inventions

MDPI

Article

Evolution of a Real-Time Markerless Kinematic Posture-Assessment System: From Multi-Participant Monitoring to Depth-Aware Angular Assessment

Jon Echeverria * and Olga C. Santos *

PhyUM Research Center, Department of Artificial Intelligence, School of Computer Science, Universidad Nacional de Educación a Distancia (UNED), 28040 Madrid, Spain
* Correspondence: jecheverria@dia.uned.es (J.E.); ocsantos@dia.uned.es (O.C.S.)

Abstract

Markerless human-motion assessment has become increasingly relevant in domains such as psychomotor learning, rehabilitation, sports training, and collaborative educational environments. While recent pose-estimation frameworks provide reliable skeletal reconstruction, translating these data into interpretable posture assessment remains challenging, particularly under varying participant-to-sensor distances and in multi-participant settings. This paper presents the technological evolution from KUMITRON, a system originally developed for synchronized multi-participant monitoring and interaction analysis in psychomotor activities, to COLLVIT, a prototype implementing configurable depth-aware joint-angle posture assessment. This evolution encompasses RGB-D integration, configurable angular templates, depth-informed adaptive tolerance mechanisms, temporal stabilization strategies, interpretable posture assessment, and multi-participant processing capabilities. The paper distinguishes between the architectures defined at the patent level, their implementation through the KUMITRON and COLLVIT prototypes, and capabilities that remain targets for future implementation and validation. Rather than reporting a controlled experimental evaluation, the contribution of this work lies in documenting the technological evolution from KUMITRON to COLLVIT, the rationale behind the patent-based design decisions, the implementation of the current prototype, and directions for its future development and validation. The resulting framework provides a flexible foundation for future research on markerless kinematic posture assessment in educational, rehabilitation, and other psychomotor application domains.

Keywords: markerless kinematic posture assessment; computer vision; pose estimation; RGB-D sensing; psychomotor learning; collaborative learning; rehabilitation; adaptive tolerance; patents; KUMITRON; COLLVIT

Check for updates

Academic Editor: Anastasios Doulamis

Received: 24 June 2026
Revised: 3 August 2026
Accepted: 18 August 2026
Published: 10 September 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Human movement analysis is relevant to numerous real-world activities, including sports, rehabilitation, physical education, and psychomotor learning. Motor activity involves different aspects that may need to be assessed, such as the quality of technical execution, postural alignment, temporal coordination, and the interaction among participants. These aspects are important for determining whether a movement has been successfully performed or whether further practice is needed. Traditionally, movement assessment has relied on expert observation by instructors, coaches, therapists, or educators.

Although expert observation remains common practice, new technologies are increasingly being incorporated to support this assessment process. Such technologies can help address limitations related to subjectivity, temporal resolution, repeatability, and the difficulty of simultaneously monitoring multiple participants during dynamic motor activities.

Emerging technologies such as computer vision, wearable sensors, inertial measurement units, physiological sensors, and artificial intelligence (AI) techniques have created new opportunities for automated human-movement analysis. These technologies enable the monitoring and tracking of one or more individuals through the analysis of their movements from various data sources, providing kinematic, physiological, and contextual information that can be processed to identify postures, detect movement patterns, estimate effort, and provide feedback during or after an activity. However, many existing approaches focus primarily on individual movement and single-person postural performance. This limits their applicability to interactive motor activities, where one participant's movement may influence another's response and where feedback must be synchronized across multiple users. In martial arts, rehabilitation exercises, technical training, choreography, or collaborative physical activities, systems may need not only to detect body positions but also to interpret temporal sequences, compare performance with reference patterns, manage multiple participants, and provide meaningful feedback in real time. Therefore, it seems evident that, in addition to estimating an individual's pose, it is necessary to analyze movement sequences involving more than one participant simultaneously.

The first architecture proposed in our previous work, described in a published patent [1], addressed the analysis of motor activity involving multiple participants by combining computer vision with physiological and inertial sensors. It supported synchronized multimodal data acquisition and the analysis of action–reaction relationships between participants. This patented architecture provided the technological foundation for KUMITRON [2,3], an AI-supported system developed for psychomotor learning in paired karate practice. In this context, KUMITRON was used to analyze the activity taking place on the tatami and to model interaction dynamics in karate kumite (i.e., paired practice or sparring involving an opponent).

KUMITRON was mainly developed between 2020 and 2022 as an early exploration of AI-supported modeling of multi-participant psychomotor activities [4]. Given the technological stage of the work at the time, these developments did not yet incorporate RGB-D (Red-Green-Blue plus Depth) sensing or the depth-informed assessment capabilities subsequently explored in the second architecture. In particular, approaches relying primarily on two-dimensional visual information and predefined angular or template-based criteria (e.g., expert posture patterns) are sensitive to perspective distortion, camera distance, partial occlusions, and temporary loss of key body points (e.g., body joints). These challenges become more demanding in multi-participant settings, where simultaneous image processing increases computational requirements and may hinder real-time feedback.

To address these limitations, our subsequent research explored technological advancements that could improve the analysis of participants performing motor activities. The second architecture, described in a patent application currently under examination [5], introduces a real-time markerless kinematic posture-assessment approach based on RGB-D acquisition, two-dimensional pose estimation, 2D/3D fusion, joint-angle calculation, dynamic adjustment of angular tolerance, kinematic posture validation, and individual memory of valid postures. In this architecture, each detected key body point is assigned a depth value to generate three-dimensional coordinates. Then, joint angles are calculated from these coordinates and compared with parameterized kinematic ranges. At the patent-design level, the effective angular tolerance can depend on the distance between the sensor and the joint being evaluated. As detailed later in the paper, the current COLLVIT prototype implements a simplified frame-level depth mechanism. This approach is intended to provide a spatially informed assessment that may reduce sensitivity to variations associated with perspective and participant-to-sensor distance, although its effect on accuracy requires future empirical validation.

This second solution builds on our previous work extending multi-participant psychomotor analysis toward neural-network-based posture classification and the analysis of group activity in paired karate practice [6]. Beyond the technical evolution, its development was motivated by pedagogical perspectives from computer-supported collaborative learning (CSCL), including sociocultural perspectives such as Vygotsky's zone of proximal development [7,8]. In this paper, however, these educational perspectives provide motivation for technological development and hypotheses for future research rather than experimentally validated learning outcomes. The second architecture provides the technological foundation for COLLVIT, which implements the depth-informed kinematic posture-assessment stage of the technological evolution described in this work.

Thus, this paper presents the evolution of a patent-based technological framework for real-time markerless kinematic posture assessment, comprising two successive architectures and their implementation through KUMITRON and COLLVIT. Hereafter, this evolving technological framework is referred to as the KUMITRON–COLLVIT framework. The contribution is twofold. First, the paper analyzes the technological progression from synchronized multi-participant motor-activity monitoring and interaction analysis toward configurable, depth-informed joint-angle posture assessment intended to account for projection-induced geometric uncertainty without modifying the underlying pose-estimation pipeline. Second, it examines how this evolution provides a foundation for future applications in sports training, rehabilitation-oriented monitoring, technical instruction, and collaborative psychomotor learning.

Accordingly, this work should be interpreted as a patent-based technical system-evolution study rather than as a controlled empirical validation study. Throughout the paper, we distinguish between patent-level architectural contributions, implemented prototype capabilities, evidence from previous developments, and aspects requiring future empirical validation.

1.1. Related Work

Accurate pose estimation and reliable assessment mechanisms are critical for real-time markerless kinematic posture-assessment systems to determine whether the observed movements meet task-specific execution criteria. However, reliable posture assessment remains challenging under variations in participant-to-sensor distance, perspective distortion, occlusion, and multi-participant settings, particularly when fixed angular thresholds are used.

Combat sports and martial arts provide a suitable domain for investigating these challenges, as they involve both structured movement execution and action–reaction dynamics. This section reviews the main technical areas relevant to the evolution of the KUMITRON–COLLVIT technological framework examined in this work, including markerless movement analysis, psychomotor feedback, multi-person assessment, RGB-D sensing, uncertainty-aware assessment, and interpretability in movement assessment.

1.1.1. Markerless Pose Estimation and Joint-Angle-Based Movement Assessment

Traditional human movement analysis has often relied on expensive and restrictive lab equipment. Recent advances in AI-based computer vision provide markerless alternatives that enable skeletal reconstruction and the calculation of kinematic variables, such as joint angles [9,10]. Such systems have been applied in areas including fitness, physical therapy, and real-time feedback [11–13].

These systems, however, remain sensitive to environmental conditions, occlusions, and the camera viewpoint. Inaccuracies in landmark localizations can propagate to joint-angle calculations and, consequently, to posture assessment and the feedback derived from it. These limitations highlight the need for mechanisms that can account for variability in real-time markerless movement assessment [14].

These challenges motivate the technological evolution examined in this work. Within the KUMITRON–COLLVIT framework, the initial developments addressed synchronized multi-participant movement monitoring and interaction analysis, whereas subsequent developments introduced depth-informed kinematic posture assessment and adaptive angular tolerance mechanisms to address variability in sensing conditions.

1.1.2. AI-Supported Feedback in Psychomotor Learning

For movement analysis to directly support psychomotor learning, the information derived from it needs to be transformed into meaningful feedback for learners, instructors, coaches, therapists, or other stakeholders. Intelligent systems that use AI in conjunction with extended reality (XR), pose estimation, user modeling, feedback design, and instructional planning can support the learning of complex psychomotor skills [15]. Feedback can be visual, auditory, haptic, verbal, kinesthetic, or multimodal, and can be delivered at different times relative to the execution of the movement (i.e., while performing, immediately following, or after the session) [16].

AI-supported feedback has also been investigated in rehabilitation and adapted sports. These systems can support individualized monitoring, longitudinal assessment, adaptive task difficulty, progress tracking, and multiple forms of feedback (e.g., via video display, audio output, tactile stimulation, etc.) [17,18].

However, regardless of how the feedback is created or presented, its value depends on the quality of the movement data upon which it is based. Instability in pose estimation resulting from occlusions, camera viewpoint changes, variations in lighting conditions, missing landmarks, or uncertainty in depth estimation can propagate to movement assessment and, consequently, to the feedback generated from it. This problem is further complicated in multi-participant settings, where feedback must be correctly associated with each participant and synchronized with the participant's specific movement sequences.

Within the KUMITRON–COLLVIT framework, the evolution of the assessment capabilities provides a basis for AI-supported psychomotor feedback. KUMITRON primarily addressed synchronized analysis of paired activities, whereas COLLVIT provides interpretable depth-informed joint-angle assessment that can serve as a basis for designing future feedback strategies.

1.1.3. Martial Arts, Combat Sports, and AI-Based Movement Assessment

By combining technical execution, biomechanics, anticipation, timing, distance management, and interaction with opponents, karate and other combat sports provide a rich environment for AI-supported psychomotor learning systems. While isolated techniques primarily depend on the accuracy of individual execution, performance in martial arts and combat sports additionally relies on defensive responses, tactical adaptation, and action–reaction dynamics.

Recent studies have applied computer vision, pose estimation, wearable sensors, and deep learning to karate and combat sports analysis. Many of these systems focus on kata (a predefined sequence of defense and attack techniques executed against one or more imaginary opponents), isolated techniques, movement classification, performance scoring, or the distinction between correct and incorrect executions [19–22]. Prior research on KUMITRON explored computer-vision-based approaches both (i) to support the training of action–reaction dynamics [23] and (ii) to use pose estimation for modeling psychomotor performance during paired karate practice, extending the focus from individual kata-based assessment toward the analysis of interaction between two participants [3]. Further research on real kumite has also shown the relevance of detecting the onset of body-part movements before an attack, as these early movement cues can support attack anticipation and reinforce the need for temporal analysis of interactive motor actions [24].

Similar challenges are present in other combat sports. In jiu-jitsu, close body contact creates occlusion and complex spatial configurations, which require pose-tracking strategies to classify combat positions and support automatic scoring [25]. In wushu, pose estimation combined with dynamic time warping has been used to compare participants' movements with professional references and identify body segments responsible for incorrect execution [26]. Wearable sensors have also been used in combat sports to quantify movement-related variables such as acceleration, velocity, orientation, impact parameters, and joint-related features [27].

In this context, the KUMITRON–COLLVIT framework extends the focus beyond the classification of isolated techniques, evolving from interaction-aware multi-participant movement analysis toward configurable, depth-informed kinematic posture assessment in interactive martial arts settings.

1.1.4. Multi-Person Pose Estimation and Real-Time Markerless Motion Capture

Multi-participant psychomotor activities, including some combat-sport and rehabilitation settings, require the estimation of keypoints for multiple participants while preserving their identities despite occlusions. Pose estimation is usually classified by dimensionality (2D or 3D) and number of participants [28]. Current approaches for working with multiple people (i.e., top-down or bottom-up approaches) can struggle when bodies overlap or interact at close range. Markerless real-time capture introduces additional challenges, including identity association, consistency over time, and latency outside controlled environments [29]. Multi-view systems can improve 3D reconstruction, but they also require greater computational resources and more complex calibration [30]. Hence, in multi-participant learning settings, technical failures such as missing keypoints or identity switches can propagate to movement assessment and compromise the reliability of the feedback derived from it.

Within this context, the KUMITRON–COLLVIT framework encompasses complementary aspects of multi-person movement assessment. KUMITRON addressed synchronized motor activity and reciprocal actions, whereas COLLVIT introduces depth-informed joint-angle assessment, 2D/3D fusion, and adaptive angular tolerance. Rather than aiming primarily at full 3D reconstruction, the technological evolution focuses on deriving interpretable kinematic assessment information from markerless pose data that can support future feedback strategies in multi-participant settings.

1.1.5. Depth-Aware Tolerance Adaptation and Uncertainty in 3D Kinematic Posture Assessment

RGB-D sensing and 3D pose estimation enable anatomically defined landmarks to be represented in three-dimensional space, allowing kinematic variables, such as joint positions and joint angles, to be estimated rather than only from two-dimensional image projections. Accordingly, systems incorporating depth sensors and/or 3D pose estimators can support spatially informed movement analysis across applications such as gait analysis and human–machine interaction [31–33].

Although depth sensing may provide additional spatial information, errors in landmark identification can still affect joint-angle computation. Because many systems use fixed angular tolerances, posture-assessment decisions may be affected by geometric uncertainty: valid executions may be rejected when capture conditions increase angular deviation, while incorrect executions may be accepted when tolerance ranges are too broad.

Research has shown that RGB-based systems may show lower reliability for parameters that depend on movement along the depth dimension [34]. More broadly, 3D pose sequences can represent both the spatial structure and temporal dynamics of movement [35]. Thus, providing kinematically interpretable feedback under variable sensing conditions requires explicit mechanisms for handling uncertainty.

The concept of depth-aware tolerance adaptation involves using depth not only for three-dimensional reconstruction, but also as contextual information for accounting for measurement uncertainty. This approach moves away from purely static threshold values toward tolerance logic that can account for variable sensing conditions. At the design level, uncertainty-aware approaches may consider factors such as sensor-to-joint distance, keypoint confidence, and other uncertainty indicators. However, in the current COLLVIT implementation, confidence values are used for keypoint filtering rather than as direct weights in the angular-tolerance function.

Accordingly, the depth-aware tolerance approach described in this work should be viewed as a design and implementation strategy for addressing depth-related uncertainty, not as an empirical demonstration of improved performance. This distinction is important because this paper describes design and implementation evolution rather than a validated uncertainty model. The evolution represented by the KUMITRON–COLLVIT framework should therefore be understood as a technological progression from synchronized multi-participant movement monitoring toward depth-informed angular-template assessment under variable sensing conditions.

1.1.6. Interpretability in Movement Assessment

In addition to technical performance, AI systems supporting movement assessment should provide interpretable assessment information. Deep-learning-based black-box approaches may be insufficient for educational and rehabilitation uses when stakeholders need to understand which body part, joint, or movement component contributed to an incorrect execution [36]. End-to-end classifiers may provide global labels, but these labels are often insufficiently informative for instructional feedback.

Assessment through joint angles provides an example of how an assessment strategy can produce transparent results. Computed joint angles can be compared with predefined acceptable ranges, allowing the assessment to identify which joint or joints fall outside the configured range. The resulting assessment information can therefore be directly related to the kinematic variables being evaluated, providing a basis for feedback that identifies which aspects of the execution deviate from the configured reference. This rule-based assessment layer is consistent with principles of interpretability, as the relationship between the computed kinematic variables, configured thresholds, and resulting assessment can be explicitly traced [37].

Interpretability in multi-participant settings requires the system to distinguish between participants and attribute posture errors to the appropriate individual. Within the KUMITRON–COLLVIT framework, KUMITRON focused on synchronized movement and interaction analysis, whereas COLLVIT introduces an explicit rule-based assessment layer based on joint-angle assessment and configurable posture templates, allowing assessment outcomes to be traced to configured kinematic criteria. Participant handling and temporal processing further support the association of assessment information with detected individuals. In addition, the current implementation uses frame-level depth information to adapt angular tolerance, providing a simplified depth-informed mechanism for accounting for variations along the depth axis. However, these technical mechanisms alone do not demonstrate the pedagogical effectiveness of the resulting assessments, which must be established through further empirical evaluation.

1.2. Related Patents and Technological Gap

In addition to the scientific and technical literature reviewed above, patent documents provide a complementary perspective for positioning the KUMITRON–COLLVIT framework with respect to existing technological solutions. This subsection therefore examines representative patent documents related to the main technological components of the framework. It first describes the patent search and screening strategy, then organizes the retained documents according to relevant technological areas and finally compares these solutions with the successive developments represented by KUMITRON and COLLVIT to identify the technological gap addressed by the patent-based evolution examined in this work.

1.2.1. Patent Search Strategy

A structured patent search was conducted to contextualize the technological evolution represented by the KUMITRON–COLLVIT framework with respect to representative patent documents related to markerless pose estimation, RGB-D/depth-based movement analysis, multi-person human-motion assessment, rehabilitation feedback, sport/fitness/martial arts training, and adaptive thresholds or uncertainty compensation.

The search was primarily conducted using Google Patents. Six thematic search blocks were defined: (A) pose estimation, joint-angle assessment, and feedback; (B) RGB-D/depth-camera-based joint-angle and posture assessment; (C) multi-person or multi-user pose estimation; (D) rehabilitation or physical therapy feedback; (E) sport, fitness, and martial arts movement assessment; and (F) adaptive tolerance, dynamic thresholds, or uncertainty compensation. Because the objective was to support a structured technological comparison rather than to conduct a freedom-to-operate analysis or an exhaustive patent-landscape review, only the first 50 records returned for each search block were screened.

The searches returned highly variable results. The resulting search and screening information for the six thematic blocks is summarized in Table 1. Documents could be retained in more than one thematic block; therefore, block-level counts are not additive. Broad searches, particularly those related to sport, fitness, martial arts, rehabilitation, and multi-user pose estimation, retrieved many documents concerning robotics, industrial automation, object pose estimation, AR/VR interfaces, generic user interfaces, or non-motor feedback. Therefore, patent documents were retained only when they explicitly involved human movement assessment, posture correction, joint-angle analysis, skeleton tracking, rehabilitation training, sport/fitness/martial arts movement assessment, multi-person exercise analysis, or feedback-oriented motor training. No geographical or jurisdictional restrictions were applied during the search or screening process.

WIPO Patentscope and Espacenet were used as selective verification sources for bibliographic or patent-family information for the most relevant retained documents. The resulting patent search should therefore be interpreted as a representative technological review intended to support technological positioning, not as a comprehensive patentability, validity, or freedom-to-operate analysis.

Table 1. Patent search strategy and screening summary.

Search Block	Search String Summary	Results Returned	Results Screened	Main Inclusion Focus	Documents Retained
A. Pose estimation + joint angle + feedback	Pose estimation/human pose estimation + joint angle/posture assessment/movement assessment + feedback/training/rehabilitation	1399	50	Human pose estimation, joint-angle/posture assessment, and feedback/training/rehabilitation	8
B. RGB-D/depth camera + joint angle	RGB-D/depth camera/Kinect/RealSense + pose estimation/skeleton tracking + joint angle/movement analysis/posture assessment	494	50	Depth/RGB-D sensing and human movement/posture assessment	2
C. Multi-person/multi-user pose estimation	Multi-person/multi-user/multiple participants + pose estimation/motion analysis + feedback/activity recognition/posture assessment	2507	50	Multi-person or multi-user human pose/motion/activity analysis	2
D. Rehabilitation/physical therapy + feedback	Rehabilitation/physical therapy/exercise therapy/motor rehabilitation + pose estimation/depth camera/joint angle/skeleton tracking + feedback/correction/assessment	5788	50	Rehabilitation, therapy, posture correction, joint-angle analysis, skeleton tracking, or movement feedback	6
E. Sport/fitness/martial arts	Sport training/fitness/martial arts/karate/kumite/combat sports/wushu + pose estimation/motion recognition/movement assessment/posture assessment	14,856	50	Sport, fitness, martial arts, posture correction, movement assessment, and motor fitness testing	8
F. Adaptive tolerance/dynamic threshold/uncertainty	Adaptive tolerance/dynamic threshold/adaptive threshold/uncertainty compensation/error compensation + joint angle/posture assessment/pose estimation/movement assessment	3084	50	Adaptive/dynamic thresholds, uncertainty compensation, error compensation, or adaptive control in movement/posture assessment	3

1.2.2. Prior Patent Work and Technological Grouping

The review of prior patent work identified several groups of technological developments related to movement analysis, posture assessment, rehabilitation feedback, multi-user interaction, and template-based movement evaluation.

A first group of patents focuses on movement analysis using sensor-based or camera-based acquisition. These inventions use wearable sensors, inertial measurement units, physiological measures, cameras, or combinations of these devices to acquire video streams, motion signals, or body-related data and provide users with performance-relevant information [38–40]. These systems are relevant because they demonstrate the use of automatic acquisition pipelines for movement monitoring and performance feedback. However, they are generally designed to assess individual movement execution rather than collaborative or paired motor activity.

A second group includes camera-based and sensor-based systems designed to support rehabilitation, physical therapy, and posture correction [41–43]. These inventions are closer to COLLVIT because they address automatic assessment, corrective feedback, and movement-quality analysis in therapeutic or corrective contexts. Nevertheless, they are also primarily oriented toward individual rehabilitation tasks, physical therapy guidance, or posture correction. They do not usually include explicit representations of collaborative learning, paired interaction, or action–reaction relationships between participants.

A third group includes inventions related to joint-point detection, three-dimensional posture estimation based on depth information, color-depth feature combination, skeleton reconstruction, and posture evaluation [44,45]. These inventions are particularly relevant to COLLVIT because they show the importance of incorporating spatial information and joint-level representations into movement assessment. However, their main focus is usually individual posture estimation, skeleton reconstruction, recovery tracking, or generic posture correction. Within this group, depth information is generally used to improve reconstruction or spatial representation, rather than being explicitly used to dynamically adapt angular-validation tolerances according to sensor-to-joint distance.

A fourth group includes inventions addressing multi-user, multi-entity, or dynamic movement analysis [46–48]. These systems acknowledge that movement analysis becomes more complex when several individuals or entities move simultaneously. This group is relevant to KUMITRON because KUMITRON specifically addresses multi-participant motor activity analysis. However, the synchronization of typified movement identification and the modelling of action–reaction relationships between participants in interactive practice, such as karate kumite, represent a distinctive aspect of KUMITRON when compared with more generic multi-user or multi-entity tracking systems.

A fifth group includes inventions based on movement templates, recurrent movement patterns, posture guidelines, or skeleton-based recognition [49–51]. These approaches compare movement signals, captured posture images, skeletal configurations, or joint-angle trajectories against predefined reference patterns or standard movements. They are technically relevant to the technological evolution examined in this work because reference movement representations are relevant to KUMITRON, whereas configurable angular templates constitute a specific component of COLLVIT's posture-assessment approach. However, template-based systems can lose stability when capture distance, viewpoint, occlusion, body morphology, or execution variability change. This limitation is directly related to the motivation for COLLVIT's depth-aware angular-validation approach.

Taken together, these prior patent groups cover important technological components, including motion capture, wearable sensing, camera-based movement analysis, rehabilitation feedback, posture correction, depth-based reconstruction, joint-angle calculation, movement-template comparison, and multi-user movement analysis. However, these components are often addressed separately or combined for technological purposes that differ from those examined in the present work, motivating the more detailed patent-based technological comparison presented below.

1.2.3. Patent-Based Technological Comparison

The patent documents retained through the structured search were further examined to compare their technical characteristics with those of the KUMITRON–COLLVIT framework. As summarized in Table 2, the comparison focused on the technological dimensions most relevant to the evolution examined in this work, including movement sensing and assessment, rehabilitation and posture correction, depth-based 3D analysis, multi-participant processing, template-based assessment, and adaptive-threshold or uncertainty-compensation mechanisms. The following groups synthesize the main similarities and differences identified across the retained documents. Because individual patent documents could satisfy the inclusion criteria of more than one search block, the numbers of documents retained across blocks are not additive. After removing overlaps and consolidating the screened records, 16 representative patent documents were retained for the comparative analysis presented in Table 2.

A first group includes systems for markerless or sensor-based movement analysis using cameras, wearable sensors, inertial measurement units, or multimodal sensing. These systems are relevant because they provide mechanisms for capturing movement data, estimating skeletal configurations, or extracting joint-related variables. However, they are usually focused on individual movement monitoring or performance assessment.

Table 2. Representative patent documents retained for technological comparison.

Patent/Publication Number	Title	Main Domain	Relevant Technical Features	Relevance to the KUMITRON–COLLVIT Framework	Main Distinction from the Present Work
CN121919786A [52]	A method, system, and storage medium for comparing and scoring standard movements	Movement scoring/rehabilitation	Compares measured joint-angle trajectory sequences with standard joint-angle trajectory templates using dynamic time warping	Highly relevant to angular-template comparison	Focuses on trajectory-template scoring rather than depth-aware adaptive angular tolerance in real-time markerless posture assessment
CN121686571A [53]	A method and system for deadlift posture detection based on mobile video	Fitness/deadlift posture detection	Uses trunk-angle and knee-joint-angle threshold ranges for posture detection	Relevant to fixed-threshold angular posture validation	Uses fixed threshold ranges; no explicit depth-aware tolerance adaptation
CN121281141A [54]	Physical training posture correction method based on machine vision	Physical training posture correction	Machine-vision-based posture correction with angular-deviation assessment	Highly relevant to posture correction and visual movement assessment	Uses multi-view RGB-D posture assessment and personalized adaptation of standard-parameter tolerances based on historical training data, but does not describe depth-dependent angular-tolerance adaptation based on sensor-to-joint distance.
CN121868837A [55]	An interactive martial arts immersive teaching system	Martial arts immersive teaching	Uses joint angles, angular velocities, and angular accelerations for martial arts teaching	Relevant to the KUMITRON application domain	Focuses on coupling respiratory-phase estimation with biomechanical simulation and immersive feedback for martial-arts teaching, rather than on multi-participant action–reaction analysis or depth-dependent angular-tolerance adaptation for markerless posture validation.
CN121858916A [56]	A method and quantitative evaluation system for verifying the structure of Tai Chi	Tai Chi/martial arts movement evaluation	Refers to depth camera for skeletal-joint capture	Relevant to martial arts/body-structure assessment	Focuses on martial arts quantitative evaluation; the screened document does not clearly describe the combination of multi-participant analysis and depth-aware adaptive angular assessment
CN121839020A [57]	Intelligent rehabilitation training system and method for lower limb dyskinesia	Lower limb rehabilitation	Human body modelling, skeleton tracking, motion recognition, and numerical extraction	Relevant to the rehabilitation application domain considered for COLLVIT	Rehabilitation-oriented; does not specifically describe depth-aware adaptive angular tolerance for posture assessment
CN121812169A [58]	A sports injury risk screening system based on motion capture and biomechanical analysis	Sports injury risk screening	Uses a depth camera to acquire 3D motion data during standard movements	Relevant to RGB-D/depth-based movement assessment	Risk screening orientation rather than real-time depth-aware angular posture assessment with adaptive tolerance
CN121904837A [59]	Rehabilitation training action evaluation method based on computer vision	Rehabilitation action evaluation	Computer-vision-based rehabilitation movement assessment with visual pose-estimation model	Relevant to rehabilitation and pose-estimation assessment	Does not clearly address adaptive angular tolerance based on depth uncertainty

https://doi.org/10.3390/inventions11050094

Table 2. *Cont.*

Patent/Publication Number	Title	Main Domain	Relevant Technical Features	Relevance to the KUMITRON–COLLVIT Framework	Main Distinction from the Present Work
CN121904845A [60]	Bodybuilding action recognition, counting, and quality assessment method based on machine vision	Exercise recognition/counting/quality assessment	Machine-vision-based exercise recognition, counting, quality assessment, and visual skeleton feedback	Relevant to feedback and joint-angle thresholding	Uses standard threshold-based quality assessment rather than depth-aware tolerance adaptation
CN121861719A [61]	Push-up training multi-person test result data identification method and system	Multi-person push-up testing	Simultaneous image acquisition of multiple users during push-up training cycles	Highly relevant to multi-person exercise assessment	Multi-person testing/scoring, but does not clearly combine multi-participant assessment with configurable angular templates and depth-aware tolerance adaptation
CN121982644A [62]	Human track prediction method and monitoring system for human–computer interaction	Multiview human tracking	Real-time 3D reconstruction of multiple human joint points, velocities, and angular speeds from multiview images	Relevant to multi-person 3D tracking	Tracking/prediction focus rather than template-based posture assessment
CN121862308A [63]	Personalized exercise training evaluation method based on multimodal data	Personalized exercise evaluation	Calculates joint-angle change sequences and aligns them with angular velocity/acceleration from inertial data	Relevant to multimodal movement evaluation	Focuses on multimodal temporal evaluation; depth-aware markerless adaptive angular validation is not clearly described.
CN121839147A [64]	Multimode hierarchical fusion network-based motion function multidimensional evaluation method	Multimodal rehabilitation evaluation	Uses inertial measurement units, surface electromyography, electroencephalography, vision, and multimodal fusion for rehabilitation evaluation	Relevant to multimodal rehabilitation assessment	Multimodal clinical evaluation rather than depth-aware markerless angular-template assessment with adaptive tolerance
CN121946517A [65]	Variable impedance control method for wrist rehabilitation training robot	Wrist rehabilitation robot	Uses dynamic safety threshold and self-adaptive rate threshold in rehabilitation robot control	Relevant to dynamic thresholds in rehabilitation	Adaptive thresholds are used for robot-control safety, not depth-aware markerless angular tolerance
CN121893289A [66]	Design method of human-fuzzy-decision-fused lower limb exoskeleton human–machine system	Lower limb exoskeleton	Uses uncertainty compensation in a human–machine/exoskeleton context	Relevant to uncertainty compensation	Robotic-control uncertainty rather than depth-related uncertainty in markerless human posture assessment
CN121960630A [67]	An embodied intelligent architecture and an environmentally adaptive learning system	Adaptive AI/robot learning	Uses dynamic threshold based on historical reconstruction errors	Relevant to generic dynamic-threshold mechanisms	Generic dynamic thresholding, not human joint-angle posture validation

A second group includes rehabilitation, physical therapy, and posture-correction systems. These inventions commonly provide automated guidance, corrective feedback, difficulty adjustment, or movement-quality assessment. Although they are relevant to COLLVIT because they address motor rehabilitation and movement evaluation, they are generally oriented toward individual therapy or device control rather than real-time multi-participant movement assessment.

A third group includes RGB-D, depth-camera, stereo-vision, and 3D posture-evaluation systems. These systems are especially relevant to COLLVIT because they demonstrate the value of spatial information, depth sensing, and three-dimensional skeletal reconstruction for human movement assessment. However, within the screened documents, depth information is generally used for reconstruction, tracking, risk screening, or movement evaluation rather than being explicitly used to dynamically adapt angular-validation tolerances according to sensor-to-joint distance.

A fourth group includes multi-person or multi-user movement-analysis systems. These technologies are relevant to KUMITRON because they address the additional complexity of detecting or evaluating more than one user. However, most of these systems focus on detection, tracking, testing, or monitoring and do not explicitly model paired action–reaction relationships.

A fifth group includes template-based posture or movement-assessment systems. These approaches compare captured joint angles, skeletal configurations, movement trajectories, or posture images against reference templates or standard movements. They are technically relevant to the KUMITRON–COLLVIT framework because they illustrate the use of reference movement representations and, more specifically, the template-based posture-assessment approach implemented in COLLVIT. However, template-based approaches are sensitive to distance, viewpoint, occlusion, morphology, and execution variability when they do not include mechanisms to adjust validation tolerance.

Finally, a smaller group includes adaptive-threshold, dynamic-threshold, or uncertainty-compensation mechanisms. These mechanisms are relevant to COLLVIT because they show that adaptive thresholds and uncertainty compensation are established design strategies. Nevertheless, in the screened documents, such mechanisms were primarily used for robot control, safety constraints, reconstruction-error filtering, exoskeleton impedance control, or generic adaptive systems rather than for depth-aware angular-tolerance adaptation in markerless human posture assessment.

Overall, the retained patent documents demonstrate that the individual technological components relevant to the KUMITRON–COLLVIT framework—including markerless movement analysis, depth sensing, joint-angle assessment, multi-person processing, template-based evaluation, and adaptive thresholding—are represented in existing technological solutions. However, the comparison also shows that these components are generally applied independently or combined for purposes that differ from the technological evolution examined in this work. These findings provide the basis for identifying the specific technological gap addressed by the KUMITRON–COLLVIT framework, as discussed in the following subsection.

1.2.4. Technological Gap with Respect to the KUMITRON–COLLVIT Framework

As summarized in Tables 1 and 2, the reviewed patent documents cover several components related to the present work, including motion capture, wearable sensing, rehabilitation feedback, posture evaluation, depth-based reconstruction, joint-angle calculation, movement-template comparison, multi-user movement analysis, and adaptive thresholding. However, these elements are generally addressed separately or combined for technological purposes that differ from those of the KUMITRON–COLLVIT framework. KU-

MITRON and COLLVIT should therefore not be interpreted as claiming novelty in generic pose estimation, skeleton tracking, RGB-D sensing, or joint-angle computation alone.

Instead, the technological contribution lies in the patent-based evolution from multi-participant markerless motor-activity analysis toward configurable angular-template validation, RGB-D/depth-aware kinematic posture assessment, and adaptive tolerance mechanisms intended to account for depth-related uncertainty. The first architecture contributes the multi-participant and interactive dimension through synchronous identification of typified movements and action–reaction modelling between participants, which were subsequently explored through KUMITRON in paired karate practice. The subsequent architecture, which provides the basis for COLLVIT, introduces RGB-D acquisition, 2D/3D fusion, joint-angle calculation, configurable angular templates, depth-aware tolerance adaptation, and participant-specific valid-posture memory.

Within the screened patent documents, no single retained document was found that clearly combines all these elements into an integrated architecture for multi-participant, real-time, markerless kinematic posture assessment. At the same time, this paper does not claim that COLLVIT has already demonstrated superior robustness, accuracy, or learning effectiveness. The current implementation should be interpreted as a technical prototype implementing the second architecture. Controlled benchmarking and empirical validation are required to determine whether depth-informed adaptive tolerance improves measurement stability, feedback quality, or educational outcomes in real learning scenarios.

1.3. Pedagogical Implications for Collaborative Psychomotor Learning

Technical limitations in movement assessment may affect the quality and interpretability of the feedback available to support psychomotor learning and peer interaction. From a sociocultural perspective, learning is mediated by tools, peers, and expert guidance, aligning with Vygotsky's zone of proximal development, where learners progress through appropriate external support [68,69]. From this perspective, collaborative motor learning can extend beyond individual repetition to incorporate observation, mutual regulation, and peer feedback in situated environments [70,71].

Within this theoretical perspective, the KUMITRON–COLLVIT framework provides a technical foundation for exploring feedback strategies in collaborative psychomotor learning. KUMITRON models action–reaction relationships in paired practice, such as karate, while COLLVIT introduces depth-informed, joint-angle-based posture assessment based on configurable angular templates. Importantly, the assessment information produced by these technological capabilities does not itself constitute collaborative feedback; rather, feedback remains a pedagogical design decision concerning how movement information is represented, delivered, and used to support interaction between learners.

At this stage, the pedagogical implications of these technical capabilities should be regarded as hypotheses for future research rather than as demonstrated learning effects. Further empirical studies are necessary to evaluate whether feedback strategies built upon these assessment capabilities effectively support learner comprehension, peer regulation, and instructional decision-making.

2. Materials and Methods

This section describes the design rationale underlying the two successive patent-based architectures and their implementation through the KUMITRON and COLLVIT prototypes. The first architecture addresses synchronized multi-participant motor-activity monitoring, multimodal sensing, movement identification, and action–reaction modelling, and provides the technological basis for KUMITRON. The second architecture extends this technological development toward depth-informed, joint-angle-based kinematic posture assessment,

https://doi.org/10.3390/inventions11050094

incorporating RGB-D sensing, 2D/3D fusion, configurable angular templates, adaptive angular tolerance, and posture-stability mechanisms, and provides the technological basis for COLLVIT. The following subsections distinguish between the architectural design defined at the patent level and the corresponding prototype implementations. Together, these successive architectures and their implementations constitute the technological evolution described in this paper as the KUMITRON–COLLVIT framework.

2.1. Architecture for Multi-Participant Motor Activity Analysis

The first architecture was designed to provide synchronized analysis of multiple participants' motor activity data [1]. It supported both independent and interactive movement, enabling synchronized analysis of both individual performance and interactions between participants. The architecture was therefore specifically suited to motor practices involving individual and paired activity, such as combat sports, in which participants may perform movements independently or in interaction with one another.

One of the most important design decisions underlying this architecture was the integration of heterogeneous data streams within a common processing flow. These data streams could include images acquired from one or more cameras together with inertial, physiological and contextual measurements. The resulting multimodal data were provided to a processor for joint analysis of movement initiation, identification of potential motor actions, and comparison with predefined reference models. At the architectural level, complementary sensor data could be used to constrain the set of candidate movements prior to and/or during the processing of visual information, thereby reducing the movement-recognition search space.

The architecture was designed to support interactive psychomotor practices by explicitly considering temporal relationships between participants. The patent defines an approach in which motor activity is analyzed as sequences of predefined movements performed by each participant, as well as action–reaction interactions between participants. Thus, it supports the identification of not only a movement performed by one participant but also the expected response of another participant to that movement, based on previously defined pairs or sequences of movements. This architecture provided the technological foundation for KUMITRON, which implemented these concepts primarily for the synchronized analysis of paired motor activity during kumite (sparring) between two karate participants. The resulting approach therefore extends the focus from the recognition of individual movements toward the analysis of synchronized and interactive motor activity.

2.1.1. System Components

The first patent-based architecture integrates sensing components, visual acquisition devices, communication units, a central processing unit, repositories of reference information, and feedback interfaces. The sensing components include inertial sensors, such as accelerometers and gyroscopes, and physiological sensors, such as heart rate, temperature, respiration, or other biometric sensors. These sensors are intended to capture physical and physiological variables associated with the execution of motor actions. In addition, contextual sensors may be incorporated into the environment, for example, to measure external variables such as room conditions, floor pressure, or spatial information.

The architecture incorporates one or more cameras. In one embodiment, the cameras may be installed on drones in order to follow each participant while maintaining a stable distance and viewing angle. In another embodiment, the system may use fixed cameras to monitor the practice area. The patent describes both possibilities, indicating that the camera stream is used to acquire image sequences of the participants and to detect changes in body joints between consecutive images. This visual information is combined with the measured sensor values to identify the start of a movement and to determine its compatibility with a "typified movement", a patent-specific concept referring to a predefined motor action represented through reference information.

The architecture also includes wireless units associated with the participants and, when necessary, with contextual sensors. These units periodically transmit the measured values to the processing unit. The processing unit, described as a computer in the patent, receives the video streams and sensor messages, synchronizes them over time, and performs the multimodal analysis. The system also includes repositories containing reference information, such as pattern images, expected physical-variable ranges, previous participant data, and typified movements. These repositories allow the system to compare the current execution with predefined or previously stored motor patterns.

Finally, the architecture includes feedback and visualization components. The processed information may be shown on a screen for the instructor, participants, or audience, and it may also be transmitted directly to participants through actuators. The patent explicitly contemplates visual, auditory, or tactile feedback, allowing the system to provide immediate recommendations or alerts derived from the movement analysis. These feedback and visualization components extend the architecture beyond movement monitoring by enabling the resulting information to be used for purposes such as learning support, performance analysis, injury prevention, and strategy development during motor practice.

2.1.2. Synchronized Movement Identification

The first patent-based architecture considers movement identification as a synchronized multimodal process. Video and sensor data are processed jointly within a common temporal structure rather than as independent information streams. At each time interval, the processing unit receives images from the camera system and measured values from inertial, physiological, and contextual sensors. Temporal synchronization allows changes visible in the participant's body to be correlated with the physical variables measured at the same instant, supporting the identification of typified movements compatible with the multimodal evidence.

Movement identification begins with the determination of the point in time at which the participant initiates the movement. Successive images of each participant are analyzed to identify body joints and detect changes between consecutive images, thereby indicating the onset of a potential motor action before the action is completed. The patent specifically identifies early movement-onset detection as relevant to interactive motor activities, since detecting movement initiation before it is readily perceived can support anticipation of another participant's subsequent action. For example, in karate kumite, recognizing the early phases of an opponent's attack could enable an instructor or the system to provide strategic advice regarding a likely counterattack before the ongoing interaction is completed.

After detecting movement onset, the collected data can be compared with a repository of typified movements. Typified movements are represented in the patent through pattern image sequences and expected ranges for physical variables such as inertial measurements, physiological variables, contextual information, and/or other sensor-derived values. Thus, the architecture does not rely solely on visual information for movement identification. Instead, synchronized sensor values can be used to constrain or confirm the set of candidate typified movements compatible with the observed data. By reducing the set of candidate movements compatible with the sensor data, the architecture is intended to support faster and more precise movement identification while reducing the computational load associated with image processing.

https://doi.org/10.3390/inventions11050094

The patent illustrates this multimodal characterization through the example of a side kick. The corresponding typified movement may be represented by a sequence of body postures in the image stream, while complementary sensor information may characterize features such as backward body inclination, a change in floor pressure from two feet to one, and a period of held breath. By combining these sources of evidence, candidate movements that are incompatible with the observed sensor data can be discarded, reducing the search space for movement identification. This illustrates the role of multimodal information in the architecture for interpreting motor activity in real time.

Movement identification is performed for all participants. Accordingly, the architecture supports the synchronized identification, at each time interval and for each participant, of a typified movement compatible with both the detected movement onset and the corresponding sensor measurements. This participant-specific yet temporally synchronized analysis provides the basis for subsequently examining relationships between the movements performed by different participants. The identified movements can subsequently be associated with each participant within the same temporal frame, enabling simultaneous monitoring while preserving the correspondence between participants and their respective sensor and image data. The resulting information can then be used for visualization, annotation, or feedback through the visual, auditory, or tactile interfaces contemplated by the architecture.

Overall, synchronized movement identification in the first patent-based architecture can be summarized as a three-stage process: i) detecting movement onset through changes between consecutive images; ii) constraining candidate typified movements using synchronized sensor data; and iii) identifying, for each participant and within a common temporal frame, the typified movement compatible with the available multimodal evidence.

2.1.3. Action–Reaction Modeling

The architectural design supports the analysis of a participant's movement in relation to the movements of other participants. In this respect, the architecture can consider the potential influence of a participant's movements on another participant. The action–reaction modeling defined in the patent represents these relationships through associations between typified movements performed by different participants. Observed movements can therefore be related to predefined combinations of movements that represent recurring sequences of interactions between participants.

In this architecture, each participant's movement is initially identified as part of a time-synchronized frame. Once the movements of the participants have been identified within this time frame, they can then be related to pre-existing associations between typified movements. These associations may reflect, for example, an offensive move followed by a defensive move, an offensive move followed by a counterattack, or other movement combinations represented during paired motor activity. Hence, the architecture supports the analysis of not just what each participant is doing at a particular point in time, but also how each participant's movements relate to the movements of others across time.

This aspect is especially pertinent to martial arts-based practices such as karate kumite. In this context, the relevant unit of analysis is not simply an individual stance or technique, but the temporal relationship between the offensive and defensive actions of the participants. For example, a lateral kick may be represented as a sequence of postures and associated sensor measurements while also being related to a predefined defensive reaction by the opponent. Thus, by recording and analyzing such action–reaction relationships, the architecture provides a basis for interpreting interactive motor behavior rather than treating each participant as a separate source of movement information.

Additionally, the patented architecture enables the incorporation of past participant data into the analytical process. Consequently, a participant's current movement can be compared with prior instances of that same movement performed by that same participant as well as with one or more reference patterns stored in the repository. This capability provides a basis for longitudinal comparison of motor performance. From a learning perspective, such information may be relevant because psychomotor learning and performance in interactive activities involve not only the execution of individual motor actions but also the ability to respond appropriately to the actions of other participants.

Furthermore, action–reaction modeling can also constrain the movement-identification search space. Once a movement has been identified for one participant, predefined action–reaction associations can be used to restrict the set of candidate movements considered for another participant. Movements that are incompatible with the expected interaction can therefore be excluded or deprioritized, reducing the candidate space and supporting real-time processing.

These action–reaction principles were also explored through KUMITRON in paired karate practice, including movement anticipation and interaction-aware feedback [23], as described in Section 2.2. Overall, action–reaction modeling introduces an interaction-aware representation of motor activity that extends beyond traditional independent single-participant movement identification.

While the first architecture primarily addressed synchronized multimodal movement identification and action–reaction relationships between participants, it did not address depth-informed joint-angle-based posture assessment or depth-dependent angular tolerance. These capabilities became the focus of the subsequent patent-based architecture described in Section 2.3, representing a further stage in the technological evolution of the KUMITRON–COLLVIT framework.

2.2. KUMITRON Implementation

KUMITRON [2] was developed as the first prototype implementing the multi-participant motor activity analysis architecture described in Section 2.1. Its initial application focused on karate kumite, integrating synchronized video acquisition, inertial and physiological sensing, movement analysis, and feedback-related information within a common training environment. The name KUMITRON combines "kumite" with "drone," reflecting the original concept of using aerial video acquisition to monitor paired karate practice. This initial concept was inspired by previous work exploring drone-based monitoring of human movement in martial arts practice [72]. During the subsequent development of KUMITRON, conventional cameras were found to provide a more practical solution for the intended indoor training scenarios, and the visual-acquisition setup evolved accordingly.

Figure 1 provides a functional synthesis of the processing workflow that guided the development and subsequent evolution of KUMITRON, integrating concepts originating in the first patent-based architecture with capabilities incorporated in later implementation stages.

KUMITRON implemented the first architecture through a real-time monitoring workflow for karate kumite that combined synchronized video data with wearable inertial and physiological signals, enabling instructors to monitor movement-related indicators during kumite practice sessions and use this information to support movement interpretation and instructor-mediated feedback regarding combat strategy and opponent movement anticipation [2]. Movement anticipation was one of the earliest specific applications explored within this technological line. A subsequent study investigated whether computer-vision processing could facilitate the anticipation of an opponent's attack by highlighting early movement cues [23].

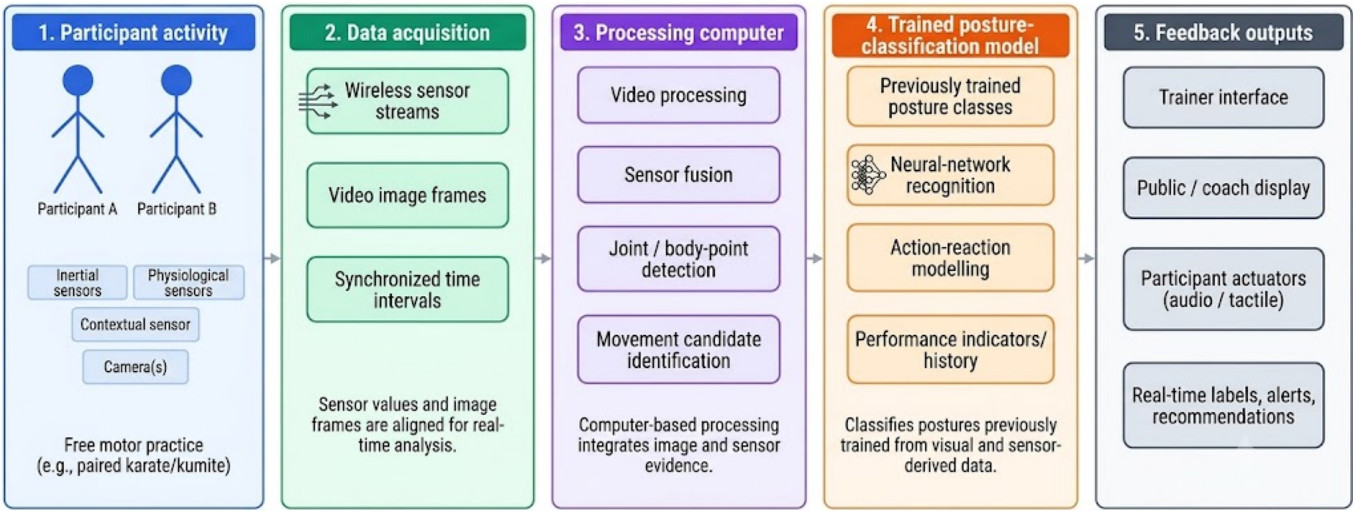

Figure 1. Functional synthesis of the processing workflow guiding the development and evolution of KUMITRON, combining technological concepts from patent ES 2 890 715 B2 [1] with capabilities incorporated in subsequent implementation stages: (1) capture of participant activity; (2) acquisition of video and sensor streams; (3) processing of synchronized evidence; (4) movement and interaction analysis; and (5) generation of feedback-related information.

Through this paired monitoring approach, KUMITRON explored the practical use of interaction-related information for movement anticipation and tactical interpretation during karate kumite [23]. Rather than treating each participant independently, the prototype provided synchronized information about both participants that could be used by the instructor to interpret their movements in relation to the ongoing interaction.

The KUMITRON interface integrated these heterogeneous sources of information to support live observation and analysis of karate combats. As illustrated in Figure 2, the interface brought together video streams, physiological and inertial information, movement-related indicators, and computer-vision outputs incorporated across successive development stages, providing the instructor with a synchronized view of the participants' activity during kumite practice.

KUMITRON's overall system design followed the four-stage SMDD model of intelligent psychomotor systems introduced in [73]: sensing movement, modeling movement, designing feedback, and delivering feedback. At the sensing level, the first KUMITRON prototype gathered data using an Arduino MKR WiFi 1010 board (Arduino S.r.l., Monza, Italy), an AZDelivery GY-521 MPU-6050 inertial sensor (AZ-Delivery Vertriebs GmbH, Deggendorf, Germany), a MakerHawk MAX30102 heart-rate sensor (MakerHawk, Dongguan, China) worn by each participant. In a subsequent hardware iteration, the Arduino-based setup was replaced by an M5Stick device (M5Stack Technology Co., Ltd., Shenzhen, China), providing a more compact platform for participant sensing and data transmission. The visual information was initially obtained using the camera of a DJI Tello drone (SZ DJI Technology Co., Ltd., Shenzhen, China). Subsequently, an Angetube 862Pro webcam (Angetube, Shenzhen, China) was also used. The acquired data streams were synchronized in real time to enable multimodal analysis at the modeling level.

The signals generated by the sensing devices were sent via wireless communication to a responsive application. Therein, those signals were displayed to instructors who monitored the participants during the kumite session, using the interface in Figure 2. The synchronized information provided the instructor (sensei) with complementary information for interpreting the ongoing interaction and determining when guidance or feedback was appropriate. KUMITRON comprised two software modules: a training module and a web module. The training module was responsible for receiving and displaying video and sensor data during a kumite session. The web module supported user account management, session storage, and visualization of recorded information.

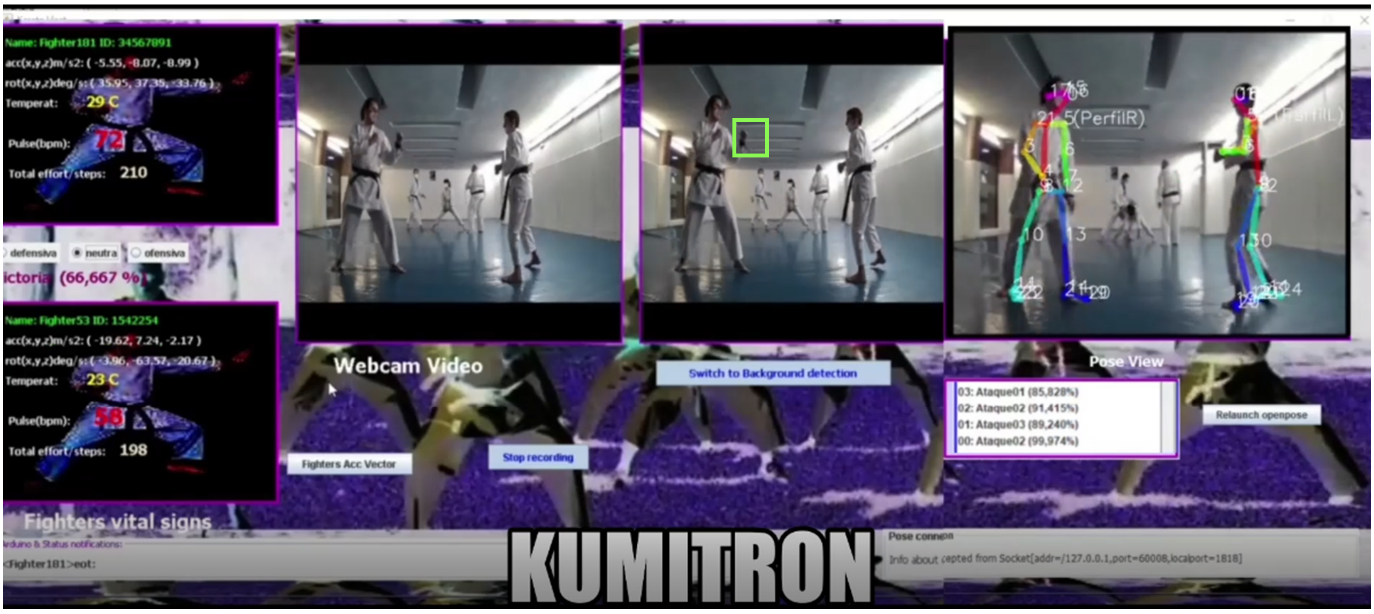

Figure 2. KUMITRON application interface during a karate kumite session, illustrating the integration of video streams (top center), physiological and inertial information (on the left), and movement sequences (bottom right). The three images at the top represent three stages of video processing: the original video stream (left); the output of the OpenCV Track Motion filter (center), in which rectangular boxes highlight image regions where movement is detected between frames, making early body-motion cues visible for the punch-anticipation approach evaluated in [23]; and the output of OpenPose-based pose estimation (right), which reflects a subsequent pose-estimation-based development [3] in which anatomical keypoints (represented as numbers) and skeletal connections (represented as colored lines) were used to derive joint-angle features for posture and movement modelling.

Initially, computer vision was implemented using JavaCV/OpenCV motion-detection algorithms (version OpenCV 4.4.0-1.5.4). The motion-detection algorithm highlighted image regions in which movement was detected between frames, making early body-motion cues visible before an attack was completed. This approach was experimentally explored for punch anticipation in a two-participant user study [23], as discussed in Section 3.

Inertial data were used to construct a real-time position and orientation vector for each participant. By displaying the vectors of both participants simultaneously, KUMITRON provided the instructor with information about their relative movement directions, such as whether a participant was advancing toward or retreating from the opponent. For instance, one participant may have been deemed defensive and moving backward, while another participant was offensive and moving forward, thus enabling instructors to recognize potential tactical advantages/strategies and offer strategy-related feedback. This information complemented the computer-vision-based anticipation cues and could inform tactical interpretation and instructor-mediated feedback [23].

During the initial stages of KUMITRON development, feedback was primarily provided through a human-in-the-loop approach. A Wizard of Oz technique was used: the sensei evaluated the information collected by the system regarding both participants and verbally provided feedback to the participants. Verbal comments, together with associated system parameters, were recorded so that they could subsequently be analyzed to elicit explicit feedback rules. This approach was intended to ground subsequent feedback automation in pedagogically meaningful rules derived from expert instruction.

KUMITRON was also conceptualized to support future automation of feedback. Under this design, once sensed information matched modelled movement information according to predetermined rules, corresponding recommendations could be generated automatically by the system without need for human intervention. Envisioned feedback modalities included auditory feedback (e.g., voice instructions concerning strategies for combat), visual feedback (e.g., video, filtered video utilizing computer vision techniques, sensor data, and directional-movement vectors), and multimodal feedback modes for real-time guidance or post-kumite review.

At a later stage of the research, KUMITRON incorporated pose-based movement modelling using OpenPose v1.7.0 (CMU Perceptual Computing Lab). Anatomical keypoints from video were grouped into 26 triplets from which joint-angle features were computed and used as inputs to movement-classification algorithms. The study focused on predefined Ippon Kihon kumite attack and defense postures and compared different data-mining algorithms for their classification. This represented an important transitional phase from general movement monitoring toward structured psychomotor performance modelling [3].

Across these successive developments, KUMITRON provided a technological testbed for synchronized multimodal monitoring of paired karate activity, movement-direction analysis, computer-vision-supported anticipation, instructor-mediated feedback, and, subsequently, pose-based modeling of predefined karate movements. However, the pose-based stage remained predominantly two-dimensional and relied on image-derived keypoints and joint-angle features. The work in [3] explicitly identified the need to progress toward complete movement sequences and noted that applying 3D pose estimation would require depth-capable cameras. Subsequent research extended this line toward the analysis of psychomotor group activity using neural-network-based movement modelling, further emphasizing the need to represent coordinated activity involving more than one participant [6]. Together, these developments exposed open technical challenges related to spatial representation, sensing variability, and multi-participant posture assessment, contributing to the subsequent exploration of depth-informed kinematic posture assessment within the KUMITRON–COLLVIT technological framework.

2.3. Architecture for Depth-Aware Adaptive Markerless Kinematic Posture Assessment

The second architecture, described in the patent application currently under examination [5], was designed to address key limitations observed in template-based and predominantly two-dimensional posture-validation systems. While KUMITRON established the basis for synchronized multi-participant motor activity analysis, the predominantly two-dimensional pose-based developments associated with KUMITRON were subject to limitations associated with camera viewpoint, perspective distortion, participant-to-camera distance, occlusion, and keypoint-detection instability. These limitations become critical when kinematic assessment must be performed in real time because small landmark errors can propagate into joint-angle estimation and produce inconsistent validation decisions.

To overcome these constraints, the second architecture introduces a real-time markerless kinematic posture assessment system based on RGB-D acquisition, 2D pose estimation, 2D/3D fusion, joint-angle computation, adaptive angular tolerance, kinematic posture validation, and posture-stability mechanisms. The architecture is designed for single-person operation and can be extended to multi-person operation through participant management and stabilization mechanisms. Its objective is not only to reconstruct a skeleton, but to determine whether a detected posture conforms to configured angular-template criteria under variable sensing conditions.

At a functional level, the system includes four main units: (i) a capture unit, (ii) a 2D pose-estimation unit, (iii) a 2D/3D fusion unit, and (iv) a kinematic assessment unit. The capture unit acquires RGB images and depth data. The 2D pose-estimation unit detects body keypoints in RGB frames. The 2D/3D fusion unit assigns depth to each detected 2D keypoint and generates 3D landmark coordinates. The kinematic assessment unit computes joint angles, compares observations against parameterized angular ranges, applies adaptive tolerance criteria, assesses conformity with angular templates, and emits results for visualization and subsequent feedback generation.

A key design principle is the explicit separation between pose estimation and kinematic posture validation. The pose-estimation layer detects keypoints; the kinematic assessment layer interprets them through angles, validation rules, adaptive tolerances, and posture templates. This separation enables updates to templates, decision rules, tolerance parameters, and validation criteria without retraining the underlying keypoint-detection model. Consequently, new postures or movement patterns can be introduced by configuring angular-template criteria rather than rebuilding the neural network.

The main technical contribution of this architectural approach is depth-aware angular tolerance. Instead of relying only on fixed thresholds, the architecture allows validation strictness to be adjusted using depth information associated with the participant or the evaluated joints. At the architectural level, the effective angular tolerance may be defined globally or on a joint-specific basis through a compensation function that depends on the distance to the sensor. This provides a mechanism for incorporating depth-related geometric uncertainty into the posture-validation decision boundary.

The architecture also supports confidence-aware preprocessing and validation. When the pose-estimation module provides keypoint confidence values, low-confidence landmarks may be filtered, down-weighted during 2D/3D fusion, or incorporated into the validation process. These mechanisms are intended to mitigate the influence of unreliable detections caused, for example, by occlusion or noisy keypoint localization. However, in the current COLLVIT implementation (detailed in Section 2.4), confidence is used for keypoint filtering rather than for confidence-dependent tolerance adaptation.

Another relevant mechanism is posture-stability control across frames. The architecture associates each participant with an individual valid-posture memory that can preserve the most recent valid state when temporary keypoint loss or detection instability occurs. In multi-person scenarios, this participant-specific memory can be combined with participant-management mechanisms to support temporal continuity and reduce ambiguity between simultaneously detected individuals.

The second architecture also defines optional modules for temporal sequencing, multi-person management, metric computation, persistence, angular-template generation, and AI-based parameter adjustment. The sequencing module validates progression through predefined scripts (e.g., kata, rehabilitation routines, choreography, and guided technical drills). The template-generation module can automatically generate reusable templates from 3D skeletal data, including joint-angle ranges, acceptance tolerances, orientation metadata, and minimum validation criteria, enabling reuse without retraining the pose-estimation model. The architectural diagram is depicted in Figure 3.

This second architecture provides the technical basis for moving from interactive movement recognition toward depth-informed kinematic posture assessment. It is implemented through the COLLVIT prototype, described in Section 2.4, which operationalizes these architectural principles through RGB-D fusion, angular-template-based assessment, depth-aware tolerance adaptation, configurable posture templates, and participant-specific stabilization mechanisms. Within the KUMITRON–COLLVIT framework, this second architecture therefore represents a shift from the interaction-oriented multi-participant monitoring addressed by the first architecture toward configurable, depth-informed kinematic posture assessment. The resulting assessment information can subsequently inform the design of interpretable feedback strategies in individual and multi-participant psychomotor learning scenarios.

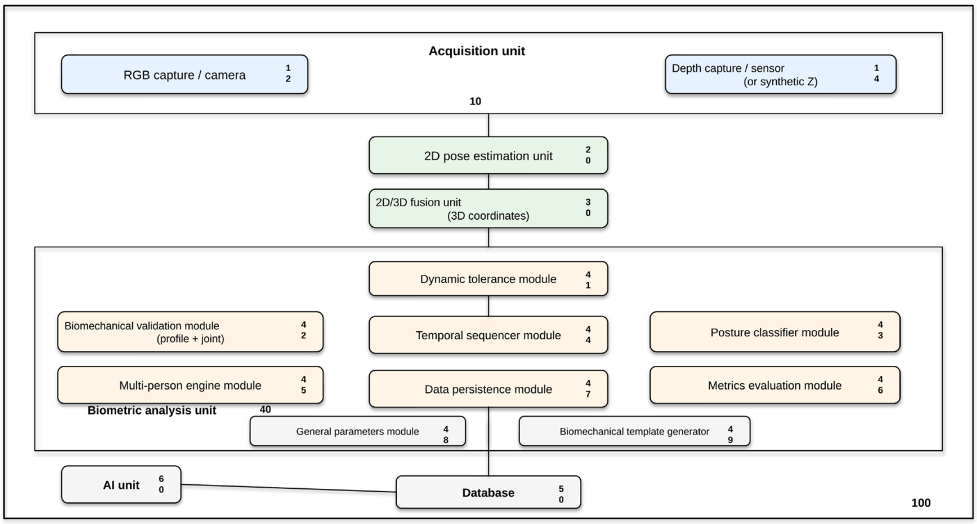

Figure 3. Architectural diagram corresponding to the patent application currently under examination [5].

2.3.1. RGB-D Acquisition and 2D/3D Fusion

The second architecture uses two complementary streams: RGB images and depth data. RGB provides the input for 2D pose detection, while depth provides the spatial component required for 3D kinematic assessment. The processing chain starts with 2D keypoint detection on RGB frames using a pose-estimation model. Detected landmarks can subsequently be processed according to confidence and validity criteria before 2D/3D fusion and kinematic posture assessment.

After 2D detection, fusion is performed by assigning a depth value to each valid keypoint. At the architectural level, depth may be obtained from an RGB-D camera or other depth-sensing technologies, such as time-of-flight, stereo, or LiDAR sensors, and may alternatively be inferred from RGB images using monocular depth estimation. The resulting depth information is spatially aligned with the RGB image before being associated with the corresponding 2D keypoints. This fusion process produces a set of 3D keypoints (x, y, z) used to compute 3D joint angles. This 3D landmark representation provides the spatial basis for the depth-aware angular-tolerance model described in the following subsection.

2.3.2. Theoretical Basis of Depth-Aware Angular Tolerance

The depth-aware angular tolerance model is based on the relationship between perspective projection, depth-dependent spatial uncertainty, and joint-angle validation. In a camera-based capture system, a 3D point $P = [X, Y, Z]^T$ is projected onto the image plane [74] according to the pinhole camera model:

$$x = f_x\left(\frac{X}{Z}\right) + c_x, y = f_y\left(\frac{Y}{Z}\right) + c_y$$

where x and y are the image coordinates, f_x and f_y are the focal lengths expressed in pixel units, c_x and c_y define the principal point, and Z represents the depth of the point with respect to the camera. This formulation shows that image measurements are inherently depth-dependent: the same physical displacement in 3D space produces different projected displacements depending on the distance between the body point and the camera.

Within the second architecture, RGB keypoints are extended into 3D by associating each valid 2D keypoint with a depth value z. A simplified back-projection model can be expressed as

$$X = \frac{(x - c_x)z}{f_x}, Y = \frac{(y - c_y)z}{f_y}, Z = z$$

Thus, each anatomical keypoint is represented as a 3D point:

$$P = [X, Y, Z]^{\mathrm{T}}$$

Once the anatomical keypoints have been represented in 3D space, joint angles can be computed from triplets of body landmarks [75]. For a joint j defined by three keypoints, p_a, p_b, and p_c, the two body-segment vectors are

$$\mathrm{v}_1 = \mathrm{p}_a - \mathrm{p}_b, \mathrm{v}_2 = \mathrm{p}_c - \mathrm{p}_b$$

The observed joint angle θ_j is then calculated as

$$\theta_j = \arccos\left(\frac{\mathrm{v}_1 \cdot \mathrm{v}_2}{\|\mathrm{v}_1\|\|\mathrm{v}_2\|}\right)$$

This angle-based formulation is useful because it preserves kinematic interpretability at the articulation level and is aligned with standard clinical kinematic descriptions [76]. However, the estimated 3D keypoints are affected by uncertainty. This uncertainty may be produced by RGB keypoint localization error, depth noise, RGB-D misalignment, occlusion, motion blur, partial body visibility, or reduced spatial resolution at larger capture distances [77]. Therefore, the issue should not be understood only as sensor error, but also as a geometric effect associated with perspective projection and depth-dependent spatial sampling.

This uncertainty propagates to angular validation. If a body segment has length L_j and the estimated keypoints have an effective spatial uncertainty $\Delta p(z)$, the induced angular uncertainty can, as a first-order approximation, be related to the spatial uncertainty as

$$\Delta\theta_j(z) \approx \frac{\Delta p(z)}{L_j}$$

If effective spatial uncertainty increases with depth within the operating range of the sensing configuration, the corresponding angular uncertainty may also increase as the participant moves farther from the sensor:

$$z_2 > z_1, \Delta p(z_2) > \Delta p(z_1) \Rightarrow \Delta\theta_j(z_2) > \Delta\theta_j(z_1)$$

This relationship motivates the use of depth-aware rather than exclusively fixed angular thresholds under variable capture conditions. A tolerance that is appropriate at close range may be too strict when the participant is farther from the sensor because the observed angular deviation may include a larger geometric uncertainty component. Conversely, using an excessively broad fixed tolerance for all distances may reduce validation precision at close range [78].

For this reason, kinematic posture validation can be expressed using an effective tolerance rather than a fully static threshold:

$$|\,\theta_j - \theta_{\mathrm{ref},j}\,| \leq \tau_j(z)$$

where θ_j is the observed joint angle, $\theta_{\text{ref},j}$ is the reference or expected angle for joint j, and $\tau_j(z)$ is the active angular tolerance used during validation.

A general joint-specific formulation of depth-aware tolerance can be written as

$$\tau_j(z) = \tau_{\text{base},j} + f_j(z)$$

where $\tau_{\text{base},j}$ is the base tolerance for joint j and $f_j(z)$ is a depth-dependent compensation function. This formulation represents a general form of the architectural model, in which each joint may have a specific compensation behavior according to its kinematic role, segment length, expected range of motion, and sensitivity to depth-related uncertainty.

In this theoretical formulation, $f_j(z)$ should be a non-negative and monotonically increasing function of depth:

$$f_j(z) \geq 0, \frac{\partial f_j(z)}{\partial z} \geq 0$$

This condition ensures that tolerance is not reduced below the base tolerance and that the compensation increases only when depth-related uncertainty is expected to increase. Different mathematical forms may be used depending on the sensor, task, and empirical calibration process. For example, a linear model can be expressed as

$$f_j(z) = k_j(z - z_0), z > z_0$$

where z_0 is a reference depth and k_j controls the compensation rate for joint j.

A quadratic model can be expressed as

$$f_j(z) = k_j(z - z_0)^2, \ z > z_0$$

This formulation may be useful when uncertainty grows nonlinearly with distance.

A bounded piecewise model can also be defined as

$$f_j(z) = \begin{cases} 0, & z \leq z_0 \\ k_j(z - z_0), & z_0 < z < z_{\max} \\ f_{\max,j}, & z \geq z_{\max} \end{cases}$$

This last formulation is especially relevant for real-time kinematic posture validation because it allows compensation to increase with depth while preventing unbounded tolerance expansion.

2.3.3. Valid-Posture Memory and Multi-Person Stabilization

The second architecture incorporates temporal continuity mechanisms intended to stabilize posture assessment when temporary degradations affect keypoint detection. Such degradations may arise from occlusion, motion blur, participant overlap, partial detection failure, or other transient sensing conditions. Without continuity mechanisms, temporary keypoint loss may lead to abrupt changes in posture-assessment outputs.

At the architectural level, valid-posture memory provides a temporal continuity mechanism by associating each participant with the most recently validated posture. When the current observation cannot be reliably validated because of temporary keypoint loss or detection instability, the previous valid state may be retained to preserve short-term continuity. This mechanism should be understood as a stabilization strategy rather than as a temporal prediction model or as evidence of improved classification accuracy.

For multi-person scenarios, valid-posture memory can be combined with participant-management mechanisms so that posture states and assessment criteria are maintained separately for simultaneously detected individuals. The architecture also allows participant-specific profiles to be associated with different angular templates or posture-validation rules, supporting scenarios in which participants perform different roles or are assessed according to different configured criteria.

The architecture further supports orientation and lateralization criteria as part of posture validation. These criteria can use anatomical landmarks and geometric consistency rules to distinguish postures that may exhibit similar joint-angle configurations but differ in the expected body side, profile, or orientation.

2.3.4. Configurable Posture Templates and Sequence Evaluation

Another relevant contribution of the second architecture is the reuse of posture templates without retraining the pose-estimation model. In many posture-recognition systems, adding a new posture requires collecting new data and retraining the neural model. In the second architecture, pose estimation and kinematic posture validation are separated, so new postures can be configured through template parameters while keeping the same keypoint detector.

In the architecture, the pose-estimation layer provides body keypoints, and the kinematic assessment layer validates them using configurable posture definitions (reference angle ranges, active-angle sets, tolerance percentages, profile constraints, and minimum valid-angle criteria). This allows new postures to be incorporated by updating configuration and template data rather than modifying the pose-estimation model weights.

Angular templates in the second architecture are configurable reference structures used to assess whether an observed posture conforms to predefined joint-angle criteria. They should not be interpreted as universal representations of biomechanical correctness. Instead, they encode task-specific, population-specific, and context-dependent posture criteria. Depending on the application, template criteria may be configured from domain-expert specifications, reference examples, or demonstration data.

Unlike a fixed posture classifier, this approach supports the configuration of new posture definitions at the template layer. When the relevant angular ranges and validation criteria are specified, the same pose-estimation pipeline can be retained, while kinematic assessment rules are modified through configuration.

The architecture also supports guided sequence evaluation. Multiple templates can be arranged as ordered expected states, and the sequence logic can validate progression and completion over time. This enables the representation of structured movement sequences, such as kata-like drills or other scripted psychomotor activities.

2.4. COLLVIT Implementation

COLLVIT implements the second architecture and focuses on real-time kinematic posture assessment using 2D and RGB-D processing configurations. The name COLLVIT combines the notions of collaboration and Vygotskian learning, reflecting the pedagogical motivation underlying its development. In particular, the system was conceived as a technological basis for exploring feedback strategies in collaborative psychomotor learning, where movement-assessment information can support interaction, peer regulation, and guided learning. However, COLLVIT's current implementation, which is described in this paper for the first time, focuses on the technical assessment layer; the design and evaluation of collaborative feedback strategies constitute a subsequent pedagogical layer and an area for further research.

The development of COLLVIT retains SMDD as the conceptual basis already adopted in the KUMITRON research line, while extending primarily its sensing and modeling stages. RGB-D acquisition and 2D/3D kinematic processing enrich the movement information captured and modeled by the system, while configurable posture assessment transforms this information into interpretable assessment results that can subsequently inform the later SMDD stages of designing and delivering feedback.

Whereas the KUMITRON implementation of the first architecture was developed primarily between 2020 and 2022 and has been documented across several previous publications [2,3], COLLVIT represents a subsequent implementation stage of the framework. Earlier work reported intermediate developments toward 3D and group-based psychomotor assessment [6], while the present paper provides the first integrated technical description of the current COLLVIT prototype. Accordingly, this section distinguishes explicitly between capabilities currently implemented in COLLVIT and broader capabilities defined at the architectural level.

2.4.1. Implementation Scope and Processing Pipeline

Consistent with the second architecture, COLLVIT maintains a functional separation between pose estimation and kinematic posture validation. The pose-estimation component uses the Ultralytics YOLO11n-pose model to detect body keypoints from RGB images, while the kinematic assessment component evaluates those identified keypoints by calculating joint angles, applying constraints based on the configured participant profile, applying tolerance parameters, and evaluating posture according to predefined rules or templates. Because these two components are functionally separate, validation criteria can be adapted without having to retrain the underlying pose-estimation model.

In the 3D processing pipeline, RGB and depth data are processed through successive stages before kinematic posture validation. COLLVIT detects 2D keypoints in RGB frames, applies filtering to invalidate low-confidence or invalid keypoints, and then associates each valid keypoint with a depth value sampled from a local region of interest (ROI). Once the 2D and depth information have been fused, a 3D coordinate (X, Y, Z) is obtained for each valid keypoint, which is then used to compute 3D joint angles and evaluate posture.

In the current implementation, depth is sampled from a 5×5-pixel region centred on each keypoint, and the median of finite depth values greater than zero is assigned to that keypoint. If no valid depth samples are available within the local window, no reliable real-depth value is assigned.

When a real depth stream is unavailable, the individual 3D classifier can generate a synthetic z-axis approximation to preserve software operation. This fallback does not provide metric depth and is therefore not equivalent to RGB-D-based 3D assessment.

2.4.2. Depth-Aware Tolerance and Posture Validation

In the current COLLVIT implementation, tolerance adaptation is applied in the individual 3D pipeline using a frame-level mean depth. Instead of computing a specific compensation function for each joint, the system applies a global scaling factor to the configured 3D base tolerance. This provides a computationally lightweight approximation of the depth-aware tolerance principle defined by the second architecture.

Specifically, the current implementation computes d using np.nanmean(depth) over the available depth frame when finite values are present. Thus, d is a frame-level depth mean rather than a participant-specific or joint-specific depth estimate. The current computation does not explicitly exclude depth values less than or equal to zero before calculating this mean.

The effective percentage tolerance is computed as

$$T_{e}ff(d) = T_{3}D,base\ d \cdot \alpha(d)$$

where d is the frame-level mean depth, $T_3D{,}base$ is the configured base percentage tolerance for individual 3D posture assessment, and $\alpha(d)$ is a bounded depth-dependent scaling factor.

In the current implementation, $\alpha(d)$ is defined as

$$\alpha(d) = 1 + min(0.5, max(0{,}0.3(d - 1.0)))$$

Therefore, the implemented effective tolerance can be written as

$$T_eff(d) = T_3D{,}base \cdot [1 + min(0.5{,}max(0{,}0.3(d - 1.0)))]$$

The resulting $T_eff(d)$ therefore represents a percentage rather than an angular margin in degrees; this effective percentage is subsequently converted into the angular margin applied to the posture-template range, subject to the minimum padding of 2°.

This formulation preserves the base 3D percentage tolerance for depths at or below 1.0 m. When the frame-level mean depth is greater than 1.0 m, the tolerance increases gradually with distance. The increase is capped at 0.5, meaning that the effective tolerance can reach at most 1.5 times the configured base tolerance. This prevents unbounded tolerance expansion.

The numerical factors used in this formulation are configurable engineering parameters rather than empirically calibrated biomechanical or sensor-error parameters. The reference depth of 1.0 m, linear scaling coefficient of 0.30, and maximum scaling factor of 1.50 were selected to provide bounded depth-dependent adaptation. These values have not yet been established through formal sensor-calibration or measurement-error studies and should therefore be configured according to the RGB-D sensor, acquisition environment, movement characteristics, and operational range.

The tolerance parameter is applied differently depending on the processing mode. In individual 2D and individual 3D modes, the configured tolerance percentage is applied as an additional margin over the angular range defined in the posture template, with a minimum padding of 2°. In individual 3D mode, the base tolerance is first adjusted using the depth-aware global scaling factor described above. In group 3D mode, the configured tolerance percentage is instead applied around the midpoint of the template angular range. Thus, the tolerance parameter is mode-dependent and should be interpreted according to the corresponding posture-assessment logic.

Joint-angle-based kinematic posture assessment is then performed by comparing the observed joint angles with the configured template ranges using the mode-specific angular margin derived from the active tolerance percentage. Joint-level checks are subsequently aggregated into posture-level decisions through the configured minimum-valid-angle criteria. This maintains an interpretable relationship between joint-level angular criteria and posture-level assessment decisions.

Accordingly, the current implementation represents a frame-level global approximation of the more general depth-dependent compensation model defined by the architecture; joint-specific depth compensation is not currently implemented. Moreover, this depth-dependent tolerance scaling is currently applied only in the individual 3D pipeline. In group 3D mode, depth is used to compute 3D joint angles, but posture validation uses the configured group tolerance rather than the individual 3D depth-scaling function.

2.4.3. Multi-Person Handling, Temporal Stabilization, and Template Configuration

In group mode, detected participants are ordered according to their horizontal position in the image and assigned from left to right to participant slots. Predefined participant profiles can be associated with these slots, allowing different angular templates or posture-validation rules to be applied to different participants.

This mechanism constitutes ordered slot assignment rather than persistent identity tracking or re-identification. Consequently, participant identities may be exchanged when individuals cross positions, overlap, become occluded, leave and re-enter the capture area, or otherwise alter their relative horizontal ordering. The current implementation should therefore not be interpreted as providing persistent identity tracking in unconstrained multi-person interactions.

Temporal stabilization mechanisms differ across operating modes. Frame-stride inference reuse is applied according to the mode-specific configuration, while group mode additionally stores the last valid posture associated with each ordered participant slot and may reuse it when the current posture is classified as unknown. In the current configuration, the individual 3D mode uses a frame stride of 1, whereas group 3D processing enforces a minimum stride of 2, resulting in an effective frame stride of 2. This valid-posture memory is intended to reduce abrupt output changes caused by transient keypoint loss or unstable detections, but retention of a previous state should not be interpreted as evidence that the posture remains correctly executed in the current frame.

Profile inference uses facial landmarks, including the ears and nose, while configurable knee-side checks provide additional orientation constraints, including a 3D-assisted fallback in the 3D classifiers. These mechanisms are used to distinguish configurations that may exhibit similar angular relationships but differ in the expected side, profile, or orientation.

In the current prototype, angular templates are implemented as configurable parameters and may specify reference angular ranges per joint, profile or orientation constraints, posture-specific knee-forward rules, active-angle subsets for the different operating modes, minimum-valid-angle criteria, and tolerance values. Template persistence is supported through posture-data and runtime-configuration files, including custom posture storage for 2D and 3D modes.

The templates used in the current prototype were not derived from a statistically validated template-generation procedure. COLLVIT therefore evaluates conformity with configured angular templates rather than universal biomechanical correctness.

COLLVIT also implements sequence-aware assessment for predefined movement sequences. In the current implementation, the expected sequence is represented as an ordered list of posture states. During execution, the posture produced by the kinematic assessment layer is compared with the posture expected at the current sequence step. When the assessed posture satisfies the expected state, the sequence controller marks the step as completed and advances to the next expected posture. If the current posture is invalid, unknown, missing, or does not match the expected state, the sequence does not advance, and the current step remains active until a valid matching posture is identified. The sequence can also be reset from the interface. This sequence-aware functionality is implemented in both individual and group variants.

The same criterion-based progression principle is used in individual and group sequence variants. In individual mode, one sequence detector is used and the sequence type is initialized from the detected profile. In group mode, separate sequence detectors are maintained for the ordered participant slots, with the first slot configured for the attack sequence and the second slot configured for the defense sequence. Figure 4 illustrates the current COLLVIT prototype interface in the individual karate assessment mode, integrating markerless pose tracking, joint-angle information, posture assessment, and system-level metrics.

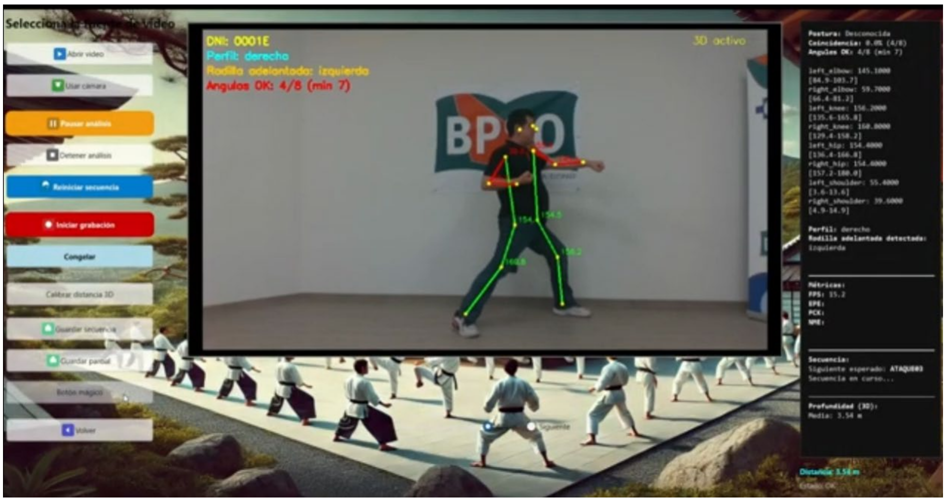

Figure 4. Prototype interface of COLLVIT for individual-based kinematic posture assessment in karate. The interface displays the menu options on the left, markerless skeleton tracking for one participant in the center, and posture assessment, angular information, and system-level metrics on the right. Green lines indicate body segments and joints that satisfy the configured posture criteria, whereas red lines indicate segments associated with angular deviations outside the accepted tolerance range. Yellow dots represent detected anatomical keypoints. The figure illustrates the implemented user-interface components and should not be interpreted as empirical validation of system performance.

2.4.4. Technical Configuration and Implementation Boundaries

In this paper, real-time refers to live-stream processing during system operation rather than to a formally benchmarked latency or frame-rate guarantee.

To improve reproducibility and clarify the current software implementation, Table 3 summarizes the main technical parameters of the COLLVIT prototype. Parameters that were not fixed or not systematically benchmarked in the current study are explicitly reported as configuration-dependent or pending future measurement rather than inferred.

Table 3. Technical configuration of the current COLLVIT prototype.

Implementation Aspect	**Current COLLVIT Prototype Status**
RGB-D input	Real RGB-D input is supported through an Intel RealSense D435 camera (Intel Corporation, Santa Clara, CA, USA) using pyrealsense2 (version 2.56.5.9235). If no real depth stream is available, the individual 3D classifier can generate a synthetic z-axis fallback only to preserve software continuity.
RGB and depth resolution	The RealSense source first attempts 424 × 240 at 15 FPS. If this profile fails, it attempts 640 × 480 at 30 FPS. The depth stream follows the same attempted profiles as the RGB stream.
Frame rate	Camera acquisition profiles are configured as 15 or 30 FPS, depending on the selected RealSense profile. The software property is currently set to 30.0 after initialization; therefore, measured end-to-end FPS should still be reported through future benchmarking.
Sensor-to-participant distance	Variable in the current prototype. No systematic operating-distance calibration is implemented or reported.
Depth-to-color alignment	Depth frames are aligned to the RGB/color stream using the Intel RealSense SDK 2.0 (version 2.56.5). Formal intrinsic/extrinsic calibration reporting and RGB-D registration-error evaluation remain future work.
Pose-estimation model	Ultralytics YOLO pose model (Ultralytics, version 8.3.153) using models/yolo11n-pose.pt.
YOLO inference size	416 px in 2D mode; 384 px in individual 3D mode and group 3D mode.
Inference device	Configured to use CUDA 12.3 on cuda:0 when available, with CPU fallback if CUDA inference fails or CUDA is unavailable.
Half precision	Enabled when CUDA is available and disabled automatically when falling back to CPU.

Table 3. *Cont.*

Implementation Aspect	Current COLLVIT Prototype Status
Confidence threshold	0.5. keypoints with confidence $\leq$ 0.5 are invalidated before angle computation.
Invalid 2D keypoint handling	Keypoints are invalidated when confidence $\leq$ 0.5, x $\leq$ 1, y $\leq$ 1 or coordinates are non-finite.
Depth window size	5 $\times$ 5 pixels around each keypoint, corresponding to a radius of 2 pixels.
Depth assignment strategy	For each valid 2D keypoint, depth is sampled from the local 5 $\times$ 5 region and assigned using the median of finite depth samples greater than 0.
Invalid or missing depth values	Non-finite values and values $\leq$ 0 are discarded for keypoint-level depth assignment. If no valid depth samples exist in the local window, no reliable real depth is assigned to that keypoint.
Mean depth used for tolerance scaling	In the individual 3D pipeline, tolerance scaling uses np.nanmean(depth) over the available depth frame when finite values exist. This is a frame-level depth mean, not participant-specific and not joint-specific. The current computation does not explicitly filter depth values $\leq$ 0 before computing the mean.
Depth-aware tolerance scaling	Implemented only in the individual 3D pipeline as $T_eff(d)$ = $T_3D,base$ $\times$ [1 + min(0.5, max(0, 0.3(d $-$ 1.0)))].
Base angular tolerance	2D mode: 15% additional margin over the template angular range (minimum padding: 2°); individual 3D mode: 10% additional margin over the template angular range (minimum padding: 2°), adjusted by depth-aware global scaling; group 3D mode: 15% around the midpoint of the template angular range.
Minimum valid angles	Seven valid angles are required in 2D, individual 3D, group 2D, and group 3D configurations.
Active angles	Left elbow, right elbow, left knee, right knee, left hip, right hip, left shoulder, right shoulder.
3D angle computation	3D angles are computed from keypoint triplets after extending 2D keypoints with z values from the depth frame. The calculation uses vector dot product, vector norms, clipping to [$-$1, 1], arccosine, and conversion to degrees.
Joint-specific depth compensation	Patent/design-level or future formulation. It is not implemented as independent per-joint tolerance compensation in the current prototype.
Confidence-weighted tolerance	Not implemented. Confidence is currently used for keypoint filtering, not as a direct tolerance-weighting factor.
Synthetic z-axis fallback	Used only as a fallback approximation when real depth is unavailable. It is generated from normalized keypoint vertical coordinates and is not equivalent to real RGB-D-based 3D validation.
Group 3D behavior	Group 3D computes 3D angles using depth per keypoint and uses the configured group 3D tolerance. It does not apply the individual 3D alpha(d) tolerance-scaling function.
Temporal stabilization	Temporal stabilization is intended to reduce abrupt output changes caused by transient detection instability. Frame-stride inference reuse is applied according to the mode-specific configuration, while group mode additionally stores the last valid posture per ordered participant slot and may reuse it when the current classification becomes "unknown."
Frame stride	2D mode: 2; individual 3D mode: 1; group 3D mode: 2 (effective minimum enforced by the group-processing logic).
Participant handling in group mode	Detections are ordered from left to right and can use predefined participant profiles. This is slot assignment, not persistent identity tracking or reidentification.
Sequence-aware assessment	Predefined ordered posture sequences with criterion-based progression; available in individual and group modes.
Software configuration	Python $\geq$ 3.11, <3.13; NumPy $\geq$ 2.3.0; OpenCV-Python $\geq$ 4.11.0.86; Ultralytics $\geq$ 8.3.153; PyQt6 $\geq$ 6.6; Torch $\geq$ 2.5.1; TorchVision $\geq$ 0.20.1; TorchAudio $\geq$ 2.5.1; Joblib $\geq$ 1.5.1; Pandas $\geq$ 2.2; Scikit-learn $\geq$ 1.7, <1.8; pyrealsense2 $\geq$ 2.56.5.9235; CUDA 12.3.
Hardware configuration	Intel RealSense D435 RGB-D camera (Intel Corporation, Santa Clara, CA, USA) and a laptop equipped with an NVIDIA GeForce RTX 3050 6 GB GPU (NVIDIA Corporation, Santa Clara, CA, USA).
Latency	Not systematically benchmarked in this paper.

The synthetic z-axis fallback should not be interpreted as equivalent to RGB-D-based three-dimensional validation. It does not provide metric depth information and therefore cannot support the same distance-aware tolerance interpretation as real depth data. Its role is limited to maintaining software compatibility, visualization, or fallback operation when RGB-D data are unavailable. Consequently, results obtained with the synthetic z-axis fallback should not be used as evidence of depth-aware validation accuracy.

3. Results

This section presents the main technical results associated with the successive stages of the KUMITRON–COLLVIT framework. Rather than presenting a new controlled experimental evaluation, it distinguishes among evidence obtained in previous evaluations of KUMITRON-related developments, capabilities implemented in the current COLLVIT prototype, and architectural capabilities that remain to be implemented or empirically validated. As such, this section explains what has already been implemented, what has been evaluated, and what still needs to be validated. Patent-level characteristics are reported as architectural contributions unless they have also been implemented or evaluated.

Table 4 is included to avoid conflating implemented software functionality with patent-level or future-development mechanisms. In particular, the current COLLVIT prototype implements frame-level depth-aware global scaling based on the mean depth of the available depth frame, whereas joint-specific depth compensation is defined at the architectural level but is not yet implemented in the current prototype, and confidence-weighted tolerance adjustment remains a possible future extension.

Table 4. Architectural, implemented, and future components of the KUMITRON–COLLVIT framework.

Component	Patent/Theoretical Architecture	Current Implementation	Future Work/Validation Need
Multi-participant motor-activity analysis	Included in the first patent-based architecture	Implemented in the KUMITRON prototype through paired karate activity analysis	More robust identity tracking, scalability analysis, and evaluation in larger groups
Action–reaction modelling	Included in the first patent-based architecture	Explored in KUMITRON through paired movement analysis, anticipation, and interaction-related information	Validation in open-ended interactive scenarios
RGB-D acquisition	Included in the second patent-based architecture	Implemented in the COLLVIT prototype in 3D processing pipelines	Extended evaluation across sensors, distances, and environments
2D/3D fusion	Included in the second patent-based architecture	Implemented in the COLLVIT prototype through depth assignment to detected 2D keypoints	Calibration and RGB-D registration refinement
Joint-angle-based kinematic posture assessment	Included in the second patent-based architecture	Implemented in the COLLVIT prototype through configurable angular templates	Validation against reference measurements or expert-labelled postures
Frame-level depth-aware global scaling	Simplified implementation of the architecture's depth-dependent tolerance mechanism	Implemented in the COLLVIT prototype using frame-level mean computed from the available depth frame	Sensitivity analysis, parameter calibration, and comparison with non-depth-aware tolerance
Joint-specific depth compensation	Included in the second patent-based architecture	Not implemented as independent per-joint compensation in the current prototype	Implementation and empirical validation of joint-specific tolerance adaptation
Confidence-dependent tolerance adjustment	Possible future extension	Confidence currently used for keypoint filtering before validation	Future confidence-weighted tolerance function
Valid-posture memory	Included in the second patent-based architecture	Implemented in the prototype as a temporal stabilization/fallback mechanism	Evaluation of false persistence, latency, and recovery after tracking loss
Sequence-aware assessment	Included in the second patent-based architecture	Implemented in the COLLVIT prototype through predefined ordered posture sequences and criterion-based progression	Empirical evaluation of sequence progression, error handling, and applicability to guided psychomotor tasks
Persistent multi-person identity tracking	Not a claimed capability of the current framework	Current implementation uses ordering/profile assignment, not full reidentification tracking	Future reidentification and tracking module

3.1. Capabilities Associated with the First Architecture and KUMITRON

The first stage of the technological evolution established a multimodal architecture for synchronized analysis of motor activity involving multiple participants. As described in Section 2.1, this architecture integrated visual, inertial, physiological, and contextual information to support movement identification and the analysis of temporal relationships between participants.

A key architectural contribution of this first stage was action–reaction modelling. This enabled the representation of paired or chained movements, such as attack–defense relationships, thereby extending the analytical unit beyond individual movement execution toward temporally related movements performed by different participants.

KUMITRON operationalized this architecture in a karate training scenario. The prototype synchronized video, inertial signals, physiological data, and instructor-mediated feedback during kumite practice. It used visual tracking, Arduino-based wearable sensors, OpenCV-based movement detection, and a visual interface through which the instructor could observe the activity and provide real-time guidance. The prototype, therefore, demonstrated the feasibility of monitoring paired karate practice through synchronized multimodal data.

The implementation also introduced a first feedback model based on a Wizard of Oz approach. In this model, the sensei observed the processed information and provided feedback to the participant through voice instructions. This allowed the system to record relationships between detected parameters, expert interpretation, and instructional feedback, creating a basis for later rule elicitation and future automation of feedback delivery.

This approach was experimentally explored for punch anticipation in a two-participant user study [23]. Participants identified the side of an incoming punch more accurately when viewing the OpenCV-filtered video than when viewing the unfiltered video (82.91% vs. 64.58% correct predictions), while mean response time was also lower (0.8594 s vs. 0.9253 s). These preliminary results supported the potential of computer-vision filtering for facilitating movement anticipation, although the small number of participants limits their generalizability.

A subsequent stage extended KUMITRON from movement monitoring toward pose-based kinematic modelling. In the study reported in [3], OpenPose was used to extract anatomical keypoints from karate movements, group them into triplets, calculate joint angles, and train classifiers for the identification of predefined kihon kumite movements. This provided evidence that pose-estimation-derived skeletal features and joint angles could support the modelling and classification of predefined karate postures.

Taken together, the first architecture and KUMITRON established a multimodal capture and synchronization infrastructure for paired motor activity, introduced real-time support for interaction-aware karate training, connected visual analysis with wearable sensing and expert feedback, and opened the path toward pose-based kinematic modelling through keypoint extraction and angular feature calculation. The subsequent pose-based developments remained predominantly two-dimensional, however, and [3] identified 3D pose estimation using depth-capable cameras as a relevant direction for extending the approach. These developments therefore provided a technological bridge toward the depth-informed kinematic posture-assessment capabilities addressed by the second architecture and subsequently implemented through COLLVIT.

3.2. Capabilities Associated with the Second Architecture and COLLVIT

The second stage of the technological evolution introduced a real-time markerless kinematic posture-assessment architecture aimed at addressing limitations of fixed-threshold and predominantly 2D posture assessment. In its current implementation, COLLVIT combines RGB-D acquisition, 2D pose estimation, 2D/3D keypoint fusion, 3D joint-angle computation, adaptive tolerance logic, configurable angular templates, and multi-person stabilization mechanisms.

The first capability is the transformation of 2D pose detection into depth-informed 3D kinematic posture assessment. The system captures RGB and depth streams, detects 2D keypoints, filters invalid or low-confidence landmarks, assigns depth values to keypoints through local depth sampling, and builds 3D keypoint representations used for joint-angle computation. This extends the earlier pose-based developments associated with KUMITRON toward spatially informed posture assessment.

The second capability is depth-aware tolerance adaptation. In the current individual 3D implementation, tolerance is adjusted using frame-level mean depth through a bounded global scaling factor. This mechanism is intended to account for distance-related variation in geometric uncertainty and keypoint stability. However, it should not be interpreted as empirical evidence of improved accuracy or robustness over fixed-threshold alternatives.

The third capability is confidence-aware input-quality control. Keypoint confidence is used mainly during preprocessing, where low-confidence or unreliable points can be filtered before angle computation. This may reduce the risk of unstable kinematic assessment decisions caused by occlusion, noise, or transient detector failures, but confidence is not yet used as a direct weighting factor in the tolerance function.

The fourth capability is temporal stabilization. Frame-stride inference reuse is applied according to the mode-specific configuration. In group mode, the system additionally maintains participant-slot-specific valid-posture history and may temporarily reuse the last valid posture when the current classification becomes unstable. These mechanisms are intended to reduce abrupt output fluctuations, but their effect on false persistence and detection delay still requires systematic evaluation.

The fifth capability is constrained multi-person processing. In the current COLLVIT implementation, simultaneous detections are handled through left-to-right participant ordering, optional profile assignment in guided activities, and per-slot result processing. This should be interpreted as constrained participant assignment rather than robust persistent identity tracking.

The sixth capability is configurable posture assessment without retraining the pose-estimation model. Posture assessment is driven by angular templates and configurable criteria, including angle ranges, active-angle sets, tolerance parameters, profile or orientation constraints, and minimum valid-angle rules. This enables extension of the posture catalogue without modifying the underlying neural pose-estimation model.

The seventh capability is sequence-aware assessment. The current implementation also includes temporal sequence logic in individual and group variants to verify progression through expected posture states in guided tasks, such as kata-like drills. This extends assessment from isolated postures toward ordered psychomotor execution, although its educational and performance effects remain to be validated.

Taken together, the current COLLVIT prototype provides capabilities for (i) depth-informed 3D kinematic posture assessment from RGB-D input, (ii) bounded adaptive tolerance behaviour in individual 3D mode, (iii) confidence-aware keypoint filtering, (iv) temporal stabilization mechanisms, (v) constrained multi-person processing, (vi) configurable angular-template-based assessment, and (vii) sequence-aware assessment. Relative to the earlier KUMITRON implementation stages, COLLVIT therefore represents a technical extension from synchronized multimodal interaction monitoring toward depth-informed, joint-angle-based kinematic posture assessment.

3.3. Comparative Evolution from KUMITRON to COLLVIT

The comparison between KUMITRON and COLLVIT can be summarized as a transition from synchronized multi-participant movement monitoring toward depth-informed, joint-angle-based kinematic posture assessment. KUMITRON focused on interactive motor-activity monitoring in karate kumite by synchronizing video, external sensor data, and instructor-mediated feedback. COLLVIT represents the subsequent implementation stage, adding RGB-D processing, 3D joint-angle computation, configurable angular templates, bounded depth-aware tolerance scaling in individual 3D mode, and temporal stabilization through mode-dependent frame-stride reuse, complemented in group mode by participant-slot-specific valid-posture history. The main differences between both stages are summarized in Table 5.

Table 5. Summary of the main technological differences between KUMITRON and COLLVIT.

Dimension	KUMITRON	COLLVIT (Current Implementation)
Main objective	Monitor paired motor activity and support strategy-oriented feedback	Assess posture with depth-aware kinematic criteria
Core scenario	Karate kumite and paired psychomotor interaction	Karate-focused psychomotor assessment, with potential applicability to other training and rehabilitation contexts
Main data sources	Video and external multimodal signals, including inertial, physiological, and contextual data	RGB images, depth frames, 2D keypoints, and 3D keypoint-derived angles
Main analytical unit	Typified movement and action–reaction relationship	Joint-angle conformity, posture constraints, and template-based rules
Validation logic	Pattern and signal interpretation with instructor support	Kinematic posture assessment with configurable tolerances and depth-adaptive global scaling in individual 3D mode
Multi-person handling	Synchronized paired monitoring	Multi-person detection, ordered participant handling, and group-mode valid-posture memory
Assessment information supporting feedback	Anticipation, strategy, movement direction, and sensor indicators	Posture-assessment information based on angular conformity, sequence progression, and stabilized outputs
Primary technical motivation	Need for synchronized analysis of paired interactive motor activity	Need to extend predominantly 2D, fixed-threshold posture assessment toward depth-informed kinematic assessment
Main technological contribution	Interaction-aware multimodal monitoring	Depth-informed kinematic posture assessment with configurable angular-template layer

The first major difference concerns analytical focus. KUMITRON primarily targeted interactive movement interpretation (e.g., anticipation and tactical cues in paired practice). COLLVIT emphasizes kinematic posture-assessment based on configurable joint-angle criteria and depth-informed processing.

A second difference concerns data integration. KUMITRON integrated heterogeneous multimodal inputs (video plus external inertial/physiological/contextual signals in its implementation of the first architecture). COLLVIT, in its current implementation, centers on RGB-D integration for posture assessment: depth is assigned to detected 2D keypoints, enabling 3D landmark construction and angle-based validation.

A third difference concerns validation logic. KUMITRON relied on typified movement interpretation and expert-supported analysis. COLLVIT implements explicit kinematic posture validation based on joint angles and posture constraints. In current software, tolerance adaptation is depth-aware but implemented as global scaling in individual 3D mode, rather than as independent joint-specific depth compensation.

A fourth difference concerns the mechanisms introduced to support depth-informed assessment and temporal continuity. COLLVIT incorporates:

1. Depth-informed 3D angle computation and tolerance adaptation in individual 3D mode.
2. Temporal stabilization through mode-dependent frame-stride reuse, complemented in group mode by participant-slot-specific valid-posture history.

A fifth difference concerns configurability. KUMITRON emphasized movement interpretation and expert feedback workflows. COLLVIT adds a configurable angular-template layer allowing new posture definitions through validation parameters (angles, profile constraints, active-angle sets, minimum criteria, and tolerances) without retraining the pose-estimation model.

Overall, KUMITRON and COLLVIT should be understood as successive implementation stages within the same technological framework rather than as competing or replacement systems. KUMITRON established an interaction-aware monitoring basis, whereas COLLVIT implements the subsequent architectural shift toward a depth-informed kinematic posture-assessment layer that can provide assessment information for angular-template-based feedback under variable sensing conditions. This layered evolution connects interaction monitoring, posture assessment, sequencing, and the generation of assessment information that can support feedback within a broader framework for collaborative psychomotor learning.

3.4. Evidence of Technological Transfer: From KUMITRON to ReBrain

The technological concepts underlying the first architecture and its implementation through KUMITRON were subsequently transferred beyond the original karate-training context through the ReBrain project. ReBrain was developed by Kapres Technology (https://kapres.es/inteligencia-artificial/ (accessed on 17 August 2026) with funding from the Spanish Centre for the Development of Industrial Technology (CDTI) and with direct involvement of the first author of this paper. ReBrain reused concepts from this technological line, including computer-vision-based movement tracking, motor activity analysis, and real-time feedback, adapting them to a rehabilitation-oriented therapy-support scenario for people recovering from stroke. ReBrain incorporates computer-vision-based analysis, AI-supported processing, real-time feedback, and data-driven monitoring tools to support rehabilitation activities and their follow-up. In this application context, the primary objective shifts from supporting paired psychomotor training in kumite to movement assessment and longitudinal monitoring in rehabilitation, providing information that can assist therapists during therapy sessions.

ReBrain provides an example of the transfer of this technological line from karate-based training to a real-world neuro-rehabilitation context. However, ReBrain should not be interpreted as clinical evidence of therapeutic effectiveness. Rather, it illustrates how concepts originating from the first architecture and its implementation in KUMITRON could be adapted from sport-training scenarios to rehabilitation-oriented movement monitoring, kinematic analysis, feedback generation, and data-driven follow-up.

4. Discussion

The evolution from KUMITRON to COLLVIT reflects a transition from synchronized multi-participant monitoring toward configurable, depth-informed kinematic posture assessment. The results presented above document the implemented capabilities of both technological stages, while also distinguishing them from capabilities specified at the architectural level and mechanisms that still require empirical validation. This section discusses the technological significance of this evolution, its potential implications for collaborative psychomotor learning, the main limitations of the current implementation, and the future research required to assess robustness, accuracy, assessment stability, and educational effectiveness.

From an SMDD perspective, the evolution from KUMITRON to COLLVIT can thus be understood as a technological refinement within a common conceptual basis. Across the KUMITRON–COLLVIT framework, successive developments progressively enrich the sensing and modeling stages, providing increasingly rich and interpretable movement information on which subsequent feedback design and delivery can be based.

An important distinction between the two stages of the KUMITRON–COLLVIT framework concerns their different levels of technical and empirical maturity. The first stage represents a more mature technological line: the first architecture has been implemented through KUMITRON, its underlying approaches have been progressively investigated in several scientific studies, and related technological concepts have subsequently been transferred to the ReBrain rehabilitation context. By contrast, the second stage is more recent. Although the second architecture has been formalized in a patent application and several of its core capabilities have been implemented through COLLVIT, these developments have not yet been supported by dedicated peer-reviewed empirical studies or technological-transfer experiences. Consequently, the evidence reported for the two stages should not be interpreted as equivalent in maturity: the results associated with COLLVIT primarily document the current implementation and establish the basis for the systematic empirical validation and transfer studies required in subsequent work.

4.1. Technological Contribution

The main technical contribution of this work is the definition and implementation of an evolving patent-based technological framework that connects interaction-aware multi-participant motor-activity monitoring with configurable, depth-informed kinematic posture assessment.

First, KUMITRON provides an interaction-aware implementation for monitoring motor activity involving multiple participants. By synchronizing video and external sensor data, KUMITRON shifts the analytical focus from isolated individual movement toward relational action–reaction dynamics between participants.

Second, COLLVIT implements the second architecture, introducing a depth-informed kinematic layer for posture assessment under variable capture distances. In the current implementation, 3D joint angles are calculated using depth-enriched keypoints. COLLVIT also includes adaptive tolerance mechanisms designed to account for distance-related variability during capture. In the current implementation, these mechanisms are approximated through frame-level depth-aware global scaling, while the joint-specific depth compensation specified at the architectural level remains to be implemented and empirically validated. COLLVIT also incorporates temporal continuity mechanisms and separates pose detection from kinematic assessment. This separation allows angular constraints and posture templates to be configured independently of the underlying pose-estimation model, providing a basis for their definition by instructors, therapists, or other domain experts.

Across its two successive stages, the KUMITRON–COLLVIT framework connects interaction-aware multi-participant monitoring with depth-informed 3D joint-angle assessment and configurable assessment logic. This evolution moves the technological approach beyond generic movement recognition toward interpretable, template-based kinematic posture assessment. However, the degree to which these technologies improve robustness, assessment stability, or measurement accuracy must be determined through future controlled research.

4.2. Potential Educational and Collaborative Implications

Beyond its technological contribution, the KUMITRON–COLLVIT framework may provide a basis for future research on collaborative psychomotor learning. Psychomotor skill acquisition often depends not only on individual motor execution, but also on social interaction, observation, correction, coordination, and contextualized practice.

KUMITRON contributes to this perspective by focusing on paired motor activity and action–reaction relationships, such as those occurring in karate kumite, rather than on isolated execution alone. From a sociocultural perspective on learning, the prototype can be understood as a human-in-the-loop tool that supports, but does not replace, the pedagogical role of instructors.

COLLVIT extends this direction by providing interpretable depth-informed kinematic assessment information that can serve as a basis for feedback design. Since the prototype separates pose estimation from assessment logic, posture templates and assessment criteria can be configured independently of the underlying pose-estimation model, providing a basis for their definition by instructors, therapists, or other domain experts. In addition, multi-person processing may support per-participant-slot assessment and stabilization in constrained guided scenarios involving more than one learner.

However, the educational implications of this framework should be treated as hypotheses for future investigation rather than as validated learning outcomes. Future studies should examine whether feedback strategies derived from interpretable kinematic assessment information, shared visualization, and social interaction mechanisms improve reflection, coordination, peer correction, and psychomotor learning processes.

Beyond these prospective educational uses, the KUMITRON–COLLVIT research line has also been progressively disseminated beyond specialist research settings. Following the corresponding patent filings, KUMITRON- and COLLVIT-related developments were presented between 2020 and 2026 through demonstrations, talks, videos, questionnaires, and activities for school groups and the general public at educational and science-communication events, including AULA, Madrid es Ciencia, and Madrid Science Week. These activities provided opportunities to communicate the role of AI, computer vision, and kinematic analysis in psychomotor training and movement assessment to non-specialist audiences. Although such activities do not constitute evidence of technological performance or learning effectiveness, they illustrate the broader dissemination and public-engagement dimension of this research line.

4.3. Future AI-Based Extensions

Although the current COLLVIT prototype already incorporates AI-based pose estimation, future versions could extend the role of AI beyond keypoint detection to complement the current rule-based and angular-template-based assessment mechanisms. Beyond pose estimation, kinematic posture assessment currently relies mainly on configurable templates, predefined tolerance values, confidence-based keypoint filtering, and depth-aware global scaling. However, future developments could use machine learning models to support individual calibration, adaptive tolerance learning, uncertainty modelling, and hybrid rule-based and data-driven assessment.

One possible extension is personalized calibration. Instead of relying only on predefined angular templates and manually configured tolerance values, future models could learn participant-specific tolerance ranges from repeated executions, expert-labelled examples, morphology-related variables (i.e., individual biomechanical characteristics), range-of-motion constraints, and sensor-specific uncertainty profiles. This could allow the system to adapt angular thresholds to individual users while preserving expert control over the interpretation of posture conformity.

A second extension is adaptive tolerance learning. Machine learning models could be trained to estimate the probability that a detected posture conforms to an expert-defined reference under varying depth, viewpoint, occlusion, confidence, and keypoint-stability conditions. Such models could help determine when tolerance should be expanded, reduced, or kept unchanged. However, these learned tolerance functions would require systematic calibration, interpretability, and validation against expert assessment and reference measurements before being used in instructional, sport-training, or rehabilitation contexts.

A third extension is the development of hybrid rule-based and data-driven assessment. In this approach, configurable angular templates would remain interpretable and available for definition or adjustment by instructors, therapists, or other domain experts, while learned models could estimate uncertainty, detect atypical movement patterns, or recommend user-specific tolerance adjustments. This hybrid design could preserve the transparency of rule-based assessment while improving adaptability to individual users and variable sensing conditions.

https://doi.org/10.3390/inventions11050094

Finally, AI-based extensions should be evaluated carefully. Future studies should assess model generalizability across body morphologies, skill levels, cameras, lighting conditions, movement speeds, and occlusion scenarios. They should also evaluate fairness, interpretability, expert agreement, false positive and false negative rates, and the extent to which AI-supported adjustments improve assessment quality and, when incorporated into pedagogical feedback, contribute to feedback quality or learning outcomes. Therefore, AI-based calibration and adaptive tolerance learning are presented here as future research directions, not as capabilities already implemented in the current COLLVIT prototype.

4.4. Limitations

Despite the technological contribution of the KUMITRON–COLLVIT evolution, several limitations should be acknowledged. First, the empirical evidence supporting the framework remains limited and uneven across its two technological stages. Previous studies provide exploratory evidence for developments associated with KUMITRON, whereas the more recent COLLVIT implementation has not yet undergone dedicated empirical evaluation. Broader validation is therefore needed across larger populations, different skill levels, body morphologies, movement repertoires, and application contexts.

Second, COLLVIT remains dependent on sensing conditions. Its depth-aware operation relies on RGB and depth quality, and its performance may be affected by occlusion, distance variation, lighting changes, incomplete depth observations, and keypoint instability. Current fusion and filtering strategies are intended to mitigate some of these sources of uncertainty, but their effect on robustness has not yet been systematically quantified.

A related limitation concerns the synthetic z-axis fallback used when valid depth information is unavailable. This mechanism can preserve processing continuity, but it does not provide measured depth information and should therefore not be interpreted as an alternative to RGB-D-based 3D posture assessment. Its use may also affect assessment stability when real depth information is intermittently unavailable. The frequency of fallback activation and its effects on posture-assessment outputs, false positive and false negative rates, temporal stability, and latency remain to be empirically characterized.

Third, tolerance adaptation is currently implemented in a simplified form. In individual 3D mode, the current prototype applies bounded global scaling based on the mean depth of the available depth frame rather than independent joint-specific depth compensation. This approach does not account for differences in sensor-to-joint distance or uncertainty across individual joints within the same posture. Moreover, the present study does not provide a controlled comparison against fixed-threshold 2D or 3D assessment. Consequently, depth-aware tolerance adaptation should be interpreted as an implemented strategy for addressing distance-related variation, rather than as empirical evidence of improved accuracy, robustness, or temporal stability. Joint-specific depth compensation, which is specified at the architectural level, remains to be implemented and compared with the current global approach.

Fourth, the numerical parameters of the current depth-aware global scaling function have not yet been independently calibrated. The reference distance, linear scaling factor, and the maximum scaling factor currently operate as configurable implementation parameters rather than empirically derived values validated against an external reference system. Their sensitivity to participant characteristics, sensing conditions, and camera configuration therefore remains to be established. Accordingly, these parameters should be calibrated using a sufficiently large and diverse sample of participants and a variety of camera configurations.

Fifth, confidence is currently integrated mainly as an input-quality control mechanism for filtering unreliable keypoints rather than as a confidence-dependent tolerance formulation. Consequently, confidence information does not yet contribute directly to tolerance adaptation, leaving room for more advanced uncertainty-aware assessment strategies.

Sixth, temporal stabilization differs across processing modes. Frame-stride reuse is mode-dependent, while group operation additionally uses participant-slot-specific valid-posture continuity mechanisms. This asymmetry may produce different stability profiles across scenarios. Moreover, valid-posture persistence may introduce false persistence or detection delay when a previous valid posture is retained despite an actual movement transition. The current implementation does not yet incorporate systematically evaluated safeguards such as maximum persistence duration, timeout conditions, or confidence-based reset criteria.

Seventh, multi-person handling remains limited in highly dynamic interactions. The current implementation relies on left-to-right participant ordering and optional profile assignment rather than persistent identity tracking or re-identification. Identity continuity may therefore degrade under severe overlap, crossing trajectories, or rapid out-of-frame transitions. Current multi-person handling should consequently be interpreted as constrained participant assignment rather than robust persistent identity tracking.

Eighth, the validity of joint-angle-based kinematic posture assessment depends on template design. Although templates can be configured without retraining the pose-estimation model, their validity depends on expert-defined angular ranges, orientation constraints, and acceptance criteria adapted to the target activity and population. The resulting assessment should therefore be interpreted as conformity with a configured reference template rather than as a universal criterion of movement correctness.

Ninth, technical posture assessment does not by itself establish pedagogical effectiveness. The transformation of assessment information into useful feedback depends on factors such as timing, granularity, modality, interpretability, and instructional context, and its effects on psychomotor learning remain to be evaluated. Similarly, the transfer of related technological concepts to rehabilitation contexts should not be interpreted as evidence of clinical effectiveness.

Finally, this paper is primarily an evolutionary and architecture-oriented contribution. Its central purpose is to describe and position the transition from interaction-aware monitoring to depth-informed kinematic posture assessment, rather than to demonstrate superior performance over alternative approaches. Claims concerning improved robustness, accuracy, or stability should therefore be regarded as design objectives until supported by controlled comparative evidence.

4.5. Future Work

Future work should combine systematic empirical validation of COLLVIT with further technical development and user-centred evaluation of the broader KUMITRON–COLLVIT framework. Given the different maturity of its two technological stages, the immediate priority is to validate the capabilities implemented in COLLVIT before drawing conclusions regarding improvements in robustness, accuracy, temporal stability, or educational effectiveness.

First, COLLVIT should be evaluated through a controlled technical benchmark comparing three assessment configurations: fixed-threshold 2D assessment, fixed-threshold 3D assessment, and depth-adaptive 3D assessment. The benchmark should systematically vary participant-to-camera distance, camera viewpoint, lighting conditions, degree of occlusion, movement speed, body orientation, and type of activity. Metrics for comparison may include mean absolute angular error (MAE), root mean squared error (RMSE), posture-assessment accuracy, sensitivity, specificity, F1-score, false positive rate, false negative rate, missing keypoint rate, frame-to-frame output jitter, temporal stability, computational latency, and frame rate. Wherever possible, system outputs should be compared against appropriate reference measurements obtained from optical motion capture systems, validated inertial measurement units (IMUs), calibrated goniometers, or expert-annotated reference data, depending on the evaluation scenario.

Second, the depth-aware tolerance mechanism requires systematic calibration and sensitivity analysis. Future studies should examine how the reference distance, linear scaling factor, and maximum scaling factor affect angular error, posture acceptance and rejection, false positive and false negative rates, and temporal stability across participants and camera configurations. The current frame-level global scaling approach should also be compared with the joint-specific depth compensation specified in the second architecture. Such work could investigate whether joint-specific compensation benefits from incorporating factors such as sensor-to-joint distance, keypoint confidence, segment characteristics, or joint-specific uncertainty. In parallel, confidence information could be extended beyond keypoint filtering to investigate confidence-dependent or uncertainty-aware tolerance adaptation.

Third, temporal stabilization and multi-person handling require dedicated evaluation and further development. The effects of valid-posture persistence and frame-stride reuse should be quantified in terms of false persistence, detection delay, output stability, and recovery after missing or unreliable keypoints. Safeguards such as maximum persistence duration, timeout conditions, and confidence-based reset criteria should also be investigated. For multi-person scenarios, future versions should incorporate and evaluate persistent identity-tracking or re-identification mechanisms, particularly under participant overlap, crossing trajectories, occlusion, and temporary out-of-frame transitions. Relevant measures should include identity-switch rate, tracking loss and recovery, reassociation success, scalability with increasing numbers of participants, computational latency, and frame rate.

Fourth, the agreement between COLLVIT assessment and expert judgement should be evaluated. In sports training or rehabilitation scenarios, system outputs could be compared with assessments provided by coaches, therapists, or domain experts to quantify agreement and identify joints, postures, or movement conditions for which automated assessment is less reliable. This research direction is also complementary to related research from our research group on computer-vision-based physiotherapy assessment, which explores the transformation of pose-estimation and biomechanical information into higher-level, clinically interpretable parameters [79]. Both lines emphasize, from complementary perspectives, the need to connect movement sensing and analysis with domain-specific assessment criteria and expert interpretation. These studies should also support the empirical refinement and validation of angular templates, orientation constraints, and acceptance criteria for specific activities and target populations.

Fifth, user-based studies should assess usability, perceived usefulness, understanding of assessment information and feedback, and subsequent potential effects on psychomotor learning or rehabilitation-related outcomes. In karate, for example, participants with different skill levels could perform predetermined kihon kumite sequences while the system evaluates their execution. In rehabilitation-oriented scenarios, evaluation should take place under appropriate professional supervision and distinguish technical assessment performance from clinical effectiveness. Such studies should investigate not only whether the system produces sufficiently accurate and stable assessment information, but also whether learners, instructors, therapists, or other intended users can understand and use that information appropriately. Collaborative psychomotor learning represents a particularly relevant research direction. Future studies should investigate whether feedback strategies derived from the framework's assessment information can support peer observation and correction, coordination, synchronization, motivation, shared regulation, or instructor-guided decision making. Different feedback modalities and representations, including visual overlays, color-coded joint indicators, auditory or haptic cues, post-session reports, and instructor dashboards, could be compared in terms of usability, interpretability, timing, cognitive demands, and effects on learning processes and outcomes.

Finally, future research should examine personalization, domain transfer, and longitudinal analytics. Building on the AI-based extensions discussed in Section 4.3, adaptive calibration could account for individual characteristics and movement variability, including user-specific joint range limits, fatigue-related changes, repetitive errors, flexibility, or recurring movement patterns. The framework could also be investigated beyond karate in domains involving structured psychomotor assessment, such as physical education, physiotherapy, adapted sport, ergonomics, dance, fitness training, and other forms of technical skill learning. However, transfer to each domain would require domain-specific templates, expert validation, and evaluation with the corresponding target population. Longitudinal collection of posture assessment outputs, angular deviations, temporal sequences, confidence values, and feedback events could additionally support learning-analytics capabilities for modeling participant progression and examining intervention effects over time.

Together, these research directions would enable the second stage of the KUMITRON–COLLVIT framework to progress from its current implementation-focused maturity toward greater empirical validation and application-oriented evidence. Such developments would provide richer and more interpretable movement information that can subsequently support the later SMDD stages of designing and delivering feedback, whose pedagogical effectiveness remains to be empirically evaluated.

5. Conclusions

This paper presented the evolution of a patent-based technological framework for real-time markerless kinematic posture assessment, comprising two successive architectures and their implementation through KUMITRON and COLLVIT. The evolution shows how an initial architecture for synchronized multi-participant motor-activity monitoring was extended through a second architecture toward configurable, depth-informed kinematic posture assessment.

The first architecture provided the foundation for monitoring multi-participant activities involving coordinated motor activity. The architecture included video acquisition, inertial sensors, physiological sensors, contextual information, wireless transmission, synchronization of processing, typified movement identification, and feedback delivery. A distinguishing feature of this architecture was its focus not only on individual actions but also on action–reaction relationships between participants. KUMITRON implemented this architecture for karate kumite training. Previous KUMITRON-related studies provided exploratory evidence that multimodal information could be synchronized and used to support the analysis of paired motor activity and strategy-oriented feedback.

The second architecture, formalized in a patent application currently under examination, addresses the limitations of fixed-threshold and primarily two-dimensional posture assessment. COLLVIT implements core capabilities of this second architecture, incorporating RGB-D acquisition, 2D/3D fusion, joint-angle computation, configurable angular templates, temporal stabilization, and multi-person processing. In the current implementation, depth information contributes to 3D joint-angle computation and to bounded global tolerance scaling based on frame-level mean depth, whereas the joint-specific depth compensation specified at the architectural level remains to be implemented and empirically validated.

The transition from KUMITRON to COLLVIT, therefore, represents a shift from multimodal movement monitoring toward depth-informed markerless kinematic posture assessment. KUMITRON represents the interaction-aware multi-participant implementation stage, while COLLVIT represents the subsequent depth-informed kinematic posture-assessment stage. Together, these successive architectures and their implementation constitute a framework in which movement interaction, posture assessment, temporal continuity, and assessment information supporting feedback can be integrated for collaborative learning environments.

From a technological perspective, the main novelty lies in the integration of interaction-aware analysis with depth-aware kinematic posture assessment. From an educational perspective, the framework may provide a basis for future studies by making movement-related kinematic information more visible, measurable, and interpretable for participants, learners, peers, instructors, and therapists. The framework does not aim to replace expert guidance, but to provide additional assessment information that may support correction, reflection, progression, and personalized feedback.

Nevertheless, further empirical validation is required. KUMITRON has already been implemented and explored through prototype development and published studies, whereas COLLVIT requires systematic testing in real scenarios. Future work should compare depth-aware dynamic tolerance with fixed-threshold assessment, evaluate agreement with expert assessment, test the system with users of different skill levels and morphologies, and analyze its potential effects on psychomotor learning and rehabilitation-related outcomes, with particular attention to multi-participant scenarios.

In conclusion, the evolution from KUMITRON to COLLVIT defines an extensible patent-based technological framework that connects interaction-aware multi-participant monitoring with configurable, depth-informed kinematic posture assessment. Rather than demonstrating superior performance or learning effectiveness, the present work establishes the technological basis for future empirical research on robust and personalized posture assessment, interpretable feedback design, and AI-supported collaborative psychomotor learning.

Author Contributions: Conceptualization, J.E. and O.C.S.; methodology, J.E. and O.C.S.; software, J.E.; formal analysis, J.E.; investigation, J.E.; writing—original draft preparation, J.E.; writing—review and editing, J.E. and O.C.S.; supervision, O.C.S.; funding acquisition, J.E. and O.C.S. All authors have read and agreed to the published version of the manuscript.

Funding: The development of KUMITRON was framed within the research projects INT2AFF (PGC2018-102279-B-I00), funded by MCIU/AEI and ERDF, and HUMANAID-Sens (TED2021-129485B-C41), funded by MCIN/AEI and EU NextGenerationEU/PRTR. KUMITRON also received funding in 2021 through the Ekintzaile programme. The development of COLLVIT received no external funding. No specific funding was received for the preparation of this manuscript.

Data Availability Statement: The original contributions presented in this study are included in the article material. Further inquiries can be directed to the corresponding authors. The technical information supporting the reported system architectures is provided within the paper and the cited patent documents. Additional implementation materials are not publicly available because they are subject to intellectual property and potential commercialization restrictions.

Acknowledgments: During the preparation of this manuscript, the authors made limited use of ChatGPT (OpenAI, GPT-5.6 Sol) for language editing, proofreading, and refinement of the written English. The tool was not used to generate research data, analyze data, interpret results, or formulate scientific conclusions. All AI-assisted content was critically reviewed and revised by the authors, who take full responsibility for the accuracy, originality, and integrity of the final manuscript. The first author would also like to thank Hospital de Mendaro (OSI Debabarrena) for facilitating access to its facilities for preliminary technical testing of the COLLVIT prototype. The BPSO® (Best Practice Spotlight Organization) programme material visible in the background of Figure 4 reflects the hospital's participation in this programme promoting excellence in evidence-based nursing care and is unrelated to the development or evaluation of COLLVIT.

Conflicts of Interest: The authors are named inventors on patents and/or patent applications related to the KUMITRON and COLLVIT technologies discussed in this manuscript. These intellectual-property relationships are disclosed for transparency. The authors declare no conflicts of interest.

References

1. Santos, O.C.; Echeverria, J.; Sánchez García, P.; Martín Hernández, M.; Universidad Nacional de Educación a Distancia (UNED). System for Analysing the Practice of Motor Activity with Participants/Sistema Para Analizar Una Práctica de Actividad Motora Con Participantes. Spanish Patent ES2890715B2, 31 January 2024. Available online: https://worldwide.espacenet.com/patent/search/family/079552828/publication/ES2890715B2?q=pn%3DWO2022008768A1 (accessed on 17 August 2026).
2. Echeverria, J.; Santos, O.C. KUMITRON: Artificial Intelligence System to Monitor Karate Fights that Synchronize Aerial Images with Physiological and Inertial Signals. In *Companion Proceedings of the 26th International Conference on Intelligent User Interfaces, College Station, TX, USA, 13–17 April 2021*; Association for Computing Machinery: New York, NY, USA, 2021; pp. 37–39. [CrossRef]
3. Echeverria, J.; Santos, O.C. Toward Modeling Psychomotor Performance in Karate Combats Using Computer Vision Pose Estimation. *Sensors* **2021**, *21*, 8378. [CrossRef] [PubMed]
4. Casas-Ortiz, A.; Echeverria, J.; Santos, O.C. Chapter 18: Intelligent systems for psychomotor learning: A systematic review and two cases of study. In *Handbook of Artificial Intelligence in Education*; Edward Elgar Publishing: Cheltenham, UK, 2023; pp. 390–421. [CrossRef]
5. Santos, O.C.; Echeverria, J. Universidad Nacional de Educación a Distancia (UNED). System And Method For Biomechanical Validation. Spanish Patent Application P202630575; Currently under examination,
6. Echeverria, J.; Santos, O.C. Towards Analyzing Psychomotor Group Activity for Collaborative Teaching Using Neural Networks. In *Artificial Intelligence in Education. Posters and Late Breaking Results, Workshops and Tutorials, Industry and Innovation Tracks, Practitioners, Doctoral Consortium and Blue Sky*; Wang, N., Rebolledo-Mendez, G., Dimitrova, V., Matsuda, N., Santos, O.C., Eds.; Springer Nature: Cham, Switzerland, 2023; pp. 403–408.
7. Anwar, M.; Mushtaq, N.; Mubeen, A.; Iqbal, M. The Power of ZPD: Enhancing Teaching and Learning. *J. Educ. Soc. Stud.* **2024**, *5*, 396–405. [CrossRef]
8. Ness, I.J. Zone of Proximal Development. In *The Palgrave Encyclopedia of the Possible*; Palgrave Macmillan: Cham, Switzerland, 2023; pp. 1781–1786. [CrossRef]
9. Edriss, S.; Romagnoli, C.; Caprioli, L.; Bonaiuto, V.; Padua, E.; Annino, G. Commercial vision sensors and AI-based pose estimation frameworks for markerless motion analysis in sports and exercises: A mini review. *Front. Physiol.* **2025**, *16*, 1649330. [CrossRef] [PubMed]
10. Ullah, R.; Asghar, I.; Nawaz, R.; Stacey, C.; Akbar, S.; Bishop, P. A real time action scoring system for movement analysis and feedback in physical therapy using human pose estimation. *Sci. Rep.* **2025**, *15*, 44784. [CrossRef] [PubMed]
11. Vu-Minh, A.; Nguyen-Quy, D.; Nguyen-Khanh, H.; Nguyen, T. AI Fitness Trainer Using 3D Human Pose Estimation and Dynamic Time Warping. In *Proceedings of the 13th Conference on Information Technology and Its Applications, Da Nang/Hoi An, Vietnam, 19–20 July 2024*; Nguyen, N.T., Huynh, C.-P., Nguyen, T.T., Le-Khac, N.-A., Nguyen, Q.-V., Eds.; Springer Nature: Cham, Switzerland, 2024; pp. 50–62. [CrossRef]
12. Benharkat, N.E.H.; Bentaalla Kaced, S.; Chakhrit, A.; Chergui, A. Automatic real-time ergonomic posture assessment using digital models: A case study of manual handling tasks. *Int. J. Adv. Manuf. Technol.* **2026**, *142*, 963–980. [CrossRef]
13. Núñez, G.A.Z.; Norambuena, N.; Cuevas Quezada, I.; Valín Rivera, J.L.; Narea Olmos, J.; Galleguillos Ketterer, C. Real-Time Automated Ergonomic Monitoring: A Bio-Inspired System Using 3D Computer Vision. *Biomimetics* **2026**, *11*, 88. [CrossRef] [PubMed]
14. Surya Abhishek, R.V.S.S.; Sridevi, T. AI and Computer Vision Techniques for Fitness Training and Form Analysis: A Comprehensive Review. In *Smart Trends in Computing and Communications*; Senjyu, T., So-In, C., Joshi, A., Eds.; Springer Nature: Singapore, 2026; pp. 327–336.
15. Casas-Ortiz, A.; Jimenez-Tellez, N.; Portaz, M.; Santos, O.C. A holistic framework to create intelligent systems for psychomotor learning: A systematic review and insights from the virtual reality Kenpo learning simulator. *Virtual Real.* **2026**, *30*, 72. [CrossRef]
16. Sigrist, R.; Rauter, G.; Riener, R.; Wolf, P. Augmented visual, auditory, haptic, and multimodal feedback in motor learning: A review. *Psychon. Bull. Rev.* **2013**, *20*, 21–53. [CrossRef] [PubMed]
17. Park, S.-W.; Ha, Y.-K. Measuring Assistive Technology Outcomes via AI-Based Kinematic Modeling of Individualized Routine Learning in Elite Boccia Athletes with Severe Cerebral Palsy: A Longitudinal Case Series. *Bioengineering* **2026**, *13*, 261. [CrossRef] [PubMed]

18. Kopalli, S.R.; Shukla, M.; Jayaprakash, B.; Kundlas, M.; Srivastava, A.; Jagtap, J.; Gulati, M.; Chigurupati, S.; Ibrahim, E.; Khandige, P.S.; et al. Artificial intelligence in stroke rehabilitation: From acute care to long-term recovery. *Neuroscience* **2025**, *572*, 214–231. [CrossRef] [PubMed]
19. Elkess, G.; Elmoushy, S.; Atia, A. Karate First Kata Performance Analysis and Evaluation with Computer Vision and Machine Learning. In *Proceedings of the 2023 International Mobile, Intelligent, and Ubiquitous Computing Conference (MIUCC), Cairo, Egypt, 27–28 September 2023*; IEEE: Piscataway, NJ, USA, 2023; pp. 1–6. [CrossRef]
20. Mostafa, K.; Hany, M.; Atia, A. Automated Evaluation of Karate Practitioners' Style during the First Kata Performance using Deep Learning and Pose Estimation. In *Proceedings of the 2023 5th Novel Intelligent and Leading Emerging Sciences Conference (NILES), Giza, Egypt, 21–23 October 2023*; IEEE: Piscataway, NJ, USA, 2023; pp. 238–241. [CrossRef]
21. Aju, A.; Mathew, C.; Prakasi, O. PoseNet based Model for Estimation of Karate Poses. *J. Innov. Image Process.* **2022**, *4*, 16–25. [CrossRef]
22. Ait-Bennacer, F.-E.; Aaroud, A.; Akodadi, K.; Cherradi, B. Applying Deep Learning and Computer Vision Techniques for an e-Sport and Smart Coaching System Using a Multiview Dataset: Case of Shotokan Karate. *Int. J. Online Biomed. Eng.* **2022**, *18*, 35–53. [CrossRef]
23. Echeverria, J.; Santos, O.C. Punch Anticipation in a Karate Combat with Computer Vision. In *Adjunct Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization (UMAP '21)*; Association for Computing Machinery: New York, NY, USA, 2021; pp. 61–67. [CrossRef]
24. Petri, K.; Lichtenstein, M.; Bandow, N.; Campe, S.; Wechselberger, M.; Sprenger, D.; Kaczmarek, F.; Emmermacher, P.; Witte, K. Analysis of anticipation by 3D motion capturing—A new method presented in karate kumite. *J. Sports Sci.* **2016**, *35*, 130–135. [CrossRef] [PubMed]
25. Hudovernik, V.; Skocaj, D. Video-Based Detection of Combat Positions and Automatic Scoring in Jiu-jitsu. In *Proceedings of the 5th International ACM Workshop on Multimedia Content Analysis in Sports, Lisbon, Portugal, 14 October 2022*; Association for Computing Machinery: New York, NY, USA, 2022; pp. 55–63. [CrossRef]
26. Wu, G.; Wen, C.; Jiang, H. Wushu Movement Recognition System Based on DTW Attitude Matching Algorithm. *Entertain. Comput.* **2025**, *52*, 100877. [CrossRef]
27. Blanco Ortega, A.; Godoy, J.I.; Szwedowicz Wasik, D.S.; Martínez Rayón, E.; Cortés García, C.; Azcaray Rivera, H.R.; Gómez Becerra, F.A. Biomechanics of the Upper Limbs: A Review in the Sports Combat Ambit Highlighting Wearable Sensors. *Sensors* **2022**, *22*, 4905. [CrossRef] [PubMed]
28. Katović, D.; Bronzin, T.; Horvat, M.; Prole, B.; Stipić, A.; Jelača, N.; Pavlović, I.; Pap, K. The use of AI in Human Pose Estimation Applications in Kinesiology: Taxonomy of Algorithms, Models, and Evaluation Methods. In *Proceedings of the 2024 47th MIPRO ICT and Electronics Convention (MIPRO), Opatija, Croatia, 20–24 May 2024*; IEEE: Piscataway, NJ, USA, 2024; pp. 1163–1166. [CrossRef]
29. Nagorny, P.; Kevelham, B.; Chagué, S.; Charbonnier, C. A Comprehensive Review of Real-Time Multi-View Multi-Person Markerless Motion Capture. *ACM Comput. Surv.* **2025**, *58*, 75. [CrossRef] [PubMed]
30. Chen, L.; Wu, B.; Zhao, Y. A Real-Time Photogrammetric System for Monitoring Human Movement Dynamics. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2020**, *43*, 561–566. [CrossRef]
31. Ribeiro, T.; Fernandes, E.; Ribeiro, A.; Lopes, C.; Ribeiro, F.; Lopes, G. Teleoperation System for Service Robots Using a Virtual Reality Headset and 3D Pose Estimation. *Sensors* **2026**, *26*, 471. [CrossRef] [PubMed]
32. Espitia-Mora, L.A.; Vélez-Guerrero, M.A.; Callejas-Cuervo, M. Development of a Low-Cost Markerless Optical Motion Capture System for Gait Analysis and Anthropometric Parameter Quantification. *Sensors* **2024**, *24*, 3371. [CrossRef] [PubMed]
33. Jha, P.; Yadav, G.P.K.; Bandhu, D.; Hemalatha, N.; Mandava, R.K.; Adin, M.Ş.; Saxena, K.K.; Patel, M. Human–machine interaction and implementation on the upper extremities of a humanoid robot. *Discov. Appl. Sci.* **2024**, *6*, 152. [CrossRef]
34. Ferraris, C.; Amprimo, G.; Olmo, G.; Ghislieri, M.; Patera, M.; Suppa, A.; Gallo, S.; Imbalzano, G.; Lopiano, L.; Artusi, C.A. From RGB-D to RGB-Only: Reliability and Clinical Relevance of Markerless Skeletal Tracking for Postural Assessment in Parkinson's Disease. *Sensors* **2026**, *26*, 1146. [CrossRef] [PubMed]
35. Qi, L.; Sun, H. Multistage fall detection framework via 3D pose sequences and TCN integration. *Sci. Rep.* **2025**, *15*, 27832. [CrossRef] [PubMed]
36. Barredo Arrieta, A.; Díaz-Rodríguez, N.; Del Ser, J.; Bennetot, A.; Tabik, S.; Barbado, A.; Garcia, S.; Gil-Lopez, S.; Molina, D.; Benjamins, R.; et al. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Inf. Fusion* **2020**, *58*, 82–115. [CrossRef]
37. Rudin, C. Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead. *Nat. Mach. Intell.* **2019**, *1*, 206–215. [CrossRef] [PubMed]
38. Dowling, A.V.; Favre, J.; Andriacchi, T.; Leland Stanford Junior University. Systems and Methods for Measuring, Analyzing, and Providing Feedback for Movement in Multidimensional Space. U.S. Patent US20130244211A1, 19 September 2013.

39. Mettler May, B.; ICUEMOTION LLC. Systems and Methods for Data-Driven Movement Skill Training. U.S. Patent US20190009133A1, 10 January 2019.
40. French, B.J.; Ferguson, K.R.; Impulse Technology Ltd. System and Method for Tracking and Assessing Movement Skills in Multidimensional Space. U.S. Patent US8861091B2, 14 October 2014.
41. Virkar, H.; Kaplan, L.R.; Furlani, S.; Borgman, J.; Bhave, A.; Mirrorar LLC. Systems and Methods for Assessing Balance and form During Body Movement. U.S. Patent US11633659B2, 25 April 2023.
42. Akdogan, E.; Aktan, M.E.; Yildiz Teknik Universitesi; Bartin Universitesi. Artificial Intelligence-Based Algorithm for Physiotherapy and Rehabilitation Robots for Diagnosis and Treatment Purposes. U.S. Patent US20220068458A1, 3 March 2022.
43. Song, A.; Tang, X.; Shi, K.; Chen, D.; Li, H.; Zeng, H.; Xu, B. Human Body Posture Estimation Method Based on OpenPose and Kinect and Rehabilitation Training System. Chinese Patent CN109003301B, 15 March 2022.
44. Kong, D.; Wu, Y.; Wang, S.; Li, J.; Wang, L.; Beijing University of Technology. Three-Dimensional Human Body Posture Estimation Method. Chinese Patent CN110097639B, 18 April 2023.
45. Yang, X.; Xie, C.; Jia, X.; Zhang, W.; Zhejiang University of Technology. Human Body Posture Estimation Method Based on Depth and Color Image Feature Fusion. Chinese Patent CN111241936A, 5 June 2020.
46. Cook, D.J.; Crandall, A.S.; Washington State University. Activity Recognition in Multi-Entity Environments. U.S. Patent US9460350B2, 4 October 2016.
47. Amini, A.A. System and Method for Motion Analysis and Feedback with Ongoing Dynamic Training Orientation Determination. U.S. Patent US20130018494A1, 17 January 2013.
48. Rosenberg, E. Quantified Movement Feedback System. International Patent WO2021062330A1, 1 April 2021.
49. Zhao, Q.; Wu, W.; Hao, A.; Zhao, Y.; Beihang University. Human Skeleton Extraction and Orientation Judging Method Based on Geodesic Model. Chinese Patent CN102622606B, 31 July 2013.
50. Canavan, C.; Hughes, G.; Focus Ventures Inc. System and Method for Identifying and Interpreting Repetitive Motions. U.S. Patent US20140270375A1, 18 September 2014.
51. Schwalb, E. Method and System for Movement Training. International Patent WO2010084490A1, 29 July 2010.
52. Chen, W. A Method, System, and Storage Medium for Comparing and Scoring Standard Movements. Chinese Patent CN121919786A, 24 April 2026.
53. Xiong, J. A Method and System for Deadlift Posture Detection Based on Mobile Video. Chinese Patent CN121686571A, 17 March 2026.
54. Zhang, D. Physical Training Posture Correction Method Based on Machine Vision. Chinese Patent CN121281141A, 6 January 2026.
55. Huang, Y. An Interactive Martial Arts Immersive Teaching System. Chinese Patent CN121868837A, 17 April 2026.
56. Wang, W. A Method and Quantitative Evaluation System for Verifying the Structure of Tai Chi. Chinese Patent CN121858916A, 14 April 2026.
57. An, T. Intelligent Rehabilitation Training System and Method for Lower Limb Dyskinesia. Chinese Patent CN121839020A, 10 April 2026.
58. Wang, W. A Sports Injury Risk Screening System Based on Motion Capture and Biomechanical Analysis. Chinese Patent CN121812169A, 7 April 2026.
59. Wu, W. Rehabilitation Training Action Evaluation Method Based on Computer Vision. Chinese Patent CN121904837A, 21 April 2026.
60. Guo, J. Body-Building Action Recognition, Counting and Quality Assessment Method Based on Machine Vision. Chinese Patent CN121904845A, 21 April 2026.
61. Liao, Z. Push-Up Training Multi-Person Test Result Data Identification Method and System. Chinese Patent CN121861719A, 14 April 2026.
62. Xu, X. Human Track Prediction Method and Monitoring System for Human-Computer Cooperation. Chinese Patent CN121982644A, 5 May 2026.
63. Ge, F. Personalized Exercise Training Evaluation Method Based on Multi-Modal Data. Chinese Patent CN121862308A, 14 April 2026.
64. Guo, L. Multi-Mode Hierarchical Fusion Network-Based Motion Function Multi-Dimensional Evaluation Method. Chinese Patent CN121839147A, 10 April 2026.
65. Liu, X. Variable Impedance Control Method for Wrist Rehabilitation Training Robot. Chinese Patent CN121946517A, 1 May 2026.
66. Huang, K. Design Method of Human-Fuzzy-Decision-Fused Lower Limb Exoskeleton Human-Machine System. Chinese Patent CN121893289A, 21 April 2026.
67. Li, E. An Embodied Intelligent Architecture and an Environmentally Adaptive Learning System. Chinese Patent CN121960630A, 1 May 2026.
68. Soto-Cabán, S.; Avila-Medina, F. Design and Implementation of an Industry Mentorship Program in a First-Year Bachelor of Science in Engineering Technology Course. In *Proceedings of the 2025 ASEE Annual Conference & Exposition, Montreal, QC, Canada, 22–25 June 2025*; American Society for Engineering Education: Washington, DC, USA, 2025. [CrossRef]
69. Cai, L.; Msafiri, M.M.; Kangwa, D. Exploring the impact of integrating AI tools in higher education using the Zone of Proximal Development. *Educ. Inf. Technol.* **2025**, *30*, 7191–7264. [CrossRef]

70. Cao, S.; Zhou, S.; Luo, Y.; Wang, T.; Zhou, T.; Xu, Y. A review of the ESL/EFL learners' gains from online peer feedback on English writing. *Front. Psychol.* **2022**, *13*, 1035803. [CrossRef] [PubMed]
71. Pritchard, R.; Dockerty, F. Game on! Enhancing primary physical education through a Rosenshine-inspired approach. *Curric. Stud. Health Phys. Educ.* **2024**, *16*, 82–96. [CrossRef]
72. Han, P.H.; Chen, Y.S.; Zhong, Y.; Wang, H.L.; Hung, Y.P. My Tai-Chi coaches: An augmented-learning tool for practicing Tai-Chi Chuan. In *Proceedings of the 8th Augmented Human International Conference (AH '17)*; Association for Computing Machinery: New York, NY, USA, 2017. [CrossRef]
73. Santos, O.C. Training the Body: The Potential of AIED to Support Personalized Motor Skills Learning. *Int. J. Artif. Intell. Educ.* **2016**, *26*, 730–755. [CrossRef]
74. Hartley, R.; Zisserman, A. *Multiple View Geometry in Computer Vision*, 2nd ed.; Cambridge University Press: Cambridge, UK, 2004.
75. Ciardiello, L.; Agnello, P.; Petyx, M.; Martinelli, F.; Cesarelli, M.; Santone, A.; Mercaldo, F. A Method for Human Pose Estimation and Joint Angle Computation Through Deep Learning. *J. Imaging* **2026**, *12*, 157. [CrossRef] [PubMed]
76. Wu, G.; Siegler, S.; Allard, P.; Kirtley, C.; Leardini, A.; Rosenbaum, D.; Whittle, M.; D'Lima, D.D.; Cristofolini, L.; Witte, H.; et al. ISB recommendation on definitions of joint coordinate system of various joints for the reporting of human joint motion—Part I: Ankle, hip, and spine. *J. Biomech.* **2002**, *35*, 543–548. [CrossRef] [PubMed]
77. Khoshelham, K.; Elberink, S.O. Accuracy and Resolution of Kinect Depth Data for Indoor Mapping Applications. *Sensors* **2012**, *12*, 1437–1454. [CrossRef] [PubMed]
78. Lafayette, T.B.D.G.; Kunst, V.H.D.L.; Melo, P.V.D.S.; Guedes, P.D.O.; Teixeira, J.M.X.N.; Vasconcelos, C.R.D.; Teichrieb, V.; Da Gama, A.E.F. Validation of Angle Estimation Based on Body Tracking Data from RGB-D and RGB Cameras for Biomechanical Assessment. *Sensors* **2023**, *23*, 3. [CrossRef] [PubMed]
79. García, V.; Santos, O.C. Towards Intelligent Assessment in Personalized Physiotherapy with Computer Vision. *Sensors* **2025**, *25*, 3436. [CrossRef] [PubMed]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.