

# 選手の姿勢類似度に基づくスポーツ動作照合

大澤遼平<sup>1</sup> 渡辺裕<sup>1</sup>

**概要:** アマチュアのスポーツ選手が効率的に上達するための練習方法の一つとして、複数映像を用いて熟練者の動作と自身の動作を比較し、熟練者の動きを模倣することが挙げられる。映像を用いて複数の動作を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重要となる。ただ、選手によって動作の速度は異なるため、細部のタイミングを手動ですべて合わせることは困難である。そこで、本研究では2本の映像に含まれる類似動作について、動作のタイミングを高精度で同期させる手法を提案する。評価実験により、提案手法は動作中の細かな動きや姿勢の変化を区別するうえで従来手法よりも有効であることを確認した。

## Sports Motion Matching Based on Athletes' Pose Similarity

RYOHEI OSAWA<sup>†1</sup> HIROSHI WATANABE<sup>†1</sup>

### 1. はじめに

近年、多くのスポーツで選手の動作解析や技術向上のために、試合中や練習中の選手を撮影した映像データが利用されている。撮影した映像は繰り返し視聴できるだけでなく、拡大したりコマ送りで再生したりすることも可能である。そのため、肉眼では認識し難い動作の特徴や上達のために改善すべき点などを発見することが容易となる。また、複数の映像を利用し動作を見比べることで、動作間の相違部分を明らかにすることも可能となる。

アマチュアのスポーツ選手が効率的に上達するための練習方法の一つとして、熟練者の動作と自身の動作を比較し、熟練者の動きを模倣することが挙げられる。映像を用いて複数の動作を細部まで比較する場合、動作のタイミングを同期させることが重要となる。ここで動作のタイミングとは、動作の開始時および終了時だけでなく、野球の投球動作におけるボールがリリースされる瞬間や、ゴルフのスイング動作におけるクラブがトップの位置に上がった瞬間などを含む。複数の映像を単純に動作開始時から再生するだけでは、選手による動作速度等の違いが考慮されないため、動作のタイミングが徐々にずれてしまう可能性がある。

そこで、本研究では2本の映像に含まれる類似動作について、選手の姿勢類似度に基づき、動作のタイミングを高精度で同期させる手法を提案する。

### 2. 関連研究

#### 2.1 OpenPose

OpenPose[1]とは、動画のみから複数人物の骨格を推定可能な手法である。OpenPoseは画像あるいは映像フレーム内の各人物について、肩や膝などの関節や目、鼻等の顔の部位を検出する。検出された身体部位ごとに、画像上の2次元

座標値および推定の信頼度が出力される。

#### 2.2 始点終点自由な Dynamic Time Warping

始点終点自由な Dynamic Time Warping (以下、始点終点自由な DTW)[2]とは、2系列の類似度が最も高くなるように、系列内の要素を対応付ける手法である。ここで、要素とは対象となる系列に応じて異なり、例として映像データに適用する場合は映像フレームが要素に該当する。

始点終点自由な DTW ではまず、一方の系列に含まれる任意の要素ともう一方の系列に含まれる任意の要素について総当たりで要素間の差を算出する。その後、差の平均値が最小となるように、要素間の対応付けを決定する。ただし、片方の系列について、系列の先頭から $m$ 個の要素および終端から $n$ 個の要素は、もう一方の系列のいずれの要素とも対応付けされない。ここで、 $m$ と $n$ は要素間の差に基づき決定される値であり、系列に含まれる総要素数未満の自然数もしくは0である。また、 $m$ と $n$ の和も同様の条件を満たす。

#### 2.3 Temporal Cycle-Consistency Learning

複数の動画に含まれる類似動作を同期させる手法として、Temporal Cycle-Consistency Learning (以下、TCC)[3]が存在する。TCCでは、自己教師あり学習により作成された特徴空間を用いて、フレーム間の類似度を算出する。その後、得られたフレーム間の類似度に基づき、2動画に含まれるすべてのフレームを対応付ける。なお、TCCは入力する2動画の開始および終了が、同期対象である動作の開始時および終了時とそれぞれ一致していることを想定した手法である。

#### 2.4 横井手法

2本の映像に含まれるスポーツ動作のタイミングを同期させる手法として、横井手法[2]が存在する。横井手法は第2.2節で述べた始点終点自由な DTW を利用し、2本の映像

<sup>1</sup> 早稲田大学大学院基幹理工学研究科  
Graduate School of Fundamental Science and Engineering, Waseda University

系列に含まれるフレームを対応付ける。なお、対応付けは選手の骨格情報を用いて算出されるフレーム間の類似度に基づき決定される。横井手法は、一方の映像について、動作の開始時および終了時が映像の開始および終了と一致していない場合でも動作の同期が可能である。

### 3. 提案手法

#### 3.1 骨格情報の取得

同期対象のスポーツ動作が含まれる映像に対して第2.1節で述べた OpenPose を適用し、映っている選手の身体部位の位置を推定する。その結果、各映像フレームにおいて検出された身体部位ごとに、画像上での 2 次元座標値および推定の信頼度が取得される。本研究では、OpenPose により検出可能な身体部位のうち、両肩、両肘、両手首、両腰、両膝、両足首の 12 部位を用いる。

#### 3.2 ベクトルデータの作成

OpenPose により得られた選手の 12 部位の推定結果を用いて、フレームごとに空間ベクトルと時間ベクトルの 2 種類のベクトルデータを作成する。

空間ベクトルデータとは、12 点の身体部位のうち、任意の 2 部位（以下、構成部位）の推定結果に基づき作成されるベクトル  $vs$  の集合である。各空間ベクトル  $vs$  の要素には、長さ  $r_{vs}$  および角度  $\theta_{vs}$  が含まれる。長さ  $r_{vs}$  は構成部位の 2 次元座標値の距離である。また、角度  $\theta_{vs}$  は構成部位の OpenPose による推定位置を結ぶベクトルと、画像の横軸に平行な右向きの直線とのなす角とする。構成部位が右肘と右手首である空間ベクトルを例に、長さ  $r_{vs}$  および角度  $\theta_{vs}$  を図 1 に示す。

時間ベクトルデータとは、近接する 2 フレームにおける各身体部位の動きを表すベクトル  $vt$  の集合である。ここで、近接するとは、フレームの間隔が 3 であることを指す。各時間ベクトル  $vt$  の要素には、移動距離  $r_{vt}$  および移動方向  $\theta_{vt}$  が含まれる。フレーム間隔が 3 である過去フレームと対象フレームにおける左手首を例に、移動距離  $r_{vt}$  および移動方向  $\theta_{vt}$  を図 2 に示す。

#### 3.3 姿勢類似度の算出

作成したベクトルデータに基づき、2 本の映像に含まれるフレーム間での選手の姿勢類似度を算出する。姿勢類似度は対応する空間ベクトルの差分および対応する時間ベクトルの差分の和で定義される。ここで、対応する空間ベクトルおよび対応する時間ベクトルは、2 選手の動作中の利き手が同一かによって異なる。本研究で用いた映像に含まれる野球の投球動作およびゴルフのスイング動作における動作中の利き手は、右投げまたは右打ちの場合は右手、左投げまたは左打ちの場合は左手と定義する。

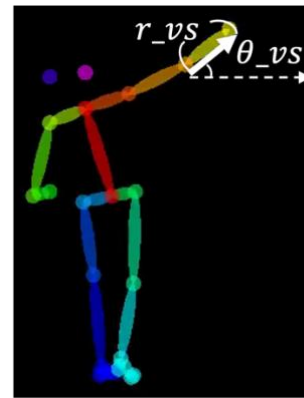


図 1 空間ベクトルの例

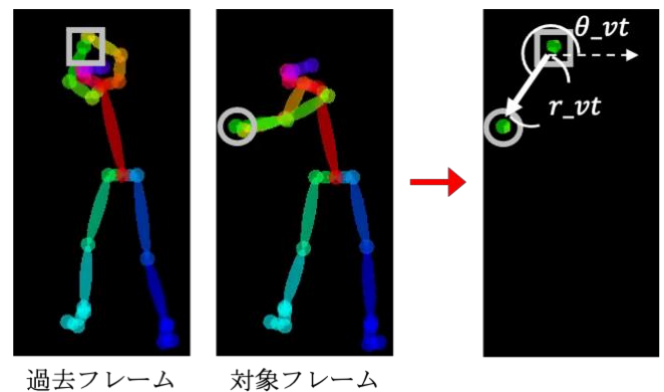


図 2 時間ベクトルの例

#### 3.4 始点終点自由な DTW による対応付け

第 2.2 節で述べた始点終点自由な DTW を用いて、フレーム間の選手の姿勢類似度に基づき、2 本の映像に含まれるフレームの対応付けを決定する。この際、一方の映像については、最初のフレームおよび最後のフレームがそれぞれ動作の開始時および終了時と一致していると仮定し、すべてのフレームを対応付ける。それに対し、もう一方の映像については、動作の開始時および終了時を選手の姿勢類似度に基づき推定し、推定された開始時から終了時までのフレームを対応付ける。

### 4. 評価実験

#### 4.1 実験手法

本実験で用いたデータセットは野球の投球動作が 1 回分含まれる映像 124 本と、ゴルフのスイング動作が 1 回分含まれる映像 85 本である。動作の開始時および終了時が映像の開始および終了とそれぞれ一致しているモデル動画と、一致しているとは限らない対象動画について提案手法および従来手法を適用した。なお、モデル動画は各動作につき 4 本用意した。モデル 1 およびモデル 2 は右投げまたは右打ち、モデル 3 およびモデル 4 は左投げまたは左打ちの選手の動作を含む。一方、対象動画に含まれる動作はすべて右投

げもしくは右打ちの選手によるものである。なお、横井手法は右投げと左投げもしくは右打ちと左打ちの2選手の動作を同期させることは研究目的として想定されていないため、モデル1およびモデル2についてのみ実験した。

#### 4.2 ケンドールの順位相関係数による評価

各手法におけるフレーム間の類似度算出処理をケンドールの順位相関係数[3]を用いて評価した。ケンドールの順位相関係数の算出方法を以下で述べる。

まず、モデル動画に含まれる任意の2フレーム $a_i, a_j (1 \leq i < j \leq M)$ について、対象動画内で最も類似度が高いフレーム $b_p, b_q (1 \leq p, q \leq N)$ をそれぞれ求める。本研究では、最も類似度が高いフレームを第3.3節で述べた選手の姿勢類似度に基づき決定する。なお、 $M$ はモデル動画のフレーム数、 $N$ は対象動画のフレーム数である。

続いて、フレーム $b_p, b_q$ の順序を比較する。対象動画において、 $p < q$ すなわちフレーム $b_p$ よりもフレーム $b_q$ の方が時間的に後のフレームである場合は、順序一致数 $U$ に1が加算される。一方、 $p \geq q$ である場合は順序不一致数 $V$ に1が加算される。

最後に、順序一致数 $U$ および、順序不一致数 $V$ を用いて、ケンドールの順位相関係数 $\tau$ を式(1)により算出する。なお、ケンドールの順位相関係数は1に近いほど評価が高いことを示す。

$$\tau = \frac{U - V}{M C_2} \quad (1)$$

投球動作映像に対する結果を表1に、スイング動作映像に対する結果を表2に示す。

表1および表2より、各モデル動画での実験において提案手法の値が最も大きく、1に近いことが確認できる。したがって、提案手法における選手の姿勢類似度を算出する処理が、動作中の細かな動きや姿勢の変化を区別するうえで、従来手法の類似度算出処理よりも有効であるといえる。

表1 ケンドールの順位相関係数 (投球動作)

手法	モデル1	モデル2	モデル3	モデル4
TCC	0.620	0.567	0.605	0.419
横井手法	0.738	0.711	—	—
提案手法	0.855	0.831	0.838	0.816

表2 ケンドールの順位相関係数 (スイング動作)

手法	モデル1	モデル2	モデル3	モデル4
TCC	0.576	0.473	0.465	0.582
横井手法	0.687	0.705	—	—
提案手法	0.882	0.897	0.911	0.917

表3 フレーム誤差 (投球動作)

手法	開始 or 終了	フレーム誤差 [フレーム]			
		モデル1	モデル2	モデル3	モデル4
横井手法	開始	6.12	8.83	—	—
	終了	28.54	10.22	—	—
提案手法	開始	6.35	4.77	7.21	5.05
	終了	7.06	8.27	6.75	4.92

表4 フレーム誤差 (スイング動作)

手法	開始 or 終了	フレーム誤差 [フレーム]			
		モデル1	モデル2	モデル3	モデル4
横井手法	開始	12.35	10.23	—	—
	終了	16.71	14.59	—	—
提案手法	開始	4.93	3.94	3.35	6.70
	終了	7.57	6.33	9.28	9.83

#### 4.3 フレーム誤差による評価

各対象動画について、始点終点自由なDTWにより求めた動作の開始時および終了時と、目視により決定した動作の開始時および終了時のフレーム誤差を算出した。比較として、始点終点自由なDTWを用いる横井手法でも同様に誤差を求めた。投球動作映像およびスイング動作映像に対するフレーム誤差の平均値を表3および表4にまとめる。

表3および表4より、多くのケースにおいて提案手法のフレーム誤差が横井手法と比べ小さいことが確認できる。また、右投げと左投げや右打ちと左打ちの2選手の動作を同期させる場合でも、提案手法は有効であるといえる。

### 5. おわりに

本研究では2本の映像に含まれる類似動作のタイミングを高精度で同期させる手法を提案した。評価実験により、動作の同期精度において提案手法が従来手法よりも有効であることを確認した。

**謝辞** 本研究はJSPS 科研費20K11344の助成を受けたものである。

#### 参考文献

- [1] Z. Cao, T. Simon, S. Wei, and Y. Sheikh. Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields. IEEE Conference on Computer Vision and Pattern Recognition, pp. 1302-1310, July 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ンメント. 電子情報通信学会総合大会, D-12-59, Feb. 2019.
- [3] D. Dwibedi, Y. Aytar, J. Tompson, P. Sermanet and A. Zisserman. Temporal Cycle-Consistency Learning. IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 1801-1810, June 2019.