



The feasibility and validity of a single camera deep learning-based method for 3D biomechanical analysis in strength training: proof-of-concept

For correspondence:
l.noteboom@vu.nl

L. Noteboom¹, M.J.M Hoozemans¹, H.E.J. Veeger² and F.C.T. van der Helm²

¹Faculty of Behavioral and Movement Sciences, Vrije Universiteit Amsterdam, the Netherlands,

²Department of Biomechanical Engineering, Delft University of Technology, the Netherlands

Please cite as: Noteboom, L., Hoozemans M.J.M., Veeger H.E.J. & van der Helm F.C.T. (2022). The feasibility and validity of a single camera deep learning-based method for 3D biomechanical analysis in strength training: proof-of-concept. *SportRxiv*.

ABSTRACT

Biomechanical analysis is valuable for injury risk and performance assessment in sports, but the application is limited due to restrictions in costs, set-up time and accuracy of available motion capture methods. Therefore, the present proof-of-concept study evaluated the feasibility and validity of a more suitable method, based on a single camera combined with a deep-learning algorithm, by comparing obtained biomechanical parameters with those obtained by the state-of-the-art optoelectronic measurement system (OMS) and the marker-less Kinect during upper-body strength exercises. Results from five athletes revealed strong to

excellent correlations for most parameters and root-mean-square deviations of 4-8 degrees for angles and 0.9-1.4Nm for moments, but insufficient ICCs compared to the OMS, and partly better performance than the Kinect. In conclusion, the present study showed that the single camera deep learning-based method is feasible for biomechanical analysis of strength exercises and provides limited evidence that some parameters can be estimated with reasonable accuracy. However, the accuracy of peak angle and moment estimations should be improved before this method can be applied for injury prevention, i.e. by training the deep-learning model on a larger variety of subject anthropometries. Furthermore, future research should investigate the validity for larger sample sizes and multiple exercises.

INTRODUCTION

Biomechanical analysis is an essential part of understanding and optimizing human motion and is often employed in the fields of sports and rehabilitation. In sports, biomechanical analysis can be used for performance improvement and to study and identify sport-related risk factors for musculoskeletal injuries [1]. The current state-of-the-art method to apply biomechanics in sports is the use of an optoelectronic measurement system (OMS) in combination with a biomechanical model imposing constraint equations to the marker trajectories [2]. An OMS consists of multiple cameras that detect light from either active or passive markers placed on an athlete's skin, usually at or related to predetermined bony landmarks, to determine the three-dimensional (3D) location of those markers by time-of-flight triangulation [2]. However, the OMS has several limitations. Firstly, the markers on the skin can move with respect to the underlying bony landmarks, causing soft-tissue artifacts [3]. Secondly, measurements are labor intensive and time consuming [4]. Thirdly, measurements are restricted to a controlled laboratory setting [5]. These circumstances are not ideal to capture an athlete's natural sports movement and only allow for analysis of a limited number of athletes.

A potential solution to make biomechanical analysis applicable in the sports practice and on a larger scale is by using markerless pose estimation, as markerless measurements are advantageous in terms of costs, set-up time and not restricted to a laboratory setting. One example of a markerless pose estimation method is the Microsoft Kinect, which uses an infrared depth-sensing camera and a random forest algorithm to estimate 3D joint positions of subjects [2]. Although much more feasible than the OMS, observed low accuracies of the Kinect's estimations in situations of body segment occlusions or axial rotations and the low

robustness of the Kinect against environmental light conditions, limit the application of this system for biomechanical analysis in a sports environment [6-8].

Interestingly, the field of pose estimation is developing rapidly and more sophisticated algorithms than used by the Microsoft Kinect may be employed to improve motion capture accuracies. Specifically, convolutional neural networks (CNNs) may be well-suited for this task because their multi-layer structure allows for incorporation of a lot of data to optimize estimations while remaining efficient [9, 10]. For instance, temporal information of previous joint position solutions, and information of other joint positions can all be used in the estimation, whereas the random forest approach makes estimates separately for each joint and based on single images [10]. CNNs could therefore potentially be more accurate and more robust in situations like temporary segment occlusion. Remarkably, some of these networks can even estimate 3D joint positions from 2D videos, which means that only a single standard camera is required. Standard cameras are widely available and could make measurements in a sports environment even more feasible. Therewith, a combination of a standard camera with a CNN could potentially be well-suited to capture motions and apply biomechanical analysis in a sports environment.

However, as stated by [11], the transfer of deep-learning based pose estimation methods towards application in biomechanics has been slow, potentially due to the requirement of advanced coding skills and in-depth computer science knowledge. As such, the accuracy of biomechanical parameters obtained by this type of methods are largely unknown. However, well-documented open-source Github repositories including pretrained CNN models are available [9] and can be applied by biomechanical researchers and clinicians without extensive computer science skills. Therefore, the objective of the present study is to assess the feasibility of single camera deep learning-based biomechanical analysis and provide some first clues regarding the validity by comparing biomechanical parameters with the the Kinect and an optoelectronic measurement system (state-of-the-art). This study aimed to assess biomechanical parameters within a sports environment. Strength training exercises were chosen for evaluation of the camera-based method. Strength training is one of the fastest-growing sport domains [12], but deals with a high amount of musculoskeletal injuries [13], from which the causes remain largely unknown. This lack of knowledge warrants the search for valid, reliable and feasible methods that can assess biomechanical load in a strength training environment, i.e. in the gym or at home. It was hypothesized that the single camera deep learning-based method is a feasible and valid way to estimate biomechanical parameters during strength exercises.

METHOD

Participants

Five healthy male participants (mean age 16.8 SD 1.3 years, body mass 80.4 SD 4.2 kg, body height 1.84 SD 0.07 m) with experience in strength training were included in this study. The study was approved by the local ethics committee of the Faculty of Behavioural and Movement Sciences, Vrije Universiteit Amsterdam (VCWE-2019-033). All participants provided written informed consent.

Procedure

During preparation, participants were equipped with a set of 12 reflective markers on the thorax and dominant arm (Table 1). Subsequently, participants performed a 10-minute warm-up protocol. During the actual measurements, participants performed two upper extremity dumbbell exercises: the lateral fly and the biceps curl (Fig. 1). These exercises were chosen in order to evaluate the ability of the motion capture systems to capture upper extremity movements in varying planes. The lateral fly movement occurs predominantly in the frontal plane, whereas the biceps curl movement occurs mainly in the sagittal plane. In addition, axial rotation of the upper arm can be expected during the lateral fly, and segment occlusion of the upper arm can be expected during the biceps curl, allowing for evaluation of the motion capture systems during these challenging conditions. In total, each exercise was performed for 3 sets of 5 repetitions, at a self-selected pace. In-between each set participants rested for 30 seconds. The mass of each dumbbell was 5 kg for the biceps curl and 3 kg for the lateral fly.

Table 1. Anatomical locations of segment markers/coordinates used per system

Segment	Marker locations per system		
	OMS	Kinect	Camera

Torso	Incisura Jugularis	Mid spine	Mid spine
	Processus Xiphoideus	Mid thorax	Mid thorax
	Cervical Vertebrae 7		
	Thoracic Vertebrae 10		
Upper arm (dominant side)	Acromion	Shoulder joint center	Shoulder joint center
	Epicondylus Medialis		
	Epicondylus Lateralis		
	Upper arm (tracking marker)		
Lower arm (dominant side)	Head of the Ulna	Elbow joint center	Elbow joint center
	Styloid Processes of Radius		
	Lower arm (tracking marker)		
Hand (dominant side)	Interphalangealis proximal III	Wrist joint center	Wrist joint center



Figure 1. Pictures of the start (left) and mid (right) pose of one repetition of the biceps curl (BC) (top) and lateral fly (LF) (bottom) exercise

Hardware and software

The performed strength exercises were simultaneously measured by three motion capture systems: an opto-electrical measurement system (OMS; markerbased), a Kinect depth sensing camera (markerless) and a standard camera built-in in the Kinect (markerless). The OMS (Vicon Motion Systems, Oxford, United Kingdom) consisted of eight infrared cameras that registered the 3D coordinates of the reflective markers with a sample frequency of 400hz. Coordinates were expressed in the laboratory's coordinate system and recordings were processed with the Vicon Nexus software (Vicon Motion Systems, Oxford, United Kingdom). The Kinect (v2, Microsoft Corporation, Redmond, WA, USA) was placed 1.5 meters in front of the subject at a height of 1.5 meters and a downwards angle of 15 degrees. The combination of an infrared sensor and standard camera within the Kinect generated 3D depth videos. From these depth videos, a built-in random forest algorithm in the Kinect software development kit was employed to extract 19 joint coordinates [14], from which five were used in the present

study (Table 1). These coordinates were expressed in the Kinect's coordinate system. Additionally, the standard camera in the Kinect captured normal 2D videos. From these videos, a pretrained 3D pose estimation model, specified in Pavlo, Feichtenhofer [9], was used to generate 3D coordinates. In their approach, first a pretrained convolutional neural network (CNN) is used to detect a person in an image, and detect their 2D joint center locations, which are subsequently lifted to 3D using a newly developed temporal dilated convolutional model. The second step additionally allows for modeling of temporal relations between individual poses. This pretrained model was available from an open-source GitHub project [15], which also included a clear step-by-step documentation on how to use the model on 2D videos [16]. Therewith, this method can also be employed by researchers or clinicians without extensive deep-learning knowledge. The model has a fully convolutional architecture that takes a sequence of 2D poses obtained from videos as input and estimates the 3D positions of joint centers with respect to the root joint (pelvis). A semi-supervised method was used by Pavlo, Feichtenhofer [9] to train the model, meaning that the model learned to make the 3D pose prediction based on both ground truth labeled data, and based on unlabeled data via a back-projection method. The pretrained model was used in the present study to generate the 3D coordinates of 17 joint centers, from which five were used (Table 1). The sample frequencies of both the camera and the Kinect were 30Hz.

Data cleaning and preparation

Coordinate data from all systems were imported in MATLAB (2020a, The MathWorks, Inc., Natick, Massachusetts, United States). The coordinates from the marker-based system were expressed in a different coordinate system than the coordinates from the markerless system. Therefore, all coordinates were first converted to the right-handed coordinate system recommended by the International Society of Biomechanics (ISB) [17]. In addition, the data from the marker-based system were corrected for switched or missing markers, under the assumption of rigid bodies. Some trials had to be excluded due to problems with one of the three systems involved, for instance due to too many missing markers (marker-based system; a minimum of three markers per segment is required) or completely insufficient joint tracking (markerless systems).

Biomechanical model definitions

Whereas the marker-based system generated coordinates of bony landmarks, the markerless systems generated coordinates of estimated joint centers. These different

kinematic outputs required two distinct local coordinate systems, one for the marker-based system and one for the two markerless systems. For the marker-based system, bony landmarks were used to define segment coordinate systems of the thorax, humerus and forearm according to ISB recommendations [17]. For the markerless systems, an alternative method was required since only joint centers were available instead of anatomical landmarks. Therefore, the same segment definitions as specified in Plantard, Muller [18] were employed. Similar to the study of Plantard, Muller [18], hand positions were not captured accurately in the present study. For that reason, the elbow was considered a hinge joint, allowing only for measurement of elbow flexion. This meant that pronation/supination could not be assessed.

Inverse kinematics and dynamics

Inverse kinematics and dynamics were used to obtain joint angles and joint moments. This step was the same for both the marker-based and the markerless systems. Thoracohumeral ('shoulder') angles were calculated by the Euler method, with an Y-X-Y decomposition order, as recommended by the ISB [17]. Elbow flexion/extension angles were calculated between the long axes of the humerus and forearm. The joint angles for the following movements were selected for analysis: shoulder elevation, shoulder plane of elevation, shoulder internal and external rotation, and elbow flexion/extension. In addition, joint moments were calculated by the moment-at-once inverse dynamics method [19]. The moments were first calculated around the global axes, and subsequently reoriented to represent anatomical meaningful moments. For the elbow, the flexion/extension moment was calculated around the z-axis of the humerus coordinate system. For the shoulder, the moments were expressed in the plane of elevation, as the arm elevation moment and the moment around the long axis of the humerus. The external force caused by the dumbbell was applied to the center of mass of the hand segment. However, since the hand position and length could not be measured accurately, the location of the hand's center of mass was estimated at 6cm from the wrist center in extension of the forearm. This estimation was based on the average hand palm length (hand length minus finger length) of 12 cm, obtained from male Dutch students in the DINED database [20]. Body segment inertial parameters were estimated with regression equations of De Leva [21]. Calculated joint angles and moments were low-pass filtered using a 4th-order Butterworth filter with a cutoff frequency of 3Hz.

Data analysis and statistics

The times series of the joint angles and moments obtained from the markerless systems were up sampled and synchronized to the time series of the OMS. Synchronization was done separately for each movement repetition, at the peak shoulder elevation angle for the lateral fly exercise and at the peak elbow flexion angle for the biceps curl exercise. One second before and after the instant of the peak angles were selected for analysis. For all synchronized time series of joint angles and moments, root-mean-square-deviations (RMSD) and Pearson's correlation coefficient (r) were calculated between the camera and the OMS, and between the Kinect and the OMS. The mean and standard deviation (or median and interquartile range (IQR) for non-normal distributions revealed by the Shapiro-Wilk test) of the RMSD and r over subjects were calculated per joint angle and per moment. The absolute values of r were categorized as weak, moderate, strong and excellent for $r \leq 0.35$, $0.35 < r \leq 0.67$, $0.67 < r \leq 0.90$ and $0.90 < r$, respectively [22]. In addition, the intra-class correlation coefficient (ICC) was calculated for the joint angles and moment values at the instant of the peak shoulder elevation angle for the lateral fly, and at the instant of the peak elbow flexion angle for the biceps curl. For the ICC values, >0.7 was considered acceptable [23]. An agreement definition, as opposed to a consistency definition, was used for the ICC, because potential systematic between-system differences were of relevance.

Results

Due to missing markers of the OMS, 12% of the lateral fly trials and 23% of the biceps curl trials had to be excluded. Based on the standard camera and deep-learning method, almost all trials could be included despite three trials (2%) in which joint estimations were completely insufficient because there was a second person visible in the background. For the Kinect, 10% of the trials had to be excluded due to completely insufficient joint tracking. In total, 64 lateral fly repetitions and 58 biceps curl repetitions were included for analysis. Resulting shoulder and elbow joint angles and moments by the different systems are shown in Figures 2,3,4 and 5. RMSDs and Pearson's correlations (r) of the whole movement cycles are represented as mean ± 1 SD over all trials and participants in table 2. ICCs of the peak values are represented in table 2 as well.

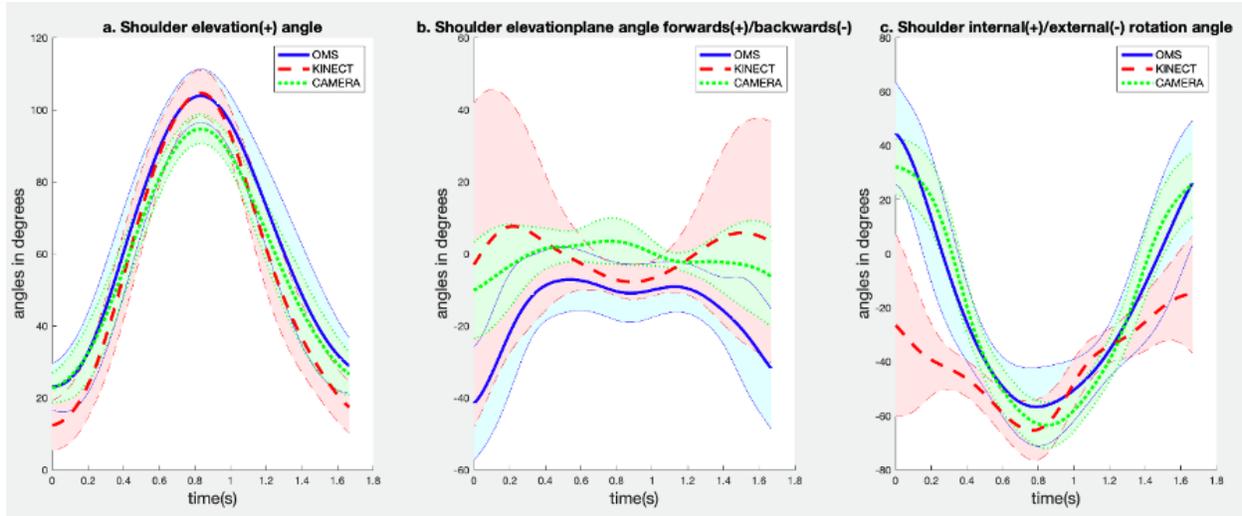


Figure 2. Results for a. shoulder elevation angle, b. shoulder elevation plane angle and c. shoulder internal/external rotation angle during the lateral fly exercise. Results obtained from OMS (blue solid), Kinect (red dashed) and camera (green dotted) are presented as mean \pm 1SD (shaded) over all trials and all participants.

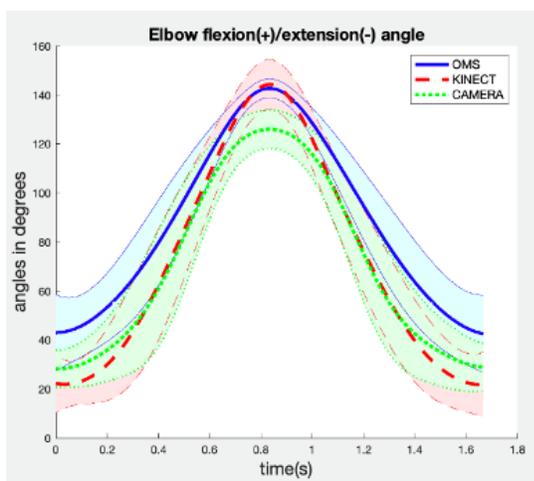


Figure 3. Results for elbow flexion/extension angle during the biceps curl exercise. Results obtained from OMS (blue solid), Kinect (red dashed) and camera (green dotted) are represented as mean \pm 1SD (shaded) over all trials and all participants.

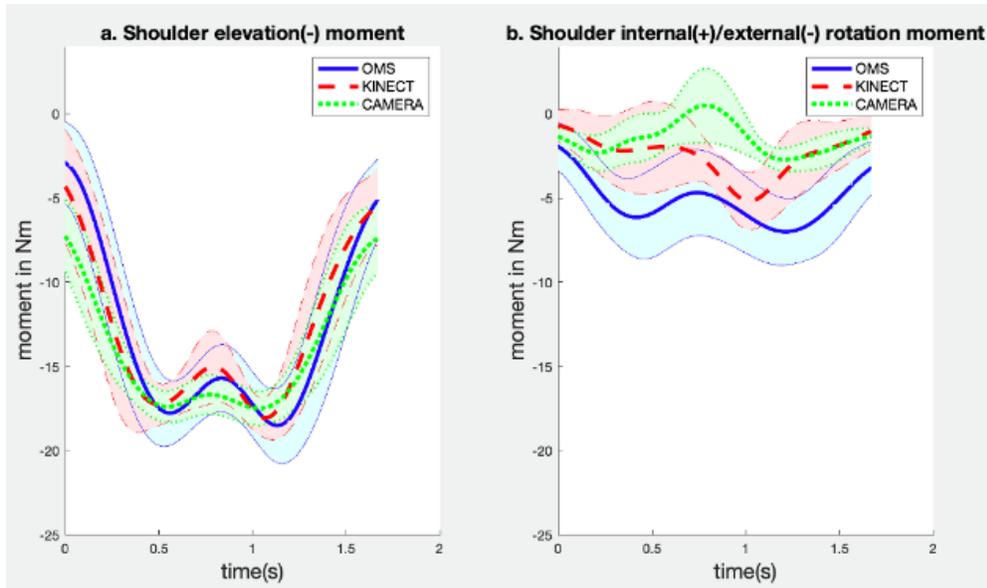


Figure 4. Results for a. shoulder elevation moment and b. shoulder internal/external rotation moment during the lateral fly exercise. Results obtained from OMS (blue solid), Kinect (red dashed) and camera (green dotted) are represented as mean \pm 1SD (shaded) over all trials and all participants.

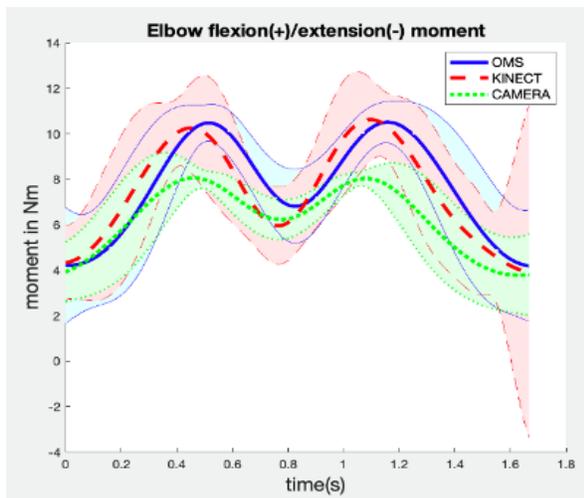


Figure 5. Results for elbow flexion/extension moment during the biceps curl exercise. Results obtained from OMS (blue solid), Kinect (red dashed) and camera (green dotted) are represented as mean \pm 1SD (shaded) over all trials and all participants.

Table 1. Root-Mean-Square-Deviations (RMSD), Pearson's correlation (r) and ICCs between the Kinect and the OMS and between the camera and the OMS of selected variables. RMSDs and correlations were determined over the whole movement cycle and averaged over the means of participants. The ICC agreement was determined from the values at the instant of peak shoulder elevation angle during the lateral fly, and at the instant of peak elbow flexion angle during the biceps curl. For normal distributions the mean and SD were reported, for non-normal distributions the median and IQR were reported and indicated with an asterisk.

Variable	<i>Kinect-OMS</i>		<i>Camera-OMS</i>	
	RMSD (Mean \pm SD or Median \pm IQR*)	r (Mean \pm SD or Median \pm IQR*)	RMSD (Mean \pm SD or Median \pm IQR*)	r (Mean \pm SD or Median \pm IQR*)
<i>Lateral fly</i>				
Shoulder elevation angle (deg)	4.79 \pm 3.88	0.99 \pm 0.00	3.92 \pm 4.12	1.00 \pm 0.00
Shoulder elevationplane angle (deg)	9.86 \pm 9.22	0.31 \pm 0.72*	6.00 \pm 0.99	0.58 \pm 0.37
Shoulder internal/external rotation angle (deg)	13.40 \pm 9.03	0.80 \pm 0.14	7.81 \pm 5.69	0.94 \pm 0.02
Shoulder elevation moment (Nm)	0.74 \pm 1.10	0.99 \pm 0.00	0.99 \pm 1.47	0.99 \pm 0.01*
Shoulder internal/external moment (Nm)	1.08 \pm 1.61	0.85 \pm 0.16	1.41 \pm 2.20	0.76 \pm 0.25

Biceps curl				
Elbow flexion/extension angle (deg)	7.68 ± 9.67	0.98±0.01	7.24 ± 9.34	0.99±0.01
Elbow flexion/extension moment (Nm)	0.78 ± 1.14	0.98±0.01	0.87 ± 1.13	0.99±0.01

Camera-OMS

When comparing the outcome parameters assessed using the camera with those assessed using the OMS, strong to excellent correlations were found for most parameters, including the shoulder elevation angle, the shoulder internal/external rotation angle, the elbow flexion/extension angle, the shoulder elevation moment, the shoulder internal/external moment and the elbow flexion/extension moment, with coefficients ranging from 0.76 to 1.00, and RMSDs ranging from 3.9 to 7.8 degrees for angles and 0.87 to 1.41 Nm for moments (Table 2). The corresponding figures of these parameters (Fig.2-Fig.5) show that shapes and magnitudes over the movement cycle were largely similar for outcome parameters based on the camera and the OMS, although the camera results seemed to underestimate peak angles and moments. In line with this, the ICCs for these variables were all found to be poor (<0.47), indicating poor between-system agreement in the detection of peak values. For the other parameter, the shoulder elevation plane angle, a moderate correlation was found, with a coefficient of 0.58 (Table 2). However, the RMSDs of six degrees of this parameter seemed comparable to the aforementioned parameters. The corresponding figure of this parameters (Fig.2) shows differences in magnitude, but largely similar shapes over the complete movement cycle. Again, the ICCs for this variable was found to be poor (<0.21).

Kinect-OMS

When comparing the parameters assessed using the Kinect with the parameters assessed using the OMS, strong to excellent correlations were found for most parameters,

including the shoulder elevation angle, the shoulder internal/external rotation angle, the elbow flexion/extension angle, the shoulder elevation moment, the shoulder internal/external moment and the elbow flexion/extension moment, with coefficients ranging from 0.80 to 0.99, and RMSDs ranging from 4.79 to 13.40 degrees for angles and 0.74 to 1.08 Nm for moments (Table 2). The corresponding figures of these parameters (Fig.2-Fig.5) show that shapes and magnitudes over the movement cycle were largely similar between the Kinect and the OMS. For the other parameter, the shoulder elevation plane angle, a weak correlation was found, with a coefficient of 0.31 and RMSD of 9.86 degrees (Table 2). When the corresponding figure (Fig.2) is inspected, differences in shape and magnitude can be detected. The ICCs between the Kinect and the OMS are acceptable for two parameters (>0.8): the shoulder elevation angle and the shoulder elevation moment. For the other parameters, the ICC was not acceptable (<0.8), indicating poor between-system agreement in detection of those peak parameters.

Discussion

The hypothesis of the present study was that the single camera deep learning-based method is a feasible and valid way to estimate biomechanical parameters during strength exercises. In line with the expectations, measurements were highly feasible because only one standard camera was required, and post-processing was less effort-consuming than the OMS method because instead of marker trajectory cleaning and gap filling, the only task was to run pre-made scripts with the recorded videos as input to obtain the estimated 3D joint center locations. In addition, results showed that significantly less trials had to be excluded based on missing data for the camera-based method compared to the OMS method. Regarding the validity, for both the camera and the Kinect, most biomechanical parameters were found to have strong to excellent correlations and reasonable RMSDs. However, results were not consistent for all parameters, and the ICCs showed poor absolute agreement on peak values for most parameters, especially for the camera.

The comparison between the camera-based system and the state-of-the-art system revealed strong to excellent correlations and reasonable RMSDs for most joint angles and moments, but also showed an exception. A lower correlation, though similar RMSD, was found for the shoulder elevation plane angle. These low correlations may be the result of the relatively small range in these angles and moments within the strength exercises studied. Correlations are sensitive to data distributions and tend to be weaker when calculations are

performed on a smaller range of values [24]. Nevertheless, it is important that the validity for a larger variety of angles and moments is investigated in future research.

When the RMSDs and correlations of the camera vs OMS based parameters were compared to those of the Kinect vs OMS, it appeared that the camera and Kinect based parameters showed similar agreements to the state-of-the-art on most counts, although the camera appeared to outperform the Kinect for two parameters. These parameters were the shoulder elevation plane angle and the internal/external rotation angle, which yielded higher correlations and lower RMSDs. This result is interesting since the Kinect has the benefit of a depth sensor whereas the camera has not. As aforementioned, a possible explanation may be provided by the different applied pose estimation algorithms. Whereas the Kinect extracted pose information based on a random forest approach from single images, the camera-based method used convolutional neural networks (CNNs) which can incorporate more temporal and spatial information [9, 25]. Therefore, CNNs can provide more accurate and smoother results than random forest approaches which may explain the better and smoother results regarding part of the camera-based parameters.

Unlike the reasonable RMSDs and correlations for whole movement cycles, the agreement between the camera-based parameters and those determined by the state-of-the-art at peak instants were found to be poor, as indicated by the low ICCs in Table 2 and the typical discrepancy around peak values in Fig. 2-5. A potential explanation that may have contributed to this result are the bone length constraints that were incorporated by Pavllo, Feichtenhofer [9] in the camera pose estimation method to improve predictions. These 'soft' constraints included that bone lengths had to remain similar during the trials and had to be similar to the bone lengths of the subjects on which the model was trained [9]. The participants in the present study were relatively tall (mean \pm SD body height: 1.84 ± 0.07 m) and their bone lengths appeared to be underestimated by the model. When bone lengths that are too small must be fitted to data of larger subjects, this may increase errors and decrease consistency in the estimation of the 3D poses and subsequently in the estimation of angles and moments, especially around peak values (i.e. peak elbow flexion), thus potentially leading to lower ICCs. In contrast, the Kinect was not restricted by such a bone length constraint in estimation of 3D joint centers, which could explain why the Kinect seemed to perform more consistently throughout whole movement cycles and has higher ICCs (Table 2). For the future, it is important that the deep-learning model is trained on a larger variety of subjects, to improve accuracy and consistency of obtained biomechanical parameters, especially around

peak values, because valid estimation of absolute peak angles and moments are important in for instance the assessment of injury risk [26, 27].

A strength of the present study is that the employed method to obtain biomechanical parameters from simple videos has not been investigated before, making this study unique. However, a number of methodological limitations apply to the present study. First of all, the present study had a small sample size and only two strength exercises were examined, since it concerned an exploratory study into a new type of technology. Future research should investigate the camera-based motion capture method in a larger sample and during a larger variety of exercises, to obtain reliable validities. Secondly, the OMS was regarded as state-of-the-art, whereas it is known that soft tissue artifacts are associated with this system [3]. Thirdly, the time series of the parameters resulting from the employed systems were synchronized at the instant of peak angles, while for instance a pulse synchronization might have been preferable. The correctness of the synchronization procedure was checked by calculating the delays corresponding to cross-correlations of the time series. The delays were found to be minimal (<0.001 s), indicating that the performed synchronization method was sufficient.

The findings of the present study may have practical implications in the field of sports biomechanics. The camera-based method might be applicable in the future to assess injury risk and performance in a strength training environment, as it is cheap, widely available, not restricted to a controlled environment, and the first results regarding the validity showed reasonable accuracies for a part of the parameters. However, it is evident that improvements are warranted in measurement of single (peak) angles and moments, and more research is needed to investigate this method in a larger variety of exercises and subjects. In addition, it would be interesting to evaluate the method in other sports, in which other challenges like high movement speed and varying distance to the camera may play a role. Moreover, it should be noted that if this method is to be applied in the future to provide direct feedback to athletes on their performance and/or injury risk, a number of practical issues will need to be addressed. Firstly, a considerable amount of computation time was required by the employed deep-learning algorithm to estimate 3D poses. This could be an issue if this method were to be used to provide direct feedback to athletes, in which case it is important that computations can be performed in (close to) real-time. Secondly, during the measurements of the present study, it appeared that the estimations were strongly disturbed when another person was visible in the background. Therefore, future research should focus on improving both the computation time of 3D pose estimation algorithms, as well as the ability to deal with background disturbances.

Conclusion

Biomechanical analysis based on data from a single standard camera combined with a deep-learning algorithm is a feasible method to obtain biomechanical parameters during strength exercises, as it is low-cost, easily applicable in a gym environment, and open-source software is available. In addition, the results provide some first clues regarding the accuracy, showing high correlations and acceptable RMSDs for part of the joint angles and moments cycles compared to the state-of-the-art, but poor agreement on peak angles and moments, which are especially important in assessment of performance and injury risk. Accuracies may be improved by training the model on subjects with a larger variety of anthropometries. It is evident that future research should validate this method in a larger sample and for other exercises and/or other sports situations, and focus on improving the computation time and robustness against background disturbances, if this method is to be employed for a real-time feedback application.

Contributions

Contributed to conception and design: LN, MH, HV, FV

Contributed to acquisition of data: LN

Contributed to analysis and interpretation of data: LN, MH, HV, FV

Drafted and/or revised the article: LN, MH, HV, FV

Approved the submitted version for publication: LN, MH, HV, FV

Funding information

This work is funded by the Dutch Research Council (NWO) under the Citius Altius Sanius perspective program P16-28 Project 4.

Data and Supplementary Material Accessibility

The datasets presented in this article are not readily available because it concerns videos of participants which cannot be shared because of privacy reasons. Requests to access the datasets should be directed to LN, l.noteboom@vu.nl

REFERENCES

- [1] Bini, R.R. and F.P. Carpes, Introduction to biomechanical analysis for performance enhancement and injury prevention, in *Biomechanics of Cycling*. 2014, Springer. p. 1-11. doi: http://dx.doi.org/10.1007/978-3-319-05539-8_1
- [2] van der Kruk, E. and M.M. Reijne, Accuracy of human motion capture systems for sport applications; state-of-the-art review. *European journal of sport science*, 2018. 18(6): p. 806-819. doi: <http://dx.doi.org/10.1080/17461391.2018.1463397>
- [3] Stagni, R., et al., Quantification of soft tissue artefact in motion analysis by combining 3D fluoroscopy and stereophotogrammetry: a study on two subjects. *Clinical Biomechanics*, 2005. 20(3): p. 320-329. doi: <http://dx.doi.org/10.1016/j.clinbiomech.2004.11.012>
- [4] Bonnechere, B., et al., Validity and reliability of the Kinect within functional assessment activities: comparison with standard stereophotogrammetry. *Gait & posture*, 2014. 39(1): p. 593-598. doi: <http://dx.doi.org/10.1016/j.gaitpost.2013.09.018>
- [5] Begon, M., et al., Computation of the 3D kinematics in a global frame over a 40 m-long pathway using a rolling motion analysis system. *Journal of biomechanics*, 2009. 42(16): p. 2649-2653.
- [6] Skals, S., et al., A musculoskeletal model driven by dual Microsoft Kinect Sensor data. *Multibody System Dynamics*, 2017. 41(4): p. 297-316. doi: <http://dx.doi.org/10.1007/s11044-017-9573-8>
- [7] Plantard, P., et al., Pose estimation with a kinect for ergonomic studies: Evaluation of the accuracy using a virtual mannequin. *Sensors*, 2015. 15(1): p. 1785-1803. doi: <http://dx.doi.org/10.3390/s150101785>
- [8] Dutta, T., Evaluation of the Kinect™ sensor for 3-D kinematic measurement in the workplace. *Applied ergonomics*, 2012. 43(4): p. 645-649. doi: <http://dx.doi.org/10.1016/j.apergo.2011.09.011>
- [9] Pavlo, D., et al. 3d human pose estimation in video with temporal convolutions and semi-supervised training. in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2019. doi: <http://dx.doi.org/10.1109/CVPR.2019.00794>
- [10] Luvizon, D.C., D. Picard, and H. Tabia. 2d/3d pose estimation and action recognition using multitask deep learning. in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2018. doi: <http://dx.doi.org/10.1109/CVPR.2018.00539>

- [11] Wade, L., et al., Applications and limitations of current markerless motion capture methods for clinical gait biomechanics. PeerJ, 2022. 10: p. E12995. doi: <http://dx.doi.org/10.7717/peerj.12995>
- [12] Spiers, A., S. Osborne, and M. Li, Engagement in Fitness Activities in England, in The Rise and Size of the Fitness Industry in Europe. 2020, Springer. p. 157-175. doi:
- [13] Kemler, E., H. Valkenberg, and E. Verhagen, More people more active, but there is a counter site. Novice athletes are at highest risk of injury in a large population-based retrospective cross-sectional study. BMJ Open Sport & Exercise Medicine, 2022. 8(1): p. E001255. doi: <http://dx.doi.org/10.1136/bmjsem-2021-001255>
- [14] Shotton, J., et al. Real-time human pose recognition in parts from single depth images. in CVPR 2011. 2011. IEEE. doi: <http://dx.doi.org/10.1109/CVPR.2011.5995316>
- [15] Pavllo, D., et al. 2019; <https://github.com/facebookresearch/VideoPose3D>.
- [16] Pavllo, D., et al. 2019; <https://github.com/facebookresearch/VideoPose3D/blob/main/INFERENCE.md>
- [17] Wu, G., et al., ISB recommendation on definitions of joint coordinate systems of various joints for the reporting of human joint motion—Part II: shoulder, elbow, wrist and hand. Journal of biomechanics, 2005. 38(5): p. 981-992. doi: <http://dx.doi.org/10.1016/j.jbiomech.2004.05.042>
- [18] Plantard, P., et al., Inverse dynamics based on occlusion-resistant Kinect data: Is it usable for ergonomics? International Journal of Industrial Ergonomics, 2017. 61: p. 71-80. doi: <http://dx.doi.org/10.1016/j.ergon.2017.05.010>
- [19] Hof, A.L., An explicit expression for the moment in multibody systems. Journal of biomechanics, 1992. 25(10): p. 1209-1211. doi: [http://dx.doi.org/10.1016/0021-9290\(92\)90076-D](http://dx.doi.org/10.1016/0021-9290(92)90076-D)
- [20] Molenbroek, J., DINED, Anthropometric database. Delft University of Technology, 2004. doi:
- [21] De Leva, P., Adjustments to Zatsiorsky-Seluyanov's segment inertia parameters. Journal of biomechanics, 1996. 29(9): p. 1223-1230. doi: [http://dx.doi.org/10.1016/0021-9290\(95\)00178-6](http://dx.doi.org/10.1016/0021-9290(95)00178-6)
- [22] Taylor, R., Interpretation of the correlation coefficient: a basic review. Journal of diagnostic medical sonography, 1990. 6(1): p. 35-39. doi: <http://dx.doi.org/10.1177/875647939000600106>
- [23] Nunnally, J.C. and I.H. Bernstein, Psychometric theory. 1994. doi:
- [24] Zaki, R., et al., Statistical methods used to test for agreement of medical instruments measuring continuous variables in method comparison studies: a systematic review. PloS one, 2012. 7(5): p. E37908. doi: <http://dx.doi.org/10.1371/journal.pone.0037908>

- [25] Chen, X., et al. Hand pose estimation in depth image using CNN and random forest. in MIPPR 2017: Pattern Recognition and Computer Vision. 2018. International Society for Optics and Photonics. doi: <http://dx.doi.org/10.1117/12.2288114>
- [26] Norman, R., et al., A comparison of peak vs cumulative physical work exposure risk factors for the reporting of low back pain in the automotive industry. *Clinical biomechanics*, 1998. 13(8): p. 561-573. doi: [http://dx.doi.org/10.1016/S0268-0033\(98\)00020-5](http://dx.doi.org/10.1016/S0268-0033(98)00020-5)
- [27] Hansson, G.-Å., et al., Physical workload in various types of work: Part II. Neck, shoulder and upper arm. *International Journal of Industrial Ergonomics*, 2010. 40(3): p. 267-281. doi: <http://dx.doi.org/10.1016/j.ergon.2009.11.002>