

# OpenPose を用いた投球シーン照合の一検討

大澤 遼平<sup>†</sup> 石川 孝明<sup>‡</sup> 渡辺 裕<sup>‡</sup>

早稲田大学 基幹理工学部<sup>†</sup> 早稲田大学国際情報通信研究センター<sup>‡</sup>

## 1. はじめに

近年、野球選手の技術向上のために、選手の動作を撮影した映像データが利用されている。映像は繰り返し視聴できるため、動作を細部まで分析することが可能である。また、肉眼では認識し難い特徴や癖などを発見することが容易になる。

動作解析における映像データの活用方法として、複数映像の比較がある。同一の動作を含む映像を見比べることにより、動作の相違点を発見することが可能になる。野球技術が優れている選手の動作と自身の動作を比べることで、上達するために改善すべき箇所や取り組むべき課題を明確にすることができる。また、練習時と試合時のように異なる状況での自身の映像を比較することで、緊張感や疲労などが動作に影響を及ぼすかを確認できる。

動作の細部まで比較するためには、動作のタイミングを同期させることが必要となる。しかし、動作の速度や身体の使い方は選手や状況によって異なるため、複数動作のタイミングを正確に合わせることは容易ではない。

そこで、本研究では野球の投球映像を使用し、二つの投球動作のタイミングを照合する手法について検討する。まず、OpenPose により投球映像から投手の身体座標を取得する。得られた身体座標データから投手の姿勢の類似度を算出し、Dynamic Time Warping を用いて動作の各タイミングを対応付ける。

## 2. 関連研究

### 2.1. OpenPose

映像中に存在する人物の身体座標を取得する手法として Cao らが提案した OpenPose[1]がある。OpenPose は人間の身体にセンサーなどを装着することなく、映像内の複数人の身体部位をリアルタイムに検出可能な手法である。使用するモデルによって、15, 18 もしくは 25 部位の位置を推定できる。顔や手が鮮明に映っている場合には、顔の輪郭や指の関節など、より細かい部位も検出可能である。映像のフレームごとに、各部位の 2 次元座標が取得され、推定の信頼度を示す確度が部位ごとに算出される。

### 2.2. Dynamic Time Warping

Dynamic Time Warping(DTW)は二つの系列が最も類似するように、系列内の要素を対応付ける手法である。二つの系列の要素間の類似度を総当たりで算出し、類似度の合計が最小となるように対応付けを決定する。DTW は二つの系列の要素数が異なる場合でも、系列同士の類似度を求めることができる。

DTW の特徴として、二つの系列のすべての要素を対応付けることが挙げられる。一方の系列のすべての要素は、もう一方の系列に類似する要素が存在しない場合でも、もう一方の系列の少なくとも一つの要素と対応付けされる。そのため、動作照合に DTW を用いる場合、映像内の動作の始点と終点を主観で付与する必要がある。そこで、要素間の類似度に基づき、系列の中から動作の始点および終点を自動で抽出する DTW が提案されている[2]。

### 2.3. 投球シーン照合

投球シーン照合として、文献[2]で提案されている動作の始点終点を自動で抽出可能な DTW を用いた手法[3]が存在する。関連手法では、投球動画に対して OpenPose を適用し、得られた身体座標データからフレーム間の類似度を算出し、対応付けを決定している。しかし、OpenPose による身体部位の検出が失敗した場合は、類似度の算出が不正確になり、誤った対応付けが生じることが確認された。

そこで、本論文では身体部位の検出失敗の影響を減らす処理を加えることで、対応付け精度の改善を図る。

## 3. 提案手法

### 3.1. フレーム数削減

本手法では、DTW を用いて映像フレーム間の投手の姿勢の類似度を算出し、投球動作を照合する。DTW は総当たりで要素間の類似度を算出するため、系列の長さに比例して類似度算出の計算量は増加する。そこで、対応付けの精度に影響を与えない範囲で DTW の計算に使用するフレームを削減する。

今回使用した動画の投球シーンでは、投手だけでなく捕手や打者、審判が映る。そのため、投球シーンでは OpenPose により検出される人物が 4 人以上であると考えられる。しかし、審判が打者や捕手と重なって映っている場合には、審判が検出されない可能性がある。よって、OpenPose により検出された人物が 2 人以下のフレームは DTW の計算に使用しない。審判が検出されているフレームと検出されていないフレームをそれぞれ図 1, 図 2 に示す。

A Study on Pitching Scene Matching Using OpenPose

<sup>†</sup>Ryohei Osawa <sup>‡</sup>Takaaki Ishikawa <sup>†‡</sup>Hiroshi Watanabe

<sup>†</sup>School of Fundamental Science and Engineering, Waseda University

<sup>‡</sup>Global Information and Telecommunication Institute, Waseda University



図1 審判が検出されているフレーム



図2 審判が検出されていないフレーム

また、本研究で使用した投球映像において、時間的に連続する複数フレームで投手の姿勢がほとんど変化しないケースが確認された。この場合、姿勢が酷似している複数フレームが DTW によって対応付けされるフレームは同一であると予想される。そのため、DTW の計算では、姿勢が似ている複数フレームのうち 1 フレームのみを使用する。

### 3.2. 検出失敗部位の座標補正

OpenPose による検出結果として、各部位の 2 次元座標および検出の信頼度を示す確度が得られる。確度は 0 から 1 の範囲の値であり、1 に近いほど検出の信頼度が高いことを示す。今回用いた投球動画について、OpenPose による検出が失敗している部位の多くは確度が 0.3 未満であることが確認された。そこで、身体部位の確度が 0.3 未満である場合は、前後 2 フレームにおける同一部位の検出結果を利用して補正値を算出する。なお、 $k$  フレームの座標値を補正する場合、前後 2 フレームは「 $k-2$  フレームと  $k-1$  フレーム」、「 $k-1$  フレームと  $k+1$  フレーム」、「 $k+1$  フレームと  $k+2$  フレーム」のいずれかである。

身体部位の確度が 0.3 未満である際、常に補正するのではなく、前後 2 フレームにおいて同一部位の確度が共に 0.5 以上である場合にのみ補正する。補正後の座標値は、身体部位が等速直線運動をしているという仮定に基づき算出する。

### 3.3. ベクトルデータの作成

本研究では、野球の投球映像から OpenPose により得られる身体座標データのうち、首、両肩、両肘、腰の中央、両腰、両膝、両足首の 12 点の身体部位の検出結果を用いる。対象とする 12 点のうち、任意の二つの部位を結ぶベクトルについて長さおよび角度を OpenPose により得られる各部位の 2 次元座標値から算出する。また、OpenPose から得られる各部位の確度を用いてベクトルの確度を求める。ベクトルを構成する二つの部位の確度のうち、最小値

をベクトルの確度とする。

### 3.4. DTW による対応付け

本研究は、二つの映像においてフレームの対応付けを決定するものである。対応付けの際には文献 [2] で提案されている動作の始点および終点を自動で抽出可能な DTW を用いる。本手法では異なる 2 つの部位を結ぶ 66 本のベクトルを用いてフレーム間の姿勢の類似度を算出する。ここで、類似度は対応するベクトルの角度差と長さの差の積とする。類似度の平均値が最小となるように対応付けを決定する。

## 4. 実験

### 4.1. 実験概要

本実験で用いたデータセットは野球の右投手の投球映像 100 本である。各映像には右投手の投球シーンが 1 球分含まれる。投球映像のうち、1 本をモデル系列とし、入力系列はモデル系列を含む全 100 本とした。モデル系列は目視により投球動作の始点と終点を決定し、入力系列は動作の始点と終点を付与しない。得られた対応付けについて、より対応付けに適したフレームが存在するかに基づき、主観で成否を判断した。

### 4.2. 実験結果

OpenPose による検出が失敗した部位の座標値が補正されたフレームについて、正確に対応付けされている割合は約 8 割であることを確認した。不正確な対応付けに関しては、身体部位が等速直線運動をするという仮定に基づき、前後 2 フレームのみで座標の補正値を算出したことにより補正が不正確になったことが大きな要因であると推察した。正確に補正するためには、より多くの前後フレームにおける身体座標データを用いて、補正値を算出する必要があると考えられる。

## 5. まとめ

本研究では、OpenPose と DTW を用いた投球動作の照合手法を提案した。今後は補正手法を改良し、対応付けの精度向上を図る。

## 参考文献

- [1] Z. Cao, T. Simon, S. Wei, and Y. Sheikh. Realtime Multi-Person 2D Pose Estimation using Part Affinity Field. In CVPR, pp7291-7299 (2017)
- [2] 横井真也, 石川孝明, 渡辺裕. スポーツ映像から取得した骨格座標データに対するアライメント. 電子情報通信学会総合大会, D-12-59 (2019)
- [3] 大澤遼平, 石川孝明, 渡辺裕. スポーツシーン照合に関する一検討, 映像情報メディア学会年次大会, 31C-3 (2019)