

深層学習を用いたウェイトトレーニング動画の ポーズ推定とパフォーマンス評価

Pose estimation and performance evaluation of weight training videos using deep learning

李 映弦（中央大） ○学 三浦 一真（中央大） Sarthak Pathak（中央大学）
正 梅田 和昇（中央大学）

Younghyun LEE, Chuo University, a19.agfb@g.chuo-u.ac.jp
Kazuma MIURA, Chuo University, miura@sensor.mech.chuo-u.ac.jp
Sarthak PATHAK, Chuo University, pathak@mech.chuo-u.ac.jp
Kazunori UMEDA, Chuo University, umeda@mech.chuo-u.ac.jp

In this study, we perform form evaluation for weight training based on skeletal point information obtained from a camera. Performing weight training in the correct form is important for improving muscle strength and reducing the risk of injury. However, weight training is usually performed without an instructor. Therefore, we propose a system that can evaluate trainee's own weight training form using a camera. In existing systems, the positional relationship between the camera and the person is fixed. On the contrary, our system acquires 3D skeleton point information from the video, and thus the system is robust to variations in the positional relationship between the camera and the person.

Key Words: Deep Learning, Pose Estimation, Motion Analysis, Weight Training

1 緒言

近年、健康志向が高まり、フィットネスへの関心が増しており、ジムや自宅での筋トレが一般的に行われている[1]。しかし、個別のトレーニング環境によって、トレーニングフォームの正確性を確認することが難しい場合がある。特に、自宅で一人でトレーニングする際や、ジムにトレーナーが常駐していない場合、正しいフォームを保つことが難しいという課題がある。負傷した人の約 40%は自宅で運動した場合であった[2]。

正確なトレーニングフォームは、トレーニングの効果や安全性に大きな影響を与える。適切なフォームを実践することで、特定の筋肉を効果的に鍛え、けがのリスクを低減させることができる。しかし、鏡がない状況や、トレーナーが不在の場合、自分のフォームを確認しにくいという問題が存在する。先行研究として、Kwon らは OpenCV[3]と Media Pipe[4]を利用して、スクワットの姿勢を評価した[5]。しかし、カメラキャリブレーションの実験的検証が行われておらず、また動画は一定の角度でしか検証されていない。Jacob らは OpenPose[6]を利用して、ダンベルカールのフォームを評価した[7]。ダンベルカールにはこの手法が有効なことが分かったが、他の種目に対しては有効性がわかっていない。そこで、道脇らの動画の撮られた角度にロバストな、StridedTransformer-Pose3D[8]を利用したアーチェリーの姿勢評価の研究[9]に注目した。この手法を元に、筋トレのフォーム評価システムを構築する。

本研究は、ジムで一人でトレーニングする人や、自宅で鏡がない状態で筋トレを行う人々に対して、スマートフォンを活用して動画を撮り、その動画を通じて正しい姿勢でトレーニングを評価する仕組みを提供することを目的とする。動画解析を通じて、ユーザーは自分のトレーニングフォームを確認し、必要に応じて修正できるようになり、効果的で安全なトレーニング環境を構築することが期待される。具体的には、本研究ではダンベルベンチプレス [10] をターゲットにする。

2 提案方法

2.1 システムの概要

単眼のウェブカメラを用いて撮影した運動映像から、姿勢推定モデルによる骨格点の抽出を行う。評価対象となるフレームを抽出して、パフォーマンス評価を行う。その結果を表示する手法を本研究では提案する。

2.2 姿勢推定モデル

動画からの骨格点の検出には、StridedTransformer-Pose3D と呼ばれる手法を利用する。StridedTransformer-Pose3D は、4 台のカメラから 11 人を撮影した約 360 万フレームの動画からなる Human3.6M [11]というデータセットを学習データとし、Vanilla Transformer Encoder[12]と呼ばれるモデルを基にして設計された姿勢推定手法である。2D ポーズシーケンスを取得し、これを活用して 3D ポーズシーケンスを生成する方法を採用している 3 次元姿勢推定手法である。具体的な検出結果を Fig.1 に示す。

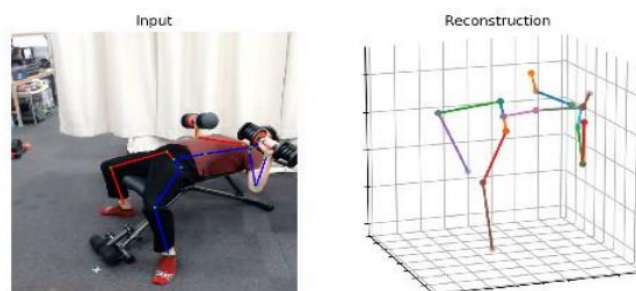


Fig.1 Appearance of skeletal point extraction

2.3 評価対象フレーム抽出

映像から肩、肘、手首が一直線上に一番近くなるフレームを基準として、その以前をダンベルプレスの押し上げるフェーズと、その後を下げるフェーズと呼ぶ。押し上げるときと下げるときで肩、肘、手首が成す角度が 90 度に一番近いフレームを抽出してパフォーマンス評価に使用する。Fig.2 にダンベルプレスの段階別動きを示す。

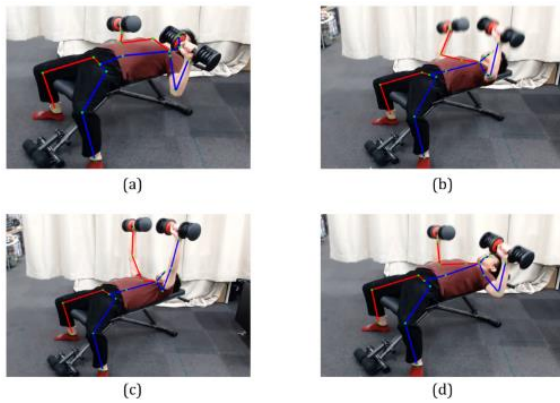


Fig.2 One set of bench press operations

2.4 評価対象

本研究で具体的なウエイトトレーニングとしてダンベルプレスを取り上げる。ダンベルプレスの1セットの動作を Fig.3 に示す。本研究のシステムでは、基本となるフォームの中で以下の姿勢を基準としてパフォーマンス評価を行う。

- (i) 体の中心から両腕が左右対称であるか
- (ii) 腕と体の中心が成す角度が適切であるか
- (iii) ダンベルを押す時と下げる時の軌跡が同じであるか

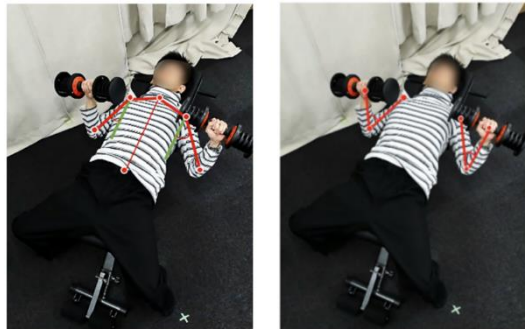


Fig 3. Angle of assessment

- (i) は、右腕の角度と左腕の角度の差の絶対値が小さいほど高得点とする。
- (ii) は、腕の胴体の角度が 60 度に近いほど高得点とする。
- (iii) は、押し上げる時の腕の角度と下げる時の腕の角度の差が低いほど高得点とする。

2.5 パフォーマンス評価

ダンベルプレスを行う映像から姿勢推定モデルを用いて三次元骨格点情報を取得する。提案する評価モデルを適用することでパフォーマンス評価を行う。頭、首、左右の肩・肘・手首の計 8 点の骨格点座標を使用してパフォーマンス評価を行う。

2.4 節で述べた 3 つのポイントのスコアを 0 から 10 の 11 段階に算出し、総合スコアと改善方法を提示する。それぞれのスコアの設定を Table 1 に示す。

Table 1 Threshold setting

点数	評価手法(1)	評価手法(2)	評価手法(3)
0	$21 < \theta$	$\theta < 15 \text{ or } 85 < \theta$	$20 < \theta$
1	$19 < \theta \leq 20$	$15 \leq \theta < 20$ or $80 < \theta \leq 85$	$18 < \theta \leq 20$
2	$18 < \theta \leq 19$	$20 \leq \theta < 25$	$16 < \theta \leq 18$
3	$17 < \theta \leq 18$	$25 \leq \theta < 30$ or $75 < \theta \leq 80$	$14 < \theta \leq 16$
4	$16 < \theta \leq 17$	$30 \leq \theta < 35$	$12 < \theta \leq 14$
5	$15 < \theta \leq 16$	$35 \leq \theta < 40$ or $70 < \theta \leq 75$	$10 < \theta \leq 12$
6	$12 < \theta \leq 15$	$40 \leq \theta < 45$	$8 < \theta \leq 10$
7	$9 < \theta \leq 12$	$45 \leq \theta < 50$ or $65 < \theta \leq 70$	$6 < \theta \leq 8$
8	$6 < \theta \leq 9$	$50 \leq \theta < 55$	$4 < \theta \leq 6$
9	$3 < \theta \leq 6$	$55 \leq \theta < 59$ or $61 < \theta \leq 65$	$2 < \theta \leq 4$
10	$\theta \leq 3$	$59 \leq \theta \leq 61$	$\theta \leq 2$

3 実験

提案手法を用いることで評価対象者を撮影する角度が変化しても本システムのパフォーマンス評価が可能であることを確認するため、検証実験を行った。

3.1 実験環境

実験環境の模式図を Fig.4 に示す。三脚の高さを調節し、ウエイトトレーニングの経験がある 20 代男性がベンチでダンベルプレスを行う様子を、ベンチの角度を 45 度ずつ回転して複数の高さから USB カメラで撮影した。

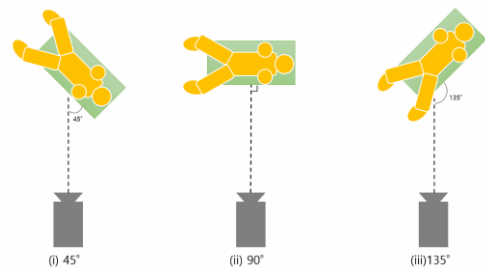


Fig.4 Schematic diagram of the experimental environment

3.2 実験方法

カメラの角度によって画像に写っていない部分が異なるため、カメラの位置は固定して、ベンチを回転して複数の角度から動画の撮影を行った。被験者とカメラの距離は 170cm に固定し、カメラとベンチの角度は 45°, 90°, 135°, 高さは 80cm, 95cm, 170cm の範囲で 9 回撮影を行った。

3.3 実験結果

本システムによって得られた結果を Table 2 に示す。ベンチからカメラまでの距離は一定な状態で、角度と高さを変えたが、本システムを用いることで、カメラの位置によらずおおよそ同じスコアが得られることを確認した。

Table 2 Performance assessment results

カメラ の高さ [cm]	ベンチ 角度 [°]	i の スコア	ii の スコア	iii の スコア	総合 スコア
170	45	4	2.5	9	4.75
170	90	0	6.5	9	5.95
170	135	0	4.5	8	4.65
95	45	0	3.5	8	4.15
95	90	7	7	8	7.30
95	135	0	6	7	5.10
80	45	0	5	9	5.20
80	90	9	6	10	7.80
80	135	5	4	4	4.50

4 結言

本研究では、深層学習による三次元姿勢推定を利用して三次元骨格点情報を抽出し、カメラを固定する必要がなく、改善すべき点が容易に認識できるパフォーマンス評価システムを構築した。具体的には、骨格点情報からベクトルを用いて関節の角度を算出し、評価に用いる2つのフレームを抽出し、関節の角度と骨格点の座標の比較によって0から10の11段階でスコアを算出する手法を提案した。

実験では、カメラと評価対象者の位置関係が変化しても角度と高さに関わらず同じ結果を期待したが、隠れた部分が多くなることによって正しい評価ができない場合があった。高さや角度によっては骨格点情報が正確に取れない映像が多く、検出が上手く行われてないためその情報を3Dに変換する過程で誤りが発生すると思われる。

今後は、評価対象ごとのしきい値の改善、撮影する距離と位置の変更でパフォーマンス評価の精度向上を目指す。また、YOLO[13]でダンベルを検出し、ダンベルの回転角も考慮する。加えて、データセットを取り、物理エンジンを活用して運動部位の負荷計算のシステム構築することも目指す。

参考文献

- [1] Statista. “Revenue of the fitness, health and gym club industry in the United States from 2010 to 2022, with a forecast for 2023”, <https://www.statista.com/statistics/605223/us-fitness-healthclub-market-size-2007-2021/>, Accessed January 2024.
- [2] Chester S. Jones, Carin Christensen and Michael Young, “Weight training injury trends: a 20-year survey”, *Phys Sportsmed*, vol 28, no. 7, pp61-72, 2000.
- [3] Open Source Computer Vision Library, <https://github.com/opencv/opencv>, Accessed January 2024
- [4] Camillo Lugaresi, Jiuqiang Tang, Hadon Nash, et al., “Mediapipe: A framework for perceiving and processing reality”, Third workshop on computer vision for AR/VR at IEEE computer vision and pattern recognition (CVPR), 2019.
- [5] Kwon Yejin and Kim Dongho, “Real-Time Workout Posture Correction using OpenCV and MediaPipe”, *The Journal of Korean Institute of Information Technology*, 2022.
- [6] Mart'inez and Gines Hidalgo. “Openpose: Whole-body pose estimation”, Carnegie Mellon University, 2019.
- [7] Jacob Simon Bernardo, Emanuel Franz Divinagrancia and Kendra Kirsten, “Determining Exercise Form Correctness in Real Time using Human Pose Estimation”, *IVSP*, 2023.
- [8] Wenhao Li, Hong Liu, Runwei Ding, et al., “Exploiting Temporal Contexts with Strided Transformer for 3D Human Pose Estimation”, *IEEE Transactions on Multimedia*, 2021.
- [9] 道脇 拓真, Sarthak Pathak, 梅田和昇: “深層学習による三次元姿勢推定を用いたアーチェリーの射型教示システム”, 日本機械学会ロボティクス・メカトロニクス講演会 2023 講演論文集, 1A2-F07, 2023.6.
- [10] Graham F John, “Dumbbell Bench Press. Strength and Conditioning” vol. 22, no. 4, p. 71, 2000.
- [11] Ionescu Catalin, Papava Dragos, Olaru Vlad, et al., “Human3.6m: Large scale datasets and predictive methods for 3d human sensing in natural environments”, *IEEE Transactions on Pattern Analysis and*

Machine Intelligence, vol. 36, no. 7, pp. 1325–1339, 2013.

- [12] Zhou Haoyi, Zhang Shanghang, Peng Jieqi, et al., “Informer: Beyond Efficient Transformer for Long Sequence Time-Series Forecasting”, *arXiv:2012.07436*, 2020.
- [13] Ultralytics. YOLOv8, <https://github.com/ultralytics/ultralytics>.