Applications and limitations of current markerless motion capture methods for clinical gait biomechanics (#68362)

First submission

Guidance from your Editor

Please submit by **20 Dec 2021** for the benefit of the authors (and your $200 publishing discount) .

Literature Review article
This is a Literature Review article, so the review criteria are slightly different.
Please write your review using the criteria outlined on the 'Structure and Criteria' page.

Author notes
Have you read the author notes on the guidance page?

Image check
Check that figures and images have not been inappropriately manipulated.

Privacy reminder: If uploading an annotated PDF, remove identifiable information to remain anonymous.

Files
Download and review all files from the materials page.

6 Figure file(s)

Structure and Criteria

Structure your review

The review form is divided into 5 sections. Please consider these when composing your review:

1. **BASIC REPORTING**
2. **STUDY DESIGN**
3. **VALIDITY OF THE FINDINGS**
4. General comments
5. Confidential notes to the editor

You can also annotate this PDF and upload it as part of your review

When ready submit online.

Editorial Criteria

Use these criteria points to structure your review. The full detailed editorial criteria is on your guidance page.

BASIC REPORTING

- Clear, unambiguous, professional English language used throughout.
- Intro & background to show context. Literature well referenced & relevant.
- Structure conforms to PeerJ standards, discipline norm, or improved for clarity.
- Is the review of broad and cross-disciplinary interest and within the scope of the journal?
- Has the field been reviewed recently? If so, is there a good reason for this review (different point of view, accessible to a different audience, etc.)?
- Does the Introduction adequately introduce the subject and make it clear who the audience is/what the motivation is?

STUDY DESIGN

- Article content is within the Aims and Scope of the journal.
- Rigorous investigation performed to a high technical & ethical standard.
- Methods described with sufficient detail & information to replicate.
- Is the Survey Methodology consistent with a comprehensive, unbiased coverage of the subject? If not, what is missing?
- Are sources adequately cited? Quoted or paraphrased as appropriate?
- Is the review organized logically into coherent paragraphs/subsections?

VALIDITY OF THE FINDINGS

- Impact and novelty not assessed. Meaningful replication encouraged where rationale & benefit to literature is clearly stated.
- Conclusions are well stated, linked to original research question & limited to
- Is there a well developed and supported argument that meets the goals set out in the Introduction?
- Does the Conclusion identify unresolved questions / gaps / future directions?

supporting results.

Standout reviewing tips

The best reviewers use these techniques

Tip	*Example*
Support criticisms with evidence from the text or from other sources	*Smith et al (J of Methodology, 2005, V3, pp 123) have shown that the analysis you use in Lines 241-250 is not the most appropriate for this situation. Please explain why you used this method.*
Give specific suggestions on how to improve the manuscript	*Your introduction needs more detail. I suggest that you improve the description at lines 57- 86 to provide more justification for your study (specifically, you should expand upon the knowledge gap being filled).*
Comment on language and grammar issues	*The English language should be improved to ensure that an international audience can clearly understand your text. Some examples where the language could be improved include lines 23, 77, 121, 128 – the current phrasing makes comprehension difficult. I suggest you have a colleague who is proficient in English and familiar with the subject matter review your manuscript, or contact a professional editing service.*
Organize by importance of the issues, and number your points	*1. Your most important issue* *2. The next most important item* *3. …* *4. The least important points*
Please provide constructive criticism, and avoid personal opinions	*I thank you for providing the raw data, however your supplemental files need more descriptive metadata identifiers to be useful to future readers. Although your results are compelling, the data analysis should be improved in the following ways: AA, BB, CC*
Comment on strengths (as well as weaknesses) of the manuscript	*I commend the authors for their extensive data set, compiled over many years of detailed fieldwork. In addition, the manuscript is clearly written in professional, unambiguous language. If there is a weakness, it is in the statistical analysis (as I have noted above) which should be*

improved upon before Acceptance.

Applications and limitations of current markerless motion capture methods for clinical gait biomechanics

Logan Wade Corresp., 1, 2 , **Laurie Needham** 1, 2 , **M. Polly McGuigan** 1, 2 , **James L.J. Bilzon** 1, 2, 3

1 Department for Health, University of Bath, Bath, United Kingdom

2 Centre for Analysis of Motion, Entertainment Research and Applications, University of Bath, Bath, United Kingdom

3 Centre for Sport Exercise and Osteoarthritis Research Versus Arthritis, University of Bath, Bath, United Kingdom

Corresponding Author: Logan Wade
Email address: lw2175@bath.ac.uk

Background: Markerless motion capture has the potential to perform movement analysis with reduced data collection and processing time compared to marker-based methods. This technology is now starting to be applied for clinical and rehabilitation applications and therefore it is crucial that users of these systems understand both their potential and limitations. This literature review aims to provide a comprehensive overview of the current state of markerless motion capture for both single camera and multi camera systems. Additionally, this review explores how practical applications of this technology are being applied in clinical and rehabilitation settings, and examines the future challenges and directions markerless research must explore to facilitate full integration of this technology within clinical biomechanics. **Methodology**: A scoping review is needed to examine this emerging broad body of literature and determine where gaps in knowledge exist, which is key to developing motion capture methods that are cost effective and practically relevant to clinicians, coaches and researchers around the world. Literature searches were performed to examine studies that report accuracy of markerless motion capture methods, explore current practical applications of markerless motion capture methods in clinical biomechanics and determine what gaps in the knowledge exist that a relevant to the future directions and limitations of this developing technology. **Results**: Markerless methods provide improved versatility of the data, enabling datasets to be re-analyzed using updated pose estimation algorithms and may even provide clinicians with the capability to collect data while patients are wearing normal clothing. While it appears that markerless temporospatial measures generally appear to be equivalent to marker-based motion capture, joint center locations and joint angles are not yet sufficiently accurate. Current pose estimation algorithms appear to be approaching similar error rates of marker-based motion capture. However, without comparison to a gold standard, such as bi-planar videoradiography, it is unknown how these two systems truly compare.

Conclusions: Current open-source pose estimation algorithms were never designed for biomechanical applications, therefore, datasets on which they have been trained are inconsistently and inaccurately labelled. Improvements to labelling of open-source training data will be a vital next step in the development of this technology.

Applications and limitations of current markerless motion capture methods for clinical gait biomechanics

Authors

Logan Wade [a,b], Laurie Needham [a,b], M. Polly McGuigan [a,b], James L. J. Bilzon [a,b,c]

Affiliation

a) Department for Health, University of Bath, Bath UK

b) Centre for Analysis of Motion, Entertainment Research and Applications, University of Bath, United Kingdom

c) Centre for Sport Exercise and Osteoarthritis Research Versus Arthritis, University of Bath, Bath UK

Corresponding Author

Logan Wade
University of Bath, 1 West, Office 5.111, Bath, BA2 7AY, United Kingdom
Email: lw2175@bath.ac.uk

Abstract

Background: Markerless motion capture has the potential to perform movement analysis with reduced data collection and processing time compared to marker-based methods. This technology is now starting to be applied for clinical and rehabilitation applications and therefore it is crucial that users of these systems understand both their potential and limitations. This literature review aims to provide a comprehensive overview of the current state of markerless motion capture for both single camera and multi camera systems. Additionally, this review explores how practical applications of this technology are being applied in clinical and rehabilitation settings, and examines the future challenges and directions markerless research must explore to facilitate full integration of this technology within clinical biomechanics. **Methodology**: A scoping review is needed to examine this emerging broad body of literature and determine where gaps in knowledge exist, which is key to developing motion capture methods that are cost effective and practically relevant to clinicians, coaches and researchers around the world. Literature searches were performed to examine studies that report accuracy of markerless motion capture methods, explore current practical applications of markerless motion capture methods in clinical biomechanics and determine what gaps in the knowledge exist that a relevant to the future directions and limitations of this developing technology. **Results**: Markerless methods provide improved versatility of the data, enabling datasets to be re-analyzed using updated pose estimation algorithms and may even provide clinicians with the capability to collect data while patients are wearing normal clothing. While it appears that markerless temporospatial measures generally appear to be equivalent to marker-based motion capture, joint center locations and joint angles are not yet sufficiently accurate. Current pose estimation algorithms appear to be approaching similar error rates of marker-based motion capture. However, without comparison to a gold standard, such as bi-planar videoradiography, it is unknown how these two systems truly compare. **Conclusions**: Current open-source pose estimation algorithms were never designed for biomechanical applications, therefore, datasets on which they have been trained are inconsistently and inaccurately labelled. Improvements to labelling of open-source training data will be a vital next step in the development of this technology.

are

or “the true accuracy of markerless systems remains unknown.”

Possibly include mention of the further work required to assess the accuracy of markerless motion capture against gold standard bi-planar videoradiography in this conclusion.

Keywords

Marker-based, deep learning, computer vision, pose estimation, clinical gait analysis, OpenPose, DeepLabCut,

Introduction

the

Movement analysis seeks to understand the cause of altered movement patterns, assisting with prevention, identification and rehabilitation of a wide array of diseases, disabilities and injuries (Astephen et al. 2008; Franklyn-Miller et al. 2017; Hausdorff et al. 2000; Heesen et al. 2008; King et al. 2018; Pavão et al. 2013; Salarian et al. 2004; Sawacha et al. 2012; Vergara et al. 2012). In modern medicine, early identification now plays a major role in combating disease progression, facilitating interventions using precise measurements of small changes in movement characteristics (Buckley et al. 2019; Noyes & Weinstock-Guttman 2013; Rudwaleit et al. 2005; Swash 1998). Movement analysis may also assist with injury prevention in athletes (Paterno et al. 2010), improve rehabilitation treatment and adherence (Knippenberg et al. 2017), and may inform surgical intervention methods to optimize outcomes and reduce additional surgeries and healthcare costs (Arnold & Delp 2005; Jalalian et al. 2013; Lofterød et al. 2007; Wren et al. 2009).

Traditional movement analysis commonly relies on patient self-reports, along with practitioner observations and visually assessed rating scales to diagnose, monitor and treat musculoskeletal diseases (Berg et al. 1992; Jenkinson et al. 1994; Zochling 2011). Unfortunately, these measures are often subjective and prone to error, as they are based on each individual's interpretation (Muro-de-la-Herran et al. 2014). Alternatively, video-based motion capture records and processes video images to identify limb location and orientation, enabling calculation of output variables such as temporospatial measures and joint angles. Describing the position and orientation or 'pose' of body segments in three-dimensions (3D) requires calculation of the limbs' translation (sagittal, frontal and transverse position, Figure 1) and rotation (flexion/extension, abduction/adduction, rotation about the longitudinal axis, Figure 1). These three translational and three rotational descriptions of a segment are commonly referred to as six degrees of freedom (DoF). The current gold standard for non-invasive, video-based motion capture, is bi-

maybe "multiple x-ray views" or "concurrent x-ray views"

planar radiovideography, which uses multi-view x-rays to capture video of bone movement (Kessler et al. 2019; Miranda et al. 2011). Software is used to outline the bones and recreate their three-dimensional structure (Kessler et al. 2019), enabling 3D joint center locations and angles to be extracted with high precision. However, even this method has joint center translational errors of 0.3 mm and rotational errors of 0.44 °(Miranda et al. 2011). Additionally, high costs, small capture volume (single joint) and exposure to radiation make clinical or sporting applications highly impractical.

Figure 1: This figure demonstrates the six degrees of freedom needed to describe position and orientation (pose) of the human body, with the red dot indicating the location (translation) of the segment center of mass and blue arrows indicating rotation in three planes. A) The reference standing posture, B) thigh segment adduction/abduction, C) thigh segment flexion/extension, D) thigh segment rotation about the longitudinal axis.

videoradiography de

Due to bi-planar radiovideography limitations, the De facto video-based motion capture method is marker-based motion capture, which identifies human poses using near-infrared cameras and reflective markers placed on the during calibration only

skin (Figure 2). Marker locations can be detected with sub-millimeter accuracy (Buckley et al. 2019; Topley & Richards 2020) and are used to identify location and orientation of body segments for calculation of joint positions and angles. However, marker-based motion capture has significant drawbacks, requiring a controlled environment

apostrophe their awareness of being

(Buckley et al. 2019; Chen et al. 2016) that may alter participants movements due to being aware they are under observation (Robles-García et al. 2015). Marker-based systems are cheaper to acquire and run compared to biplanar videoradiography, but are generally still too expensive for many clinical applications, as highly trained personnel are required to operate them (Simon 2004). Marker-based motion capture also suffers from human error when placing markers on the participant (Gorton et al. 2009), and marker placement is very time intensive which can be a significant barrier in clinical or sporting environments, particularly with specific population groups (Whittle 1996).

Figure 2: Optoelectronic motion capture. Left - markers placed on the subject. Right - view of the markers in 3D space.

While highly popular, marker-based motion capture is not a gold standard, despite often being treated as such. Comparisons of marker-based motion capture against bi-planar videoradiography reveal joint center position errors as high as 30 mm, with averages between 9-19 mm, and joint rotation errors as high as 14 °, with averages between 2.2-5.5 ° (Miranda et al. 2013). For all motion capture methods, rotation about the longitudinal axis (Figure 1D) produces the greatest errors of all rotational planes (Kessler et al. 2019; Miranda et al. 2013) as measurement devices placed on the skin (i.e., markers) are much closer to the axis of rotation, for example hip internal-external rotational errors are possibly as high as 21.8 ° (Fiorentino et al. 2017)).

Marker-based errors are partially due to an assumption that markers on the skin represent movement of the bone, leading to soft tissue artefact errors as muscle, fat and skin move beneath markers (Camomilla et al. 2017; Cappozzo et al. 1996; Peters et al. 2010; Reinschmidt et al. 1997). Compared to bi-planar videoradiography, errors for markers placed over shank soft tissue were 5-7 mm, while markers placed over bony landmarks on the foot were 3-5 mm (Kessler et al. 2019). Soft tissue errors for hip joint range of motion may be on average between 4-8 °

during walking, stair descent and rising from a chair. (D'Isidoro et al. 2020). Procedures such as filtering the marker data can help to reduce some of this soft tissue error (Camomilla et al. 2017; Peters et al. 2010), although without invasively attaching markers to bone this error cannot be eliminated (Benoit et al. 2006) and therefore soft tissue artefact will continue to limit the accuracy of marker-based methods.

There is a need for motion capture methods that are less time intensive, do not require specialist personnel, and

"are" in place of "may be" (there is a need for methods that are less impacted by errors…)

may be less impacted by errors associated with markers-based methods (e.g., soft tissue artefact). Markerless motion capture uses standard video to record movement without markers, often leveraging deep learning-based software to identify body segment positions and orientations (pose). However, this technology has been slow to transfer to biomechanics, likely due to the requirement of advanced coding skills and in-depth computer science knowledge. As such, researchers, clinicians and coaches using this technology need to be informed of the benefits and limitations of these methods. Currently, there are no reviews targeted at applications of markerless motion capture for clinical biomechanics and sports medicine, which we aim to resolve within this review. This scoping review is intended to inform clinical biomechanical researchers, clinicians and coaches of current markerless motion capture performance, explore how this technology can be used in real world applications and discuss future directions and limitations that need to be overcome for markerless systems to become viable for clinical, rehabilitation and sporting applications.

Survey Methodology

A scoping review is needed to examine this emerging broad body of literature and determine where gaps in knowledge exist (Munn et al. 2018), which is key to developing motion capture methods that are cost effective and practically relevant to clinicians, coaches and researchers around the world. Literature searches were performed to target studies that report accuracy of markerless motion capture methods compared to marker-based motion capture or manually labelled methods. Literature searches were then performed to target current practical applications of markerless motion capture methods in clinical biomechanics. Finally, examination of markerless motion capture literature was performed to determine what gaps in the knowledge exist and discuss future directions and limitations of this developing technology. Literature was obtained using Google Scholar and Scopus, which were surveyed using different combinations of the keywords 'markerless', 'motion capture', 'pose estimation', 'gait analysis', 'clinical biomechanics', 'accuracy', '2D' and '3D', without limits on publication date. Literature was also obtained from references lists of identified articles. **Could include date(s) of database searches.**

Markerless Motion Capture

Markerless motion capture uses standard video and often relies on deep learning-based software (pose estimation algorithms) to describe human posture for each individual image within the video, or videos for multiple cameras (Figure 3). Because pose estimation algorithms are not dependent on markers attached to the skin, soft tissue
artefact (for consistency)
deformation errors may be reduced compared to marker-based methods, although this is yet to be examined experimentally. Pose estimation algorithms can be applied to new or old videos, providing sufficient image resolution, and while marker-based methods are limited by the marker-set used during data collection, old markerless video data could be reprocessed with new pose estimation algorithms to improve accuracy or extract more in-depth measures. Accurate application of this technology could therefore facilitate streamlined monitoring of changes in disease progression (Kidziński et al. 2020), rehabilitation (Cronin et al. 2019; Natarajan et al. 2017), athletic training and competition (Evans et al. 2018), and injury prevention (Zhang et al. 2018).

Figure 3: Twenty-five keypoints detected using the OpenPose (Cao et al. 2018) applied to a single image.

Hardware

The two main types of camera hardware employ either depth cameras or standard video cameras and may be used
now the Microsoft Azure Kinect
in single or multi-camera systems. Depth cameras, such as the Microsoft Kinect, record standard video and additionally also record the distance between each pixel and the camera (depth). While depth cameras are relatively cheap and accessible, research has demonstrated large differences compared to marker-based methods (Dolatabadi et al. 2016; Mentiplay et al. 2015; Natarajan et al. 2017; Otte et al. 2016; Pantzar-Castilla et al. 2018; Rodrigues et al. 2019; Tanaka et al. 2018). Additionally, depth cameras have limitations on capture rate, capture

volume and data collection may require controlled lighting conditions (Clark et al. 2019; Sarbolandi et al. 2015). There have been several in-depth reviews of these systems (Clark et al. 2019; Garcia-Agundez et al. 2019; Knippenberg et al. 2017; Mousavi Hondori & Khademi 2014; Webster & Celik 2014) and while depth cameras are still an active area of research, this review will focus on single and multi-camera markerless systems that use standard video cameras, as these systems are relatively new and have recently started to be employed for clinical, rehabilitation and injury prevention applications.

Markerless motion capture using standard video hardware does have some limitations similar to marker-based systems, as the capture volume is still limited by the number of cameras and high-speed cameras require increased lighting demands. However, compared to marker-based systems that rely on infrared cameras, markerless motion capture is not limited by sunlight or multiple systems running simultaneously. Zoom lenses or high-resolution video can enable data collection from long distances and is currently being used during sporting competitions such as tennis (Hawk-Eye) and baseball (Kinatrax) to track the ball and players. Low-cost systems could employ webcams or smartphones to record video data, facilitating motion capture by clinicians and coaches in real world applications. Higher end multi-camera systems that record synchronized video at high frame rates may be used for collection of high precision data, akin to current marker-based motion capture laboratories. However, extracting meaningful information (joint centers) from recorded images using software is a very difficult task to perform with high accuracy

missing period

Software

Once video data is collected, software in the form of pose estimation algorithms are employed to detect and extract joint center locations. Pose estimation algorithms typically use machine learning techniques that allow them to recognize patterns associated with anatomical landmarks. These algorithms are 'trained' using large scale

this is not relevant

datasets that provide many examples of the points of interest. or even estimate the temperature and time of day. However, to a computer, video data is comprised of pixels that are essentially a grid of numbers, with each number in the grid describing color and brightness in a given video frame, which makes identifying keypoints a very difficult task. Training a pose estimation algorithm generally requires the creation of a dataset containing thousands of manually labelled keypoints (Figure 4) (Chen et al. 2020; Ionescu et al. 2014; Lin et al. 2014; Sigal et al. 2010). Deep learning-based pose estimation algorithms perform mathematical calculations on each image in the training data, using a layered network (Convolutional Neural Network) where the output of one layer becomes the input of the next layer (Figure 4), and may be many layers deep (Mathis et al. 2020b). In doing this, a pose estimation algorithm learns to identify keypoints (e.g., joint centers) as patterns of pixel color, gradient and texture from the training data. Distance between the manually labelled and estimated keypoint locations are then examined by an optimization method, which updates filters within each layer of the pose estimation algorithm to reduce the distance between keypoints (Figure 4). This process is repeated using the entire training dataset until improvements between each iteration become negligible (Figure 4). The pose estimation algorithm is then tested on new images and compared to manually labelled data or marker-based joint center locations to determine how well it performs on images it has never seen. As such, deep learning-based pose estimation will only ever be as good as the training data used.

Figure 4: Training a pose estimation algorithm. Stage One: Creation of manually labelled training dataset. Stage Two: Using the unlabeled images from stage one, the pose estimation algorithm estimates the desired keypoint locations (joint centers). Estimated keypoint locations are compared to the manually labelled training data to determine the distance is between the estimated keypoint and the manually labelled keypoint. The optimization method then adjusts filters within the layers of the algorithm to try to reduce this distance and new estimated keypoints are calculated. This process is repeated until improvements to the pose estimate4ion algorithm are negligible.

Two pose estimation algorithms that have become very popular for biomechanical applications are OpenPose (Cao et al. 2018) and DeepLabCut (Insafutdinov et al. 2016; Mathis et al. 2018). OpenPose is a powerful pose estimation algorithm that can track multiple people in an image and is very easy to use. DeepLabCut enables users to retrain/refine a pre-trained pose estimation algorithm by providing the algorithm with a subset of manually labelled images that are specific to the desired task (~200 images) (Mathis et al. 2018), which can be especially useful for uncommon movements (e.g., clinical gait or sporting movements). For an in-depth review of pose estimation algorithm designs, readers are directed to numerous alternative reviews (Chen et al. 2020; Colyer et al. 2018; Dang et al. 2019; Mathis et al. 2020a; Sarafianos et al. 2016; Voulodimos et al. 2018).

While marker-based motion capture relies heavily on hardware (markers physically placed on the skin) to extract segment poses (location and orientation), markerless motion capture relies on complex software to process the complicated image data obtained by standard video hardware (as explained above). Unfortunately, current pose most, not all estimation algorithms have generally been trained to only extract two points on each segment (proximal and distal joint center locations), whilst three keypoints are required to calculate 6DoF (e.g., proximal and distal end of a segment, and a third point placed somewhere else on the segment). Two keypoints can provide information about the sagittal and coronal planes (Figure 1B and C), while the third keypoint is needed to determine rotation about the segment's longitudinal axis (Figure 1D). Thus, markerless methods that only identify joint center locations are limited to 5DoF, which only enables examination of 2D planar joint angles. This may be overcome to some degree by combining 5DoF methods with musculoskeletal modelling to constrain the movement and estimate movement in 6DoF (Chen & Ramanan 2017; Gu et al. 2018), however, manually relabeling training data with an additional third keypoint location on each segment may produce improved results with less processing of the data (Needham et al. 2021b).

Markerless motion capture has been slow in transferring to biomechanics, primarily due to inaccuracy of detecting joint center locations (Harsted et al. 2019) and requiring knowledge of computer vision and advanced programming skills. In this review we have classified markerless motion capture into two broad categories: monocular markerless motion capture which uses a single camera, and multi-camera markerless motion capture which obtains video data from two or more synchronized cameras. Despite its previously outline outlined faults, marker-based motion capture has generally been used as the reference method when assessing accuracy of markerless motion capture, and this should be kept in mind when comparing results between systems.

Performance of Current Markerless Applications

Monocular Markerless Motion Capture

2D monocular markerless motion capture obtains joint center locations from a single image or video using 2D pose estimation algorithms (Figure 5), making it cost and space efficient. However, self-occlusion errors are a major issue, often causing joint center locations to be missing for one or more frames and contribute to instances where the opposite limb is incorrectly detected (e.g., right knee labelled as the left knee) (Serrancolí et al. 2020; Stenum et al. 2021). Similar to marker-based methods, obtaining biomechanically relevant 2D planar joint angles requires an assumption that the camera is perfectly aligned with frontal or sagittal plane movements (Stenum et al. 2021). If correctly aligned with the plane of action (1DoF), the pose estimation method detects the translational horizontal and vertical joint center coordinates (2DoF), which are then combined with coordinates of neighboring joints to calculate 2D rotational segment and joint angles (3DoF).

Figure 5: Markerless motion capture examples: 2D pose estimation from monocular motion capture (2D keypoints detected using OpenPose Cao et al. (2018), 3D pose estimation from monocular motion capture (adapted from Cheng et al. (2020) with license from the Association for the Advancement of Artificial Intelligence, Copyright © 20) and 3D pose estimation from multi-camera motion capture (adapted from Sigal et al. (2010) with permission from Springer Nature).

Three studies have examined 2D monocular applications (25 - 60 Hz) of DeepLabCut against manual labelling or marker-based methods for the leg closest to the camera (sagittal view), in underwater running (Cronin et al. 2019), countermovement jumping (Drazan et al. 2021) and walking in stroke survivors (Moro et al. 2020). Markerless joint center differences were 10-20 mm greater than marker-based motion capture, but no significant differences were found between methods for temporospatial and joint angle outcome measures during walking and underwater running, and therefore this method may be a suitable alternative to 2D marker-based motion capture (Cronin et al. 2019; Moro et al. 2020). Strong correlations were found for joint angles during countermovement jumping compared to marker-based methods, however this study had to perform a knee and hip correction based on marker-based results (5.6 ˚). Therefore, it is unknown if these systematic offsets would be applicable for future applications.

While not strictly monocular, Serrancolí et al. (2020) and Stenum et al. (2021) used two video cameras (25- 60 Hz), placed on either side of a person, to extract information of the side closest to each camera without occlusion errors during walking over-ground or cycling on an ergometer. During walking, temporal differences were on average within 1 frame and spatial differences were less than 1 cm, although maximum differences were as high has 20 cm (Stenum et al. 2021). For both studies, lower limb joint angle differences were 3-11 degrees greater than marker-based methods and thus are too large to detect small changes needed for real world applications. Both studies also required additional manual input to fix incorrectly detected joints (e.g., right knee labelled as the left knee) (Serrancolí et al. 2020; Stenum et al. 2021). Therefore, some 2D monocular methods may obtain temporospatial (DeepLabCut and OpenPose) and planar 2D joint angles (DeepLabCut) with accuracy similar to marker-based motion capture (Miranda et al. 2013), but this has only been examined for the side closest to the camera. 2D motion capture will likely have the most value in general clinical or rehabilitation environments, where data collection can be tightly controlled to reduce occlusion issues and decreasing data collection and processing time is paramount.

Obtaining 3D joint center locations from monocular markerless motion capture (Figure 5) seeks to estimate joint locations in 3D using a single camera (Mehta et al. 2017). However, because the participant may move in any direction (plane), entire limbs may be occluded for significant periods. Additionally, depth must be estimated from 2D video data to determine which joints are closer to the camera (Chen et al. 2020). Obscured 3D joint locations may be estimated using past or future un-occluded frames, or from the position of un-occluded neighboring joints in the current frame (Cheng et al. 2020; Cheng et al. 2019; Khan et al. 2020; Mehta et al. 2017; Moon et al. 2019; Yang et al. 2018). Alternatively, 2D monocular methods may be combined with musculoskeletal modelling (Chen & Ramanan 2017; Gu et al. 2018) or estimation of forces (Rempe et al. 2020), to restrict the limb position in 3D and assist with unnatural leaning angles towards or away from the camera (Rempe et al. 2020). Multi-camera marker-based motion capture can be used to help train pose estimation methods in making an educated guess about where a joint is most likely to be in 3D, however due to fundamental lack of data, this will only ever be an estimate. Finally, as mentioned earlier, current pose estimation methods generally only detect two points on a segment (proximal and distal joint center locations) (Cao et al. 2018; Mathis et al. 2018), which can only measure 5DoF. Thus, manually relabeling training data to detect a third point on each segment could improve estimates of 6DoF.

3D monocular joint center location differences compared to reference methods are generally 40-60 mm (Chen et al. 2020), with some algorithms producing 30-40 mm differences when specifically trained to overcome occlusion issues (Cheng et al. 2020; Cheng et al. 2019). 3D monocular ankle joint angle differences during walking are between -10 and 10 ° for normal walking with maximal differences of 30 ° compared to marker-based methods. Two studies have examined temporospatial measures (step length, walking speed and cadence) using 2D Monocular methods combined with projection mapping (Shin et al. 2021) or a 3D musculoskeletal model (Azhand et al. 2021), finding strong correlations when compared to the GAITRite pressure walkway (Azhand et al. 2021; Shin et al. 2021). Therefore, while temporospatial measures may have sufficient accuracy for real world applications, significant improvements to identification of joint center location and angle are needed. Applications of this method will likely also require the user to minimize instances where limbs are fully occluded (e.g. setting up the camera in the frontal plane) (Shin et al. 2021).

Multi-camera Markerless Motion Capture

Multi-camera markerless motion capture is a progression of 2D monocular methods that minimizes joint occlusion errors by employing multiple cameras (Figure 5). This method combines 2D pose estimation with an additional multi-camera reconstruction step to estimate 3D joint center locations (Nakano et al. 2020; Needham et al. 2021a; Slembrouck et al. 2020). Compared to monocular systems, multi-camera systems are more costly due to additional hardware and require more space, thus this method generally seeks to replicate the results obtained from current high-end marker-based systems (e.g., Qualisys/Vicon).

Several studies have examined multi-camera markerless systems using the OpenPose pose estimation algorithm (30-120 Hz), reporting joint center location differences ranging between 10-50 mm (Nakano et al. 2020; Slembrouck et al. 2020; Zago et al. 2020) and temporospatial differences of 15 mm compared to marker-based methods(Zago et al. 2020). Slower movements performed better, with walking joint center differences compared to marker-based methods of 10-30 mm, while faster jumping and throwing movements were 20-40 mm (Nakano et al. 2020), which may be exacerbated with slow video frame rates (Slembrouck et al. 2020; Zago et al. 2020). Manual adjustments were required when OpenPose incorrectly detected joints (e.g. detects left knee as the right) for one study (Nakano et al. 2020). Needham et al. (2021b) performed a recent comparison of OpenPose (Cao et al. 2018), DeepLabCut (Mathis et al. 2018) and a third pose estimation algorithm (AlphaPose (Fang et al. 2017)) using 9 video cameras and 15 marker-based cameras, both collecting at 200 Hz. Compared to marker-based methods, 3D lower limb joint center differences were smallest for OpenPose and AlphaPose at 16-34 mm during walking, 23-48 mm during running and 14-36 mm during jumping. It should be noted that they did not retrain models using DeepLabCut and instead used the DeepLabCut standard human pose estimation algorithm (Mathis et al. 2018). While these results are now approaching error rates of marker-based motion capture identified by Miranda and colleagues (Miranda et al. 2013), Needham and colleagues demonstrated that there were systematic differences for all markerless methods, with the largest systematic differences occurred at the hip. Their paper suggested this could be the product of poorly labelled open access datasets, which may limit detection of accurate and reliable joint center locations.

While previous studies have used open-source pose estimation algorithms and therefore may be considered as standalone experimental setups, commercial systems have been developed. Joint angles were compared between an 8 camera (50 Hz) Captury markerless system (Captury) and a 16 camera marker-based system, although Captury identifies the silhouette of a person instead of using deep learning to extract joint center locations (Harsted et al. 2019). The authors stated that planar joint angles could not be considered interchangeable between motion capture systems, with lower limb joint angle differences of 4-20 °. Another commercial system (SIMI Reality Motion Systems) recorded multiple movements with 8 cameras (100 Hz) and then was processed using Simi Motion software which detects markers placed on the skin and Simi Shape 3D software, a markerless software which uses silhouette-based tracking similar to Captury (Becker 2016). Standard deviations of lower limb joint angles were between 3-10 degrees with the markerless method compared to marker-based, and correlations for hip and ankle frontal and rotation planes were poor (0.26-0.51), indicating high variability of this system. Most recently, Thiea3D

Theia3D

markerless software (Theia Markerless Inc.) which uses a proprietary pose estimation algorithm was compared between an 8 camera markerless system (85 Hz) and a 7 camera marker-based system (85 Hz) (Kanko et al. 2021b; Kanko et al. 2021c). They reported no bias or statistical difference for walking spatial measures (e.g., step length, step width, velocity) and a small difference in temporal measures (e.g., swing time and double support time) (Kanko et al. 2021c). A follow-on study using the same data found average differences of 22-36 mm for joint centers and 2.6-11 degrees for flexion/extension and abduction/adduction, although rotation about the longitudinal axis differences were 6.9-13.2 degrees compared to marker-based methods (Kanko et al. 2021b). Importantly, the lower ranges of these translational and rotational differences are within errors identified by previous research (Fiorentino et al. 2017; Kessler et al. 2019; Miranda et al. 2013). These strong results appear to be due to Theia3D having labelled their own biomechanically applicable data set which identifies 51 keypoints on the body (Kanko et al. 2021b; Kanko et al. 2021c), compared to OpenPose which only identifies 25 points (Cao et al. 2018). However, Theia3D software is somewhat of a black box, as it is unknown exactly what keypoints are being used, how much the data is being smoothed or exactly how rotations are being computed. Now that markerless systems may be approaching the accuracy of marker-based methods which have known errors discussed previously, future examination of markerless accuracy will require comparison to a gold standard method such as bi-planar videoradiography (Miranda et al. 2013).

Practical applications

While markerless systems may still be considered in their infancy, there have been several studies that demonstrate markerless potential for clinical applications. DeepLabCut was used to extract walking sagittal 2D joint angles in stroke survivors, showing significant differences between the affected and unaffected side (Moro et al. 2020). Cunningham et al. (2019) examined 2D monocular segment angles of a multi-segmented trunk and head in young children with cerebral palsy to automate application of a clinical test. Baldewijns et al. (2016) measured walking speed recorded unobtrusively in patient’s homes using a webcam, demonstrating how markerless methods could provide continuous monitoring of patients as they go about their daily lives. Martinez et al. (2018) used a 2D monocular markerless system with OpenPose to examine walking cadence and automate calculation of an anomaly score for Parkinson's disease patients, providing clinicians with an unbiased general overview of patient disease progression. Finally, Shin et al. (2021) retrospectively analyzed monocular frontal videos of Parkinson's patients for temporospatial outcome measures (step length, walking velocity and turning time) (Shin et al. 2021). They demonstrated high correlations between subjective clinical gait tests and were able to detect minor gait disturbances unnoticed by the clinician.

In one significant clinical example, Kidziński et al. (2020) analyzed 2D outcomes of cerebral palsy gait collected from a single camera (30 Hz) between 1994 and 2015 (~1800 videos). OpenPose derived 2D joint centers were used as the input for a secondary deep learning-based neural network that predicted parameters of clinical relevance, such as walking speed, cadence and knee flexion angle. However, direct comparisons to marker-based methods could not be performed due to data collection methods and therefore, new test data collected simultaneously with marker-based motion capture is needed to examine the accuracy of their system. Nevertheless, this study compiled outcome measures into a gait report that was automatically generated for the clinician, providing strong rationale for the future of clinical biomechanics and its ability to analyze gait in a cost and time efficient manner. Furthermore, the applications by Kidziński et al. (2020) and Shin et al. (2021) highlight the value of markerless motion capture to extract new information from old datasets. Without the need to place markers on participants or manually process results, quantitatively tracking patients throughout disease progression and rehabilitation becomes a much more viable option.

While some markerless systems may be approaching the accuracy of marker-based methods, some applications may not need highly accurate data and instead, numerous trials (e.g., numerous walking strides) could be averaged to obtain reliable average results (Pantzar-Castilla et al. 2018). Unfortunately, this approach may be unable to detect small changes over time and it is not always be possible to collect many trials in a clinical, rehabilitation or sport setting. Alternatively, using markerless motion capture as a motivational tool to perform rehabilitation exercises does not require highly accurate results. Markerless motion capture can be used to control a game or move around a virtual environment, which can increase adherence and motivation to perform repetitive or potentially painful rehabilitation exercises (Knippenberg et al. 2017; Vonstad et al. 2020). This could lead to improved rehabilitation methods, as interaction with virtual environments has also been shown to reduce pain felt by patients (Gupta et al. 2017; Scapin et al. 2018). While this application has been used with depth cameras (e.g., Microsoft Kinect) (Chanpimol et al. 2017; Knippenberg et al. 2017), current applications using standard cameras and pose estimation algorithms are limited (Clark et al. 2019).

Future challenges and applications

Clothing

Currently, markerless systems are assessed while participants wear tight fitting clothing, as marker-based motion capture cannot be used with normal/baggy clothing. However, normal clothing is often loose fitting and may change shape during movement, which may or may not impact a pose estimation algorithms ability to accurately extract joint center locations (Sarafianos et al. 2016). If markerless systems are resistant to this issue, it could greatly improve data collection in clinical and real-world applications. Using 8 cameras (60 Hz) with Theia3D's pose estimation, inter-trial and inter-session joint angle variability during walking was examined compared to previously reported marker-based results (Kanko et al. 2021a). Participants wore their own clothing which generally consisted of shoes, long trousers, shirt and sweater. Markerless inter-trial joint angle variability was on average 2.5 °, compared to 1.0 ° from marker-based methods (Kanko et al. 2021a; Schwartz et al. 2004), while markerless inter-session variability was on average 2.8 ° compared to 3.1 ° for marker-based methods (Kanko et al. 2021a; Schwartz

remove "of"

et al. 2004). Therefore, markerless joint angle variability of may be similar to marker-based data collected on multiple days (inter-session). Testing across multiple days or changes of clothing had no impact on the overall variability of the markerless system. However, the higher inter-trial variability suggests that markerless methods do produce greater errors during the same session. Unfortunately, because they did not examine marker-based walking variability of their participants, it is unknown if variability from previous marker-based studies was identical to the participants included within this study. Importantly, markerless data collection was able to be completed in 5-10 minutes, demonstrating the benefits of this system for applications where time is limited (Kanko et al. 2021a). Based on these results, markerless systems could one day collect data on patients at home during daily life, without the need of an operator or tight-fitting clothing. Such systems could also be set up in common areas of care homes, **that** facilitating data collection of numerous patients in an environment this is less likely to alter their gait (Robles-García et al. 2015). Additionally, applications that do not require high accuracy will likely cope better with loose clothing.

or, "Additionally, some current systems' results may be sufficient for applications that do not require high accuracy."

Diversity of human shapes and movements

While pose estimation algorithms are good at identifying keypoints from images they have been trained on, they can be poor at generalizing to identify keypoints in images that differ substantially from the training dataset (Cronin 2021; Mathis et al. 2020b; Seethapathi et al. 2019). Image databases (Chen et al. 2020; Ionescu et al. 2014; Lin et al. 2014; Sigal et al. 2010) may be biased towards humans of a certain race or a specific type of movement, and therefore, pose estimation algorithm performance may decrease when movements and people do not have sufficient representation (e.g., gymnastic movements (Seethapathi et al. 2019)). Manually labelled training datasets need to be diverse to account for varied movements of daily life (e.g., walking, standing from a chair, picking up objects), sporting movements (e.g., figure skating, gymnastics and weightlifting) and clinical movements (e.g., neurological disorders and amputations), visual differences of participants (e.g., age, race, anthropometrics) and visual differences of markerless setups (e.g., lighting levels, scale of participant, camera angle). Because current pose estimation algorithms are trained to label each image in a video independently, they may perform well at detecting keypoints of patients with pathological gait abnormalities such as cerebral palsy and stroke, while physical abnormalities such as amputations will likely present a more difficult challenge. Clinical datasets could be collectively sourced from clinical research studies worldwide, however as standard video will be used to collect data, challenges in the form of patient confidentiality and ethical considerations must be overcome at the ethical application stage to achieve this.

Shortcomings of current training datasets

Currently available open-source training datasets were never designed with biomechanical applications in mind. While these datasets encompass millions of images and thousands of manually labelled poses [55-58, 102], only a subset of major joint centers have been labelled (ankle, knee, hip, shoulder, etc.), which increases errors as major joints are treated as a rigid segment (Zelik & Honert 2018). For example, when walking with a fixed ankle/toe orthosis, markerless ankle joint angle (OpenPose) differences compared to marker-based methods were reduced relative to normal walking, as toe flexion was not accounted for in normal walking by the markerless algorithm (Takeda et al. 2020). Additionally, open-source pose estimation algorithms that only detect joint centers struggle to identify more than 5DoF, as detecting rotation about the longitudinal axis requires three points on a segment.

Open-source manually labelled pose estimation training datasets (Andriluka et al. 2014; Chen et al. 2020; Ionescu et al. 2014; Lin et al. 2014; Sigal et al. 2010) have recruited labelers from the general population who likely do not possess anatomical knowledge. As such, these datasets have not been labelled with the accuracy required for biomechanical applications, leading to errors in joint center locations and angles (Needham et al. 2021b). Furthermore, joints such as the hip or shoulder may appear very different from the side compared to a frontal or 45 °angle (Figure 6). Evidence of this can be seen in the systematic offset of joint center locations and segment lengths outlined by Needham et al. (2021b). Furthermore, open-source labelled datasets generally do not require all images to pass a second verification step, therefore two people may have very different interpretations of a joint center, which may lead to inconsistency in the labelled images (Cronin 2021). It is unwise to expect pose estimation algorithms to match marker-based methods when the labelled data they are trained on is fundamentally flawed. Several commercial companies have created their own propriety datasets (Kanko et al. 2021b; Kanko et al. 2021c), with Theia3D employing trained labelers who likely have anatomical knowledge, labelling multiple points on each segment and integrating a verification step by an expert labeler (Kanko et al. 2021c). This two-step labelling process may produce a more biomechanically accurate dataset, enabling the strong results discussed previously (Kanko et al. 2021a; Kanko et al. 2021b; Kanko et al. 2021c).

Large open-source datasets have labelled keypoints even when joints are occluded (Figure 6). This is a requirement for entertainment applications as it would be unacceptable for limbs to suddenly go missing in video games or virtual reality. However, this results in occluded joints being labelled onto points that are biomechanically incorrect (Lin et al. 2014), for example, the right knee may be occluded by the left leg and thus labelled as being located somewhere on the left thigh. This results in two potential issues, firstly, the labeler must guess the location of the occluded joint, which reduces the accuracy of the dataset and secondly, the algorithm may learn that it is possible for joints to appear on locations that are biomechanically incorrect (Cronin 2021). Finally, Seethapathi et al. (2019) highlighted that training and testing datasets often do not include temporal information (sequentially labelled images) and therefore current pose estimation algorithms can vary wildly in estimation of joint center locations between consecutive frames. It is possible to reduce these differences using Kalman filtering (Needham et al. 2021a) and therefore, improving current open-source data sets (e.g., COCO (Lin et al. 2014)) may be a more viable solution to improving accurate detection of joint center locations. New open-source datasets for biomechanical applications should include at least three points for each body segment, are labelled by trained labelers who possess anatomical and biomechanical knowledge, include a verification step by a secondary subset of expert users and additionally ignore or account for occluded joints.

Figure 6: Keypoints of large open-source datasets have been labelled to estimate occluded joint centers, however this requires users to guess where these locations are, as they are not visible (adapted from (COCO 2021; Lin et al. 2014) under creative commons license by (COCO 2020)).

Evaluation

Current publicly available video datasets with synchronized marker-based motion capture, often use limited or sub-optimal marker placements, have low frame rates and camera resolution and thus may result in overestimating differences between systems compared to when run on private higher quality datasets (Colyer et al. 2018; Corazza et al. 2009). Publicly available, highspeed, high resolution evaluation datasets are needed for true comparisons between markerless and marker-based motion capture. While Needham and colleagues (Needham et al. 2021b) demonstrated that OpenPose had a greater difference on average between 16-48 mm, joint center location differences could be as high as 80 mm or even higher for some joints during running. Examining not only the accuracy, but the reliability of a system to accurately measure joint center locations is crucial, as systems are beginning to obtain average results that rival marker-based methods. However, we also need to question whether improving markerless motion capture to align closer to marker-based motion capture is the best solution. Marker-based motion capture has inherent errors discussed previously and markerless motion capture may potentially out-perform outcompete marker-based methods in some areas (e.g., soft tissue artefact). As such, markerless methods that reach a similar level of accuracy to marker-based methods next need to be assessed with bi-planar videoradiography radiovideography or similarly accurate methods, to determine the true accuracy and reliability of these methods.

Decision making

Previous work has demonstrated the potential for markerless systems to automatically process video data and report quantitative results that could be immediately used by a clinician (Kidziński et al. 2020; Martinez et al. 2018). While pose estimation algorithms are learning to detect human poses, they are not able to think on their own. Desired outcome measures (e.g., temporospatial measures and joint angles) extracted using pose estimation algorithms are still decided by humans. Emerging applications of markerless motion capture are therefore likely to require outcome measures to be chosen by the user prior to data collection, after which the markerless system will collect and process the data, similar to current implementations of commercial IMU systems (i.e., Mobility Lab ADPM Inc.). As such, the clinician is still needed to interpret the results and their applicability to the patient. Deep learning methods could potentially be applied to this problem in the future (Simon 2004), however, speculating on how this would be achieved is beyond the scope of this review.

Usability

Current applications of open-source pose estimation algorithms require in-depth knowledge of deep learning-based neural networks and computer vision methods. As such, this technology requires usability improvements for users who do not have programming or computer science backgrounds. Some commercial systems such as Theia3D have made their software highly accessible by enabling data to be collected and processed with leading video-based motion capture companies (e.g., Qualisys and Vicon). However, because they have a proprietary dataset and pose estimation algorithm, it is not possible for a user to determine what keypoints their algorithm is extracting, nor how the raw pose estimation data is being filtered and processed.

While previous pose estimation algorithms have required substantial processing power housed in high end computers, new pose estimation algorithms can run on standard computers with modest graphical processing units (Cao et al. 2018) or even smaller devices such as mobile phones (Bazarevsky et al. 2020). As pose estimation software develops, it will become more feasible to integrate both the phone camera and processor to provide compact and affordable markerless motion capture (Steinert et al. 2020). Alternatively, cloud-based computing could be harnessed to record video using a smartphone, which is then uploaded to a server for processing and results are returned to the user (Zhang et al. 2021). Clinicians, researchers and coaches could one day perform automatic markerless motion capture in real time, without large setup costs. Finally, pose estimation algorithms have the potential to be used with cameras that move freely during data collection (Elhayek et al. 2015), which could allow accurate examination of how patients move through the natural environment.

Conclusion

Markerless motion capture has the potential to perform movement analysis with decreased data collection and processing time compared to marker-based methods. Furthermore, markerless methods provide improved versatility of the data, enabling datasets to be re-analyzed using updated pose estimation algorithms and may even provide clinicians with the capability to collect data while patients are wearing normal clothing. While it appears that markerless temporospatial measures generally appear to be equivalent to marker-based motion capture, joint center locations and joint angles are not yet sufficiently accurate. Current pose estimation algorithms appear to be approaching similar error rates of marker-based motion capture. However, without comparison to a gold standard,

or “the true accuracy of these markerless systems is unknown.”

such as bi-planar videoradiography, it is unknown how these two systems truly compare. Current open-source pose estimation algorithms were never designed for biomechanical applications, therefore, datasets on which they have been trained are inconsistently and inaccurately labelled. Improvements to labelling of open-source training data will be a vital next step in the development of this technology.

Competing Interests statement

Logan Wade, Laurie Needham, M. Polly McGuigan, and James L. J. Bilzon declare they have no competing interests.

Acknowledgements

This work was funded by CAMERA, the RCUK Centre for the Analysis of Motion, Entertainment Research and Applications, Bath, United Kingdom [EP/M023281/1 and EP/T014865/1].

Author Contributions

All authors have made substantial contributions to the conception and design of the review, drafting and revising the article critically for important intellectual content and have approved the final version for submission.

Figure Captions

Figure 1: This figure demonstrates the six degrees of freedom needed to describe position and orientation (pose) of the human body, with the red dot indicating the location (translation) of the segment center of mass and blue arrows indicating rotation in three planes. A) The reference standing posture, B) thigh segment adduction/abduction, C) thigh segment flexion/extension, D) thigh segment rotation about the longitudinal axis.

Figure 2: Optoelectronic motion capture. Left - markers placed on the subject. Right - view of the markers in 3D space.

Figure 3: Twenty-five keypoints detected using the OpenPose (Cao et al. 2018) applied to a single image.

Figure 4: Training a pose estimation algorithm. Stage One: Creation of manually labelled training dataset. Stage Two: Using the unlabeled images from stage one, the pose estimation algorithm estimates the desired keypoint locations (joint centers). Estimated keypoint locations are compared to the manually labelled training data to determine the distance is between the estimated keypoint and the manually labelled keypoint. The optimization method then adjusts filters within the layers of the algorithm to try to reduce this distance and new estimated keypoints are calculated. This process is repeated until improvements to the pose estimate4ion algorithm are negligible.

Figure 5: Markerless motion capture examples: 2D pose estimation from monocular motion capture (2D keypoints detected using OpenPose Cao et al. (2018), 3D pose estimation from monocular motion capture (adapted from Cheng et al. (2020) with license from the Association for the Advancement of Artificial Intelligence, Copyright © 20) and 3D pose estimation from multi-camera motion capture (adapted from Sigal et al. (2010) with permission from Springer Nature).

Figure 6: Keypoints of large open-source datasets have been labelled to estimate occluded joint centers, however this requires users to guess where these locations are, as they are not visible (adapted from (COCO 2021; Lin et al. 2014) under creative commons license by (COCO 2020)).

References

Andriluka M, Pishchulin L, Gehler P, and Schiele B. 2014. 2d human pose estimation: New benchmark and state of the art analysis. Proceedings of the IEEE Conference on computer Vision and Pattern Recognition. p 3686-3693.

Arnold AS, and Delp SL. 2005. Computer modeling of gait abnormalities in cerebral palsy: application to treatment planning. *Theoretical Issues in Ergonomics Science* 6:305-312. 10.1080/14639220412331329636

Astephen JL, Deluzio KJ, Caldwell GE, and Dunbar MJ. 2008. Biomechanical changes at the hip, knee, and ankle joints during gait are associated with knee osteoarthritis severity. *Journal of Orthopaedic Research* 26:332-341. 10.1002/jor.20496

Azhand A, Rabe S, Müller S, Sattler I, and Heimann-Steinert A. 2021. Algorithm based on one monocular video delivers highly valid and reliable gait parameters. *Scientific Reports* 11:14065. 10.1038/s41598-021-93530-z

Baldewijns G, Claes V, Debard G, Mertens M, Devriendt E, Milisen K, Tournoy J, Croonenborghs T, and Vanrumste B. 2016. Automated in-home gait transfer time analysis using video cameras. *Journal of Ambient Intelligence and Smart Environments* 8:273-286. 10.3233/AIS-160379

Bazarevsky V, Grishchenko I, Raveendran K, Zhu T, Zhang F, and Grundmann M. 2020. BlazePose: On-device Real-time Body Pose tracking. *arXiv preprint arXiv:200610204*.

Becker L. 2016. Evaluation of joint angle accuracy using markerless silhouette-based tracking and hybrid tracking against traditional marker tracking Masters. Otto-von-Guericke-University.

Benoit DL, Ramsey DK, Lamontagne M, Xu L, Wretenberg P, and Renström P. 2006. Effect of skin movement artifact on knee kinematics during gait and cutting motions measured in vivo. *Gait & Posture* 24:152-164. https://doi.org/10.1016/j.gaitpost.2005.04.012

Berg KO, Maki BE, Williams JI, Holliday PJ, and Wood-Dauphinee SL. 1992. Clinical and laboratory measures of postural balance in an elderly population. *Archives of physical medicine and rehabilitation* 73:1073-1080. 10.5555/uri:pii:000399939290174U

Buckley C, Alcock L, McArdle R, Ur Rehman RZ, Del Din S, Mazzà C, Yarnall AJ, and Rochester L. 2019. The role of movement analysis in diagnosing and monitoring neurodegenerative conditions: Insights from gait and postural control. *Brain Sciences* 9. 10.3390/brainsci9020034

Camomilla V, Bonci T, and Cappozzo A. 2017. Soft tissue displacement over pelvic anatomical landmarks during 3-D hip movements. *Journal of Biomechanics* 62:14-20. https://doi.org/10.1016/j.jbiomech.2017.01.013

Cao Z, Hidalgo G, Simon T, Wei S-E, and Sheikh YJapa. 2018. OpenPose: realtime multi-person 2D pose estimation using Part Affinity Fields. *arXivorg (Preprint)*.

Cappozzo A, Catani F, Leardini A, Benedetti MG, and Della Croce U. 1996. Position and orientation in space of bones during movement: experimental artefacts. *Clinical Biomechanics* 11:90-100. https://doi.org/10.1016/0268-0033(95)00046-1

Chanpimol S, Seamon B, Hernandez H, Harris-Love M, and Blackman MR. 2017. Using Xbox kinect motion capture technology to improve clinical rehabilitation outcomes for balance and cardiovascular health in an individual with chronic TBI. *Archives of Physiotherapy* 7:6. 10.1186/s40945-017-0033-9

Chen C-H, and Ramanan D. 2017. 3d human pose estimation= 2d pose estimation+ matching. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. p 7035-7043.

Chen S, Lach J, Lo B, and Yang GZ. 2016. Toward Pervasive Gait Analysis With Wearable Sensors: A Systematic Review. *IEEE Journal of Biomedical and Health Informatics* 20:1521-1537. 10.1109/JBHI.2016.2608720

Chen Y, Tian Y, and He M. 2020. Monocular human pose estimation: A survey of deep learning-based methods. *Computer Vision and Image Understanding* 192:102897. https://doi.org/10.1016/j.cviu.2019.102897

Cheng Y, Yang B, Wang B, and Tan RT. 2020. 3D Human Pose Estimation Using Spatio-Temporal Networks with Explicit Occlusion Training. *Proceedings of the AAAI Conference on Artificial Intelligence* 34:10631-10638. 10.1609/aaai.v34i07.6689

Cheng Y, Yang B, Wang B, Yan W, and Tan RT. 2019. Occlusion-aware networks for 3d human pose estimation in video. Proceedings of the IEEE International Conference on Computer Vision. p 723-732.

Clark RA, Mentiplay BF, Hough E, and Pua YH. 2019. Three-dimensional cameras and skeleton pose tracking for physical function assessment: A review of uses, validity, current developments and Kinect alternatives. *Gait & Posture* 68:193-200. https://doi.org/10.1016/j.gaitpost.2018.11.029

COCO. 2020. COCO 2020 Keypoint Detection Task. https://creativecommons.org/licenses/by/4.0/legalcode.

COCO. 2021. Common Objects In Context: 2020 KeyPoint Detection Task. *Available at* *https://cocodataset.org/#keypoints-2020* (accessed 15/05/2021.

Colyer SL, Evans M, Cosker DP, and Salo AIT. 2018. A Review of the Evolution of Vision-Based Motion Analysis and the Integration of Advanced Computer Vision Methods Towards Developing a Markerless System. *Sports Medicine - Open* 4:24. 10.1186/s40798-018-0139-y

Corazza S, Mündermann L, Gambaretto E, Ferrigno G, and Andriacchi TP. 2009. Markerless Motion Capture through Visual Hull, Articulated ICP and Subject Specific Model Generation. *International Journal of Computer Vision* 87:156. 10.1007/s11263-009-0284-3

Cronin NJ. 2021. Using deep neural networks for kinematic analysis: challenges and opportunities. *Journal of Biomechanics*:110460. https://doi.org/10.1016/j.jbiomech.2021.110460

Cronin NJ, Rantalainen T, Ahtiainen JP, Hynynen E, and Waller B. 2019. Markerless 2D kinematic analysis of underwater running: A deep learning approach. *Journal of Biomechanics* 87:75-82. https://doi.org/10.1016/j.jbiomech.2019.02.021

Cunningham R, Sánchez MB, Butler PB, Southgate MJ, and Loram ID. 2019. Fully automated image-based estimation of postural point-features in children with cerebral palsy using deep learning. *Royal Society Open Science* 6:191011. doi:10.1098/rsos.191011

D'Isidoro F, Brockmann C, and Ferguson SJ. 2020. Effects of the soft tissue artefact on the hip joint kinematics during unrestricted activities of daily living. *Journal of Biomechanics* 104:109717. https://doi.org/10.1016/j.jbiomech.2020.109717

Dang Q, Yin J, Wang B, and Zheng W. 2019. Deep learning based 2D human pose estimation: A survey. *Tsinghua Science and Technology* 24:663-676. 10.26599/TST.2018.9010100

Dolatabadi E, Taati B, and Mihailidis A. 2016. Concurrent validity of the Microsoft Kinect for Windows v2 for measuring spatiotemporal gait parameters. *Medical Engineering & Physics* 38:952-958. https://doi.org/10.1016/j.medengphy.2016.06.015

Drazan JF, Phillips WT, Seethapathi N, Hullfish TJ, and Baxter JR. 2021. Moving outside the lab: Markerless motion capture accurately quantifies sagittal plane kinematics during the vertical jump. *Journal of Biomechanics* 125:110547. https://doi.org/10.1016/j.jbiomech.2021.110547

Elhayek A, Stoll C, Kim KI, and Theobalt C. 2015. Outdoor Human Motion Capture by Simultaneous Optimization of Pose and Camera Parameters. *Computer Graphics Forum* 34:86-98. 10.1111/cgf.12519

Evans M, Colyer S, Cosker D, and Salo A. 2018. Foot Contact Timings and Step Length for Sprint Training. 2018 IEEE Winter Conference on Applications of Computer Vision (WACV). p 1652-1660.

Fang H-S, Xie S, Tai Y-W, and Lu C. 2017. Rmpe: Regional multi-person pose estimation. Proceedings of the IEEE International Conference on Computer Vision. p 2334-2343.

Fiorentino NM, Atkins PR, Kutschke MJ, Goebel JM, Foreman KB, and Anderson AE. 2017. Soft tissue artifact causes significant errors in the calculation of joint angles and range of motion at the hip. *Gait & Posture* 55:184-190. https://doi.org/10.1016/j.gaitpost.2017.03.033

Franklyn-Miller A, Richter C, King E, Gore S, Moran K, Strike S, and Falvey EC. 2017. Athletic groin pain (part 2): a prospective cohort study on the biomechanical evaluation of change of direction identifies three clusters of movement patterns. *British Journal of Sports Medicine* 51:460-468. 10.1136/bjsports-2016-096050

Garcia-Agundez A, Folkerts A-K, Konrad R, Caserman P, Tregel T, Goosses M, Göbel S, and Kalbe E. 2019. Recent advances in rehabilitation for Parkinson's Disease with Exergames: A Systematic Review. *Journal of NeuroEngineering and Rehabilitation* 16:17. 10.1186/s12984-019-0492-1

Gorton GE, Hebert DA, and Gannotti ME. 2009. Assessment of the kinematic variability among 12 motion analysis laboratories. *Gait & Posture* 29:398-402. https://doi.org/10.1016/j.gaitpost.2008.10.060

Gu X, Deligianni F, Lo B, Chen W, and Yang GZ. 2018. Markerless gait analysis based on a single RGB camera. 2018 IEEE 15th International Conference on Wearable and Implantable Body Sensor Networks (BSN). p 42-45.

Gupta A, Scott K, and Dukewich M. 2017. Innovative Technology Using Virtual Reality in the Treatment of Pain: Does It Reduce Pain via Distraction, or Is There More to It? *Pain Medicine* 19:151-159. 10.1093/pm/pnx109

Harsted S, Holsgaard-Larsen A, Hestbæk L, Boyle E, and Lauridsen HH. 2019. Concurrent validity of lower extremity kinematics and jump characteristics captured in pre-school children by a markerless 3D motion capture system. *Chiropractic & Manual Therapies* 27:39. 10.1186/s12998-019-0261-z

Hausdorff JM, Lertratanakul A, Cudkowicz ME, Peterson AL, Kaliton D, and Goldberger AL. 2000. Dynamic markers of altered gait rhythm in amyotrophic lateral sclerosis. *Journal of Applied Physiology* 88:2045-2053. 10.1152/jappl.2000.88.6.2045

Heesen C, Böhm J, Reich C, Kasper J, Goebel M, and Gold SM. 2008. Patient perception of bodily functions in multiple sclerosis: gait and visual function are the most valuable. *Multiple Sclerosis Journal* 14:988-991. 10.1177/1352458508088916

Insafutdinov E, Pishchulin L, Andres B, Andriluka M, and Schiele B. 2016. DeeperCut: A Deeper, Stronger, and Faster Multi-person Pose Estimation Model. In: Leibe B, Matas J, Sebe N, and Welling M, editors. Computer Vision – ECCV 2016. Cham: Springer International Publishing. p 34-50.

Ionescu C, Papava D, Olaru V, and Sminchisescu C. 2014. Human3.6M: Large Scale Datasets and Predictive Methods for 3D Human Sensing in Natural Environments. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 36:1325-1339. 10.1109/TPAMI.2013.248

Jalalian A, Gibson I, and Tay EH. 2013. Computational Biomechanical Modeling of Scoliotic Spine: Challenges and Opportunities. *Spine Deformity* 1:401-411. https://doi.org/10.1016/j.jspd.2013.07.009

Jenkinson TR, Mallorie PA, Whitelock HC, Kennedy LG, Garrett SL, and Calin A. 1994. Defining spinal mobility in ankylosing spondylitis (AS). The Bath AS Metrology Index. *The Journal of rheumatology* 21:1694-1698.

Kanko RM, Laende E, Scott Selbie W, and Deluzio KJ. 2021a. Inter-session repeatability of markerless motion capture gait kinematics. *Journal of Biomechanics*:110422. https://doi.org/10.1016/j.jbiomech.2021.110422

Kanko RM, Laende EK, Davis EM, Scott Selbie W, and Deluzio KJ. 2021b. Concurrent assessment of gait kinematics using marker-based and markerless motion capture. *Journal of Biomechanics*:110665. https://doi.org/10.1016/j.jbiomech.2021.110665

Kanko RM, Laende EK, Strutzenberger G, Brown M, Selbie WS, DePaul V, Scott SH, and Deluzio KJ. 2021c. Assessment of spatiotemporal gait parameters using a deep learning algorithm-based markerless motion capture system. *Journal of Biomechanics* 122:110414. https://doi.org/10.1016/j.jbiomech.2021.110414

Kessler SE, Rainbow MJ, Lichtwark GA, Cresswell AG, D'Andrea SE, Konow N, and Kelly LA. 2019. A Direct Comparison of Biplanar Videoradiography and Optical Motion Capture for Foot and Ankle Kinematics. *Frontiers in Bioengineering and Biotechnology* 7. 10.3389/fbioe.2019.00199

Khan F, Salahuddin S, and Javidnia H. 2020. Deep Learning-Based Monocular Depth Estimation Methods-A State-of-the-Art Review. *Sensors (Basel, Switzerland)* 20:2272. 10.3390/s20082272

Kidziński Ł, Yang B, Hicks JL, Rajagopal A, Delp SL, and Schwartz MH. 2020. Deep neural networks enable quantitative movement analysis using single-camera videos. *Nature Communications* 11:4054. 10.1038/s41467-020-17807-z

King E, Franklyn-Miller A, Richter C, O'Reilly E, Doolan M, Moran K, Strike S, and Falvey É. 2018. Clinical and biomechanical outcomes of rehabilitation targeting intersegmental control in athletic groin pain: prospective cohort of 205 patients. *British Journal of Sports Medicine* 52:1054-1062. 10.1136/bjsports-2016-097089

Knippenberg E, Verbrugghe J, Lamers I, Palmaers S, Timmermans A, and Spooren A. 2017. Markerless motion capture systems as training device in neurological rehabilitation: a systematic review of their use, application, target population and efficacy. *Journal of NeuroEngineering and Rehabilitation* 14:61. 10.1186/s12984-017-0270-x

Lin T-Y, Maire M, Belongie S, Hays J, Perona P, Ramanan D, Dollár P, and Zitnick CL. 2014. Microsoft COCO: Common Objects in Context. In: Fleet D, Pajdla T, Schiele B, and Tuytelaars T, editors. Computer Vision – ECCV 2014. Cham: Springer International Publishing. p 740-755.

Lofterød B, Terjesen T, Skaaret I, Huse A-B, and Jahnsen R. 2007. Preoperative gait analysis has a substantial effect on orthopedic decision making in children with cerebral palsy: Comparison between clinical evaluation and gait analysis in 60 patients. *Acta orthopaedica* 78:74-80. 10.1080/17453670610013448

Martinez HR, Garcia-Sarreon A, Camara-Lemarroy C, Salazar F, and Guerrero-González ML. 2018. Accuracy of Markerless 3D Motion Capture Evaluation to Differentiate between On/Off Status in Parkinson's Disease after Deep Brain Stimulation. *Parkinson’s Disease* 2018:5830364. 10.1155/2018/5830364

Mathis A, Mamidanna P, Cury KM, Abe T, Murthy VN, Mathis MW, and Bethge M. 2018. DeepLabCut: markerless pose estimation of user-defined body parts with deep learning. *Nature Neuroscience* 21:1281-1289. 10.1038/s41593-018-0209-y

Mathis A, Schneider S, Lauer J, and Mathis MW. 2020a. A Primer on Motion Capture with Deep Learning: Principles, Pitfalls and Perspectives. *arXiv preprint arXiv:200900564*.

Mathis A, Schneider S, Lauer J, and Mathis MW. 2020b. A Primer on Motion Capture with Deep Learning: Principles, Pitfalls, and Perspectives. *Neuron* 108:44-65. https://doi.org/10.1016/j.neuron.2020.09.017

Mehta D, Rhodin H, Casas D, Fua P, Sotnychenko O, Xu W, and Theobalt C. 2017. Monocular 3D Human Pose Estimation in the Wild Using Improved CNN Supervision. 2017 International Conference on 3D Vision (3DV). p 506-516.

Mentiplay BF, Perraton LG, Bower KJ, Pua Y-H, McGaw R, Heywood S, and Clark RA. 2015. Gait assessment using the Microsoft Xbox One Kinect: Concurrent validity and inter-day reliability of spatiotemporal and kinematic variables. *Journal of Biomechanics* 48:2166-2170. https://doi.org/10.1016/j.jbiomech.2015.05.021

Miranda DL, Rainbow MJ, Crisco JJ, and Fleming BC. 2013. Kinematic differences between optical motion capture and biplanar videoradiography during a jump–cut maneuver. *Journal of Biomechanics* 46:567-573. https://doi.org/10.1016/j.jbiomech.2012.09.023

Miranda DL, Schwartz JB, Loomis AC, Brainerd EL, Fleming BC, and Crisco JJ. 2011. Static and Dynamic Error of a Biplanar Videoradiography System Using Marker-Based and Markerless Tracking Techniques. *Journal of Biomechanical Engineering* 133. 10.1115/1.4005471

Moon G, Chang JY, and Lee KM. 2019. Camera distance-aware top-down approach for 3d multi-person pose estimation from a single rgb image. Proceedings of the IEEE International Conference on Computer Vision. p 10133-10142.

Moro M, Marchesi G, Odone F, and Casadio M. 2020. Markerless gait analysis in stroke survivors based on computer vision and deep learning: a pilot study. Proceedings of the 35th Annual ACM Symposium on Applied Computing. Brno, Czech Republic: Association for Computing Machinery. p 2097–2104.

Mousavi Hondori H, and Khademi M. 2014. A review on technical and clinical impact of microsoft kinect on physical therapy and rehabilitation. *Journal of medical engineering* 2014.

Munn Z, Peters MDJ, Stern C, Tufanaru C, McArthur A, and Aromataris E. 2018. Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. *BMC Medical Research Methodology* 18:143. 10.1186/s12874-018-0611-x

Muro-de-la-Herran A, Garcia-Zapirain B, and Mendez-Zorrilla A. 2014. Gait Analysis Methods: An Overview of Wearable and Non-Wearable Systems, Highlighting Clinical Applications. *Sensors* 14. 10.3390/s140203362

Nakano N, Sakura T, Ueda K, Omura L, Kimura A, Iino Y, Fukashiro S, and Yoshioka S. 2020. Evaluation of 3D Markerless Motion Capture Accuracy Using OpenPose With Multiple Video Cameras. *Frontiers in Sports and Active Living* 2. 10.3389/fspor.2020.00050

Natarajan SK, Wang X, Spranger M, and Gräser A. 2017. Reha@Home - a vision based markerless gait analysis system for rehabilitation at home. 2017 13th IASTED International Conference on Biomedical Engineering (BioMed). p 32-41.

Needham L, Evans M, Cosker DP, and Colyer SL. 2021a. Can Markerless Pose Estimation Algorithms Estimate 3D Mass Centre Positions and Velocities during Linear Sprinting Activities? *Sensors* 21:2889.

Needham L, Evans M, Cosker DP, Wade L, McGuigan PM, Bilzon JL, and Colyer SL. 2021b. The accuracy of several pose estimation methods for 3D joint centre localisation. *Scientific Reports* 11:20673. 10.1038/s41598-021-00212-x

Noyes K, and Weinstock-Guttman B. 2013. Impact of diagnosis and early treatment on the course of multiple sclerosis. *Am J Manag Care* 19:s321-331.

Otte K, Kayser B, Mansow-Model S, Verrel J, Paul F, Brandt AU, and Schmitz-Hübsch T. 2016. Accuracy and Reliability of the Kinect Version 2 for Clinical Measurement of Motor Function. *PLoS ONE* 11:e0166532. 10.1371/journal.pone.0166532

Pantzar-Castilla E, Cereatti A, Figari G, Valeri N, Paolini G, Della Croce U, Magnuson A, and Riad J. 2018. Knee joint sagittal plane movement in cerebral palsy: a comparative study of 2-dimensional markerless video and 3-dimensional gait analysis. *Acta orthopaedica* 89:656-661. 10.1080/17453674.2018.1525195

Paterno MV, Schmitt LC, Ford KR, Rauh MJ, Myer GD, Huang B, and Hewett TE. 2010. Biomechanical Measures during Landing and Postural Stability Predict Second Anterior Cruciate Ligament Injury after Anterior Cruciate Ligament Reconstruction and Return to Sport. *Am J Sports Med* 38:1968-1978. 10.1177/0363546510376053

Pavão SL, dos Santos AN, Woollacott MH, and Rocha NACF. 2013. Assessment of postural control in children with cerebral palsy: A review. *Research in Developmental Disabilities* 34:1367-1375. https://doi.org/10.1016/j.ridd.2013.01.034

Peters A, Galna B, Sangeux M, Morris M, and Baker R. 2010. Quantification of soft tissue artifact in lower limb human motion analysis: A systematic review. *Gait & Posture* 31:1-8. https://doi.org/10.1016/j.gaitpost.2009.09.004

Reinschmidt C, van den Bogert AJ, Lundberg A, Nigg BM, Murphy N, Stacoff A, and Stano A. 1997. Tibiofemoral and tibiocalcaneal motion during walking: external vs. skeletal markers. *Gait & Posture* 6:98-109. https://doi.org/10.1016/S0966-6362(97)01110-7

Rempe D, Guibas LJ, Hertzmann A, Russell B, Villegas R, and Yang J. 2020. Contact and Human Dynamics from Monocular Video. Proceedings of the European Conference on Computer Vision (ECCV).

Robles-García V, Corral-Bergantiños Y, Espinosa N, Jácome MA, García-Sancho C, Cudeiro J, and Arias P. 2015. Spatiotemporal Gait Patterns During Overt and Covert Evaluation in Patients With Parkinson's Disease and Healthy Subjects: Is There a Hawthorne Effect? 31:189. 10.1123/jab.2013-0319

Rodrigues TB, Catháin CÓ, Devine D, Moran K, O'Connor NE, and Murray N. 2019. An evaluation of a 3D multimodal marker-less motion analysis system. Proceedings of the 10th ACM Multimedia Systems Conference. Amherst, Massachusetts: Association for Computing Machinery. p 213–221.

Rudwaleit M, Khan MA, and Sieper J. 2005. The challenge of diagnosis and classification in early ankylosing spondylitis: Do we need new criteria? *Arthritis & Rheumatism* 52:1000-1008. 10.1002/art.20990

Salarian A, Russmann H, Vingerhoets FJG, Dehollain C, Blanc Y, Burkhard PR, and Aminian K. 2004. Gait assessment in Parkinson's disease: toward an ambulatory system for long-term monitoring. *IEEE Transactions on Biomedical Engineering* 51:1434-1443. 10.1109/TBME.2004.827933

Sarafianos N, Boteanu B, Ionescu B, and Kakadiaris IA. 2016. 3D Human pose estimation: A review of the literature and analysis of covariates. *Computer Vision and Image Understanding* 152:1-20. https://doi.org/10.1016/j.cviu.2016.09.002

Sarbolandi H, Lefloch D, and Kolb A. 2015. Kinect range sensing: Structured-light versus Time-of-Flight Kinect. *Computer Vision and Image Understanding* 139:1-20. https://doi.org/10.1016/j.cviu.2015.05.006

Sawacha Z, Carraro E, Del Din S, Guiotto A, Bonaldo L, Punzi L, Cobelli C, and Masiero S. 2012. Biomechanical assessment of balance and posture in subjects with ankylosing spondylitis. *Journal of NeuroEngineering and Rehabilitation* 9:63. 10.1186/1743-0003-9-63

Scapin S, Echevarría-Guanilo ME, Boeira Fuculo Junior PR, Gonçalves N, Rocha PK, and Coimbra R. 2018. Virtual Reality in the treatment of burn patients: A systematic review. *Burns* 44:1403-1416. https://doi.org/10.1016/j.burns.2017.11.002

Schwartz MH, Trost JP, and Wervey RA. 2004. Measurement and management of errors in quantitative gait data. *Gait & Posture* 20:196-203. https://doi.org/10.1016/j.gaitpost.2003.09.011

Seethapathi N, Wang S, Saluja R, Blohm G, and Kording KP. 2019. Movement science needs different pose tracking algorithms. *arXiv preprint arXiv:190710226*.

Serrancolí G, Bogatikov P, Huix JP, Barberà AF, Egea AJS, Ribé JT, Kanaan-Izquierdo S, and Susín A. 2020. Marker-Less Monitoring Protocol to Analyze Biomechanical Joint Metrics During Pedaling. *IEEE Access* 8:122782-122790. 10.1109/ACCESS.2020.3006423

Shin JH, Yu R, Ong JN, Lee CY, Jeon SH, Park H, Kim H-J, Lee J, and Jeon B. 2021. Quantitative Gait Analysis Using a Pose-Estimation Algorithm with a Single 2D-Video of Parkinson's Disease Patients. *Journal of Parkinson's Disease* 11:1271-1283. 10.3233/JPD-212544

Sigal L, Balan AO, and Black MJ. 2010. HumanEva: Synchronized Video and Motion Capture Dataset and Baseline Algorithm for Evaluation of Articulated Human Motion. *International Journal of Computer Vision* 87:4. 10.1007/s11263-009-0273-6

Simon SR. 2004. Quantification of human motion: gait analysis—benefits and limitations to its application to clinical problems. *Journal of Biomechanics* 37:1869-1880. https://doi.org/10.1016/j.jbiomech.2004.02.047

Slembrouck M, Luong H, Gerlo J, Schütte K, Van Cauwelaert D, De Clercq D, Vanwanseele B, Veelaert P, and Philips W. 2020. Multiview 3D Markerless Human Pose Estimation from OpenPose Skeletons. Cham: Springer International Publishing. p 166-178.

Steinert A, Sattler I, Otte K, Röhling H, Mansow-Model S, and Müller-Werdan U. 2020. Using new camera-based technologies for gait analysis in older adults in comparison to the established GAITRite system. *Sensors* 20:125.

Stenum J, Rossi C, and Roemmich RT. 2021. Two-dimensional video-based analysis of human gait using pose estimation. *PLOS Computational Biology* 17:e1008935. 10.1371/journal.pcbi.1008935

Swash M. 1998. Early diagnosis of ALS/MND. *Journal of the Neurological Sciences* 160:S33-S36. https://doi.org/10.1016/S0022-510X(98)00215-9

Takeda I, Yamada A, and Onodera H. 2020. Artificial Intelligence-Assisted motion capture for medical applications: a comparative study between markerless and passive marker motion capture. *Computer Methods in Biomechanics and Biomedical Engineering*:1-10. 10.1080/10255842.2020.1856372

Tanaka R, Takimoto H, Yamasaki T, and Higashi A. 2018. Validity of time series kinematical data as measured by a markerless motion capture system on a flatland for gait assessment. *Journal of Biomechanics* 71:281-285. https://doi.org/10.1016/j.jbiomech.2018.01.035

Topley M, and Richards JG. 2020. A Comparison of Currently Available Optoelectronic Motion Capture Systems. *Journal of Biomechanics*:109820. https://doi.org/10.1016/j.jbiomech.2020.109820

Vergara ME, O'Shea FD, Inman RD, and Gage WH. 2012. Postural control is altered in patients with ankylosing spondylitis. *Clinical Biomechanics* 27:334-340. https://doi.org/10.1016/j.clinbiomech.2011.10.016

Vonstad EK, Su X, Vereijken B, Bach K, and Nilsen JH. 2020. Comparison of a Deep Learning-Based Pose Estimation System to Marker-Based and Kinect Systems in Exergaming for Balance Training. *Sensors* 20. 10.3390/s20236940

Voulodimos A, Doulamis N, Doulamis A, and Protopapadakis E. 2018. Deep Learning for Computer Vision: A Brief Review. *Computational Intelligence and Neuroscience* 2018:7068349. 10.1155/2018/7068349

Webster D, and Celik O. 2014. Systematic review of Kinect applications in elderly care and stroke rehabilitation. *Journal of NeuroEngineering and Rehabilitation* 11:108. 10.1186/1743-0003-11-108

Whittle MW. 1996. Clinical gait analysis: A review. *Human Movement Science* 15:369-387. https://doi.org/10.1016/0167-9457(96)00006-1

Wren TAL, Kalisvaart MM, Ghatan CE, Rethlefsen SA, Hara R, Sheng M, Chan LS, and Kay RM. 2009. Effects of Preoperative Gait Analysis on Costs and Amount of Surgery. *Journal of Pediatric Orthopaedics* 29:558-563. 10.1097/BPO.0b013e3181b2f8c2

Yang W, Ouyang W, Wang X, Ren J, Li H, and Wang X. 2018. 3d human pose estimation in the wild by adversarial learning. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. p 5255-5264.

Zago M, Luzzago M, Marangoni T, De Cecco M, Tarabini M, and Galli M. 2020. 3D Tracking of Human Motion Using Visual Skeletonization and Stereoscopic Vision. *Frontiers in Bioengineering and Biotechnology* 8. 10.3389/fbioe.2020.00181

Zelik KE, and Honert EC. 2018. Ankle and foot power in gait analysis: Implications for science, technology and clinical assessment. *Journal of Biomechanics* 75:1-12. https://doi.org/10.1016/j.jbiomech.2018.04.017

Zhang F, Juneau P, McGuirk C, Tu A, Cheung K, Baddour N, and Lemaire E. 2021. Comparison of OpenPose and HyperPose artificial intelligence models for analysis of hand-held smartphone videos. 2021 IEEE International Symposium on Medical Measurements and Applications (MeMeA). p 1-6.

Zhang H, Yan X, and Li H. 2018. Ergonomic posture recognition using 3D view-invariant features from single ordinary camera. *Automation in Construction* 94:1-10. https://doi.org/10.1016/j.autcon.2018.05.033

Zochling J. 2011. Measures of symptoms and disease status in ankylosing spondylitis: Ankylosing Spondylitis Disease Activity Score (ASDAS), Ankylosing Spondylitis Quality of Life Scale (ASQoL), Bath Ankylosing Spondylitis Disease Activity Index (BASDAI), Bath Ankylosing Spondylitis Functional Index (BASFI), Bath Ankylosing Spondylitis Global Score (BAS-G), Bath Ankylosing Spondylitis Metrology Index (BASMI), Dougados Functional Index (DFI), and Health Assessment Questionnaire for the Spondylarthropathies (HAQ-S). *Arthritis Care & Research* 63:S47-S58. 10.1002/acr.20575

Figure 1

Six degrees of freedom

Figure 1: This figure demonstrates the six degrees of freedom needed to describe position and orientation (pose) of the human body, with the red dot indicating the location (translation) of the segment center of mass and blue arrows indicating rotation in three planes. A) The reference standing posture, B) thigh segment adduction/abduction, C) thigh segment flexion/extension, D) thigh segment rotation about the longitudinal axis.

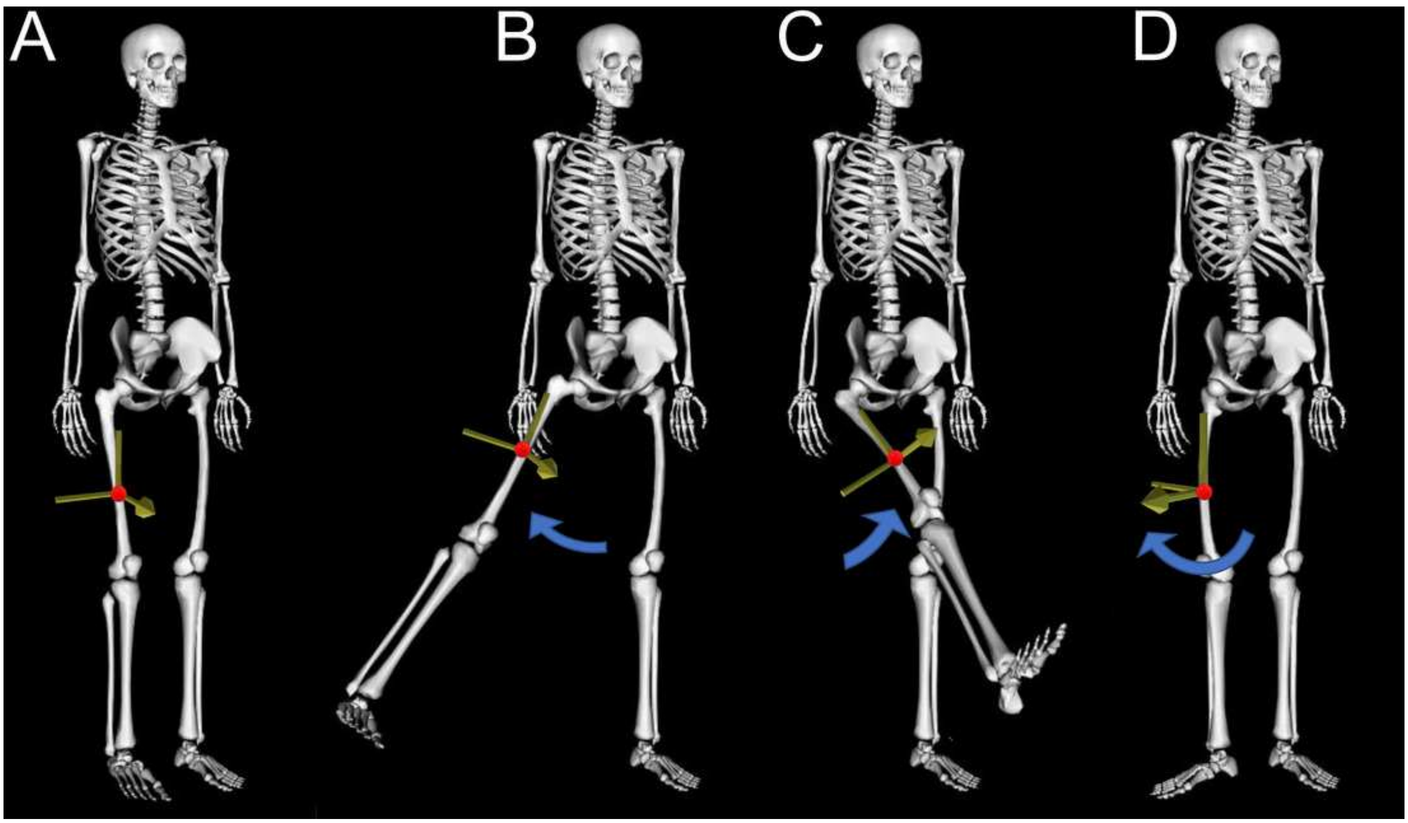

Figure 2

Optoelectronic motion capture markers

Figure 2: Optoelectronic motion capture. Left - markers placed on the subject. Right - view of the markers in 3D space.

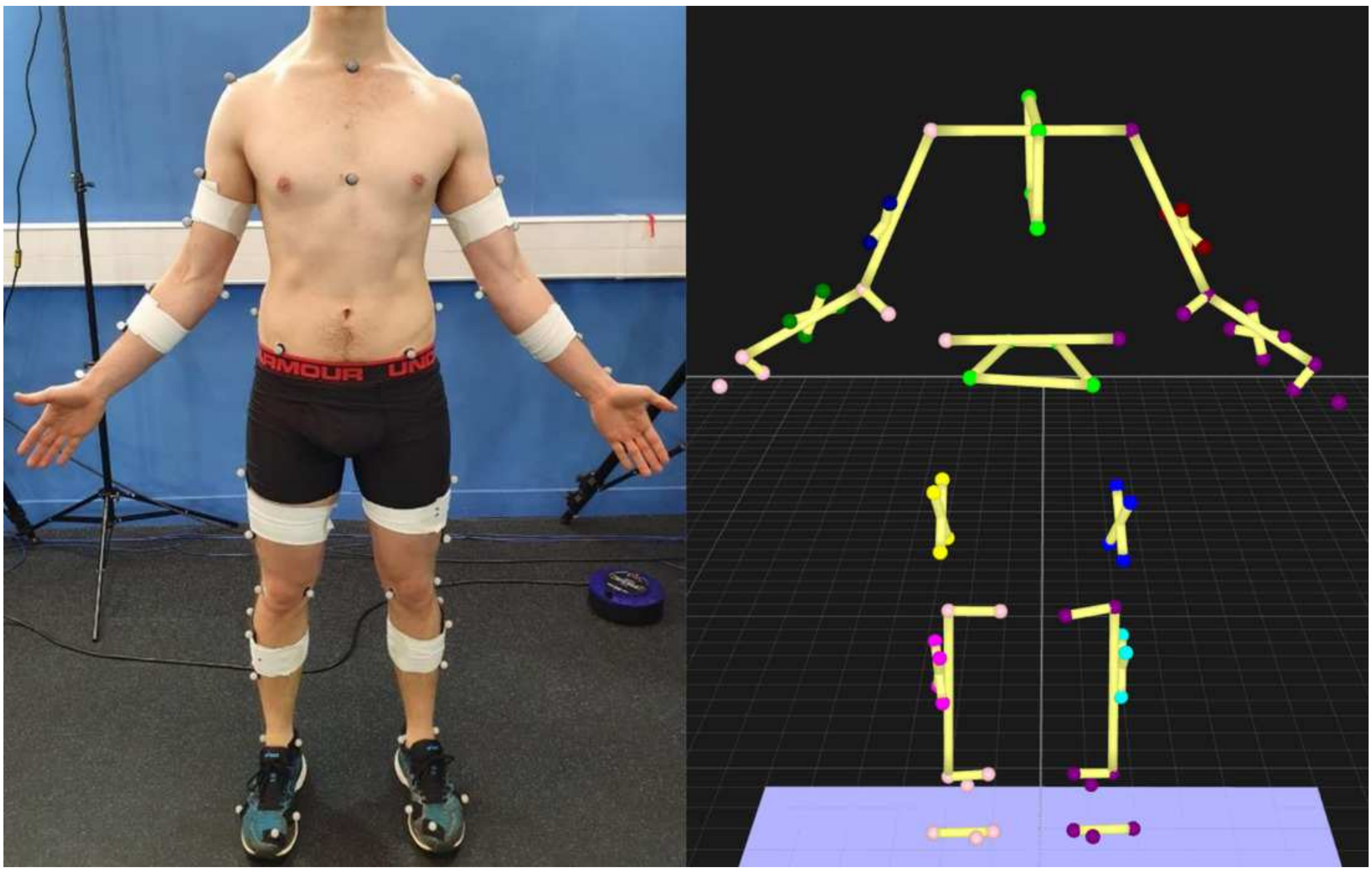

Figure 3

Figure 3

Figure 3: Twenty-five keypoints detected using the OpenPose (Cao et al. 2018) applied to a single image.

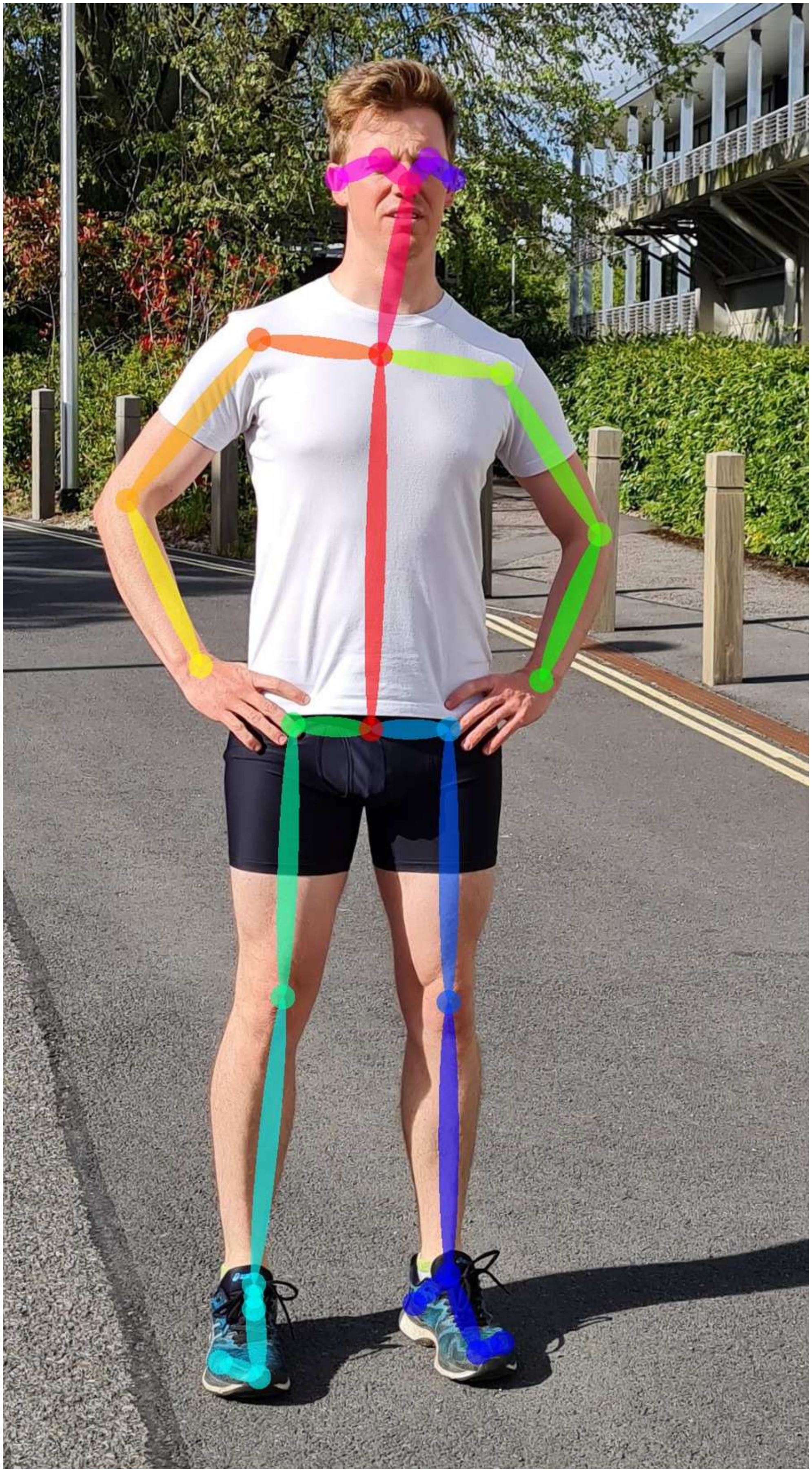

Figure 4

Pose estimation algorithm training workflow

Figure 4: Training a pose estimation algorithm. Stage One: Creation of manually labelled training dataset. Stage Two: Using the unlabeled images from stage one, the pose estimation algorithm estimates the desired keypoint locations (joint centers). Estimated keypoint locations are compared to the manually labelled training data to determine the distance is between the estimated keypoint and the manually labelled keypoint. The optimization method then adjusts filters within the layers of the algorithm to try to reduce this distance and new estimated keypoints are calculated. This process is repeated until improvements to the pose estimate4ion algorithm are negligible.

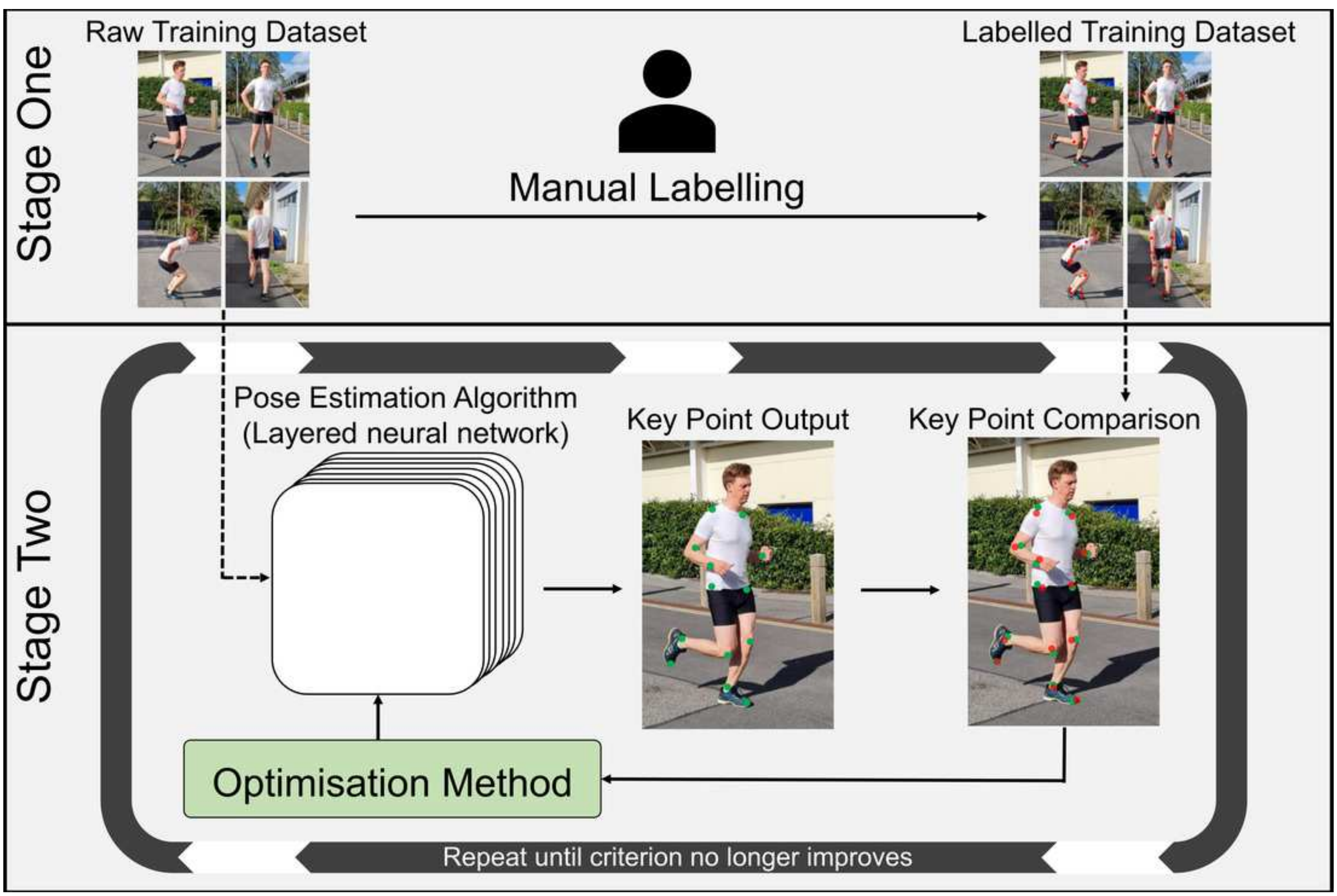

Figure 5

2D and 3D pose estimation

Figure 5: Markerless motion capture examples: 2D pose estimation from monocular motion capture (2D keypoints detected using OpenPose Cao et al. (2018) , 3D pose estimation from monocular motion capture (adapted from Cheng et al. (2020) with license from the Association for the Advancement of Artificial Intelligence, Copyright © year? 20) and 3D pose estimation from multi-camera motion capture (adapted from Sigal et al. (2010) with permission from Springer Nature).

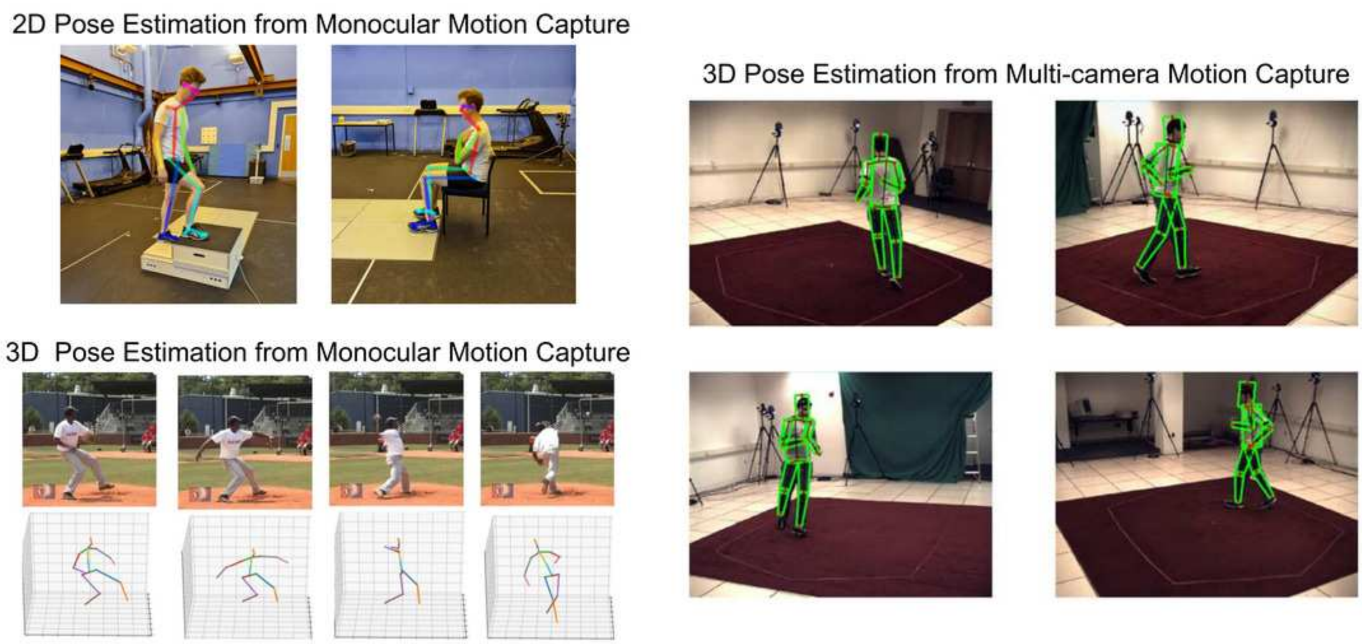

Figure 6

Current open-source labelled dataset

Figure 6: Keypoints of large open-source datasets have been labelled to estimate occluded joint centers, however this requires users to guess where these locations are, as they are not visible (adapted from (COCO 2021; Lin et al. 2014) under creative commons license by (COCO 2020)).