

RGB-Dカメラを用いた 筋力トレーニングに対する骨格点情報抽出

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

近年、スポーツやトレーニングにおいてカメラ画像から推定された姿勢を活用した研究が行われている。しかし、既存の学習済み姿勢推定モデルでは、運動の場面で起こりえる人間の姿勢に対して正しく関節の角度を評価できない場合が生じる問題が存在する。本研究では、姿勢推定モデルの再学習と RGB-D カメラを組み合わせることにより、高精度な人間の骨格点情報を抽出する手法を提案する。脇の角度の姿勢変化を 5.52° の誤差で算出することに成功した。

1. 緒言

近年、ジムや自宅で筋トレを行う人が増加している [1]。筋トレは一人で行う場合、トレーニングフォームを確認することが難しい問題が存在する。トレーニングフォームを確認することは、怪我の防止や筋肉を左右対称に発達させるために不可欠である。筋トレ中に負傷した人の約 40% が自宅で運動を行っていた場合であるというデータ [2] が存在し、一人でトレーニングを行う環境でも正確なトレーニングフォームで行えるようにする必要があると言える。安全かつ、効率的な筋トレを行うために、トレーニングフォームを画像から評価する先行研究が行われている。Kwon らは OpenCV [3] と Media Pipe [4] を利用し、スクワットの回数をカウントした [5]。リアルタイムでスクワットを行った回数を 100% の精度で数えることに成功しているが、カメラキャリブレーションが理論的にのみ行われている問題や、スクワット時の関節の角度の精度評価を行っていない課題点が存在する。Jacob らは OpenPose [6] を利用して、リアルタイムでダンベルカールのフォーム評価を行った [7]。ダンベルカール時の身体のゆれや腕の可動域等に着目し、入力されたダンベルカールの動画が良いフォームか否かを機械学習アルゴリズムと組み合わせ定量的に評価している。90% 以上の精度で良いフォームと悪いフォームの判別を行うことに成功している。しかし、OpenPose を使用している為、腰のキーポイントが評価に必要な筋トレに適応できない問題や、被験者にフィードバックを行う際に判別理由を提示できない課題が存在する。李らは、道脇らが行った StridedTransformer-Pose3D [8] を利用した撮影された動画の角度にロバストなアーチェリーの射型評価の研究 [9] に着目し、StridedTransformer-Pose3D から得た関節の角度からダンベルベンチプレスの姿勢評価の研究 [10] を行った。しかし、オクルージョンが含まれる場合、誤った 3D 骨格点が推定され、姿勢評価に利用されてしまう点や、使用した事前に学習された姿勢推定モデルは Human3.6M [11] で学習されている為、ダンベルベンチプレスのような日常生活から離れた姿勢の推定には有効的でない課題が存在する。また、骨格点の動作から深層学習を用いて動作の良し悪しを判断する場合、動きが悪いと判断した理由を人がわかる形で説明を行えない課題が存在する。そこで、本手法では少ない学習データで姿勢推定モデルを学習することで、ダンベルベンチプレスに対して関節の角度を高精度で計測できる手法の提案を行う。

2. 提案手法

2.1 システム概要

RGB-D カメラを用いて撮影した運動の動画から、評価対象となるフレームに対して姿勢推定モデルにより 2D の骨格点抽出を行う。抽出された 2D 骨格点を RGB-D カメラから取得された Depth 情報と組み合わせることで、3D の骨格点抽出を行う。抽出された 3D 骨格点から運動時の関節の角度を算出し、評価を行う。

カメラは RealSenseD435 を用いており、解像度は 1280×720 ピクセル、フレームレートは 30fps で使用する。

2.2 姿勢推定モデル

動画から骨格点を抽出する際には、DeepLabCut [12] と呼ばれる既存手法を利用する。DeepLabCut は学習済みの ResNet [13] をエンコーダーに持つことで、20 枚から 200 枚の画像を使って転移学習するだけで、動画に対して任意の対象の任意のキーポイントを推定できるシステムである。今回は一人がベンチプレスを行っている動画から 20 枚のフレームを ResNet101 で 10 万エポック学習させる。学習させるキーポイントは、図 1 に示すように、目、首、右肩、右肘、右手首、左肩、左肘、左手首、腰、骨盤、右股関節、右膝、右足首、左股関節、左膝、左足首の計 16 点を学習させた。



図 1 学習キーポイント

2.3 評価対象フレーム抽出

本研究では、筋トレの動画像処理に対する基礎検討として、動画からフレーム抽出された静止画に対し、姿勢の角度の評価を行う。評価に使用するフレームは、図 2 のように、ダンベルベンチプレスの際にダンベルを

押し上げる前の姿勢を撮っているフレームと、ダンベルを押しあげきった姿勢のフレームを抽出する。

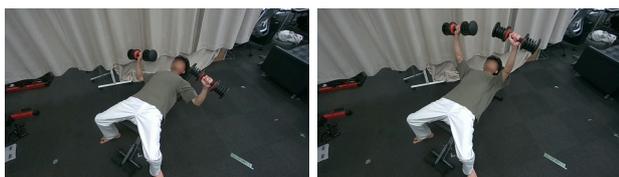


図 2 評価対象フレーム

2.4 3D 骨格点算出

Intel から提供されている RealSense の SDK を利用して、RGB 動画と Depth 画像に対して画角合わせを行う。その後、DeepLabcut を利用して、RGB 画像上で 2D の骨格点を推定する。RGB 画像上で推定された座標と対応する Depth 画像上で点を参照して、骨格点の Depth 情報を取得する。取得された Depth 情報を RealSense の SDK を利用して、カメラの内部パラメータと深度情報から $[x, y, z]$ の 3 次元座標に変換する。以上のようにして得られた 3 次元骨格点情報を姿勢の評価に使用する。

2.5 評価基準

評価対象は以下の 3 つの角度に設定する。

- (1) ダンベルを押し上げる前の脇角度
- (2) ダンベルを押し上げる前の肘角度
- (3) ダンベルを押し上げた後の腕角度

(1) は身体の首と骨盤のキーポイントを結んだ線に対して、肩と肘を結んだ線の角度を計算することで算出する。(2) は肩と肘のキーポイントを結んだ線に対して、肘と手首のキーポイントを結んだ線の角度を計算することで算出する。(3) は肩と骨盤のキーポイントを結んだ線に対して、肩と手首を結んだ線の角度を計算することで算出する。フレームごとに算出された角度は RealSense の精度を考慮して、姿勢を維持している 1 秒間の平均角度で評価する。

3. 実験

3.1 実験概要

カメラの位置を固定して、ベンチプレス時の関節の角度を変えながら実験を行った。ダンベルを持ち上げる前の状態で 9 回、持ち上げた状態で 3 回撮影した。カメラの高さは 220cm、カメラとお腹までの距離を 255cm に設定した。提案手法で算出した角度を比較するために、OptiTrack 社の V120: Trio を用いて、角度を取得した。

3.2 実験結果

ダンベルを持ち上げる前の状態での実験結果を表 1 に示す。ダンベルを持ち上げた状態での実験結果を表 2 に示す。

3.3 考察

脇の角度に関して、9 本中 8 本で左右の脇の角度の大小関係が D435 と V120: Trio で一致する結果が得られた。しかし、左脇は平均して 5.05° D435 の方が大きく

表 1 ダンベルを持ち上げる前の脇と脇の角度 $[^\circ]$

番号	カメラ	右脇	左脇	右肘	左肘
1	V120:Trio	34.4	30.2	92.5	93.8
	D435	21.1	29.0	66.3	76.8
2	V120:Trio	28.5	33.3	89.3	85.0
	D435	13.2	35.2	76.0	74.5
3	V120:Trio	31.2	35.8	87.0	90.0
	D435	17.3	32.1	68.7	74.4
4	V120:Trio	22.3	25.2	85.4	92.9
	D435	13.3	22.2	68.4	61.2
5	V120:Trio	27.6	36.1	86.7	89.3
	D435	18.8	37.7	63.3	69.2
6	V120:Trio	20.4	31.9	86.8	85.2
	D435	16.5	36.8	55.6	64.3
7	V120:Trio	22.9	28.9	79.2	77.8
	D435	11.4	46.0	54.9	61.9
8	V120:Trio	29.0	36.8	73.9	70.7
	D435	16.4	53.2	52.5	47.5
9	V120:Trio	31.0	42.3	75.2	71.2
	D435	19.7	54.6	50.7	47.7

表 2 ダンベルを持ち上げた腕の角度 $[^\circ]$

番号	カメラ	右腕	左腕
1	V120: Trio	59.8	63.2
	D435	72.0	55.3
2	V120: Trio	79.9	81.7
	D435	106.1	89.6
3	V120: Trio	90.8	90.7
	D435	122.4	108.6

算出された。原因としては、V120: Trioを使用する際に身体に取り付けた反射マーカの座標と DeepLabCut で検出されたキーポイントの座標のずれが考えられる。右脇は平均して 11.05° D435 の方が小さく算出された。右脇の方が誤差が大きい理由は、反射マーカの座標と DeepLabCut で検出されたキーポイントの座標のずれに加えて、DeepLabCut で検出されたキーポイントの座標が RGB 画像と Depth 画像間で正しく対応していないため、特に右手首の Depth 情報の誤差が大きくなっていったことが考えられる。これは D435 のキャリブレーションが正しく行われていなかったため、RGB 画像と Depth 画像で座標がずれていたと考えられる。肘の角度に関して、9 本中 6 本で左右の脇の角度の大小関係が D435 と V120: Trio で一致する結果が得られた。

また、2 本を選択して V120: Trio で算出した角度の 2 本の動画での差と D435 で算出した角度の 2 本の動画での差をそれぞれ求めた。さらに、求めた V120: Trio での角度の差と D435 での角度の差の差を求めることで、角度の変化量に関して、V120: Trio を基準として D435 の結果を比較した。9 本の中から 2 本を選択する方法は 36 通りあるので、36 通り行くと結果は表 3, 4 となる。表 3, 4 より、平均して D435 の方が右脇で 1.45° 左脇で 8.14° 小さく角度の変化が算出されることが分かった。肘の角度変化を 5.52° の誤差で算出できた。

表 3 右脇の動画間での角度の差 [°]

動画ペア	差	動画ペア	差
1 と 2	2.02196	4 と 5	-0.2124
1 と 3	0.52627	4 と 6	-5.11382
1 と 4	-4.32765	4 と 7	2.45481
1 と 5	-4.54005	4 と 8	3.57845
1 と 6	-9.44147	4 と 9	2.33675
1 と 7	-1.87284	5 と 6	-4.90142
1 と 8	-0.7492	5 と 7	2.66721
1 と 9	-1.9909	5 と 8	3.79085
2 と 3	-1.49569	5 と 9	2.54915
2 と 4	-6.34961	6 と 7	7.56863
2 と 5	-6.56201	6 と 8	8.69227
2 と 6	-11.46343	6 と 9	7.45057
2 と 7	-3.8948	7 と 8	1.12364
2 と 8	-2.77116	7 と 9	-0.11806
2 と 9	-4.01286	8 と 9	-1.2417
3 と 4	-4.85392		
3 と 5	-5.06632		
3 と 6	-9.96774		
3 と 7	-2.39911		
3 と 8	-1.27547		
3 と 9	-2.51717		

表 4 左脇の動画間での角度の差 [°]

動画ペア	差	動画ペア	差
1 と 2	-3.1074	4 と 5	-4.58381
1 と 3	2.48976	4 と 6	-7.78148
1 と 4	1.78284	4 と 7	-20.07571
1 と 5	-2.80097	4 と 8	-19.32081
1 と 6	-5.99864	4 と 9	-15.25741
1 と 7	-18.29287	5 と 6	-3.19767
1 と 8	-17.53797	5 と 7	-15.4919
1 と 9	-13.47457	5 と 8	-14.737
2 と 3	5.59716	5 と 9	-10.6736
2 と 4	4.89024	6 と 7	-12.29423
2 と 5	0.30643	6 と 8	-11.53933
2 と 6	-2.89124	6 と 9	-7.47593
2 と 7	-15.18547	7 と 8	0.7549
2 と 8	-14.43057	7 と 9	4.8183
2 と 9	-10.36717	8 と 9	4.0634
3 と 4	-0.70692		
3 と 5	-5.29073		
3 と 6	-8.4884		
3 と 7	-20.78263		
3 と 8	-20.02773		
3 と 9	-15.96433		

4. 結言

本研究では、深層学習による二次元姿勢推定モデルと Depth カメラを組み合わせることで、三次元骨格点情報の抽出を行い、筋トレ時の関節の角度を算出する手法を提案した。実験では、関節の角度を変えながら正しく人がとっている姿勢の角度を算出できるか実験を行った。結果から 5.52° の誤差範囲で角度を求めることが分かった。誤差の原因として、D435 のキャリブレーションが考えられる。また、反射マーカと DeepLabCut で検出する座標を合わせることで、角度の評価の精度が向上することが考えられる。今後は、被験者が左右対称な姿勢でトレーニングを行えるようにリアルタイムでフィードバックを行えるシステムの構築を目指す。

参考文献

- [1] “スポーツライフ・データ 2022”, https://www.ssf.or.jp/thinktank/sports_life/datalist/2022/index.html
- [2] “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-health-club-market-size-2007-2021/>
- [3] “Open Source Computer Vision Library”, <https://github.com/opencv/opencv>
- [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 ´ınez and Gines Hidalgo : “Openpose: Whole-body pose estimation”, Carnegie Mellon University, 2019.
- [7] Jacob Simon Bernardo, Emanuel Franz Divinagracia 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指示システム”, 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.
- [10] 李 映弦, 三浦 一真, Sarthak Pathak, 梅田和昇 : “深層学習を用いたウエ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”, 日本機械学会ロボティクス・メカトロニクス講演会 2024 講演論文集, 1P2-H07, 2024.
- [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] Alexander Mathis, Pranav Mamidanna, Kevin M. Cury, et al. : “DeepLabCut: markerless pose estimation of user-defined body parts with deep learning”, Nature Neuroscience, vol.21, pp.1281–1289, 2018.
- [13] Kaiming He, Xiangyu Zhang, Shaoqing Ren, et al. : “Deep residual learning for image recognition”, 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.