking allows one to improve the assessment process and individualize rehabilitation programs based on quantitative data.

In the context of human-machine interaction systems, the proposed technology allows one to implement a natural control interface without the need for specialized controllers.

Economic benefit of the markerless approach is also substantial. Elimination of specialized markers and equipment significantly reduces costs of any implementation, making it more accessible and less dependent on specific hardware [37], [38]. This approach provides flexibility and scalability – it can be implemented in different settings without the need for expensive stationary installations.

A potential social benefit of the markerless approach is that movement tracking technologies can be made more accessible (e.g., educational institutions, sports clubs, rehabilitation centers, etc.).

It should be noted that potential applications of the proposed technology are not limited to the above-mentioned areas. Other potential applications areas include ergonomic analysis of workplaces, monitoring of industrial safety, and non-verbal communication.

VI. CONCLUSIONS

In this work software technology for markerless human motion analysis is implemented. It achieves high accuracy, computational efficiency, and robustness to typical challenges of real-time video analysis.

As a basis for markerless human motion analysis, the MediaPipe Blazepose GHUM skeleton model has been selected. This model provides an optimal tradeoff between the accuracy and processing speed.

To determine similarity of human movements, an angle-based analysis method is employed. This method achieves scale invariance and robustness with respect to anatomical variations.

A modified Adaptive Dynamic Time Warping (ADTW) algorithm is developed that overcomes limitations of the classical DTW in the context of the task of real-time human motion analysis. Unlike the classical approach, ADTW is able to incrementally process streaming data with one reference sequence and another incomplete sequence. It significantly reduces computational complexity by optimizing the comparison window. Experimental results also confirm the increased robustness of the algorithm when analyzing the angular characteristics obtained using human skeleton model.

Potential directions of further research include integration of multimodal data in order to increase the robustness of the system under complex lighting conditions and in the case of occlusions, and further optimizations to allow implementation on mobile devices.

Acknowledgements

The authors would like to thank O. Vasyliev for his recommendations on mathematical reasoning and data analysis. All aspects of the software implementation were performed by the first author independently.

References

[1] S. L. Colyer, M. Evans, D. P. Cosker, A. I. Salo, "A review of the evolution of vision-based motion analysis and the integration of advanced computer vision methods towards developing a markerless system," *Sports Medicine - Open*, vol. 4, no. 24, 2018. https://doi.org/10.1186/s40798-018-0139-y.

[2] T. Hellsten, J. Karlsson, M. Shamsuzzaman, G. Pulkkis, "The potential of computer vision-based marker-less human motion analysis for rehabilitation," *Rehabilitation Process and Outcome*, vol. 10, pp. 1-12, 2021. https://doi.org/10.1177/11795727211022330.

[3] W. W. T. Lam, Y. M. Tang, K. N. K. Fong, "A systematic review of the applications of markerless motion capture (MMC) technology for clinical measurement in rehabilitation," *Journal of NeuroEngineering and Rehabilitation*, vol. 20, no. 57, 2023. https://doi.org/10.1186/s12984-023-01186-9.

[4] C. Mercadal-Baudart, C. J. Liu, G. Farrell, M. Boyne, J. G. Escribano, A. Smolic, C. Simms, "Exercise quantification from single camera view markerless 3D pose estimation," *Heliyon*, vol. 10, issue 6, 2024. https://doi.org/10.1016/j.heliyon.2024.e27596.

[5] C. Zheng, W. Wu, C. Chen, T. Yang, S. Zhu, J. Shen, N. Kehtarnavaz, M. Shah, "Deep learning-based human pose estimation: A survey," *Association for Computing Machinery*, vol. 56, issue 1, 2023. https://doi.org/10.1145/3603618.

[6] N. Ma, Z. Wu, Y. M. Cheung, Y. Guo, Y. Gao, J. Li, "A survey of human action recognition and posture prediction," *Tsinghua Science and Technology*, vol. 27, issue 6, pp. 973-1001, 2022. https://doi.org/10.26599/TST.2021.9010068.

[7] S. Verbeek, "Limb Length Discrepancy: Prevalence, Patterns, and Variation in Pre- and Post-Industrial Humans," 2022, [Online]. Available at: http://arks.princeton.edu/ark:/88435/dsp01v118rh72b.

[8] C. Wall-Scheffler, "Women have stronger legs and other side-effects of human body proportions," *American Journal of Human Biology*, vol. 36, issue 5, 2024. https://doi.org/10.1002/ajhb.24034.

[9] Z. O. Abu-Faraj, G. F. Harris, P. A. Smith, S. Hassani, *Human Gait and Clinical Movement Analysis*, in: J.G. Webster (Ed.), Wiley Encyclopedia of Electrical and Electronics Engineering, Second Edition, John Wiley & Sons, Inc., New York, USA, 2015, pp. 1-34. https://doi.org/10.1002/047134608X.W6606.pub2.

[10] T. W. Holliday, C. B. Ruff, "Relative variation in human proximal and distal limb segment lengths," *American Journal of Physical Anthropology*, vol. 116, no. 1, pp. 26-33, 2001. https://doi.org/10.1002/ajpa.1098.

[11] P. Skurowski, M. Pawlyta, "Detection and classification of artifact distortions in optical motion capture sequences," *Sensors*, vol. 22, no. 11: 4076, 2022. https://doi.org/10.3390/s22114076.

[12] X. Chen, D. James, "Camera placement considering occlusion for robust motion capture," *Computer Graphics Laboratory*, Stanford University, 2000.

[13] J. Zhang, C. Wu, Y. Wang, P. Wang, "Detection of abnormal behavior in narrow scene with perspective distortion," *Machine Vision and Applications*, vol. 30, pp. 987–998, 2019. https://doi.org/10.1007/s00138-018-0970-7.

[14] X. Jiang, Z. Hu, S. Wang, Y. Zhang, "A survey on artificial intelligence in posture recognition," *Computer Modeling in Engineering & Sciences*, vol. 137, no. 1, pp. 35-82, 2023. https://doi.org/10.32604/cmes.2023.027676.

[15] B. M. Gamra, M. A. Akhloufi, "A review of deep learning techniques for 2D and 3D human pose estimation," *Image and Vision Computing*, vol. 114, 104282, 2021. https://doi.org/10.1016/j.imavis.2021.104282.

[16] P. Nguyen Huu, N. Nguyen Thi and T. P. Ngoc, "Proposing posture recognition system combining MobilenetV2 and LSTM for medical

surveillance," *IEEE Access*, vol. 10, pp. 1839-1849, 2022. https://doi.org/10.1109/ACCESS.2021.3138778.
[17] A. Bharathi, R. Sanku, M. Sridevi, S. Manusubramanian, S. Kumar Chandar, "Real-time human action prediction using pose estimation with attention-based LSTM network," *Signal, Image and Video Processing*, vol. 18, pp. 3255-3264, 2024. https://doi.org/10.1007/s11760-023-02987-0.
[18] A. Çalışkan, "Detecting human activity types from 3D posture data using deep learning models," *Biomedical signal processing and control*, vol. 81, 104479, 2023. https://doi.org/10.1016/j.bspc.2022.104479.
[19] J. Zhao, I. Laurent, "ShapeDTW: Shape dynamic time warping," *Pattern Recognition*, vol. 74, pp. 171-184, 2018. https://doi.org/10.1016/j.patcog.2017.09.020.
[20] P. Senin, *Dynamic Time Warping Algorithm Review*, Information and Computer Science Department, University of Hawaii at Manoa, Honolulu, USA, 2008, 23 p.
[21] I. Grishchenko, V. Bazarevsky, A. Zanfir, E. G. Bazavan, M. Zanfir, R. Yee, K. Raveendran, M. Zhdanovich, M. Grundmann, C. Sminchisescu, "BlazePose GHUM Holistic: Real-time 3D Human Landmarks and Pose Estimation," *Computer Vision and Pattern Recognition*, arXiv:2206.11678, New Orleans, LA, 2022. https://doi.org/10.48550/arXiv.2206.11678.
[22] A. Grover, D. Arora, A. Grover, "Human Posture identification and recognition using deep learning techniques," In: R.P. Mahapatra, S.K. Peddoju, S. Roy, P. Parwekar (Eds.), *Proceedings of International Conference on Recent Trends in Computing*. ICRTC 2023. Lecture Notes in Networks and Systems, Springer, Singapore, vol. 954. 2023. https://doi.org/10.1007/978-981-97-1724-8_7.
[23] P. Elias, J. Sedmidubsky, P. Zezula, "Understanding the limits of 2D skeletons for action recognition," *Multimedia Systems*, vol. 27, issue 3, pp. 547-561, 2021. https://doi.org/10.1007/s00530-021-00754-0.
[24] G. Panconi, S. Grasso, S. Guarducci, L. Mucchi, D. Minciacchi, R. Bravi, "Deep-learning-based markerless pose estimation systems in gait analysis: DeepLabCut custom training and the refinement function," *Computer Vision and Pattern Recognition*, arXiv:2407.10590, 2024. https://doi.org/10.1038/s41598-025-85591-1.
[25] L. Xu, L. Zhao, X. Sun, D. Wang, G. Li, K. Yan, "A comprehensive framework for occluded human pose estimation," *Processing of the 2024 IEEE International Conference on Acoustics, Speech and Signal (ICASSP)*, Seoul, Republic of Korea, pp. 3405-3409, 2024. https://doi.org/10.1109/ICASSP48485.2024.10448372.
[26] S. K. Roy, I. Badanin, S. Honari, P. Fua, "Occlusion resilient 3D human pose estimation," *Proceedings of the 2024 IEEE International Conference on 3D Vision (3DV)*, Davos, Switzerland, pp. 1198-1207, 2024. https://doi.org/10.1109/3DV62453.2024.00099.
[27] M. Kocabas, C.-H. P. Huang, O. Hilliges, M. J. Black, "PARE: Part Attention Regressor for 3D Human Body Estimation," *Computer Vision and Pattern Recognition*, arXiv:2104.08527, 2021. https://doi.org/10.1109/ICCV48922.2021.01094.
[28] J. Huaiming, W. Li, Z. Yuwei, L. Zhi, W. Chenglong, "A review of human pose estimation methods in markerless motion capture," *Computer-Aided Design and Applications*, vol. 21, no. 3, pp. 392-423, 2024. https://doi.org/10.14733/cadaps.2024.392-423.
[29] K. Song, T. J. Hullfish, R. Scattone Silva, K. G. Silbernagel, J. R. Baxter, "Markerless motion capture estimates of lower extremity kinematics and kinetics are comparable to marker-based across 8 movements," *Journal of Biomechanics*, vol. 157, 111751, 2023. https://doi.org/10.1016/j.jbiomech.2023.111751.
[30] B. K. Lahkar, A. Muller, R. Dumas, L. Reveret, T. Robert, "Accuracy of a markerless motion capture system in estimating upper extremity kinematics during boxing," *Frontiers in sports and active living*, vol. 4, 939980, 2022. https://doi.org/10.3389/fspor.2022.939980.
[31] L. Wade, L. Needham, P. McGuigan, J. Bilzon, "Applications and limitations of current markerless motion capture methods for clinical gait biomechanics," *PeerJ*, vol. 10, e12995, 2022. https://doi.org/10.7717/peerj.12995.
[32] V. Bazarevsky, I. Grishchenko, "On-device, real-time body pose tracking with MediaPipe BlazePose," *Google Research*, August 13, 2020, [Online]. Available at: https://research.google/blog/on-device-real-time-body-pose-tracking-with-mediapipe-blazepose/.
[33] H. Sakoe, S. Chiba, "Dynamic programming algorithm optimization for spoken word recognition," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 26, no. 1, pp. 43-49, 1978. https://doi.org/10.1109/TASSP.1978.1163055.
[34] X. Song, Q. Wen, Y. Li, L. Sun, "Robust time series dissimilarity measure for outlier detection and periodicity detection," *Proceedings of the 31st ACM International Conference on Information & Knowledge Management (CIKM '22)*, 2022, pp. 4510–4514. https://doi.org/10.1145/3511808.3557686.
[35] K. Rajput, D. B. Nguyen, G. Chen, "Evaluating DTW measures via a synthesis framework for time-series data," arXiv:2402.08943, 2024. https://doi.org/10.48550/arXiv.2402.08943.
[36] Penn-Action-Dataset [Online]. Available at: https://www.kaggle.com/datasets/kaushalbora18/penn-action-dataset/data.
[37] L. A. Espitia-Mora, M. A. Vélez-Guerrero, M. Callejas-Cuervo, "Development of a low-cost markerless optical motion capture system for gait analysis and anthropometric parameter quantification," *Sensors*, vol. 24, no. 11, 3371, 2024. https://doi.org/10.3390/s24113371.
[38] M. G. Bernal-Torres, H. I. Medellín-Castillo, J. C. Arellano-González, "Development of a new low-cost computer vision system for human gait analysis: A case study," *Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine*, vol. 237, no. 5, pp. 628-641, 2023. https://doi.org/10.1177/09544119231163634.

BOHDAN MYCHKOVSKYI, a student of Taras Shevchenko National University of Kyiv, Department of Software Systems and Technologies. Scientific interests: information technology, computer vision, software systems, mathematical model.

ANASTASIIA NIKOLAIENKO, graduated National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute". Now Now she works as an Assistant at Department of Software Systems and Technologies, Taras Shevchenko National University of Kyiv. Scientific interests: information technology, mathematical model, physical activity.

EVGEN IVANOV, graduated Taras Shevchenko National University of Kyiv. Now he works as an Assistant Professor at Department of Software Systems and Technologies, Taras Shevchenko National University of Kyiv. Scientific interests: information technology, software systems