Preprints.org

Review

Not peer-reviewed version

Deep Learning-based Pose Estimation in Providing Feedback for Physical Movement: a Review

Atima Tharatipyakul and Suporn Pongnumkul *

Posted Date: 6 June 2023

doi: 10.20944/preprints202306.0395.v1

Keywords: Pose estimation; Movement assessment; Augmented feedback; Physical movement; Review

Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 6 June 2023 doi:10.20944/preprints202306.0395.v1

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.

Article

Deep Learning-Based Pose Estimation in Providing Feedback for Physical Movement: A Review

Atima Tharatipyakul and Suporn Pongnumkul *

National Electronics and Computer Technology Center (NECTEC), 112 Phahonyothin Road, Khlong Nueng, Khlong Luang District, Pathumthani 12120, Thailand
* Correspondence: suporn.pongnumkul@nectec.or.th

Abstract: Pose estimation has various applications in analyzing human movement and behavior, including providing feedback to users about their movements so they could adjust and improve their movement skills. To investigate the current research status and possible gaps, we searched Scopus and Web of Science for articles that (1) human 'body' pose estimation is used and (2) user movement is assessed and communicated. We used either a bottom-up or top-down approach to analyze 20 articles for methods used to estimate human pose, assess movement, provide feedback to users, as well as methods to evaluate them. Our review found that pose estimation systems typically used CNNs while movement assessment methods varied from mathematical formulas or models, rule-based approaches, to machine learning. Feedback was primarily presented visually in verbal forms and nonverbal forms. The experiments to evaluate each part ranged from the use of public datasets to human participants. We found that while there was an improvement, the majority of pose estimation challenges remain. The effectiveness and factors for choosing movement assessment methods for a new context are still unclear. In the end, we suggest that studies about feedback prioritization and erroneous feedback are needed.

Keywords: pose estimation; movement assessment; augmented feedback; physical movement; review

1. Introduction

Human pose estimation is the process of determining the positions and orientations of specific human body parts, such as the head, shoulders, arms, and legs. Pose estimation has a wide range of applications in fields that involve the analysis and understanding of human movement and behavior. For example, in human-computer interaction, the estimation of hand pose could enable gesture-based natural interaction [1]. In healthcare, pose estimation could help monitor and analyze the movements and posture of patients in rehabilitation or therapy settings (e.g., [2,3]). While doing a physical movement, such as rehabilitation or sports training, pose estimation could be used for various purposes [4–7], and one of them is to provide feedback about the movement to a user.

Feedback is an important aspect of physical movement learning and training, as it allows individuals to assess their performance and adjust to improve their movement skills. Feedback could be *task-intrinsic*, from the sensory system of a performer, or *augmented*, from an external source to a performer. Traditionally, a coach or instructor provides augmented feedback as verbal cues or corrections during a training session. For example, a yoga instructor may tell a learner who does an inverted U-shape downward dog pose to do a V-shape instead. With automated pose estimation, a computer system could assess user movement and provide augmented feedback. Da Gama et al. [8], for example, reviewed 31 articles that used Kinect to assess and provide feedback on motor rehabilitation. The review suggested development possibilities and further studies of using Kinect to rehabilitate at home.

While there is a large body of knowledge on using Kinect or other sensors for pose estimation and physical movement applications, recent advances in pose estimation using a web camera open new opportunities in this field. Several physical movement applications using deep learning-based

© 2023 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 6 June 2023 doi:10.20944/preprints202306.0395.v1

human pose estimation have been proposed and reviewed [5–7]. Yet, none of them focus on how it is used for assessing a movement and providing feedback to a user, which is an important aspect of physical movement applications and an indication of how pose estimation is adopted in practice.

This paper presents a review of recent articles that use deep learning-based human pose estimation to assess user movement and provide feedback on the user's physical movement. We divided the physical movement applications into three parts or modules: (1) *pose estimation* that detects keypoints of a human body; (2) *movement assessment* that uses keypoints or the motion of keypoints to evaluate quality or value of movement; and (3) *augmented feedback presentation* that communicates the results of other modules to users.

The goal of our paper is threefold:

- to investigate methods used for pose estimation, movement assessment, and augmented feedback presentation;
- to investigate how those methods were experimented or evaluated;
- to discuss the current research status of each part, possible gaps, and future research directions.

We searched for articles published between January 2017 to July 2022 in Scopus and Web of Science, then we analyzed 20 articles that fitted our inclusion criteria.

We organize this article as follows: Section 2 gives an overview of the current topic followed by Section 3 that presents the review methodology. Section 4 and 5 presents the results and discussions. We conclude this article in Section 6.

2. Background and Related Work

This section clarifies terms used in this paper and summarizes related articles on pose estimation for physical movement applications, smart technology for movement assessment, and augmented feedback and their effectiveness.

2.1. Pose estimation for physical movement applications

In this paper, we use the term *pose estimation* for a process for determining the position of key points of a person's body from a given image or video. The methods for pose estimation can be classified into two-dimensional (2D) [9] and three-dimensional (3D) [10,11], and could be further classified based on the number of people in the image (single-person or multi-person) or approaches (e.g., top-down or bottom-up) [12].

Pose estimation has a wide range of applications. For physical movement learning or training, a lot of works have been proposed, and as a result, there exist a number of articles that review them. For instance, Difini et al. [5] systematically reviewed the usage of human pose estimation for training assistance. They identified 8 articles and investigated the challenges, the technology used, the application context, and the accuracy of human pose estimation. Stenum et al. [6] reviewed the applications of pose estimation in three domains: motor and non-motor development; human performance optimization, injury prevention, and safety; and clinical motor assessment. Their identified application limitations included occlusions, limited training data, capture errors, positional errors, and limitations of recording devices. Their identified application limitations included user-friendliness (i.e., set-up time, delayed results, and programming and training requirements), outcome measure challenges, limited hardware infrastructure, technology challenges, and lack of validation and feasibility data. Badiola-Bengoa and Mendez-Zorrilla [7] reviewed 20 articles on pose estimation in sports and physical exercise. They discussed the available data, methods, performance, opportunities, and challenges. Niu et al. [13] surveyed articles that used inertial measurement units (IMU) and computer vision for human pose estimation in rehabilitation applications. They summarized the research status and challenges as well as suggested that the two methods can be combined to get better outcomes.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 6 June 2023 doi:10.20944/preprints202306.0395.v1

These reviews are useful for identifying existing methods, challenges, and future directions of using pose estimation in physical movement applications. Still, they had not reviewed how those applications assessed movement and provided feedback to users, which is our focus in this paper.

2.2. Smart technology for movement assessment

We use the term *movement assessment* as a way to evaluate or estimate the quality of movement to provide feedback to a user. According to this definition, we consider works on, for example, action recognition or movement modeling, if a user could use the result to assess their movement, such as telling whether it is correct or not. The use of Artificial Intelligence to analyze user movement and provide feedback to improve user performance, quality of life, and well-being has been a subject of research for more than a decade. Several reviews have been published to present the level of knowledge on the topic.

Caramiaux et al. [14] conducted a short review of machine learning approaches for motor learning. They focused on motor variability which requires the algorithms to differentiate between new movements and variations from known ones. They identified three types of adaptation: parameter adaptation in probabilistic models, transfer and meta-learning in deep neural networks, and planning adaptation by reinforcement learning. They also discussed challenges for applying the models, including variations of an already-trained skill, an adaptation that involves re-training procedures, and continuous evolution of motor variation patterns.

Rajšp and Fister [15] identified 109 articles related to intelligent data analysis for sports training. They focused on the competitive activity (e.g., no leisure training) in four training stages: planning, realization, control, and evaluation. They discussed the challenges of gathering data sets, working with coaches and players, and applying knowledge to practical situations.

Gámez Díaz et al. [16] focused on digital twin coaching, which collects user data and provides personalized feedback. They categorized the works into sports, well-being, and rehabilitation domain. The discussed algorithms, devices, performance, and usability feedback of users. They discussed the challenges in evaluating the user's feedback and user interface.

As these reviews did not particularly focus on pose estimation, they mostly analyzed articles that used other techniques, such as sensors, and only some or a few articles that used pose estimation. Tsiouris et al. [17], for instance, reviewed 41 articles on virtual coaching for users with morbidity, but only one article monitored 3D pose. In this paper, we focus on works that use pose estimation only, as we deem it a promising technology for its capability to work on a normal laptop or mobile phone without extra equipment needed.

2.3. Augmented feedback and their effectiveness

Augmented feedback refers to feedback from an external source to a performer. While works on smart technology for movement assessment discuss feedback, a large body of knowledge about augmented feedback comes from various fields within sport science. Several ways to provide feedback, such as knowledge of results and knowledge of performance, have been discussed and studied their effectiveness in various settings. Lauber and Keller [18], for example, reviewed studies that have applied augmented feedback in exercise and prevention settings, focusing on the positive influence of augmented feedback on motor performance. They discussed the limitations of studies, which caused difficulties for practitioners to determine the best way to provide augmented feedback.

Recent reviews of works suggested more pieces of evidence for the effectiveness of different augmented feedback methods. Zhou et al. [19], for example, investigated the data supporting the value of feedback in physical education. Based on 23 studies, the effectiveness of feedback over no-feedback had strong evidence, the effectiveness of visual feedback over verbal feedback had limited evidence, and the effectiveness of information feedback compared with praise or corrective feedback was inconsistent. Meanwhile, Mödinger et al. [20] systematically reviewed the effectiveness of video-based visual feedback in physical education in schools. They found 11 articles in total and

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 6 June 2023 doi:10.20944/preprints202306.0395.v1

suggested that visual feedback was more effective than solely verbal feedback. Still, they found its practical usage required considerations of specific conditions.

The reviewed articles range from the usage of video recording with human instructors to the usage of smart technology to provide augmented feedback. In this paper, we investigate and discuss how pose estimation applications implemented or extended existing knowledge of augmented feedback.

3. Methodology

We adopt PRISMA guidelines [21], an evidence-based minimum set of items for reporting in systematic reviews and meta-analyses.

We searched Scopus and Web of Science databases on 27 July 2022, using the following query: (“pose estimation" OR “pose tracking") AND (“exercise" OR “sport" OR “rehabilitation” OR “physical education" OR “motor" OR “movement" OR “athlet*") AND (“assistance" OR “correction" OR “guidance" OR “feedback" OR “learning" OR “coach" OR evaluat* OR assess* OR performan*). The search was in title, abstract, and keywords (i.e., “topic" in Web of Science). We limited the years to between 2017 - 2022 and the language to English only. For Scopus, we also exclude articles with the document type “Review".

We used CADIMA [22] as a tool for data collection and selection. The inclusion criteria were: (1) human ‘body’ pose estimation is used; (2) user movement is assessed and communicated. Articles that simply classify movement or count correct repetitions without giving feedback to users were excluded. The feedback must be shown, explained, and/or discussed (e.g., articles that simply mentioned giving feedback without screenshots or other details are excluded). We also excluded review articles from the analysis.

The first author performed article selection and initial analysis. We first identified methods for pose estimation, methods for assessing movement, and methods for presenting augmented feedback. Once the methods were noted, we categorized the pose estimation and movement assessment using a bottom-up approach while we largely adopt classifications of augmented feedback from Lauber and Keller [18] and Magill and Anderson [23]. As we focus on the methods, not the result, we did not assess the risk of bias or the certainty (or confidence) of experiments that were conducted to evaluate proposed methods. Instead, we reported how experiments were conducted and what data was used. Both authors discussed the results and wrote this paper.

4. Result

Figure 1 presents the article selection process. We found 104 articles from the search (81 from Scopus and 23 from Web of Science). After removing 18 duplicated records, we screened the title and abstract based on the criteria and excluded 34 articles that did not meet the criteria. In full-text screening, 6 articles were excluded because the full-text was not accessible. 7 articles (e.g., [3]) were excluded as they failed the first criterion and 25 articles (e.g., [24]) were excluded as they failed the second criterion. In the end, we found 20 articles as listed in Table 1. Human pose estimation, movement assessment, and augmented feedback presentation of each article were analyzed and are presented in the next subsections.

Table 1. List of 20 articles reviewed in this paper. The articles are listed with the context of use and grouped by publication year.

Year	Articles	Count
2018	Tennis [25]	1
2019	Dance [26], Ski [27], Tai Chi [28], Exercise [29], Rehabilitation [30]	5
2020	Rehabilitation [2]	1
2021	Ballet [31], Yoga [32–34], Tai Chi [35], Kickboxing [36], Baseball [37], Exercise [38–41]	11
2022	Baseball [42], Exercise [43]	2

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 6 June 2023 doi:10.20944/preprints202306.0395.v1

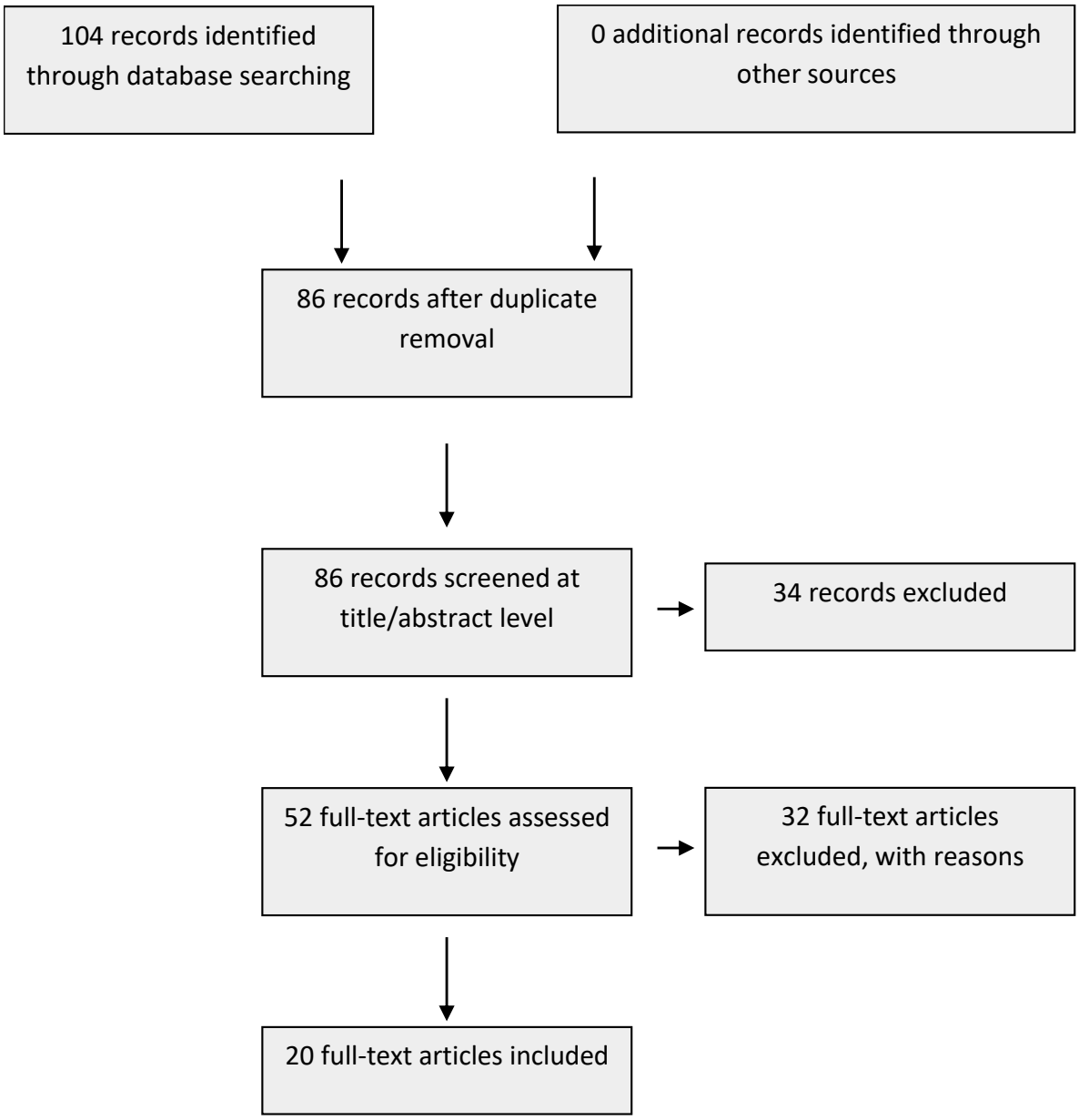

Figure 1. Paper selection flow diagram.

4.1. Human Pose Estimation

Human pose estimation module infers landmarks or keypoints of human figures, such as elbow locations, from an input. Human pose estimation could be broadly classified into 2D (X and Y coordinates) and 3D (X, Y, and Z coordinates). We discuss the techniques and studies conducted to evaluate those techniques.

4.1.1. Libraries or techniques used for pose estimation

We found the usage of OpenPose, PoseNet, Convolutional Neural Network (CNN), and other techniques for human pose estimation, as summarized in Table 2.

OpenPose is a real-time multi-person pose estimation library based on Cao et al. [44]. The library[1] used multi-stage CNN to detect 135 keypoints and reconstruct either 2D pose (e.g., [34,37,41–43]) and

[1] https://github.com/CMU-Perceptual-Computing-Lab/openpose

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 6 June 2023 doi:10.20944/preprints202306.0395.v1

3D pose (e.g., [2,31]). Though Nagarkoti et al. [29] and Kurose et al. [25] did not state using OpenPose directly, they cited Cao et al. [45], which OpenPose is based on.

Most articles used the library as it is while some extended the library for their purposes. Jan et al. [35] employed Lifting from the Deep [46] to to translate a 2D pose into a 3D pose. Similarly, [26] obtained 2D poses from multiple cameras, then implemented the binocular stereo matching principle to obtain a 3D pose. Yang et al. [38] used data augmentation, randomly cropped and rotated images, to address "abnormal human" situations and considered context information to address the uncertainty in the case of occlusion.

PoseNet is an open-sourced library [47] that uses a deep learning TensorFlow model to detect 17 2D-keypoints of human. The library supports multiple models. The lightest model could be run in realtime on modern smartphones but with lower accuracy. Articles that used PoseNet include [32,36,39].

Other CNN is used in [27,28,33,40]. Jeon et al. [40] used use Mobilenetv2 [48] to optimize an HPE model [49], which added a few deconvolutional layers over the last convolution stage in the ResNet [50]. They implemented Online Pose Distillation to minimize performance drop. Wang et al. [27] proposed a structural-aware convolution module, which concatenated spatial and temporal relation modules to go through a convolution layer to reduce dimension. Shi and Jiang [33] proposed a framework containing two branches: one used a confidential map to estimate the positions of bone joint points; another used affinity domain to predict the positions and directions of the limbs. They iterated the above two branches to construct a human skeleton based on the confidence set. Kamel et al. [28] implemented CNN with four convolutional layers, which received input from an RGB-D camera and generated a 3D skeleton model.

Other techniques, such as Kinect and other deep learning models, were experimented by Gu et al. [30].

Table 2. Methods for pose estimation. The asterisk (*) indicates that the method is further modified or extended.

Technique	2D Examples	3D Examples
OpenPose	[42], [43], [41], [34], [37], [29], [25], [38]*	[31], [2], [35]*, [26]*
PoseNet	[36], [39], [32]	NA
Other CNN	[40], [27], [33]	[28]
Others	[30]	[30]

4.1.2. Experiments

Experiments performed on the pose estimation module include accuracy and/or performance such as speed or frame rate. The accuracy is generally measured by comparing the positions of landmarks from the pose estimation module to some ground truth. Kamel et al. [28], for example, compared their pose estimation results with the results they gathered from Kinect. Similarly, Zhang et al. [26] evaluated dance movement reconstruction against previously measured actual 3D coordinates. On the other hand, some articles used publicly available datasets, including COCO [38,40], MPII [38], Penn Action, and JHMDB [27].

A few works experimented with different models. Gu et al. [30] compared 2D pose and 3D pose generated from four deep learning models [46,48,49,51] and Kinect. They selected Kinect for their system as it provided the most accurate result. Huang et al. [34] evaluated the accuracy and frame rate of different models, but there is no detail about the data used for the evaluation.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 6 June 2023 doi:10.20944/preprints202306.0395.v1

4.2. Movement Assessment

This module replaces the need for human experts, such as instructors or trainers, to oversee user movement. In order to analyze the quality of movement and give users feedback, the system may process keypoints or the motion of keypoints and/or measure the amount or degree of user attributes, then assess how well a user performs. We discuss the techniques for each part and studies conducted to evaluate those techniques.

Table 3. Movement assessment

Technique	Examples
Pre-processing	
Temporal alignment	Dynamic time warping [31], [35], [41], [30], [29]
Spatial alignment and normalization	Align using selected origin point and normalize [30], [28]; Normalization only [33]
Noise handling and smoothing	Filter [43], [31], [40]; Fill missing values [33]; Quarter-shift [40]
Selection	Spatial [35], [43]; Temporal [42]
Temporal segmentation	Peak detection [43]
Measurement	
Spatial measurement	Angle of body parts [39], [31], [35], [34], [37], [27], [28], [38], [30],
Temporal measurement	Direction/displacement [28], [42], [26]; Range of motion/rotated angle [39], [26]; Number of repetition [43]
Assessment methods	
Mathematical formula or model	Difference of a measurement [29], [26], [26]; Angular similarity/distance [42], [40], [34]; Euclidean distance [34], [33]; Others [35], [40], [38]
Rule-based method	Grading [26]; Correct posture checking [39], [36], [39]
Machine learning	Classification using SVM [25], [27], [2]; using Neural Network based on LSTM [2]; using Artificial Neural Network (ANN) [36], [32]; using time series classification [43]; Similarity score of embedding pairs from multi-stage CNN [41,52]

4.2.1. Pre-processing

Pre-processing involves translating keypoints or the motion of keypoints into a more desirable one, such as findings correct keypoints by removing wrong keypoints. Pre-processing techniques include temporal alignment, spatial alignment and normalization, noise handling and smoothing, selection, and temporal segmentation.

Temporal alignment aligns temporal sequences to cope with temporal issues. For instance, a delay typically occurs when a user is trying to imitate the model's movement. The speed of movement between a user and a model could differ. Temporal alignment attempts to minimize such temporal differences so the movement assessment can focus on comparing, for example, the posture. All papers with temporal alignment [29–31,35,41] employ dynamic time warping (DTW). The technique has been used to find patterns in time series [53] in various domains. The technique optimizes a distance metric and nonlinearly maps a frame of the model to a frame of the user, as illustrated in Figure 2. The distance metric can be customized. For instance, Nagarkoti et al. [29] used angles between the pair of limbs as the distance metric.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 6 June 2023 doi:10.20944/preprints202306.0395.v1

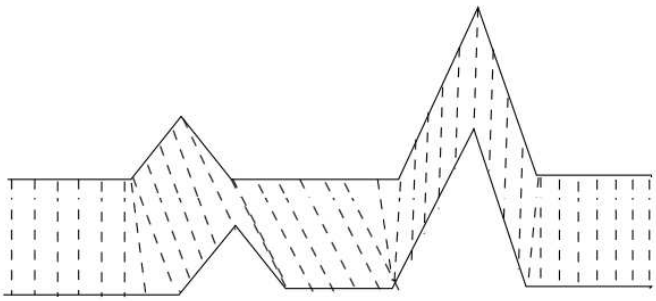

Figure 2. Dynamic time warping by Programminglinguist [CC BY-SA 4.0], via Wikimedia Commons. Two sequences (the solid lines) are matched (the dash lines) with certain rules.

Spatial alignment and normalization processes and geometrically matches points within a frame to address spatial difficulties, such as differences in physique or camera distance. The alignment and/or normalization are particularly essential when user performance is inferred from the position of human keypoints. Gu et al. [30] and Shi and Jiang [33] used a pelvis joint or a middle point to be the origin and normalized coordinates before calculating the Euclidean distance between a model and a user. On the other hand, Kamel et al. [28] only normalized coordinates to neutralize the difference in body size. They used rotations and direction of motions to evaluate the user performance, thus the alignment was not necessary.

Noise handling and smoothing processes data to minimize the effect of imperfections in pose estimation. Singh et al. [43] used SavGol filter [54] to minimize fluctuations of coordinates before detecting repetitions. [31] applied a median filter to angle sequences to prevent poor results due to noisy data. Jeon et al. [40] applied heatmap-smoothing, quarter-shift, and the one-euro filter to minimize fluctuation. Shi and Jiang [33] predicted undetected joint coordinates based on the standard movement of the limbs.

Selection chooses a subset of relevant features before processing, which could be either temporal or spatial features. Jan et al. [35] used only elbow, knee, and foot information for evaluating Tai-Chi Chuan practice. In contrast, [43] removed ankle, knee, and other points that have low variability when evaluating CrossFit workouts. Akiyama and Umezu [42] excluded redundant frames before and after a baseball pitching motion by selecting only 30 frames after the system detected a specific body part at an angle.

Temporal segmentation splits a sequence into sub-sequences. Singh et al. [43], for example, used peak detection methods to segment workout exercises into repetitions.

4.2.2. Measurement

Measurement refers to the quantification of attributes of one person. For instance, a system could identify the value of the angle between body parts or quantify the speed of a movement. Measurement could be briefly categorized into spatial and temporal ones.

Spatial measurement. The measurement involves the attributes of one person within one frame. Yang et al. [38], for example, calculated the horizontal distance between the hip and heel and the score of thigh length of a user to infer the proper form of a squat. The angle of body parts seems to be the most common spatial measurement [27,28,30,31,34,35,37–39]. Ranasinghe et al. [39], for example, calculated hip-shoulder-elbow angle to check for proper elbow lock during an arm curl exercise.

Temporal measurement. The measurement involves the attributes of one person across frames. A straightforward measurement is a difference in keypoint position over time. The direction of motion was used by Kamel et al. [28] in practicing Tai Chi. Similarly, Akiyama and Umezu [42] calculated displacement direction and distance of the body joint to provide suggestions for improving a baseball pitching form. Ranasinghe et al. [39] used the range of motion, i.e. the angular distance of the

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 6 June 2023 doi:10.20944/preprints202306.0395.v1

movement around a joint, to evaluate an arm curl exercise. Zhang et al. [26] incorporated the distance of the joint point between frames and the corresponding rotated angle as a curvature of the joint point combination movement for analyzing dance actions. Akiyama and Umezu [42] visualized the amount of joint movement between two successive frames as the indication of speed. They also mentioned the acceleration of joints and timing of the motion, but the detail of how those attributes were measured is unclear. A number of repetitions counted from a number of segments such as [43] can also be seen as a temporal measurement. However, a number of correct and/or incorrect repetitions would typically involve some assessment methods.

4.2.3. Assessment methods

Assessment methods assess how well users perform. The system could implement multiple methods or conditionally select methods, for example, based on the type of movement as seen in [34]. Note that some methods could be explained in multiple ways. We categorized the methods as the authors expressed them.

Mathematical formula or model evaluates user performance mathematically. It could be as simple as the summation or average of the absolute difference of measurement (e.g., angle [26,29] or displacement [26]). Variations of formulas have been adopted to find angular similarity/distance [34,40,42] as well as Euclidean distance [33,34]. Other mathematical formulas or models found include Gaussian function-based similarity metric [35] and the dot product between the joint dynamic vectors [40].

The mathematical formula or model typically results in a numerical value, which could be further classified into classes (e.g., good and bad) using a formula or other assessment methods. Yang et al. [38], for example, expressed the conditions of good angle and hip-heel distance in mathematical form.

Rule-based method evaluates or classifies user performance based on conditions. Zhang et al. [26], for example, graded user performance (excellent, good, pass, fail) based on the similarity score.

The rule-based method could also lead to a numerical result. Ranasinghe et al. [39] detected a number of correct arm curl repetitions when shoulder-elbow-wrist angles are less than 90 degrees, then larger than 170 degrees repeatedly. Wessa et al. [36] and [39] tracked the time taken to perform a specific action by identifying the starting and ending point according to conditions of user posture.

Machine learning leverages data to evaluate user performance. It is typically used for classifying user performance into a predefined class. The technique used ranges from well-known supervised learning, such as Support Vector Machines (SVM) or K-nearest neighbors, to deep learning. Kurose et al. [25] first created feature vectors using joint position coordinates and classified the vectors with Gaussian Mixture Model (GMM). They then used SVM to predict a tennis shot result based on the posture class, movement amount, and play area. Wang et al. [27] used SVM with radial basis kernel function (RBF) to classify good and bad poses of skiing. Chalvatzaki et al. [2] used gait parameters, such as stride length or gait speed, as a feature vector for an SVM classifier with classes from Performance Oriented Mobility Assessment [55]. They also used Neural Network based on LSTM units for human activity recognition and gait stability assessment. Wessa et al. [36] and Tarek et al. [32] used the Artificial Neural Network (ANN) of 3 layers to classify keypoints into correct and incorrect poses of kickboxing and yoga. Singh et al. [43] implemented and compared four time series classification methods, including 1-nearest neighbors dynamic time warping (1NN-DTW), ROCKET [56], Fully Convolutional Network (FCN), and Residual Network (ResNet).

In addition to classifying a pose into a class, machine learning could be used for calculating a numerical value. Zhou et al. [41], Park et al. [52] proposed a body part embedding model for motion similarity. They decomposed joint points into 5 body parts, then encoded motion classes, skeletons, and camera views of each body part as an embedding using multi-stage CNN. The similarity score was then calculated from the average cosine similarity between the embedding pairs.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 6 June 2023 doi:10.20944/preprints202306.0395.v1

4.2.4. Experiments

Studies related to movement assessment include its accuracy and/or performance. We found various ways to test how the proposed automated assessment conforms to the correct value or standard. One common way was to ask humans to put a label or value, such as a score or number of repetitions, on movement gather from themselves or other participants, then compared human annotation with the system, as seen in [32–34,40]. For user movement assessment that employed machine learning, labeled data could be also used for training. Zhou et al. [41] and Wang et al. [27] asked participants to perform the correct and incorrect movements, then split the data to train and test their classification. The ground truth could come from seen results, such as the hit distance of a baseball swing [37] or a tennis shot result [25], and the authors discussed the correlation between their result and the actual one [25,37]. Some articles used less formal methods for the study. Li and Pulivarthy [31] and Kamel et al. [28] compared the score of experienced and novice users, then inferred the effectiveness of their assessment module as the experienced users' scores were higher than the novice scores. Yang et al. [38] simply intentionally performed incorrect movement and presented the output. Lastly, we found only one article that used a publicly available dataset. Zhou et al. [41] used NTU RGB+D similarity annotations dataset to validate their results.

For the performance, Singh et al. [43] reported training time. [33] mentioned that their system satisfied realtime requirements, but no actual time was given.

4.3. Augmented Feedback Presentation

Augmented feedback (or extrinsic feedback or external feedback) refers to information about performance from others. In our scope, it is the feedback provided by the system to a user. We discuss the type of information, format, and timing and frequency of the augmented feedback. Table 4 provides examples of works that implemented each category of feedback. Note that some categories are excluded from the table due to the lack of examples, and one work could provide multiple pieces of feedback.

Table 4. Example of augmented feedback, classified by type of information and format. We note the timing Concurrent and Terminal as superscripts. We also note how the model and/or user are visualized: Model only, User only, Juxtaposition, Superposition, and Relationship encoding. The asterisk (*) indicates different media types are put, for example, in juxtaposition. NA means we could not find any example from the reviewed papers.

Category	Knowledge of Results	Knowledge of Performance	
		General	Error
Visual-verbal			
- Number	[40]^C, [30]^C, [35]^T, [37]^T, [28]^T, [41], [2]	NA	NA
- Word(s)	[42]^T, [37]^T	NA	NA
- Phase	NA	[32]^C,	[34]^C, [31]^T, [36], [38]
Visual-nonverbal			
- Video	NA	U: [40]^C, [37]^T, [26], [2]; U&R: [32]^C, [28]^T; J: [39]^C*, [36]*; J&R: [35]^T*; S: [29]^T*	M: [27]
- Image	NA	S&R: [42]^T	U: [43]^T; J: [39]^C*, [35]^T*, [36]*
- Animation	NA	U: [2], [26]; J: [30]^C; S: [29]^T*	J&R: [41]
- Other	[32]^C, [40]^C, [25]^T	NA	NA

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 6 June 2023 doi:10.20944/preprints202306.0395.v1

4.3.1. Type of information

Type of information involves what content of the feedback is communicated to a user. Literature classifies augmented feedback into two main types: knowledge of results and knowledge of performance.

Knowledge of results (KR) gives information about the outcome of the user's movement. Common feedback of this type includes an indication whether a user does a movement correctly [37] or a score relating to movement quality [2,28,35,37,40], which could be a result from assessing overall performance or individual aspects. Li et al. [37], for example, color-coded a list of body parts according to the goodness of the posture of each part. A number of repetitions, as seen in [30,40], could be another feedback that informs the successful movement.

Some knowledge of results is tied to the context of movement. Akiyama and Umezu [42], for example, compared the user's baseball swing timing with the model and displayed whether the user was faster, slower, or had similar timing. Kurose et al. [25] listed the characteristics of tennis shots for each prediction result (a score, a losing point, or a rally continuation).

Knowledge of performance (KP) gives information about the characteristics of the user's movement that lead to the result. Tarek et al. [32] informed whether and why the pose is correct or incorrect (e.g., a foot above kneecap for Yoga Tree Pose) by highlighting the related body parts over user video recordings. Video replay or live video of a user is a common approach to demonstrate user performance, and systems usually annotate them with information from pose estimation or movement assessment modules. For example, a system could show a user video, overlaid with its detected skeleton [2,37,40]. A system could highlight parts with low scores [41] or color-coded body parts according to similarity score [35], then overlaid on a user video. Other types of knowledge of performance include movement kinetics and kinematics, such as the amount of movement over time [42]

The information from an assessment module, as the annotation or other forms, could highlight correct aspects or error aspects of the performance. *Correct aspects* informs that a user is on track and encourages them to continue. *Error aspects* informs a user about their mistakes (*descriptive*) and/or directs how to correct the mistakes (*prescriptive*). Akiyama and Umezu [42] highlighted body parts that received minimum similarity score. They provided advice on form improvements by drawing arrows indicating the motion direction of the model on the user's image. Some systems, such as [27,31,33,36,38,39], suggested the correction or the model movement when the system detected the wrong movement.

In addition to displaying information from an assessment module, the system could have a dedicated module for feedback. Ranasinghe et al. [39] used a reinforcement learning model to provide correct and incorrect instruction images. [31] surveyed correction feedback from experts, then used a decision tree classifier with Gradient Boosting to classify the feedback and present it to a user.

4.3.2. Format

The format involves the ways the feedback is communicated to a user. We broadly classify format according to signs and channels [57] into audio-verbal (word heard), visual-verbal (word read), and visual-nonverbal. There could be a system that provides audio-nonverbal, such as music or sound effect, but we have not found such a system in this review.

Audio-verbal is the use of words to communicate feedback through speech. This approach is the closest to traditional motor learning, where a coach or trainer verbally gives a learner feedback. To translate an output to speech, Chalvatzaki et al. [2] used a TTS (text-to-speech) system to give verbal feedback to the user.

Visual-verbal is the use of words to communicate the feedback visually (e.g., on-screen). Quantitative information is generally presented as *Number*, as found in [28,35,37,40,41]. For qualitative information, a system could display a category as *simple words*, e.g., "faster" or "slower" [42], or "front

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 6 June 2023 doi:10.20944/preprints202306.0395.v1

arm angle" or "back arm angle" [37]). Lastly, a system could output *phases or sentences*, as seen in [31,34,36,38].

Visual-nonverbal does not use words to communicate the feedback. We found the use of *videos* (e.g., [2,34,37,40]), *images* (e.g., [26,35,36,39,43]), *2D animations of skeleton* (e.g., [26,30]), and *other graphic* (e.g., correct/incorrect symbols [32] or movement paths and colormap [25]). Some systems display multiple types of media. Jan et al. [35], for example, displayed a user video beside a 3D model of a trainer.

The system could display information about a user only, a model only, or both. Displaying user only could be used to provide KP feedback in general as well as highlight incorrect aspects of user movement. Systems, such as [2,37,40], visualized detect skeleton on a user video. Singh et al. [43] displayed frames from the user camera feed that was identified as incorrect. Displaying model only, on the other hand, is typically used to prescribe the correct movement. Wang et al. [27] offered advice to a user by sampling a related video clip with correct poses.

In case both model and user information are displayed, we could further consider visual comparison [58], as a user typically needs to identify differences or similarities of their movement in relation to a model in order to see errors and/or improve their movement. *Juxtaposition* places a user and a model in separated, nearby spaces. Juxtaposition could be side-by-side, as seen in [30,34,35,39], or picture-in-picture, as seen in [36]. *Superposition* places a user and a model in the same space. Akiyama and Umezu [42], for example, superimposed the user image on the model image. Nagarkoti et al. [29] displayed the user's video with their skeleton and overlaid the skeleton of a model on them when the error is above a threshold. Lastly, *relationship encoding* displays relationships between a user and a model, such as a score or classification results, directly. The relationship could appear as an annotation on the user's video, as seen in [32,35,41].

We also found the usage of the temporal component of a video to communicate the feedback that a user could perceive visually. Kamel et al. [28] provided motion replay, which would be suspended when the similarity score of a frame was lower than 50%, and the joints that have lower-than-average scores were highlighted with yellow.

A visual comparison could happen with a visual-verbal format. For example, a similarity score could be considered as relationship encoding since it represents a relationship between a model and a user.

4.3.3. Timing and frequency

Timing and frequency involve when and how often the feedback is communicated to a user.

Timing could be generally classified into concurrent and terminal. *Concurrent feedback* is the feedback given while a person is performing a skill or movement, as seen in [30,32,34,39]. *Terminal feedback* is the feedback given after a person has finished performing a skill or movement, as seen in [25,28,29,31,35,37,42,43].

Frequency of feedback could be on every practice trial or less. In case of reduced frequency, a system could provide feedback only when user performance meets a certain value (*performance-based bandwidths*). Nagarkoti et al. [29], for example, projected a model movement on a user only when their error is above a threshold. The frequency could be *self-selected*, or the feedback could be *summarized or averaged* after a number of practice trials. However, we have yet to find examples in this review.

4.3.4. Experiments

Ranasinghe et al. [39], which provided feedback based on the reinforcement learning model, reported the relevance and value of feedback on errors. However, most articles evaluated augmented feedback, specifically or as a whole system, through users. Straightforward ways included asking participants to use a system, then conducting interviews or using questionnaires to investigate system usability, usability issues, user preference, and/or user satisfaction [27,28,30,35]. Alternatively, user

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 6 June 2023 doi:10.20944/preprints202306.0395.v1

performance or improvement when using a system could be used to evaluate the overall system. A study could be short-term (within a day, e.g., [28,36,39,42]) or long-term (e.g., [28,30,36]).

5. Discussion

This section summarizes the findings of each module that we observed while reviewing the articles. We also suggest some future research directions for pose estimation-based physical movement applications.

5.1. Pose Estimation

Open-source libraries that support pose estimation in real-time plays an important role for research on pose estimation-based physical movement applications, as supported by the number of works that used OpenPose and PoseNet (see Table 2). The recent release of pose estimation libraries, such as AlphaPose [59], Detectron2 [60], and MediaPipe [61], could bring more researchers to work in this area.

Pose estimation libraries for a specific purpose may be needed, even though general-purpose pose estimation libraries are useful. Wang et al. [27], for example, proposed a pose estimation method that incorporates spatial and temporal relation of human keypoints to deal with the fast movement of ski technique and skiing skis. They trained the model, experimented with a dataset, and reported errors in pose estimation due to crossed snowboards that mixed with the background. As a physical movement application usually focused on one or a few specific types of movement, a pose estimation library that could be trained with a custom dataset could be useful for research in this area and possibly yield more accurate results.

Many challenges of pose estimation remains. There was a proposal to address well-known limitations of unusual poses and occlusion Yang et al. [38]. However, the challenges identified by prior works such as capture errors, positional errors, or limitations of recording devices Stenum et al. [6] were rarely addressed or discussed in the articles. This lack of discussion about pose estimation issues might be because they are not a focus of the works we reviewed. However, one possible reason is that the pose estimation result was good enough in their experiments and settings.

Pose estimation may not have to be perfect to be useful in physical movement applications. Though this point was not obvious from reviewed articles, the lack of discussion about pose estimation issues and our experience with a pose estimation-based physical movement application [24] could suggest so. We observed some inaccuracy of pose estimation results during the experiment, but our participants did not notice it because they focused on other parts. Thus, in addition to continue improving pose estimation accuracy, one future research is to study the effect of imperfect pose estimation in physical movement applications on users.

5.2. Movement Assessment

There is limited evaluation and discussion to tell which technique is better or which factors should be considered when applying a technique to a new context. For instance, differences in physique, camera angle, or camera distance could result in high Euclidean distance even if a user perfectly imitates a trainer. Though there could be articles outside our review that compare techniques, such as Srikaewsiew et al. [62], none of them was compared with proposed movement assessment methods. We thus encourage more experiments to evaluate and identify the limitations of movement assessment methods.

The lack of ready-to-use datasets could be one factor that hinders the experiment. Only one review article [41] used the public dataset, NTU RGB+D 120 Similarity Annotations, to evaluate their movement assessment module. The dataset was annotated with motion similarity score via Amazon Mechanical Turk [52]. In addition to the lack of movement quality data, the list of existing datasets provided by Zhou et al. [41] also showed that almost all datasets were gathered using Kinect, which may need some additional processing to be used with pose estimation-based applications. Similar to

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 6 June 2023 doi:10.20944/preprints202306.0395.v1

pose estimation that is generally evaluated using publicly available datasets, such as COCO [63] or MPII [64], there is a need for datasets for evaluating movement assessment.

5.3. Augmented Feedback Presentation

Relating results of automated feedback to a traditional one is difficult due to the lack of feedback description in a number of reviewed papers. We mainly adopt feedback classification from literature in motor learning Magill and Anderson [23]. The classification is well-established and includes research works where a human trainer gives feedback to a learner, which should allow us to compare research within our scope with others. However, we found difficulty in classifying feedback. For instance, Ranasinghe et al. [39] mentioned verbal encouragement phrases, but no detail about the channel was given. Chalvatzaki et al. [2] gave verbal feedback, but it is unclear what feedback information was given to users. We were unable to identify the timing of a number of works ([36], [41], [38], [2], [27], [26], [33]). This lack of description limits the replication, comparison, as well as contribution to general knowledge in motor learning. Thus, we encourage authors to always report content, format, and timing of feedback in their literature.

Researchers could experiment with different ways to provide augmented feedback in the design space. While it is not comprehensive, Table 4 can still give us an idea of trends and gaps that can be experimented with. We could observe more usage of a visual channel in communicating verbal feedback, compared to audio one. Different forms of visual objects, such as a number or a video, seem to be suitable for different types of content. Still, one could try to come up with items in NA cells or items excluded from the table. For instance, a number could be used to provide knowledge of performance by annotating angles between body parts on a user video. A system could provide audio feedback, correct aspects of knowledge of performance, kinetics, kinematics, and/or biofeedback.

Study about feedback prioritizing is still underexplored. In motor learning, selecting the skill component for knowledge of performance is recommended [23]. For instance, a trainer could tell a learner to correct arm movement first, then leg movement later. For automated feedback, one area that we found interesting to explore is the usage of artificial intelligence to provide feedback according to user needs, similar to [31,39]. Alternately, designing a user interface that allows a user to customize provided feedback could take less time to implement and easier to test the effect of priority feedback in an automated system.

Erroneous feedback should be considered and discussed, even though there is no apparent discussion about the challenges of providing augmented feedback from the review. Erroneous feedback is wrong feedback, which could hinder motor learning as a user would use the wrong feedback and ignore their (correct) sensory feedback [65]. Though the accuracy of pose estimation improves over years, it is not perfect and could lead to erroneous feedback. A system could, for example, disable augmented feedback when the confidence value of pose estimation is below a threshold. Finding a way to mitigate the erroneous feedback could be one essential future research direction for pose-estimation-based physical movement applications to be widely adopted.

5.4. Limitations

Our work has two limitations. First, there exist articles that addressed specific module(s), which may provide rich knowledge about the module, but were excluded from this review by our inclusion criteria or other reasons. Our prior works, for instance, either provided feedback with no movement assessment [24] or assess movement without providing feedback [62].

Second, as we discussed in the previous section, some articles provided very limited descriptions of the user interface and feedback. We thus used provided images as a source of information. It is possible that we could not study every aspect of the user interface. Also, it was sometimes unclear whether an image depicted concepts or actual implementation. Because the description was quite limited, we decided to consider those images as a way to communicate feedback to either users or readers and included them in our analysis.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 6 June 2023 doi:10.20944/preprints202306.0395.v1

6. Conclusion

The purpose of this study was to examine current methods for estimating human pose, assessing movement, and providing feedback using deep learning techniques. The review analyzed 20 articles on these topics. We found that systems used CNN (notably OpenPose) for pose estimation and used either mathematical formula or model, rule-based method, or machine learning for assessing movement. The feedback, including knowledge of result and knowledge of performance, was mostly presented visually in verbal forms (i.e., number, word, and phase) and nonverbal form (i.e., video, image, animation, and other graphics). Public datasets and human subjects were both used in the tests to assess each module.

We discuss the remaining gaps and areas for future research. We suggest a need for general-purpose pose estimation libraries as well as specialized ones. While there are still challenges to be addressed in pose estimation, it may not be necessary for it to be perfect to be useful in certain applications. For movement assessment, there is limited evaluation and discussion on which techniques are best and what factors should be taken into consideration when applying them to new contexts. Ready-to-use datasets should facilitate the evaluation. Still, the usefulness of pose estimation-based physical movement application in providing feedback is unclear. It is difficult to relate the results of automated feedback to traditional methods due to the lack of description. We suggest further study about feedback prioritization and erroneous feedback, which could possibly facilitate the adoption of pose-estimation-based physical movement applications.

Acknowledgments: This research is funded by the National Electronics and Computer Technology Center, Thailand.

References

1. A. Erol, G. Bebis, M. Nicolescu, R. D. Boyle, and X. Twombly, "Vision-based hand pose estimation: A review," *Computer Vision and Image Understanding*, vol. 108, no. 1-2, pp. 52–73, 2007.
2. G. Chalvatzaki, P. Koutras, A. Tsiami, C. S. Tzafestas, and P. Maragos, "i-Walk Intelligent Assessment System: Activity, Mobility, Intention, Communication," School of E.C.E., National Technical University of Athens, Athens, Greece, pp. 500–517, 2020.
3. H. S. S. Blas, A. S. Mendes, F. G. Encinas, L. A. Silva, and G. V. González, "A multi-agent system for data fusion techniques applied to the internet of things enabling physical rehabilitation monitoring," *Applied Sciences (Switzerland)*, vol. 11, no. 1, pp. 1–19, 2021.
4. H. Mousavi Hondori and M. Khademi, "A review on technical and clinical impact of microsoft kinect on physical therapy and rehabilitation," *Journal of medical engineering*, vol. 2014, 2014.
5. G. M. Difini, M. G. Martins, and J. L. V. Barbosa, "Human Pose Estimation for Training Assistance: A Systematic Literature Review," in *ACM International Conference Proceeding Series*, 2021, pp. 189–196.
6. J. Stenum, K. M. Cherry-Allen, C. O. Pyles, R. D. Reetzke, M. F. Vignos, and R. T. Roemmich, "Applications of pose estimation in human health and performance across the lifespan," *Sensors*, vol. 21, no. 21, p. 7315, 2021.
7. A. Badiola-Bengoa and A. Mendez-Zorrilla, "A systematic review of the application of camera-based human pose estimation in the field of sport and physical exercise," *Sensors*, vol. 21, no. 18, p. 5996, 2021.
8. A. Da Gama, P. Fallavollita, V. Teichrieb, and N. Navab, "Motor rehabilitation using kinect: a systematic review," *Games for health journal*, vol. 4, no. 2, pp. 123–135, 2015.
9. T. L. Munea, Y. Z. Jembre, H. T. Weldegebriel, L. Chen, C. Huang, and C. Yang, "The progress of human pose estimation: a survey and taxonomy of models applied in 2d human pose estimation," *IEEE Access*, vol. 8, pp. 133 330–133 348, 2020.
10. J. Wang, S. Tan, X. Zhen, S. Xu, F. Zheng, Z. He, and L. Shao, "Deep 3d human pose estimation: A review," *Computer Vision and Image Understanding*, vol. 210, p. 103225, 2021.
11. Y. Desmarais, D. Mottet, P. Slangen, and P. Montesinos, "A review of 3d human pose estimation algorithms for markerless motion capture," *Computer Vision and Image Understanding*, vol. 212, p. 103275, 2021.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 6 June 2023 doi:10.20944/preprints202306.0395.v1

12. M. B. Gamra and M. A. Akhloufi, "A review of deep learning techniques for 2d and 3d human pose estimation," *Image and Vision Computing*, vol. 114, p. 104282, 2021.
13. Y. Niu, J. She, and C. Xu, "A survey on imu-and-vision-based human pose estimation for rehabilitation," in *2022 41st Chinese Control Conference (CCC)*. IEEE, 2022, pp. 6410–6415.
14. B. Caramiaux, J. Françoise, W. Liu, T. Sanchez, and F. Bevilacqua, "Machine learning approaches for motor learning: A short review," *Frontiers in Computer Science*, vol. 2, 2020. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fcomp.2020.00016
15. A. Rajšp and I. Fister, "A systematic literature review of intelligent data analysis methods for smart sport training," *Applied Sciences (Switzerland)*, vol. 10, no. 9, p. 3013, 2020.
16. R. Gámez Díaz, Q. Yu, Y. Ding, F. Laamarti, and A. El Saddik, "Digital twin coaching for physical activities: A survey," *Sensors (Switzerland)*, vol. 20, no. 20, pp. 1–21, 2020.
17. K. M. Tsiouris, V. D. Tsakanikas, D. Gatsios, and D. I. Fotiadis, "A Review of Virtual Coaching Systems in Healthcare: Closing the Loop With Real-Time Feedback," *Frontiers in Digital Health*, vol. 2, p. 567502, 2020.
18. B. Lauber and M. Keller, "Improving motor performance: Selected aspects of augmented feedback in exercise and health," *European Journal of Sport Science*, vol. 14, no. 1, pp. 36–43, 2014.
19. Y. Zhou, W. De Shao, and L. Wang, "Effects of feedback on students' motor skill learning in physical education: A systematic review," *International Journal of Environmental Research and Public Health*, vol. 18, no. 12, p. 6281, 2021.
20. M. Mödinger, A. Woll, and I. Wagner, "Video-based visual feedback to enhance motor learning in physical education—a systematic review," *German Journal of Exercise and Sport Research*, pp. 1–14, 2021.
21. M. J. Page, J. E. McKenzie, P. M. Bossuyt, I. Boutron, T. C. Hoffmann, C. D. Mulrow, L. Shamseer, J. M. Tetzlaff, E. A. Akl, S. E. Brennan *et al.*, "The prisma 2020 statement: an updated guideline for reporting systematic reviews," *Systematic reviews*, vol. 10, no. 1, pp. 1–11, 2021.
22. C. Kohl, E. J. McIntosh, S. Unger, N. R. Haddaway, S. Kecke, J. Schiemann, and R. Wilhelm, "Online tools supporting the conduct and reporting of systematic reviews and systematic maps: a case study on cadima and review of existing tools," *Environmental Evidence*, vol. 7, no. 1, pp. 1–17, 2018.
23. R. Magill and D. Anderson, *Motor learning and control*. McGraw-Hill Publishing New York, 2010.
24. A. Tharatipyakul, K. T. Choo, and S. T. Perrault, "Pose estimation for facilitating movement learning from online videos," in *Proceedings of the International Conference on Advanced Visual Interfaces*, 2020, pp. 1–5.
25. R. Kurose, M. Hayashi, T. Ishii, and Y. Aoki, "Player pose analysis in tennis video based on pose estimation," in *2018 International Workshop on Advanced Image Technology, IWAIT 2018*, Matsudo Orthopedics Hospital, Chiba, Japan, 2018, pp. 1–4.
26. L. Zhang, X. Liang, W. Zhang, R. Tang, Y. Fan, Y. Nan, and R. Song, "Behavior Recognition on Multiple View Dimension," in *International Conference on Wavelet Analysis and Pattern Recognition*, vol. 2019-July, State Grid Henan Electric Power Company Kaifeng Power Supply Company, China, 2019.
27. J. Wang, K. Qiu, H. Peng, J. Fu, and J. Zhu, "AI Coach: Deep human pose estimation and analysis for personalized athletic training assistance," in *MM 2019 - Proceedings of the 27th ACM International Conference on Multimedia*, Microsoft Research Asia, China, 2019, pp. 2228–2230.
28. A. Kamel, B. Liu, P. Li, and B. Sheng, "An Investigation of 3D Human Pose Estimation for Learning Tai Chi: A Human Factor Perspective," *International Journal of Human-Computer Interaction*, vol. 35, no. 4-5, pp. 427–439, 2019.
29. A. Nagarkoti, R. Teotia, A. K. Mahale, and P. K. Das, "Realtime Indoor Workout Analysis Using Machine Learning Computer Vision," in *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, Samsung Research Institute, Bangalore, 560037, India, 2019, pp. 1440–1443.
30. Y. Gu, S. Pandit, E. Saraee, T. Nordahl, T. Ellis, and M. Betke, "Home-based physical therapy with an interactive computer vision system," in *Proceedings - 2019 International Conference on Computer Vision Workshop, ICCVW 2019*, Boston University, United States, 2019, pp. 2619–2628.
31. J. E. Li and H. Pulivarthy, "BalletNetTrainer: An Automatic Correctional Feedback Instructor for Ballet via Feature Angle Extraction and Machine Learning Techniques," in *Proceedings of the International Conference on Industrial Engineering and Operations Management*, ARQuest SSERN International, Kirkland, WA, United States, 2021, pp. 603–613.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 6 June 2023 doi:10.20944/preprints202306.0395.v1

32. O. Tarek, O. Magdy, and A. Atia, "Yoga Trainer for Beginners Via Machine Learning," in *Proceedings of the 2021 International Japan-Africa Conference on Electronics, Communications, and Computations, JAC-ECC 2021*, HCI-LAB, Faculty of Computers and Artificial Intelligence, Helwan University, Egypt, 2021, pp. 75–78.
33. D. Shi and X. Jiang, "Sport training action correction by using convolutional neural network," Xinzhou Teachers Univ, Xinzhou, Peoples R China AD, 2021.
34. X. Huang, D. Pan, Y. Huang, J. Deng, P. Zhu, P. Shi, R. Xu, Z. Qi, and J. He, "Intelligent Yoga Coaching System Based on Posture Recognition," in *Proceedings - 2021 International Conference on Culture-Oriented Science and Technology, ICCST 2021*, Communication University of China, School of Information and Communication Engineering, Beijing, China, 2021, pp. 290–293.
35. Y. F. Jan, K. W. Tseng, P. Y. Kao, and Y. P. Hung, "Augmented Tai-Chi Chuan Practice Tool with Pose Evaluation," in *Proceedings - 4th International Conference on Multimedia Information Processing and Retrieval, MIPR 2021*, Graduate Institute of Networking and Multimedia, National Taiwan University, Taipei, Taiwan, 2021, pp. 35–41.
36. E. Wessa, A. Ashraf, and A. Atia, "Can pose classification be used to teach Kickboxing?" in *International Conference on Electrical, Computer, and Energy Technologies, ICECET 2021*, Helwan University, HCI-LAB, Faculty of Computers and Artificial Intelligence, Egypt, 2021.
37. Y. C. Li, C. T. Chang, C. C. Cheng, and Y. L. Huang, "Baseball Swing Pose Estimation Using OpenPose," in *2021 IEEE International Conference on Robotics, Automation and Artificial Intelligence, RAAI 2021*, Physical Education Tunghai University, Taichung, Taiwan, 2021, pp. 6–9.
38. L. Yang, Y. Li, D. Zeng, and D. Wang, "Human Exercise Posture Analysis based on Pose Estimation," in *IEEE Advanced Information Technology, Electronic and Automation Control Conference (IAEAC)*, Sichuan Sports Industry Group Tanma Ai Technology Co. Ltd, Chengdu, China, 2021, pp. 1715–1719.
39. I. Ranasinghe, C. Yuan, R. Dantu, and M. V. Albert, "A Collaborative and Adaptive Feedback System for Physical Exercises," in *Proceedings - 2021 IEEE 7th International Conference on Collaboration and Internet Computing, CIC 2021*, University of North Texas, Computer Science and Engineering, Denton, TX, United States, 2021, pp. 11–15.
40. H. Jeon, D. Kim, and J. Kim, "Human Motion Assessment on Mobile Devices," in *International Conference on ICT Convergence*, vol. 2021-October, Electronics and Telecommunications Research Institute, Intelligent Robotics Research Division, Deajeon, South Korea, 2021, pp. 1655–1658.
41. J. Zhou, W. Feng, Q. Lei, X. Liu, Q. Zhong, Y. Wang, J. Jin, G. Gui, and W. Wang, "Skeleton-based Human Keypoints Detection and Action Similarity Assessment for Fitness Assistance," in *2021 6th International Conference on Signal and Image Processing, ICSIP 2021*, Ningbo University of Technology, Ningbo, China, 2021, pp. 304–310.
42. S. Akiyama and N. Umezu, "Similarity-based Form Visualization for Supporting Sports Instructions," in *LifeTech 2022 - 2022 IEEE 4th Global Conference on Life Sciences and Technologies*, Ibaraki University, Dept. Mechanical System Engineering, Hitachi, Japan, 2022, pp. 480–484.
43. A. Singh, B. T. Le, T. L. Nguyen, D. Whelan, M. O'Reilly, B. Caulfield, and G. Ifrim, "Interpretable Classification of Human Exercise Videos Through Pose Estimation and Multivariate Time Series Analysis," Output Sports Limited, NovaUCD, Dublin, Ireland, pp. 181–199, 2022.
44. Z. Cao, G. Hidalgo Martinez, T. Simon, S. Wei, and Y. A. Sheikh, "Openpose: Realtime multi-person 2d pose estimation using part affinity fields," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019.
45. Z. Cao, T. Simon, S.-E. Wei, and Y. Sheikh, "Realtime multi-person 2d pose estimation using part affinity fields," 2016. [Online]. Available: https://arxiv.org/abs/1611.08050
46. D. Tome, C. Russell, and L. Agapito, "Lifting from the deep: Convolutional 3d pose estimation from a single image," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 2500–2509.
47. TensorFlow, "PoseNet," 2019. [Online]. Available: https://github.com/tensorflow/tfjs-models/tree/master/posenet
48. M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 4510–4520.
49. B. Xiao, H. Wu, and Y. Wei, "Simple baselines for human pose estimation and tracking," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 466–481.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 6 June 2023 doi:10.20944/preprints202306.0395.v1

50. K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
51. D. Mehta, H. Rhodin, D. Casas, P. Fua, O. Sotnychenko, W. Xu, and C. Theobalt, "Monocular 3d human pose estimation in the wild using improved cnn supervision," in *2017 international conference on 3D vision (3DV)*. IEEE, 2017, pp. 506–516.
52. J. Park, S. Cho, D. Kim, O. Bailo, H. Park, S. Hong, and J. Park, "A Body Part Embedding Model with Datasets for Measuring 2D Human Motion Similarity," *IEEE Access*, vol. 9, pp. 36 547–36 558, 2021.
53. D. Berndt and J. Clifford, "Using dynamic time warping to find patterns in time series," in *Workshop on Knowledge Knowledge Discovery in Databases*, vol. 398, no. 16. Seattle, WA, USA:, 1994, pp. 359–370. [Online]. Available: http://www.aaai.org/Papers/Workshops/1994/WS-94-03/WS94-03-031.pdf
54. A. Savitzky and M. J. Golay, "Smoothing and Differentiation of Data by Simplified Least Squares Procedures," *Analytical Chemistry*, vol. 36, no. 8, pp. 1627–1639, 1964.
55. M. E. Tinetti, T. Franklin Williams, and R. Mayewski, "Fall risk index for elderly patients based on number of chronic disabilities," *The American Journal of Medicine*, vol. 80, no. 3, pp. 429–434, 1986.
56. A. Dempster, F. Petitjean, and G. I. Webb, "ROCKET: exceptionally fast and accurate time series classification using random convolutional kernels," *Data Mining and Knowledge Discovery*, vol. 34, no. 5, pp. 1454–1495, 2020.
57. P. Zabalbeascoa, "The nature of the audiovisual text and its parameters," *The didactics of audiovisual translation*, vol. 7, pp. 21–37, 2008.
58. M. Gleicher, D. Albers, R. Walker, I. Jusufi, C. D. Hansen, and J. C. Roberts, "Visual comparison for information visualization," *Information Visualization*, vol. 10, no. 4, pp. 289–309, 2011.
59. H.-S. Fang, J. Li, H. Tang, C. Xu, H. Zhu, Y. Xiu, Y.-L. Li, and C. Lu, "Alphapose: Whole-body regional multi-person pose estimation and tracking in real-time," 2022. [Online]. Available: https://arxiv.org/abs/2211.03375
60. Y. Wu, A. Kirillov, F. Massa, W.-Y. Lo, and R. Girshick, "Detectron2," https://github.com/facebookresearch/detectron2, 2019.
61. V. Bazarevsky and I. Grishchenko, "On-device, real-time body pose tracking with mediapipe blazepose," https://ai.googleblog.com/2020/08/on-device-real-time-body-pose-tracking.html, 2020, accessed: 2023-01-02.
62. T. Srikaewsiew, K. Khianchainat, A. Tharatipyakul, S. Pongnumkul, and S. Kanjanawattana, "A comparison of the instructor-trainee dance dataset using cosine similarity, euclidean distance, and angular difference," 2022, in press.
63. T.-Y. Lin, M. Maire, S. Belongie, L. Bourdev, R. Girshick, J. Hays, P. Perona, D. Ramanan, C. L. Zitnick, and P. Dollár, "Microsoft coco: Common objects in context," 2014. [Online]. Available: https://arxiv.org/abs/1405.0312
64. M. Andriluka, L. Pishchulin, P. Gehler, and B. Schiele, "2d human pose estimation: New benchmark and state of the art analysis," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2014.
65. M. J. Buekers, R. A. Magill, and K. G. Hall, "The effect of erroneous knowledge of results on skill acquisition when augmented information is redundant," *The Quarterly Journal of Experimental Psychology Section A*, vol. 44, no. 1, pp. 105–117, 1992.

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.