

セグメンテーションマスクを活用した器械体操の骨格推定

Skeleton Estimation in Gymnastics Using Segmentation Masks

寺本 直暉^{*1}, 越智 洋司^{*2*3}
Naoki TERAMOTO^{*1}, Youji OCHI^{*2}

^{*1} 近畿大学理工学部

^{*1} Faculty of Science and Technology, Kindai University

^{*2} 近畿大学情報学部

^{*2} Faculty of Informatics, Kindai University

^{*3} 近畿大学情報学研究所

Cyber Informatics Research Institute, Kindai University

Email: ochi@kindai.ac.jp

あらまし：近年、AI（人工知能）や画像認識を利用したデータ解析が注目されている。本研究は器械体操の骨格推定に注目した。器械体操の骨格は誤推定されることがあるため、セグメンテーションマスクを活用した骨格推定の補完を行うことにした。本研究の骨格推定の精度を仮の真データと比較を行った。精度は全体的に上がったが、一部骨格が誤って重なるなど問題点もあった。今後改良を含め、器械体操の自動採点システムに応用していきたい。

キーワード：器械体操, MediaPipe, 骨格推定, セグメンテーションマスク, 人物領域

1. はじめに

近年、AI（人工知能）や画像認識を利用したデータ解析が注目されている⁽¹⁾。中でもスポーツ界では選手にセンサーなどを装着しない非接触型センシング技術を用いた研究事例もある⁽²⁾。そこで書かれているセンシング技術は画像の情報を読み取り、骨格推定を行っている。しかし既存の機械学習のみでは、競技特有の精度が低いことが挙げられている⁽³⁾。そこで本研究では、セグメンテーションマスクを活用した器械体操の骨格推定を行うことにした。

2. 研究内容

本研究では MediaPipe を用いて骨格推定を行ったものに対して修正を行う⁽⁴⁾。MediaPipe は既に機械学習された骨格推定などを搭載したライブラリである。しかしこのライブラリを用いて実際に骨格推定を行うと、左右の手足の骨格の入れ替わりや人物領域外を骨格座標として捉えることなどの様々な誤推定があった。本研究では主に誤推定している骨格座標を人物領域内に修正するため、セグメンテーションマスクを用いた骨格推定の補完を提案する。図1はセグメンテーションマスクと MediaPipe の骨格推定を示した様子である。本研究はこれらの情報を用いて骨格推定を行う。

2.1 セグメンテーションマスクの利用

本研究で用いるセグメンテーションマスクとは人物領域の部分を二値化したマスク（データ）のことを指している。本研究では Detectron2 を用いた⁽⁵⁾。本研究では機械学習モデルの再学習を行わず、セグメンテーションマスクや MediaPipe の骨格推定を用いて修正を行う。この手法により再学習のコストなどがかからないというメリットがある。

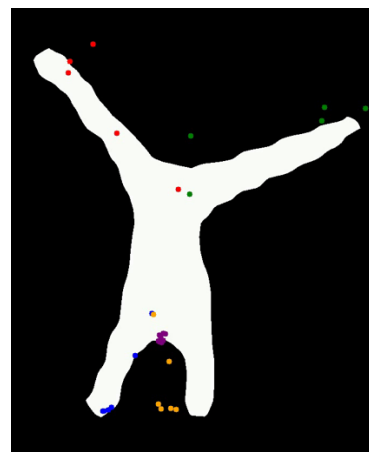


図1 セグメンテーションマスクと MediaPipe の骨格推定を示した様子

2.2 実験環境

研究対象の動画は器械体操の経験者の動画 mp4 を扱った。対象動画の撮影条件として、屋内のガラスや鏡が映らない場所で対象人物を1人とし、対象人物との距離が3m～5mの間、横1920ピクセル縦1080ピクセルの動画を扱う。動画は一般的なスマートフォンで撮影を行い、30fpsの動画を扱う。撮影の動画は器械体操の経験者の側転と前方倒立回転の2本を対象とし、主に側転の動画で評価を行った。

3. 提案手法

本研究ではまず、カラー動画から人物領域のデータを取得し保存する。次にカラー動画から MediaPipe の骨格データを取得し保存を行う。その後人物領域を小さく縮小させ、その人物領域のデータを保存する。その後骨格推定の事前修正を行いその骨格データを保

存する。事前修正後、人物領域に収まるように最近傍点探索で骨格推定を行い、データを保存する。その後骨格点が示された動画を作成し、その動画で骨格点が示された動画を観覧することができる。

4. 実験評価

本研究では評価実験を大きく4つに分けて行った。4.1節では本提案のプログラムが期待通りに動作しているかの確認、4.2節では目視でしか確認できない場面の確認、4.3節、4.4節では仮の真データとの比較した際の距離や角度で確認を行った。本研究の真のデータは把握するのが困難であるため、本研究では目視での手入力で行なった。本研究では距離や角度の結果の値は最も近い整数に丸めて比較を行った。

4.1 人物領域内に収まった骨格の数

人物領域内に収まった骨格の数については側転の動画に対して行った。この動画は1フレームあたり33個の骨格で166フレームの動画であり、計5478個を対象にした。MediaPipeでの骨格推定で312個の人物領域外の骨格が確認されたが、本提案の骨格推定では人物領域外の骨格が0個という結果になった。この結果は本提案のプログラムが期待通りに動いていることを示していると考えられる。

4.2 目視での確認

次に目視での確認を行った。目視での確認の動画は側転、前方倒立回転の動画で確認を行った。図2は実際に目視で確認した動画の一部を切り抜いた場面だが、足の骨格が全体的に収まっていることが分かる。確認しづらい点が少し重なっていることも分かった。目視での確認では、左右の骨格の入れ替わり、明らかに骨格がおかしい場面や人物領域外での骨格推定について確認した。この3つの問題は概ね改善されていることが確認できた。しかし一部フレームで人物領域内に収まっているが、点が誤って重なってしまった場面もあったという結果になった。



図2 目視での確認場面

4.3 仮の真データと比較した際の距離

仮の真データと比較した際の距離について、MediaPipeと本研究の手法の距離を表1に示した。この表は全骨格で骨格ごとに、仮の真データとの距離平均、最大距離を示したものである。表を見る限り、距離が全体的に小さくなっていることが分かる。距離が小さくなるということは正しいと思われる座

表1 仮の真データと比較した際の距離の表

	距離平均	最大距離
MediaPipeの提案手法	22	672
本研究の提案手法	18	148

標に近づいているということであるため、精度が上がっていると考えられる。

4.4 仮の真データと比較した際の角度

仮の真データと比較した際の角度について、角度は肘や膝で確認を行った。表2は特に大きく変化した肘や膝の角度である。側転の動画のフレームでAが118フレーム、Bが125フレーム、Cが108フレームの場面である。Aのように真のデータに近付くものもあれば、Bのように遠ざかってしまうフレームもあった。Cはただ角度が入れ替わったという場面だった。特に時系列的に見て、角度が急激に変化してしまうという角度があった。この角度に対して今後対策が必要と考えられる。

表2 特に大きく変化した肘や膝の角度

	MediaPipeの骨格推定	本研究の骨格推定	仮の真データ
A: Knee Left	135度	160度	172度
B: Elbow Right	91度	70度	170度
C: Knee Left	98度	171度	166度
C: Knee Right	171度	93度	165度

5. おわりに

評価実験より、本研究の骨格推定は全体的に精度が上がっていると考えられる。しかし誤った座標の重なりや、急激な角度の変化に対する改善点が必要という結果になった。今後は、初心者に対して経験者との違いを数値や、視覚的にもわかりやすい支援型システムを目指す。

謝辞

本研究はJSPS 科研費 22K12304 の助成を受けた。

参考文献

- (1) 清水千弘, et al. エディトリアル: 特集「スポーツ競技と AI」. 人工知能, 2019, 34.4: 492-496.
- (2) 田摩雅基, et al. 画像センシング技術を活用したスポーツ ICT の取り組み. O plus E: Optics+ electronics, 2018, 40.4: 420-427.
- (3) 長谷川雄大, et al. 体操競技における自動採点のための骨格推定. 第 82 回全国大会講演論文集, 2020, 539-540.
- (4) MediaPipe | Google for Developers, <https://developers.google.com/mediapipe> (参照 2024-2-7)
- (5) Detectron2, <https://github.com/facebookresearch/detectron2> (参照 2024-2-7)