

時系列相関性を用いた姿勢推定モデルの骨格推定精度改善

山川 敦也[†] 石川 孝明[‡] 渡辺 裕^{†‡}

[†]早稲田大学基幹理工学部情報通信学科 [‡]早稲田大学国際情報通信研究センター

1. はじめに

人間の姿勢推定という課題はコンピュータビジョンの分野で注目され続けてきた問題である。

近年、画像から人間の骨格を推定する技術が多く提案されている。特に 2014 年に Karen らが提案した DeepPose[1]は深層学習を活用した最初の主要モデルであり、姿勢推定技術に大きな変化を与えた。Convolutional Neural Network(CNN)を用いたアプローチは、人間が様々な姿勢をとっていても、柔軟な姿勢推定を可能にした。これらの手法はマーカや撮影スタジオなどの特殊な撮影環境が必要ないという利点があり、スポーツや介護などで活用できるアプリケーションが多数提案されている。しかし、一般的なカメラで撮影された動画に対して適用した場合は、撮影条件や撮影方法によって骨格検出精度が大きく左右される。骨格推定をより多くの現場で手軽に利用するためには、撮影環境に対して推定精度がある程度頑健である骨格推定モデルを作成する必要がある。

本研究では、既存の姿勢推定モデルを動画に適用した場合の骨格推定精度を改善する手法を提案する。

2. 関連研究

2.1. OpenPose

OpenPose[2]は入力画像に映る人物の目や関節などのキーポイントを検出できる人体骨格推定モデルである。検出できるキーポイントの数は、トレーニングデータセットの異なる 2 つのモデルによって 18 個、あるいは 25 個のキーポイントが出力値として得られる。それぞれのキーポイントは、画像上の x 座標と y 座標、すなわち 2 次元座標値として得られる。出力値のデータフォーマットは JSON, XML, YML のいずれかを指定できる。また、OpenPose 動画に対しても適用できる。この場合、動画の各フレームを静止画と見なして連続的に処理していく。

2.2. DeepSORT

DeepSORT は Simple Online Realtime Tracking(SORT)[3]に深層学習を組み込んだフレームワークである。SORT ではオブジェクトを検出すると、カルマンフィルターによって対象物の絶対座標と速度係数を用いて、線形速度モデルを想定したオブジェクトトラッキングを実現する。しかし、実際の映像ではオクルージョンが発生したり、様々な視点からオブジェクトを撮影することがある。そのような場合、SORT は高いトラッキング精度を出せない。

そこで、DeepSORT では「appearance」という特徴ベクトルを距離関数として導入し、オクルージョンや多視点映像に対するトラッキング精度を改善した。

3. 提案手法

3.1. OpenPose の問題点

OpenPose の問題点として、モーションキャプチャと比較して検出精度が低いことが挙げられる。光学式モーションキャプチャは、複数カメラを正確にキャリブレーションすることで多くの情報量を取得できる。そのため、測定対象者の細かい動きの変化を用いた骨格推定が可能である。OpenPose は 2 次元画像から骨格を推定するため、オクルージョンが発生するとキーポイントの誤検出に繋がる。

また、入力画像の空間的な解像度あるいは時間的な解像度が低い場合も誤検出が発生する。空間的な解像度が低い場合とは、入力画像内の非常に小さな領域に人物が映っている場合や、そもそも入力画像の解像度が低い場合を指す。また、時間的な解像度が低い場合とは、測定対象者の動きがシャッタースピードに対して速すぎる場合を指す。

3.2. 検出失敗フレーム

本研究では動画に対して OpenPose を適用した際、キーポイントの検出が失敗している箇所を検出失敗フレームと呼ぶことにする。

OpenPose では、検出できなかった部位の座標が x 座標、 y 座標ともに 0 となる。そこで、 $k-1$ フレーム目と $k+1$ フレーム目で座標値が取得できているが、 k フレーム目で x 座標、 y 座標ともに 0 となっている箇所を検出失敗フレームとして定義し、補間対象とする。図 1 に例を示す。図 1 では 3 フレーム分の画像を示している。このとき 2

Improving Pose Estimation Models Using Time Series Correlation

[†]Atsuya YAMAKAWA[†] Takaaki ISHIKAWA[‡] and Hiroshi WATANABE^{†‡}

[†]Waseda University

[‡]Global Information and Telecommunication Institute, Waseda University

フレーム目のフレームが検出失敗フレームとして定義される。

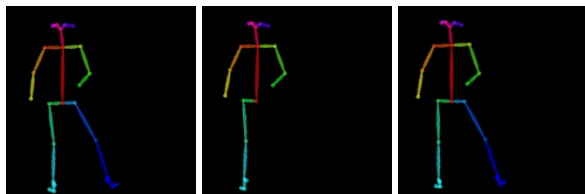


図1 検出失敗フレームの例

3.3. 誤検出フレーム

本研究では動画に対して OpenPose を適用した際、キーポイントの誤検出が発生している箇所を誤検出フレームと呼ぶことにする。

具体的には、人物ごとに毎フレーム座標値の差分を算出していく。この差分とはすなわちキーポイントの変化量である。この変化量が過去数フレーム分の変化量と比較して明らかに大きい場合、そのフレームを誤検出フレームと定義し、補正対象とする。これは、人間の動きが微小時間に急速に変化しないことを根拠としている。図2に例を示す。図2では3フレーム分の画像を示している。2フレーム目のフレームだけ、左右の足がそれぞれ逆の足として検出されている。このとき、2フレーム目のフレームは誤検出フレームとして定義される。

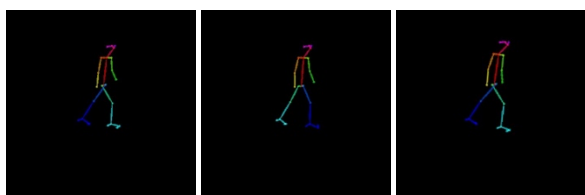


図2 誤検出フレームの例

3.4. キーポイントの座標値補間及び補正

本研究では検出失敗フレーム及び誤検出フレームに対するキーポイント座標値の補間及び補正手法を提案する。

検出失敗フレーム、誤検出フレームとして定義したフレームは正常に骨格情報を取得できなかったシーケンスである。これらに対しては線形補間を用いて座標値を補間及び補正する。

検出失敗フレームの場合、1フレームだけ x 座標、 y 座標が0となっているキーポイントは、前後フレームの座標値を参照し、その中点を算出してそれをキーポイント座標値とする。

誤検出フレームの場合、閾値を超える変化量の座標値は前後フレームの座標値を参照し、その中点を算出してそれをキーポイントの座標値とする。

キーポイントの座標値の中点を算出にあたって、画像内の人物とJSONファイル内の骨格座標値を対応させる必要がある。本研究では、OpenPoseのPython APIとDeepSORTを組み合わせることで画像内の各人物に識別

子を付与し、それを骨格座標値とともに出力する。したがって、人物ごとの識別子を参照することで、複数フレーム間で同一人物の骨格座標値を参照可能になる。

4. 実験

4.1. 実験手法

座標値の補間及び補正による骨格検出精度向上を検証するために、サッカーと野球のスポーツ動画と、定点カメラで撮影した複数人物の動画を用いて実験した。動画の補間及び補正する前の骨格推定結果では、検出失敗や誤検出が確認できた。例えば定点カメラで撮影した動画では、正立している人物の右足の座標値が1フレームだけ検出できていないフレームが存在した。また、サッカーの動画では選手がドリブルしているシーンで右足と左足がそれぞれ逆の足として検出されているフレームが存在した。

4.2. 実験結果

複数の動画に対して骨格を推定し、座標値の補間及び補正をした結果、一部の検出失敗フレーム及び誤検出フレームの座標値が正しく補正されることを確認した。補正の前後での骨格推定結果を比較すると、検出できなかったキーポイントが補間され、誤検出されたキーポイントが補正された。また、補間及び補正後の骨格座標はgroundtruthの骨格位置に近づいた。

実験結果から、動画全体としての骨格推定精度が向上したといえる。しかし、全ての検出失敗及び誤検出を補正できていない。具体的には、複数フレームにまたがる検出失敗及び誤検出は、補間及び補正対象にはならなかった。

5. 結論

本研究では、動画に対してOpenPoseを適用した場合の骨格推定精度を向上させる手法を検討した。今後は、検出失敗フレーム及び誤検出フレームの定義方法を見直すことでより多くの検出失敗と誤検出を補正し、更なる精度向上を図る。また、線形補間以外の補間方法で座標値を補間した場合を検証し、より正確に補間するための手法を検討する。

文 献

- [1] Nicolai Wojke, Alex Bewley, Dietrich Paulus: "Simple Online and Realtime Tracking with a Deep Association Metric", In ICIP, pp 3645-3649, 2017.
- [2] Z. Cao, T. Simon, S. Wei, Y. Sheikh: "Realtime Multi-Person 2D Pose Estimation using Part Affinity Field," In CVPR, pp7291-7299, 2017.
- [3] Alex Bewley, Zongyuan Ge, Lionel Ott, Fabio Ramos, Ben Upcroft: "Simple Online and Realtime Tracking" In ICIP, pp3464-3468, 2016.