


修士論文概要書

Master's Thesis Summary

Date of submission: 01/24/2022

専攻名 (専門分野) Department	情報理工・ 情報通信専攻	氏名 Name	大澤 遼平	指導 教員 Advisor	渡辺 裕 
研究指導名 Research guidance	オーディオビジュアル 情報処理研究	学籍番号 Student ID number	CD 5120F009-9		
研究題目 Title	身体部位の動きを考慮した姿勢類似度に基づく動作照合 Motion Matching Based on Pose Similarity Using Movement of Body Parts				

1. まえがき

近年,多くのスポーツで選手の動作を分析する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.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発見することが容易となる.

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合わせることが重要となる.複数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. 映像データに適用する場合, 要素とは映像フレームを指す.

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. この際, 一方の系列に含まれる要素はすべて対応付けされるが, もう一方の系列については, 常にすべての要素が対応付けされるとは限らない.

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. TCC は, 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する 2本の動画に対してのみ適用可能な手法である.

2.4. スポーツ動作照合手法

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, 選手の骨格情報に基づいて 2本の映像系列に含まれるフレームを対応付ける. 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適用可能である.

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 θ_{vs} が含まれる. 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 θ_{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 θ_{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 θ_{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.



図1 空間ベクトルの例

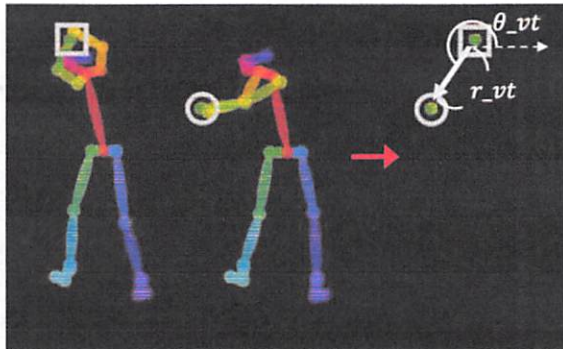


図2 時間ベクトルの例

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、2本の映像に含まれるフレームの対応付けを決定する。一方の映像において、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、もう一方の映像に含まれる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決定する。

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回含まれる映像124本と、ゴルフのスイング動作が1回含まれる映像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

ケンドールの順位相関係数[3]を用いて、各手法におけるフレーム間の類似度算出処理を評価した。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。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。

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。

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、提案手法が従来手法よりも有効であることを確認した。

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 (CVPR), pp. 1302-1310, July 2017.
- [2] 横井真也, "スポーツ映像から取得した身体座標データに対するアラインメント," 早稲田大学大学院基幹理工学研究科情報理工・情報通信専攻修士論文, 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 (CVPR), pp. 1801-1810, June 2019.

2021 年度

早稲田大学大学院基幹理工学研究科情報理工・情報通信専攻 修士論文

身体部位の動きを考慮した姿勢類似度に基づく
動作照合

Motion Matching Based on Pose Similarity
Using Movement of Body Parts

大澤 遼平

(5120F009-9)

提出日：2022.01.24

指導教員：渡辺裕教授

研究指導名：オーディオビジュアル情報処理研究

目次

第1章 序論	1
1.1 研究背景	1
1.2 研究目的	2
1.3 本論文の構成	2
第2章 関連研究	3
2.1 まえがき	3
2.2 OpenPose	3
2.2.1 概要	3
2.2.2 問題点	5
2.3 Dynamic Time Warping	5
2.3.1 概要	5
2.3.2 問題点	7
2.4 始点終点自由な Dynamic Time Warping	8
2.5 Temporal Cycle-Consistency Learning	10
2.6 スポーツ動作照合手法	10
2.7 むすび	11
第3章 提案手法	12
3.1 まえがき	12
3.2 提案手法の概要	12
3.3 OpenPose による骨格情報の取得	13
3.4 手足の左右反転検出の修正	14
3.4.1 左右反転検出の概要	14
3.4.2 左右反転検出の判定	16
3.4.3 座標値の修正	18
3.5 姿勢変化に基づくフレーム数削減	18
3.6 ベクトルデータの作成	19
3.6.1 概要	19
3.6.2 空間ベクトル	20
3.6.3 時間ベクトル	22
3.7 姿勢類似度の算出	24
3.7.1 概要	24

3.7.2	空間ベクトルの差	25
3.7.3	時間ベクトルの差	25
3.7.4	姿勢類似度	27
3.8	始点終点自由な Dynamic Time Warping による対応付け	28
3.9	むすび	30
第4章	評価実験および考察	31
4.1	まえがき	31
4.2	実験の概要	31
4.2.1	データセット	31
4.2.2	評価項目	33
4.3	左右反転検出の修正に関する評価	33
4.4	ケンドールの順位相関係数による評価	34
4.4.1	ケンドールの順位相関係数	34
4.4.2	投球動作が含まれる映像に対する評価実験および考察	35
4.4.3	スイング動作が含まれる映像に対する評価実験および考察	36
4.5	対応付け結果	37
4.6	むすび	43
第5章	結論と今後の課題	44
5.1	結論	44
5.2	今後の課題	44
謝辞		45
参考文献		46
図一覧		48
表一覧		49
研究業績		50

第1章 序論

1.1 研究背景

近年、多くのスポーツで選手の動作解析や目標達成のために、高性能機器や高解像度カメラ等を活用した多種多様なデータが利用されている。その中でも、特にデータ活用・IT化が普及しているのが日本プロ野球界である。多くの本拠地球場には、投球の回転数やボールの変化量、打球の角度や飛距離等を測定可能な高性能弾道測定器「TrackMan（トラックマン）[1]」（TRACKMAN社）が設置されている[2]。また、球場に設置されている高解像度カメラで撮影した画像を利用し、投球や打撃に加えて守備や走塁の動作をデータ化するシステムが普及している[3]。

2017年から2020年まで4年連続日本一に輝いた福岡ソフトバンクホークスは、全選手にApple社のスマートフォンiPhoneとタブレットiPadを支給し、選手自身のプレイ映像や対戦相手の様々なデータをモバイルアプリで常に確認可能な環境を整備している[3][4]。練習の前後や遠征時の移動中などのすきま時間にデータを確認・分析可能な点が選手に好評であり、トップ選手ほどモバイルアプリを活用していることが文献[3][4]で示されている。

2021年に20年ぶりの日本一を達成した東京ヤクルトスワローズは、12球団で唯一「ホークアイ[5]」（ソニー株式会社）のプレー分析サービスを導入している。神宮球場に設置された8台の専用カメラを活用して、ボールや選手の姿勢をミリ単位の正確さでトラッキングし、あらゆるプレーを正確かつ緻密に確認・解析することを可能にしている[6][7]。チームに所属している投手は、状態が良い時と悪い時の手首の角度や変化球のホップ成分等を比較したり、データから有効と判断された新球種の習得に挑戦したりするなどして、好成績を残した[8]。

TrackManやホークアイは高精度かつ多様なデータを取得するために、軍事レーダーを用いた高性能機器や複数台の高解像度カメラを必要とする。測定機器は高コストであるため、アマチュアの選手やチームにとってはTrackManやホークアイの導入は容易ではない。

これに対して、データ取得用の高価な専用機材を必要とせず、低コストかつ手軽な動作分析手法の一つとして、スマートフォン等で撮影された映像の利用が挙げられる。スマートフォンやタブレットなどの普及により、映像データを用いた分析はアマチュアの選手・チームにとっても容易に導入可能である。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。

1.2 研究目的

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。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。

比較対象となる複数の映像を単純に動作開始時から再生すると、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。その場合、動作の細部を比較することは困難になる。

そこで、本論文では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。具体的には、まず映像から選手の身体部位に関する2次元座標値を取得する。その後、得られた2次元座標値から各身体部位の時間変化を考慮した姿勢データをフレームごとに算出する。最後に、姿勢データから求めたフレーム間の類似度を基に、2本の映像のフレームを対応付ける。

1.3 本論文の構成

以下に本論文の構成を示す。

第1章は本章であり、本研究の背景および目的を述べている。

第2章では、本研究で用いる関連技術について説明する。また、関連する動作照合手法を示す。

第3章では、2本の映像に含まれる類似動作のタイミングを同期させる手法を提案する。

第4章では、提案手法における実験および結果、考察を示す。

第5章では、本研究のまとめと今後の課題を述べる。

第2章 関連研究

2.1 まえがき

本章では，本研究で用いる姿勢推定手法および対応付け手法について述べる．また，動作照合に関する従来手法を示す．

2.2 OpenPose

2.2.1 概要

OpenPose[9]とは，単眼カメラ等で撮影された画像または映像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，肩や肘，膝などの関節や目，鼻等の顔の部位を検出する．検出可能な特徴点の数は機械学習時に使用するデータセットに応じて異なり，18 部位の位置を推定できるモデルや，足の親指・小指等を含む 25 部位を検出可能なモデルが存在する．また，手と顔についてより細かい特徴点を検出可能な機能も OpenPose には含まれている．手検出機能では各指の第一関節や第二関節など片手につき 21 点の位置を推定可能である．また，目や口等の輪郭を推定できる顔検出機能では，検出可能な特徴点は 70 点を数える．

25 部位の位置を推定可能なモデルを適用した実行例の入力画像および出力画像をそれぞれ図 2.1，図 2.2 に示す．なお，図 2.2 は OpenPose により推定された特徴点の位置と隣接する特徴点同士を結ぶ骨格をそれぞれ色付きの点と線で入力画像上に描画した画像である．また，入力画像と同サイズの黒色画像上に検出された特徴点と骨格のみを描画した出力結果を図 2.3 に示す．



図 2.1 入力画像[10]



図 2.2 出力画像

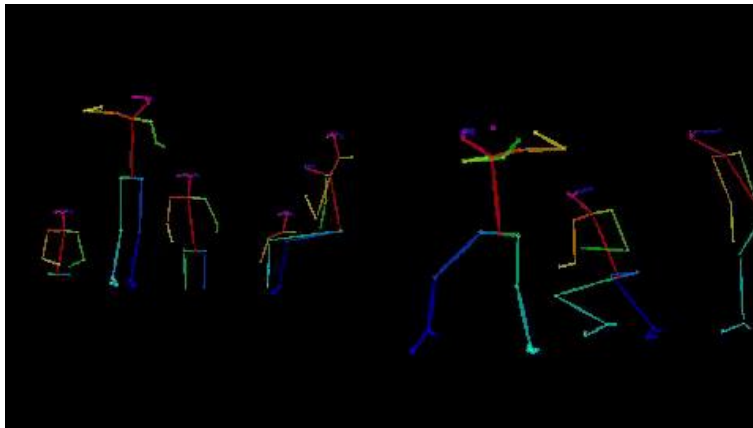


図 2.3 出力結果（特徴点と骨格のみ）

OpenPose では、図 2.2 のような画像に加え、検出された特徴点ごとに画像上の 2 次元座標値および推定の信頼度を出力することが可能である。なお、2 次元座標値の原点は画像の左上の角である。推定の信頼度は 0 から 1 の範囲の値をとり、1 に近いほど推定が正確であることを示す。検出に失敗した特徴点については、2 次元座標値と推定の信頼度は共に 0 となる。

OpenPose の最も大きな利点は、画像や映像のみから人物の骨格を推定可能な点である。特殊なセンサや複数の高性能な撮影器具等を必要としないため、スマートフォン等で撮影された動画像に対しても人物の姿勢推定が可能である。また、身体に取り付けるデータ取得用の装置を使用しない点は、特にスポーツ選手の動作解析に利用する上で大きな利点となる。カメラで選手を撮影するだけで骨格推定が可能のため、解析対象の選手に器具の装着による負担がかからない。そのため、練習中だけでなく試合中の選手の姿勢を推定し、分析することが可能となる。

2.2.2 問題点

OpenPose は 1 枚の画像や映像フレームのみから映っている人物の骨格を推定する。そのため、特徴点の位置を正確に推定するには、入力動画像内に人物が鮮明にかつ十分な大きさで映っていることが必要となる。したがって、画像内の非常に小さな領域に映っている人物に対して OpenPose を適用した場合は、人物の身体部位が検出されない可能性がある。また、人物と撮影機器の間に他の人物や物体が存在し、一部の身体部位が画像上に映らない場合には、検出精度が低下する。

さらに、OpenPose を映像に適用する際の問題点として、時系列の前後情報を考慮しない点が挙げられる。OpenPose は各映像フレームについて独立に人物の骨格を推定する。すなわち、時系列で前後の映像フレームにおける同一人物の検出結果は利用しない。そのため、肩幅ほどに開いた左右の足の位置が微小時間で複数回入れ替わるなど、人間の動作として現実的でない誤った検出が生じる場合がある。

2.3 Dynamic Time Warping

2.3.1 概要

Dynamic Time Warping (DTW) [11] 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。DTW は特に映像や音声などの時系列データに対して用いられることが多い。なお、要素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。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、差の合計値が最小となるように、要素間の対応付けを決定する。

要素数 M の系列 $A = \{a_1, a_2, \dots, a_M\}$ と要素数 N の系列 $B = \{b_1, b_2, \dots, b_N\}$ を例に DTW のアルゴリズムを説明する。まず、 $(M + 1) \times (N + 1)$ の行列 D を用意する。行列 D の初期条件は式 (2.1) に示す通りである。なお、式 (2.1) において m は 1 以上 M 以下の自然数、 n は 1 以上 N 以下の自然数である。

$$D(0, 0) = 0, \quad D(m, 0) = D(0, n) = \infty \quad (2.1)$$

次に、系列 A の要素 a_m と系列 B の要素 b_n の差 $C(a_m, b_n)$ を用いて、 $D(m, n)$ を式 (2.2) により再帰的に算出する。

$$D(m, n) = C(a_m, b_n) + \min \begin{cases} D(m - 1, n) \\ D(m, n - 1) \\ D(m - 1, n - 1) \end{cases} \quad (2.2)$$

このとき、式 (2.2) の第 2 項に示した 3 つのうち最小値として $D(m, n)$ の計算に用いられた値

の添字情報 $DI_{(m,n)}$ を記録する．例えば， $D(m-1, n)$ が最小値として計算に用いられた場合は， $DI_{(m,n)_A} = m-1, DI_{(m,n)_B} = n$ と記録される．

最後に，添字情報 $DI_{(m,n)}$ を利用し要素間の対応付けを決定する．まず，2系列の最後の要素である a_M と b_N の2要素を対応付ける．次に， $D(M, N)$ の計算に用いられた値の添字情報 $DI_{(M,N)}$ に基づき，系列Aの $DI_{(M,N)_A}$ 番目の要素と系列Bの $DI_{(M,N)_B}$ 番目の要素を対応付ける．続いて， $D(DI_{(M,N)_A}, DI_{(M,N)_B})$ の計算に用いられた値の添字情報 $DI_{(DI_{(M,N)_A}, DI_{(M,N)_B})}$ に基づき，同様に対応付けを決定する．この操作を2系列の最初の要素同士が対応付けされるまで繰り返す．

この際，任意の対応付け (a_i, b_p) と (a_j, b_q) について常に式(2.3)が成り立つ．なお， (a_i, b_p) は系列Aの i 番目の要素 a_i と系列Bの p 番目の要素 b_p が対応付けされていることを示す．

$$(j-i) \times (q-p) \geq 0 \quad (2.3)$$

DTWでは式(2.3)を常に満たすように，すなわち各系列内の要素の順序を考慮し対応付けが決定される．

DTWは一方の系列に含まれる1要素をもう一方の系列の複数の要素と対応付けることが可能である．そのため，系列の要素数が異なる場合でもDTWは適用できる．要素数が60点である系列Aと63点である系列Bに対してDTWを適用した際の対応付け結果を図2.4に示す．なお，2系列の要素は共に2次元座標値である．

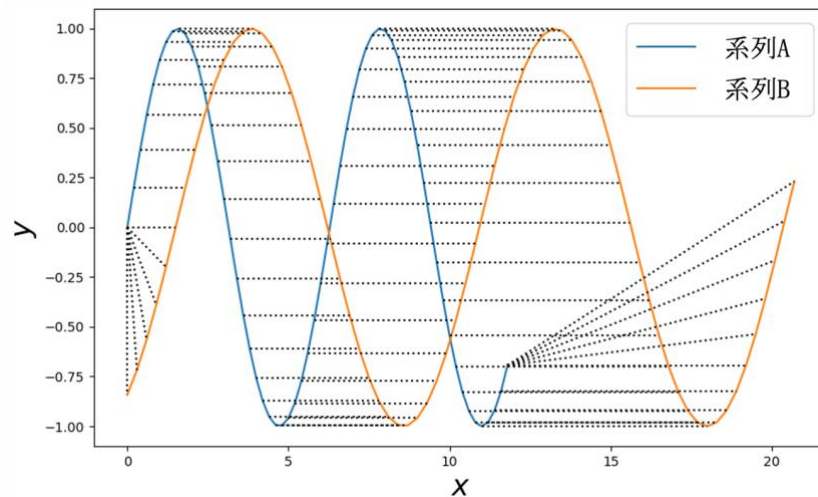


図 2.4 要素数の異なる2系列に対するDTWの適用結果

図2.4では系列Aの要素を青い線，系列Bの要素をオレンジの線でそれぞれ繋いでいる．また，DTWにより対応付けされた2点を黒の点線で結んでいる．なお，対応付けは点線で結ばれた2点のy座標差の合計値が最小となるように決定されている．

図 2.4 より、 x 軸に平行な点線、すなわち y 座標差が 0 となる 2 点が多く対応付けされていることが明らかである。また、図 2.4 の右側にて、系列 A の最後の要素が系列 B の複数要素と対応付けされていることが確認できる。以上のことから、要素数が異なる 2 系列に対しても、DTW は適用可能であり、2 系列の類似度が最も高くなるような対応付けが可能であることがわかる。

2.3.2 問題点

DTW は対象となる 2 系列に含まれるすべての要素を対応付ける。すなわち、一方の系列の任意の要素は、もう一方の系列内に類似する要素が存在しない場合でも、いずれか一つ以上の要素と対応付けされる。差が大きい 2 要素を強引に対応付けることにより、類似する要素同士が対応付けされないケースが生じる可能性がある。類似しない要素の対応付けにより、不正確な対応付けが発生した例を図 2.5 に示す。

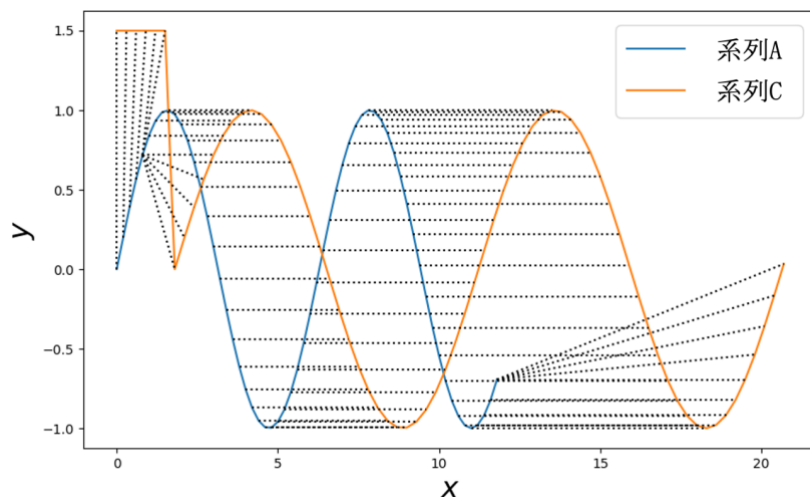


図 2.5 DTWによる不正確な対応付けの例

図 2.5 では系列 A の要素を青い線、系列 C の要素をオレンジの線でそれぞれ繋いでいる。なお、系列 C は図 2.4 で示した系列 B の先頭に y 座標が 1.5 である要素を 6 点追加した系列である。図 2.5 中の黒の点線は、要素間の y 座標差に基づき対応付けされた 2 点を結んでいる。

図 2.4 で示した結果と同様、 x 軸に平行な点線、すなわち y 座標差が 0 となる 2 点の対応付けが全体の半数以上を占めることが確認できる。一方、グラフの左側では、 y 座標が異なる 2 点が点線で結ばれている。特に系列 A の始点付近の要素は、系列 C 内に y 座標が同一である要素が存在しているにも関わらず、 y 座標差が 0.5 以上ある要素と対応付けされている。2 系列に含まれるすべての要素を対応付けるという DTW の性質により、系列 C の先頭付近に存在する要素は系列 A 内の類似しない要素と強引に対応付けられる。その結果、類

似する要素同士が対応付けされず、不正確な対応付けが発生している。

2.4 始点終点自由な Dynamic Time Warping

始点終点自由な Dynamic Time Warping (始点終点自由な DTW) [12]は、2 系列のうち片方の系列における対応付けの始点および終点を要素間の差に基づき決定する手法である。ここで、対応付けの始点および終点とは、各系列内で対応付けされる要素のうち、系列自身の先頭および終端にそれぞれ最も時間の近い要素を指す。第 2.3 節で述べた DTW は対象となる 2 系列に含まれるすべての要素を対応付ける。そのため、対応付けの始点および終点は系列の最初の要素および最後の要素と常に一致する。

それに対して始点終点自由な DTW では、一方の系列における対応付けの始点と終点を系列の最初の要素および最後の要素に限定せず、すべての要素の中から最適な位置を選択する。対応付けの始点と終点を固定しない最大の利点は、系列の先頭付近もしくは終端付近における類似しない要素同士の対応付けを減少できることである。

要素数 M の系列 $A = \{a_1, a_2, \dots, a_M\}$ と要素数 N の系列 $B = \{b_1, b_2, \dots, b_N\}$ を例に始点終点自由な DTW のアルゴリズムを説明する。ここで、系列 A の対応付けの始点および終点はそれぞれ a_1 および a_M とし、系列 B の対応付けの始点および終点は固定しない。まず、 $(M + 1) \times (N + 1)$ の行列 D を用意する。行列 D の初期条件は式(2.4)に示す通りである。なお、式(2.4)において m は 1 以上 M 以下の自然数、 n は 1 以上 N 以下の自然数である。

$$D(0, 0) = D(0, n) = 0, \quad D(m, 0) = \infty \quad (2.4)$$

次に、系列 A の要素 a_m と系列 B の要素 b_n の差 $C(a_m, b_n)$ を用いて、 $D(m, n)$ を式(2.5)により再帰的に算出する。第 2.3 節で述べた DTW が要素間の差の合計値で大小を比較するのに対し、始点終点自由な DTW では差の平均値で大小を比較し、 $D(m, n)$ を算出する。

$$D(m, n) = \min \left\{ \begin{array}{l} \frac{C(a_m, b_n) + D(m-1, n) \times L(m-1, n)}{L(m-1, n)} \\ \frac{C(a_m, b_n) + D(m, n-1) \times L(m, n-1)}{L(m, n-1) + 1} \\ \frac{C(a_m, b_n) + D(m-1, n-1) \times L(m-1, n-1)}{L(m-1, n-1) + 1} \end{array} \right. \quad (2.5)$$

ここで、 L は行列 D と同サイズの行列であり、初期条件は式(2.6)に示す通りである。

$$L(0, 0) = L(m, 0) = 0, \quad L(0, n) = 1 \quad (2.6)$$

なお、 $L(m, n)$ は式(2.5)の右辺に示した 3 つの分数のうち、最小値となる分数の分母として与えられる。

式(2.5)により行列 D の全成分を算出後、系列 B の対応付けの始点および終点を決定する。まず、 $D(M, n)$ の値をすべて比較し、最小値 $D(M, n')$ を求める。ここで、系列 B の n' 番目の要素 $b_{n'}$ が系列 B の対応付けの終点となる。ただし、全対応付け決定後、系列 A の最後の要素 a_M が系列 B の複数の要素と対応付けされた場合は、複数要素のうち a_M との差が最小である、すなわち a_M と最も類似している要素が対応付けの終点となる。

続いて、DTW と同様に添字情報 $DI(m, n)$ を利用し要素間の対応付けを決定する。系列 A の最初の要素 a_1 と対応付けされた系列 B の要素の中で、 a_1 との差が最小である、すなわち a_1 と最も類似している要素が対応付けの始点となる。

第 2.3 節で述べた DTW は要素間の差の合計値が最小となるように、対応付けを決定する。一方、始点終点自由な DTW は合計値ではなく、要素間の差の平均値に基づき、対応付けを決定する。ここで、平均値は要素間の差の合計値を対応付けされた系列 B の要素数で除算した値である。この処理により、系列の先頭付近もしくは終端付近における類似しない要素同士の間対応付けを減らすことを可能にしている。

図 2.5 にて示した系列 A と系列 C について始点終点自由な DTW を適用した際の対応付け結果を図 2.6 に示す。なお、対応付けの始点および終点は、系列 A では最初の要素および最後の要素とし、系列 C では固定せず要素間の差に基づき決定する。

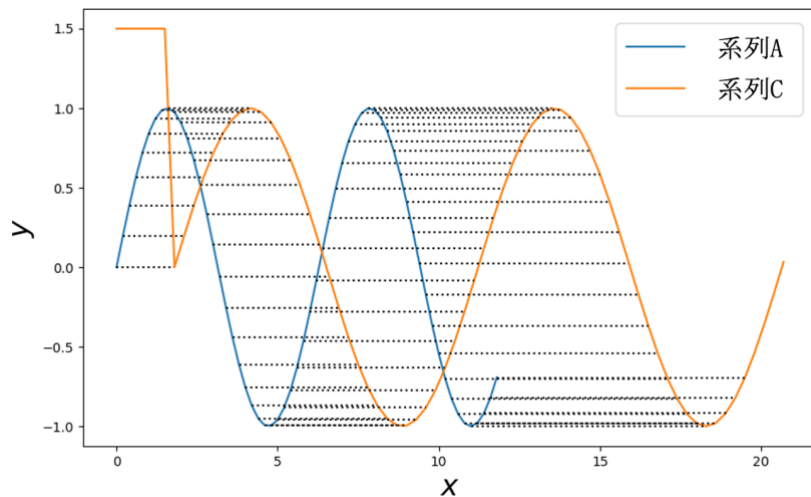


図 2.6 始点終点自由なDTWによる対応付け結果

図 2.6 では系列 A の要素を青い線、系列 C の要素をオレンジの線でそれぞれ繋いでいる。また、要素間の y 座標差の平均値に基づき対応付けされた 2 点を黒の点線で結んでいる。

図 2.5 に示した DTW の適用結果とは異なり、系列 C の始点付近および終点付近、すなわちオレンジのグラフの左端付近と右端付近に存在する一部の要素は系列 A のいずれの要素

とも対応付けされていない。また、系列 A のすべての要素は系列 C に存在する y 座標が同一の要素と対応付けされており、不正確な対応付けが存在しないことが確認できる。

2.5 Temporal Cycle-Consistency Learning

複数動画内の類似動作を同期させる手法として Temporal Cycle-Consistency Learning (TCC) [13] が存在する。TCC は動作に含まれるキーイベントに着目した手法である。ここで、キーイベントとは「ある動作において常に存在する瞬間」と論文内で述べられている。例えば、野球の投球動作の場合、膝の位置が最も高く上がったタイミングやボールをリリースする瞬間などがキーイベントに該当する。各キーイベントはカメラ視点や動作のスピードなどの違いによらず、全ての類似動作に含まれる。また、キーイベントは常に特定の順序で発生する。

TCC は Cycle-Consistency と呼ばれるサイクル一貫性を活用し、自己教師あり学習を用いて複数の類似動作間でキーイベント同士の対応を探索する。TCC の自己教師あり学習の概要を以下で述べる。

まず、学習に用いる動画の全フレームをエンコーダに通し、対応する特徴空間を作成する。続いて、2本の動画を選択し、一方の動画を参照用動画、もう一方の探索用動画とする。参照用動画の任意のフレームについて特徴空間上の対応点 u を求め、特徴空間内での近傍点 $NN(u)$ を探索する。なお、 $NN(u)$ は探索用動画に含まれるフレームに対応する特徴空間上の点である。その後、近傍点として選択された $NN(u)$ について、同様に近傍点 $NN(NN(u))$ を探索する。この時、 $NN(NN(u))$ は参照用動画のフレームに対応する特徴空間上の点である。

TCC では $u = NN(NN(u))$ となる点が最も多くなるように学習が進んでいく。これは、 $u \neq NN(NN(u))$ となった場合の、 u と $NN(NN(u))$ にそれぞれ対応するフレーム間の距離を小さくするよう学習することと同義である。学習が進むにつれて、特徴空間を作成するエンコーダは、類似動作同士が特徴空間内で近くなるように改良される。

2.6 スポーツ動作照合手法

スポーツ動作を対応付ける手法として、第 2.4 節で述べた始点終点自由な DTW を用いた手法が存在する [12]。手法の概要を以下に示す。

まず、第 2.2 節で述べた OpenPose を対応付け対象となる 2本の動画に適用する。ここで、対応付け対象となる 2本の動画には各 1回のスポーツ動作が含まれている。また、一方の映像は映像の開始および終了がそれぞれ投球動作の開始および終了と一致しているものとする。各映像のフレームごとに選手の骨格データが OpenPose により推定される。

続いて、映像に含まれる一連のスポーツ動作の中から個人依存動作を除去する。ここで個人依存動作とは、選手や状況により変化する、動作内の個人に依存する動作と論文 [12] 内で定義されている。2系列の骨格データから個人依存動作の箇所を探索し、個人依存動作が存在する場合は、該当する区間のフレームは対応付けに使用しない。

その後、推定された骨格データを用いてフレーム間での姿勢の相違度を総当たりで算出する。最後に、始点終点自由な DTW を利用し、姿勢の相違度の平均値が最小となるようにフレーム間の対応付けを決定する。

従来手法[12]では、始点終点自由な DTW を利用することにより、一方の系列内のスポーツ動作の前後に存在するフレームを対応付けず、多くのケースで正確な対応付けを実現している。ただし、系列に個人依存動作が含まれる場合には、個人依存動作の区間のフレームが対応付けされないという問題点が生じる。結果として、選手が対応付けを確認する際に、一部の動作が欠けていることで違和感を感じる。また、正確な比較が困難になる。

2.7 むすび

本章では、本研究で用いる姿勢推定手法の OpenPose と対応付け手法の始点終点自由な DTW について述べた。また、複数映像内の類似動作を対応付ける従来手法を示した。

第3章 提案手法

3.1 まえがき

本章では、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提案手法について述べる。

3.2 提案手法の概要

本手法は、人物の姿勢情報に基づき動作の類似度が最も高くなるように2系列のフレームを対応付けるものである。手法の大きな処理フローを図3.1に示す。

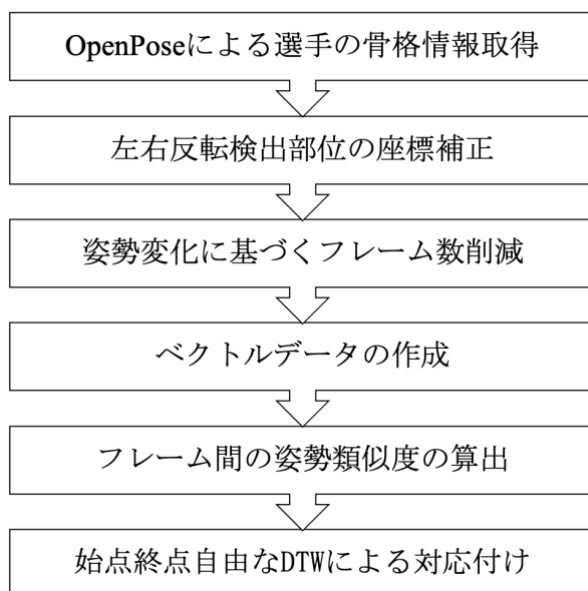


図 3.1 提案手法の処理フロー

まず、映像から人物の骨格情報を推定する。対応付け対象となるスポーツ動作を含む2動画にOpenPoseを適用し、選手の各身体部位の2次元座標データおよび検出の信頼度を取得する。なお、本研究では対応付け対象となるスポーツ動作として野球の投球動作およびゴルフのスイング動作の2種類を用いる。

次に、OpenPoseによる推定が誤っている身体部位に関して、2次元座標値を補正する。座標値の補正は前後フレームでの推定結果を利用し、人間の動作として現実的か等に基づき決定する。

続いて、得られたOpenPoseの推定結果に基づき、類似度算出に用いるフレーム数を削減する。後続の処理である類似度算出やDTWによる対応付けは動画のフレーム数に比例して

演算量が増加する．そこで，対応付け精度に影響を与えない範囲でフレーム数を減らし，後続の処理の演算量削減を図る．

その後，2次元座標値から各身体部位の時間変化を考慮した姿勢データをフレームごとに算出する．一方の系列のフレームともう一方の系列のフレームについて姿勢データを比較し，フレーム間の姿勢類似度を総当たりで求める．

最後に，姿勢類似度に基づき2本の映像のフレームを対応付ける．対応付けには始点終点自由なDTWを利用する．

3.3 OpenPoseによる骨格情報の取得

対応付け対象のスポーツ動作が含まれる映像に対して，第2.2節で述べたOpenPoseを適用し，映っている人物の身体部位の位置を推定する．OpenPoseは各映像フレームで独立に人物を検出する．その結果，各フレームにおいて検出された人物ごとに，各身体部位の画像上での位置が推定され，推定の信頼度が算出される．ここで，画像上での位置は画像の左上のコーナーを原点とする2次元座標値で表される．推定の信頼度は0から1の範囲の値をとり，1に近いほど各身体部位の推定が正確であることを示す．なお，人物の全身が画角に収まっていない等の理由により検出されなかった身体部位については，2次元座標値および推定の信頼度は共に0となる．

本研究では対応付け対象のスポーツ動作として野球の投球動作およびゴルフのスイング動作の2種類を用いた．野球の投球動作およびゴルフのスイング動作が含まれる映像に対して，OpenPoseの25部位検出可能なモデルを適用した結果の一部をそれぞれ図3.2および図3.3に示す．なお，OpenPoseにより推定された人物の骨格を視認しやすくするために，入力映像フレームと同サイズの黒色画像上に検出された特徴点と骨格を表示している．

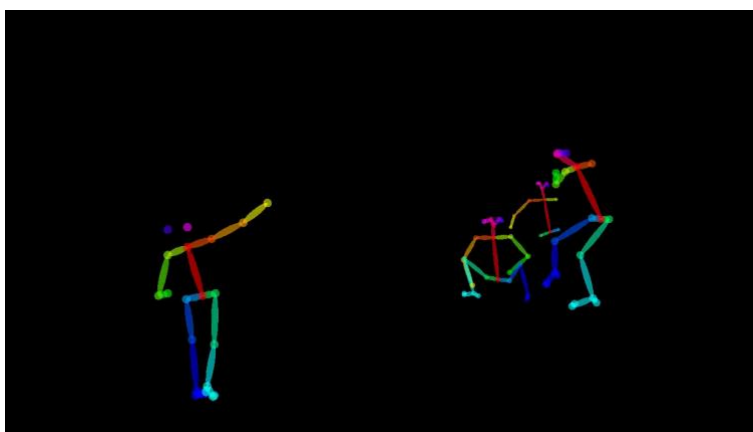


図 3.2 投球動作を含む映像に対するOpenPoseの推定結果

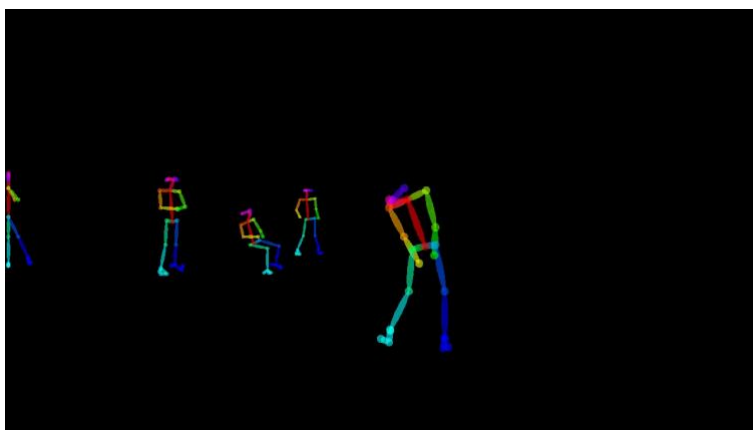


図 3.3 スイング動作を含む映像に対するOpenPoseの推定結果

図 3.2 や図 3.3 から確認できるように，本研究で使用した映像には対応付け対象のスポーツ動作を行なっている選手以外の人物が映っているケースが存在する．映像内の投球シーンでは，撮影カメラから見て左手前に投手，右奥に打者や捕手等が映る．図 3.2 では最も左の骨格が投球している選手を表す．一方，ゴルフのスイングシーンでは，画角中央にスイングしている選手が映り，選手の背中側に複数の人物が存在する．図 3.3 では最も右の骨格がスイング中の選手を表す．

本手法は，動作中の人物の姿勢類似度に基づき動作のタイミングを同期させるものである．そのため，動画内に複数人物が映っている場合は，対応付け対象のスポーツ動作を行なっている選手を特定する必要がある．

図 3.2 および図 3.3 において各人物の足首に着目すると，投球中もしくはスイング中の選手の足首が画像上の最も下にあることがわかる．本研究で使用した投球動作もしくはスイング動作を含む映像は，それぞれ図 3.2 または図 3.3 と同一のカメラアングルで撮影したものである．そこで，各映像フレームについて，OpenPose により推定された足首の座標値を比較し，足首の位置が画像上で最も下の人物を投球中もしくはスイング中の選手とする．

3.4 手足の左右反転検出の修正

3.4.1 左右反転検出の概要

本論文では OpenPose により取得された選手の各身体部位の 2 次元座標データを用いて，フレーム間の対応付けを決定する．そのため，OpenPose による推定位置が誤っている身体部位が多い場合，対応付け精度は低下する．

第 2.2.2 項で述べたように，OpenPose を映像に適用する際の問題点として，時系列の前後情報を考慮しない点が挙げられる．OpenPose は各映像フレームについて独立に人物の身体部位の位置を推定する．すなわち，時系列で前後のフレームにおける同一人物の検出結果は利用しない．そのため，時間的に連続する複数フレームにおいて，OpenPose により推定さ

れた身体部位の位置の変化が人間の動きとして現実的でない、誤った検出が生じる可能性がある。

人間の動きとして現実的でない誤検出の一例として、同一人物の腕や足が一部フレームで左右逆に検出され、腕または足の推定位置が関節の可動域を超えるほど微小時間で変化することが挙げられる。本研究で使用した一部の映像に対する OpenPose の推定結果には、選手の腕や足が左右逆に検出されているケースが含まれることを確認した。ゴルフスイング中の選手について、OpenPose による腕の検出が左右反転しているフレームを含む実行結果例を図 3.4 に示す。

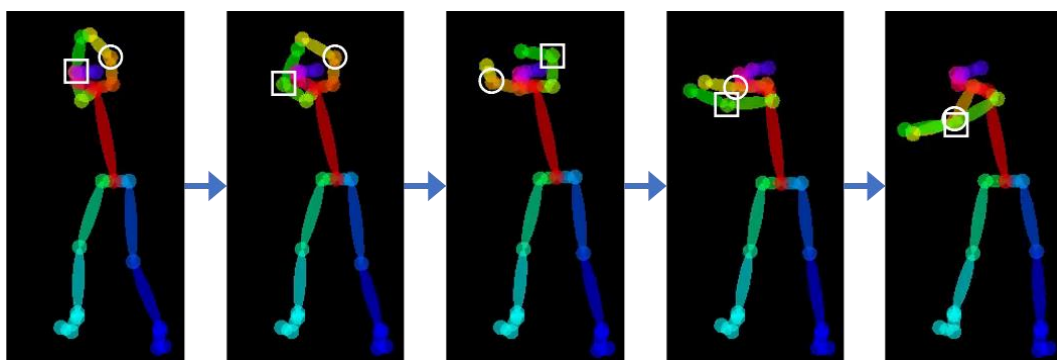


図 3.4 OpenPoseによる腕の左右反転検出例

図 3.4に示した画像は、同一映像内の時間的に連続した5フレームにおいてOpenPoseにより推定されたゴルフスイング中の選手の骨格である。なお、最も左の骨格が5フレームの中で時間的に一番前のフレームでの推定結果である。図 3.4において、色付きの点および線は OpenPoseにより検出された各身体部位の位置および身体部位同士を結ぶ骨格を示す。各身体部位のうち、右肘として推定された位置を白い丸、左肘として推定された位置を白い四角でそれぞれ囲んでいる。

図 3.4 に示した 5 つの推定結果と映像内の選手を見比べると、中央の推定結果について、右腕と左腕が左右逆に検出されていることが確認された。すなわち、選手の右肘や右手首が OpenPose により左肘、左手首として、選手の左肘や左手首が右肘、右手首として検出されていた。一方、図 3.4 の左 2 枚および右 2 枚の計 4 画像においては、腕の各部位の位置は正しく推定されていた。

腕の検出が左右反転している中央の検出結果では、白い四角で囲まれた左肘の推定位置が赤い線で示されている胴体の右上にあることが確認できる。一方、他の 4 枚の画像では、白い四角で囲まれた左肘の推定位置はいずれも赤い線で示された胴体の左側にある。図 3.4 に示した各推定結果がすべて正しいと仮定すると、左肘と胴体の左右の位置関係は中央の 3 フレームにおいて 2 度入れ替わっている。元映像は 30fps の動画のため、位置関係の 2 度の入れ替わりは約 0.07 秒の間に生じたこととなる。これは、人間の動作として現実的でない。

本論文では、時間的に連続する 2 フレームにおける各身体部位の OpenPose による推定結果を用いて、腕や足が左右逆に検出されているフレームを特定する。図 3.4 における左から 2 枚目および中央の画像での推定結果を比較すると、左膝として推定された特徴点を囲む白い四角の画像上での位置はやや異なることがわかる。それに対して、2 枚目の画像において右肘と推定された白い丸の位置と、中央の画像において左肘と推定された白い四角の位置はほぼ同一である。この例のように、時間的に連続する 2 フレームのうち、どちらか一方のフレームのみ OpenPose による人物の腕や足の検出が左右逆である場合、2 フレームで推定された同一部位の位置よりも対称部位の位置の方が近くなる。例外として、右腕と左腕もしくは右足と左足が画像上で重なるほど位置が近い場合は、同一部位の位置が対称部位の位置よりもわずかに近くなるケースが存在する。なお、同一部位とは右肘同士や右膝同士等の 2 部位を指し、対称部位とは右肘と左肘などのように、左右対称な 2 部位を表す。

3.4.2 左右反転検出の判定

本処理の概要をフレーム数 M の映像系列 $A = \{a_1, a_2, \dots, a_M\}$ を例に説明する。まず、系列 A に含まれるフレーム $a_i (2 \leq i \leq M)$ について、直前のフレーム a_{i-1} との同一部位の座標差 $D_Same(i)$ を式(3.1)により算出する。

$$D_Same(i) = \sum_{l=1}^4 d_keypoint(kp_data_{i,l}, kp_data_{i-1,l}) \quad (3.1)$$

ここで、式(3.1)の右辺に含まれる $kp_data_{i,l}$ とはフレーム a_i における選手の身体部位 l に関する OpenPose の推定結果である。なお、式(3.1)において身体部位 l は OpenPose により検出可能な 25 部位のうち、腕の 4 部位もしくは足の 4 部位のいずれかである。腕の左右反転を判定する場合は腕の 4 部位、足の左右反転を判定する場合は足の 4 部位を使用する。各 l の値に対応する身体部位を表 3.1 にまとめる。

表 3.1 各 l の値と身体部位の対応関係

l の値	腕の部位	足の部位
1	右肘	右膝
2	右手首	右足首
3	左肘	左膝
4	左手首	左足首

OpenPose による各身体部位 kp の推定結果には、画像上の 2 次元座標値 x, y および推定の信頼度 c_score が含まれる。フレーム a_m における選手の身体部位 l に関する OpenPose の推定結果 $kp_data_{m,l}$ を式(3.2)に示す。なお、 m は 1 以上 M 以下の自然数である。

$$kp_data_{m,l} = (x_{m,l}, y_{m,l}, c_score_{m,l}) \quad (3.2)$$

また、式(3.1)における $d_keypoint$ は2個の kp_data を引数にとり、2次元座標値 x, y の座標差を算出する関数である。ただし、少なくともどちらか一方の kp_data においてOpenPoseによる検出が失敗している、すなわち2次元座標値および推定の信頼度がすべて0の場合は、関数 $d_keypoint$ は0を返す。

続いて、系列 A に含まれるフレーム $a_i (2 \leq i \leq M)$ について、直前のフレーム a_{i-1} との対称部位の座標差 $D_Symmetry(i)$ を式(3.3)により算出する。

$$D_Symmetry(i) = \sum_{l=1}^4 d_keypoint(kp_data_{i,l}, kp_data_{i-1, Rev(l)}) \quad (3.3)$$

ここで、 $Rev(l)$ は式(3.4)により定義される。

$$Rev(l) = \begin{cases} l+2 & \text{if } 1 \leq l \leq 2 \\ l-2 & \text{if } 3 \leq l \leq 4 \end{cases} \quad (3.4)$$

式(3.3)において、座標差を算出する対称部位の組み合わせを表 3.2 に示す。

表 3.2 対称部位の組み合わせ

l の値	腕の左右反転検出		足の左右反転検出	
	フレーム a_i	フレーム a_{i-1}	フレーム a_i	フレーム a_{i-1}
1	右肘	左肘	右膝	左膝
2	右手首	左手首	右足首	左足首
3	左肘	右肘	左膝	右膝
4	左手首	右手首	左足首	右足首

式(3.1)および式(3.3)により算出した $D_Same(i)$ および $D_Symmetry(i)$ を用いて、式(3.5)により $D_Judge(i)$ を求める。

$$D_Judge(i) = D_Same(i) - D_Symmetry(i) \quad (3.5)$$

OpenPose による腕もしくは足の検出がフレーム a_i では左右逆であり、かつ直前のフレーム a_{i-1} では正しい場合、画像上における左右の腕もしくは足の距離が離れているほど同一部位の座標差 $D_Same(i)$ は大きくなる。それに伴い、 $D_Judge(i)$ の値も大きくなる。そこで、

$D_Judge(i)$ が正の値かつ、一定以上の値となった場合には、フレーム a_i において腕もしくは足の検出が左右逆であると判定する。

3.4.3 座標値の修正

第 3.4.2 項で述べた処理により腕もしくは足の検出が左右反転していると判定されたフレーム s_i について、各身体部位の座標値を修正する。腕の検出が左右逆であると判定された場合は、右肘と左肘および右手首と左手首のそれぞれの組み合わせについて、OpenPoseにより推定された2次元座標値および推定の信頼度を入れ替える。足の検出が左右逆であると判定された場合は、右膝と左膝および右足首と左足首のそれぞれの組み合わせについて、推定された2次元座標値および推定の信頼度を入れ替える。

3.5 姿勢変化に基づくフレーム数削減

本手法は、第 3.3 節で示した処理により取得された選手の骨格情報を利用して、フレーム間の姿勢類似度を算出する。その後、姿勢類似度に基づき、始点終点自由な DTW を用いて2系列のフレームを対応付ける。姿勢類似度の算出や始点終点自由な DTW による対応付けは、2本の映像系列のフレーム数に比例して計算量が増加する処理である。そこで、対応付け精度に影響を与えない範囲で、類似度算出や対応付け処理に使用するフレーム数を削減する。

本手法は2系列において選手の姿勢が類似しているフレーム同士を対応付ける。そのため、一方の系列 $A = \{a_1, a_2, \dots, a_M\}$ に含まれる連続フレーム a_{i-1} および $a_i (2 \leq i \leq M)$ において選手の姿勢が酷似している場合、もう一方の系列における a_{i-1} および a_i の対応付け先フレームは同一となることが予想される。よって、時間的に連続な複数フレームにおいて選手の姿勢が酷似している場合は、1フレームのみを後続の類似度算出や対応付け処理で使用し、残りのフレームは後続の処理で使用せずに対応付け先フレームを決定することが可能となる。

削減するフレームは直前フレームからの選手の姿勢変化量に基づき決定される。映像系列 A に含まれるフレーム $a_i (2 \leq i \leq M)$ における、フレーム a_{i-1} からの選手の姿勢変化量 $D_Pose(i)$ は式(3.6)により算出される。

$$D_Pose(i) = \sum_{kp=1}^{12} d_keypoint(kp_data_{i,kp}, kp_data_{i-1,kp}) \quad (3.6)$$

ここで、式(3.6)の右辺に含まれる $kp_data_{i,kp}$ とはフレーム a_i における選手の身体部位 kp に関する OpenPose の推定結果である。なお、式(3.6)において身体部位 kp は OpenPose により検出可能な 25 部位のうち、両肩、両肘、両手首、両腰、両膝、両足首の 12 部位のいずれか

である。目や耳、足のつま先等の 13 部位は上記の 12 部位よりも選手の姿勢情報を評価する上で重要度が低いと推測したため、姿勢変化量の算出には使用しない。OpenPose による各身体部位 kp の推定結果には、画像上の 2 次元座標値 x, y および推定の信頼度 c_score が含まれる。フレーム a_m における選手の身体部位 kp に関する OpenPose の推定結果 $kp_data_{m,kp}$ は第 3.4.2 項の式(3.2)に示した通りである。

また、式(3.6)における $d_keypoint$ は 2 個の kp_data を引数にとり、2 次元座標値の座標差を算出する関数である。ただし、少なくともどちらか一方の kp_data において OpenPose による検出が失敗している、すなわち 2 次元座標値および推定の信頼度がすべて 0 の場合は、関数 $d_keypoint$ は 0 を返す。

式(3.6)に示したように、 a_1 を除く各フレームでの姿勢変化量は、肩や肘などの 12 部位における直前のフレームとの座標差に基づき算出される。姿勢変化量が 0 に近い、すなわち直前のフレームとの座標差が小さいほど、選手の姿勢が酷似していることを表す。よって、姿勢変化量 $D_Pose(i)$ が 0 に近い場合は、類似度算出や対応付け処理においてフレーム a_i は使用しない。例外として、 i が 5 以上、かつ直前の 4 フレーム $\{a_{i-4}, a_{i-3}, a_{i-2}, a_{i-1}\}$ がすべて削減すると判定されている場合は、フレーム a_i は削減しない。これは、連続して 5 フレーム以上削減した場合に対応付け精度が低下することを予備実験で確認したためである。

3.6 ベクトルデータの作成

3.6.1 概要

第 3.3 節で述べた処理により取得した選手の骨格情報からベクトルデータを作成する。本論文では、OpenPose により検出可能な 25 部位のうち、両肩、両肘、両手首、両腰、両膝、両足首の 12 点の推定結果を使用する。これは、目や耳、足のつま先等の他の 13 部位は上記の 12 部位よりも選手の姿勢情報を評価する上で重要度が低いと推測したためである。以降の説明で使用するため、本論文で用いる 12 部位それぞれに身体部位番号 $kp(1 \leq kp \leq 12)$ を割り当てる。なお、身体部位番号 kp は対応付け対象のスポーツ動作を行う選手の動作中の利き手によって異なる。ここで、投球動作とゴルフ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。本論文で用いる 12 部位と身体部位番号 kp の対応関係を表 3.3 にまとめる。

表 3.3 各身体部位と kp の値の対応関係

身体部位	kp の値	
	利き手：右	利き手：左
右肩	1	4
右肘	2	5
右手首	3	6
左肩	4	1
左肘	5	2
左手首	6	3
右腰	7	10
右膝	8	11
右足首	9	12
左腰	10	7
左膝	11	8
左足首	12	9

対応付け対象のスポーツ動作を行なっている選手の 12 部位の推定結果を用いて、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は各フレームにおける選手の姿勢情報を、時間ベクトルデータは各身体部位の微小時間での動きをそれぞれ表すベクトルの集合である。

各ベクトルデータについて、フレーム数 M の映像系列 $A = \{a_1, a_2, \dots, a_M\}$ から第 3.5 節で述べた処理により削減されたフレームを除いた、 $A' = \{a'_1, a'_2, \dots, a'_P\} (P \leq M)$ を例に、詳細を第 3.6.2 項および第 3.6.3 項で述べる。

3.6.2 空間ベクトル

系列 $A' = \{a'_1, a'_2, \dots, a'_P\}$ に含まれるフレーム $a'_j (1 \leq j \leq P)$ における空間ベクトルデータ $V_Spatial(j)$ を式(3.7)に示す。

$$V_Spatial(j) = \{vs_{j,1}, vs_{j,2}, \dots, vs_{j,66}\} \quad (3.7)$$

ここで、空間ベクトルデータ $V_Spatial(j)$ を構成する $vs_{j,v} (1 \leq v \leq 66)$ は、フレーム a'_j における OpenPose の推定結果から作成されるベクトルである。表 3.3 に示した 12 点の身体部位のうち、任意の 2 部位の推定結果を用いて各空間ベクトルの長さ r_vs 、角度 θ_vs 、信頼度 c_score_vs を算出する。空間ベクトル $vs_{j,v}$ を作成する際に用いる 2 部位（以下、構成部位）と添字 v の値の対応関係を表 3.4 にまとめる。なお、構成部位は表 3.3 に示した各部位に対

応する身体部位番号 kp を用いて表す.

表 3.4 構成部位の身体部位番号と添字 v の値の対応関係

構成部位 1 の 身体部位番号	構成部位 2 の身体部位番号											
	1	2	3	4	5	6	7	8	9	10	11	12
1	—	1	2	3	4	5	6	7	8	9	10	11
2	—	—	12	13	14	15	16	17	18	19	20	21
3	—	—	—	22	23	24	25	26	27	28	29	30
4	—	—	—	—	31	32	33	34	35	36	37	38
5	—	—	—	—	—	39	40	41	42	43	44	45
6	—	—	—	—	—	—	46	47	48	49	50	51
7	—	—	—	—	—	—	—	52	53	54	55	56
8	—	—	—	—	—	—	—	—	57	58	59	60
9	—	—	—	—	—	—	—	—	—	61	62	63
10	—	—	—	—	—	—	—	—	—	—	64	65
11	—	—	—	—	—	—	—	—	—	—	—	66
12	—	—	—	—	—	—	—	—	—	—	—	—

表 3.4 における 1 列目および 2 行目の数字は構成部位に対応する身体部位番号 kp を表す. その他のマスに記述されている数字は空間ベクトル $vs_{j,v}$ に含まれる添字 v の値を示す. 空間ベクトル $vs_{j,v}$ の構成部位にそれぞれ対応する身体部位番号は, 表 3.4 において添字 $v = v'$ が記述されているマスと同じ行の 1 列目の数字, および同じ列の 2 行目の数字である. 例として, $v = 33$ の場合, 空間ベクトル $vs_{j,33}$ の構成部位に対応する身体部位番号は 4 と 7 である. 表 3.3 より, 空間ベクトル $vs_{j,33}$ の構成部位は選手の動作中の利き手が右である場合は左肩と右腰, 利き手が左の場合は右肩と左腰となる.

空間ベクトルの長さ r_{vs} は構成部位の 2 次元座標値の距離である. 角度 θ_{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. この時, ベクトルの始点は, 構成部位のうち, 表 3.3 に示した対応する身体部位番号 kp が小さい部位の推定位置とする. 例として, 右肘と右手首の 2 部位を結ぶ空間ベクトルの長さ r_{vs} および角度 θ_{vs} を図 3.5 に示す. 表 3.3 において身体部位番号 kp が小さい右肘の推定位置が空間ベクトルの始点となる. なお, OpenPose により推定された骨格を視認しやすくするために, 右肘と右手首を結ぶ空間ベクトルを表す白の矢印は骨格と重ねずに下へずらして表示している.

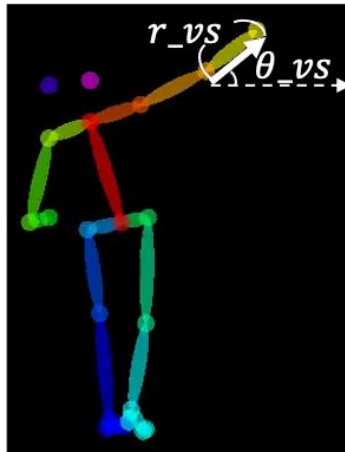


図 3.5 空間ベクトルの長さおよび角度

空間ベクトルの信頼度 c_score_vs は OpenPose によって得られる身体部位の検出の信頼度から算出される。フレーム a'_j において OpenPose により推定された 2 部位 $kp1$, $kp2$ の位置を結ぶ空間ベクトルの信頼度 $c_score_vs_{j,(kp1,kp2)}$ は、各部位の推定の信頼度 $c_score_{j,kp1}$ および $c_score_{j,kp2}$ を用いて、式(3.8)により算出される。

$$c_score_vs_{j,(kp1,kp2)} = \min(c_score_{j,kp1}, c_score_{j,kp2}) \quad (3.8)$$

なお、空間ベクトルを構成する 2 部位のうち、少なくとも一方の部位が OpenPose により検出されていない場合は、ベクトルの長さ、角度を -1 、信頼度を 0 とする。

3.6.3 時間ベクトル

系列 $A' = \{a'_1, a'_2, \dots, a'_P\}$ に含まれるフレーム $a'_j (1 \leq j \leq P)$ における時間ベクトルデータ $V_Temporal(j)$ を式(3.9)に示す。

$$V_Temporal(j) = \{vt_{j,1}, vt_{j,2}, \dots, vt_{j,12}\} \quad (3.9)$$

ここで、時間ベクトルデータ $V_Temporal(j)$ を構成する $vt_{j,kp} (1 \leq kp \leq 12)$ は、フレーム a'_j を含む 3 フレームでの OpenPose の推定結果を基に算出される。なお、 kp は表 3.3 に示した各身体部位に対応する番号である。式(3.10)に示す通り、 $vt_{j,kp}$ は時間ベクトル $vt1_{j,kp}$ および $vt2_{j,kp}$ により構成される。

$$vt_{j,kp} = \{vt1_{j,kp}, vt2_{j,kp}\} \quad (3.10)$$

ここで、時間ベクトル $vt1_{j,kp}$ および $vt2_{j,kp}$ は式(3.11)および式(3.12)により定義される。なお、本論文では系列 A' に含まれるフレーム数 P は6以上であると仮定している。

$$vt1_{j,kp} = \begin{cases} (r_vt_{(j-3,j),kp}, \theta_vt_{(j-3,j),kp}, c_score_vt_{(j-3,j),kp}) & \text{if } 3 < j \leq P \\ (-1, -1, 0) & \text{if } 0 \leq j \leq 3 \end{cases} \quad (3.11)$$

$$vt2_{j,kp} = \begin{cases} (r_vt_{(j,j+3),kp}, \theta_vt_{(j,j+3),kp}, c_score_vt_{(j,j+3),kp}) & \text{if } 0 \leq j \leq P - 3 \\ (-1, -1, 0) & \text{if } P - 3 < j \leq P \end{cases} \quad (3.12)$$

時間ベクトル $vt1_{j,kp}$ および $vt2_{j,kp}$ は系列 A' 内で近接する2フレームにおける同一部位の推定結果に基づき作成される。ここで、近接するとは系列 A' 内においてフレーム間隔が3であることを指す。なお、式(3.11)、式(3.12)において下の条件に該当する場合、すなわち3フレーム前の a'_{j-3} もしくは3フレーム後のフレーム a'_{j+3} が存在しない場合は、ベクトルの各要素は定数値とする。

式(3.11)、式(3.12)における c_score_vt は時間ベクトルの信頼度を表す。信頼度 c_score_vt はOpenPoseにより得られる身体部位の推定の信頼度 c_score を用いて、式(3.13)により算出される。

$$\begin{aligned} c_score_vt_{(j-3,j),kp} &= \min(c_score_{j-3,kp}, c_score_{j,kp}) \\ c_score_vt_{(j,j+3),kp} &= \min(c_score_{j,kp}, c_score_{j+3,kp}) \end{aligned} \quad (3.13)$$

ただし、式(3.13)の右辺に含まれる二つの推定の信頼度のうち、少なくとも一つの信頼度が閾値未満である場合は、 c_score_vt は0とする。なお、 $c_score_{j,kp}$ はフレーム a'_j においてOpenPoseにより推定された、 kp に対応する身体部位の検出の信頼度である。

式(3.11)、式(3.12)に含まれる r_vt および θ_vt は、近接する2フレームにおける同一部位の移動距離および移動方向を表す。フレーム a'_{j-3} とフレーム a'_j における左手首を例に、移動距離 r_vt および移動方向 θ_vt を図3.6に示す。なお、図3.6では $3 < j$ が満たされているものとする。

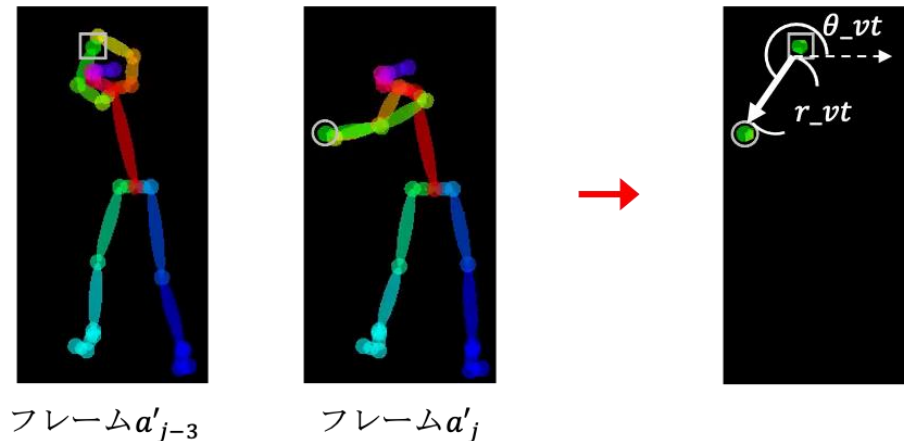


図 3.6 時間ベクトルの長さおよび角度の例

図 3.6 の左 2 枚の画像には、それぞれフレーム a'_{j-3} およびフレーム a'_j において OpenPose により推定された選手の骨格を表示している。最も右の画像における緑の点は 2 フレームにおいてそれぞれ推定された左手首の画像上での位置を示す。各フレームにおいて推定された左手首の位置を灰色の四角および丸でそれぞれ囲んでいる。

移動距離 r_{vt} および移動方向 θ_{vt} は近接する 2 フレームにおいて OpenPose により取得される同一部位の 2 次元座標値を用いて算出される。図 3.6 に示した例では、フレーム a'_{j-3} で推定された左手首の 2 次元座標値を始点、フレーム a'_j で推定された左手首の 2 次元座標値を終点とする白色ベクトルの長さが移動距離 r_{vt} 、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が移動方向 θ_{vt} である。なお、少なくとも一方のフレームにおいて OpenPose により得られる身体部位の推定の信頼度が閾値未満である場合は、移動距離 r_{vt} および移動方向 θ_{vt} は共に -1 とする。

3.7 姿勢類似度の算出

3.7.1 概要

第 3.6 節で述べた処理により作成され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求める。姿勢類似度は第 3.6.2 項で述べた空間ベクトルおよび第 3.6.3 項で述べた時間ベクトルを用いて算出される。系列 $A' = \{a'_1, a'_2, \dots, a'_P\}$ に含まれるフレーム $a'_j (1 \leq j \leq P)$ と系列 $B' = \{b'_1, b'_2, \dots, b'_Q\}$ に含まれるフレーム $b'_k (1 \leq k \leq Q)$ を例に、姿勢類似度の算出処理の概要を以下で述べる。なお、 A' および B' は、それぞれフレーム数が M 、 N の映像系列 $A = \{a_1, a_2, \dots, a_M\}$ および $B = \{b_1, b_2, \dots, b_N\}$ から、第 3.5 節の処理により削除対象としたフレームを取り除いて得られる系列である。系列 A' のフレーム数 P は M 以下の自然数、系列 B' のフレーム数 Q は N 以下の自然数である。

3.7.2 空間ベクトルの差

系列 A' および B' に含まれるフレーム a'_j および b'_k における空間ベクトルデータを式(3.14)に示す.

$$\begin{aligned} V_Spatial_A'(j) &= \{vs_A'_{j,1}, vs_A'_{j,2}, \dots, vs_A'_{j,66}\} \\ V_Spatial_B'(k) &= \{vs_B'_{k,1}, vs_B'_{k,2}, \dots, vs_B'_{k,66}\} \end{aligned} \quad (3.14)$$

空間ベクトルデータ $V_Spatial_A'(j)$ および $V_Spatial_B'(j)$ にはそれぞれ 66 次元の空間ベクトルが含まれる. 第 3.6.2 項で述べたように, 各空間ベクトルは, 表 3.3 に示した 12 点の身体部位のうち任意の 2 部位の推定結果に基づき算出された長さ, 角度, 信頼度の 3 要素を持つ.

フレーム a'_j および b'_k における空間ベクトルの差 $D_Spatial(j, k)$ は式(3.15)により定義される.

$$D_Spatial(j, k) = \frac{d_vs(j, k)}{66 - zero_count_vs} \quad (3.15)$$

ここで, 式(3.15)に含まれる $d_vs(j, k)$ は 66 次元の空間ベクトルを用いて, 式(3.16)により算出される.

$$d_vs(j, k) = \sum_{v=1}^{66} \left(d_length(vs_A'_{j,v}, vs_B'_{k,v}) \times d_angle(vs_A'_{j,v}, vs_B'_{k,v}) \right) \quad (3.16)$$

式(3.16)における d_length および d_angle は対応する空間ベクトルを引数にとり, 空間ベクトル間の長さの差, および角度の差をそれぞれ算出する関数である. ただし, 対応する空間ベクトルのうち, 少なくとも一方のベクトルの信頼度が閾値未満である場合は長さの差, および角度の差は計算されない. これは, OpenPose による検出が失敗した身体部位が姿勢類似度に与える影響を減らすためである. また, ベクトルの信頼度が低い, すなわち OpenPose により得られる構成部位の推定の信頼度が低い場合, 推定された身体部位の位置が不正確であるケースが多いことを予備実験により確認したためである. 少なくとも一方のベクトルの信頼度が 0 である場合は, 関数 d_length および d_angle は 0 を返し, 式(3.15)における $zero_count_vs$ に 1 が加算される. それに対して, 少なくとも一方のベクトルの信頼度が 0 より大きく閾値未満である場合は, 関数 d_length および d_angle は定数値を返す.

3.7.3 時間ベクトルの差

系列 A' および B' に含まれるフレーム a'_j および b'_k における時間ベクトルデータを式(3.17)

に示す.

$$\begin{aligned} V_Temporal_A'(j) &= \{vt_A'_{j,1}, vt_A'_{j,2}, \dots, vt_A'_{j,12}\} \\ V_Temporal_B'(k) &= \{vt_B'_{k,1}, vt_B'_{k,2}, \dots, vt_B'_{k,12}\} \end{aligned} \quad (3.17)$$

式(3.17)に含まれる $vt_A'_{j,kp}$ および $vt_B'_{k,kp}$ ($1 \leq kp \leq 12$)はそれぞれ 2 次元の時間ベクトルを用いて, 式(3.18)により定義される.

$$\begin{aligned} vt_A'_{j,kp} &= \{vt1_A'_{j,kp}, vt2_A'_{j,kp}\} \\ vt_B'_{k,kp} &= \{vt1_B'_{k,kp}, vt2_B'_{k,kp}\} \end{aligned} \quad (3.18)$$

第 3.6.3 項で述べたように, 時間ベクトルデータ $V_Temporal_A'(j)$ および $V_Temporal_B'(j)$ を構成する各時間ベクトルは, 近接する 2 フレームにおける同一部位の推定結果に基づき算出された移動距離, 移動方向, 信頼度の 3 要素を持つ. なお, 同一部位とは表 3.3 に示した 12 部位のいずれかである.

フレーム a'_j および b'_k における時間ベクトルの差 $D_Temporal(j, k)$ は式(3.19)により定義される.

$$D_Temporal(j, k) = \frac{d_vt1(j, k) + d_vt2(j, k)}{24 - zero_count_vt} \quad (3.19)$$

ここで, 式(3.19)に含まれる $d_vt1(j, k)$ および $d_vt2(j, k)$ は時間ベクトルを用いて, 式(3.20)および式(3.21)により算出される.

$$\begin{aligned} d_vt1(j, k) &= \sum_{kp=1}^{12} \left(d_distance(vt1_A'_{j,kp}, vt1_B'_{k,kp}) \right. \\ &\quad \times d_direction(vt1_A'_{j,kp}, vt1_B'_{k,kp}) \\ &\quad \left. \times max_weight(vt1_A'_{j,kp}, vt1_B'_{k,kp}) \right) \end{aligned} \quad (3.20)$$

$$\begin{aligned}
d_vt2(j, k) = & \sum_{kp=1}^{12} \left(d_distance(vt2_A'_{j,kp}, vt2_B'_{k,kp}) \right. \\
& \times d_direction(vt2_A'_{j,kp}, vt2_B'_{k,kp}) \\
& \left. \times max_weight(vt2_A'_{j,kp}, vt2_B'_{k,kp}) \right)
\end{aligned} \tag{3.21}$$

式(3.20)および式(3.21)における $d_distance$ および $d_direction$ は対応する時間ベクトルを引数にとり、時間ベクトルの要素である移動距離の差、および移動方向の差をそれぞれ算出する関数である。ただし、対応する時間ベクトルのうち、少なくとも一方のベクトルの信頼度が0である場合は移動距離の差、および移動方向の差は計算されない。これは、第3.7.2項と同様、OpenPoseによる検出が失敗した身体部位が姿勢類似度に与える影響を減らすためである。少なくとも一方のベクトルの信頼度が0である場合は、関数 $d_distance$ および $d_direction$ は0を返し、式(3.19)における $zero_count_vs$ に1が加算される。

また、式(3.20)および式(3.21)に含まれる max_weight は対応する時間ベクトルの重みのうち、より大きい値を返す関数である。ここで、時間ベクトル $vt1_A'_{j,kp}$ および $vt1_B'_{k,kp}$ の重み $weight_vt1_A'_{j,kp}$ 、 $weight_vt1_B'_{k,kp}$ は式(3.22)および式(3.23)により定義される。

$$weight_vt1_A'_{j,kp} = \frac{r_vt1_A'_{j,kp}}{\sum_{kp'=1}^{12} r_vt1_A'_{j,kp'}} \tag{3.22}$$

$$weight_vt1_B'_{k,kp} = \frac{r_vt1_B'_{k,kp}}{\sum_{kp'=1}^{12} r_vt1_B'_{k,kp'}} \tag{3.23}$$

ここで、 $r_vt1_A'_{j,kp}$ および $r_vt1_B'_{k,kp}$ はそれぞれ時間ベクトル $vt1_A'_{j,kp}$ および $vt1_B'_{k,kp}$ の移動距離である。式(3.23)により移動距離の値が大きいベクトルほど、重みの値も大きくする。これは、移動距離の値が大きい、すなわち近接する2フレームにおいて画像上での位置が離れている身体部位ほど、姿勢類似度における重要度が高いと推察したためである。

3.7.4 姿勢類似度

系列 A' および B' に含まれるフレーム a'_j および b'_k におけるフレーム間の姿勢類似度は、第3.7.2項、第3.7.3項により算出された $D_Spatial(j, k)$ と $D_Temporal(j, k)$ の和で定義される。

姿勢類似度は 0 以上の数値であり, 0 に近いほど類似していることを表す.

3.8 始点終点自由な Dynamic Time Warping による対応付け

本論文では, 第 2.4 節で述べた始点終点自由な DTW[12]を用いて, 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. フレーム数がそれぞれ M , N の映像系列 $A = \{a_1, a_2, \dots, a_M\}$ および $B = \{b_1, b_2, \dots, b_N\}$ を例に, 対応付け処理の概要を以下で述べる. なお, 2 本の映像系列には対応付け対象となるスポーツ動作が 1 回分含まれているものとする. 映像系列 A において, 対応付け対象の動作の開始フレームおよび終了フレームはそれぞれ a_1 と a_M であると仮定し, 始点終点自由な DTW により系列 A のすべてのフレームを系列 B の少なくとも一つ以上のフレームと対応付ける.

系列 A および B から, 第 3.5 節の処理により削除対象としたフレームを取り除いて得られる系列を $A' = \{a'_1, a'_2, \dots, a'_P\}$ ($P \leq M$) および $B' = \{b'_1, b'_2, \dots, b'_Q\}$ ($Q \leq N$) とする.

式(3.24)にサイズを $(P+1) \times (Q+1)$ とする行列 D の初期条件を示す. なお, 式(3.24)において j は 1 以上 P 以下の自然数, k は 1 以上 Q 以下の自然数である.

$$D(0,0) = D(0,k) = 0, \quad D(j,0) = \infty \quad (3.24)$$

行列 D の残りの要素については, 第 3.7 節の処理により算出したフレーム間の姿勢類似度を用いて再帰的に求める. 系列 A' のフレーム a'_j ($1 \leq j \leq P$) と系列 B' のフレーム b'_k ($1 \leq k \leq Q$) における姿勢類似度 $C(a'_j, b'_k)$ を用いて, $D(j,k)$ は式(3.25)により定義される.

$$D(j,k) = \min \left\{ \begin{array}{l} \frac{C(a'_j, b'_k) + D(j-1, k) \times L(j-1, k)}{L(j-1, k) + 1} \\ \frac{C(a'_j, b'_k) + D(j, k-1) \times L(j, k-1)}{L(j, k-1) + 1} \\ \frac{C(a'_j, b'_k) + D(j-1, k-1) \times L(j-1, k-1)}{L(j-1, k-1) + 1} \end{array} \right. \quad (3.25)$$

ここで, L は行列 D と同サイズの行列であり, 初期条件は式(3.26)に示す通りである.

$$L(0,0) = L(j,0) = 0, \quad L(0,k) = 1 \quad (3.26)$$

なお, $L(j,k)$ は式(3.25)の右辺に示した 3 つの分数のうち, 最小値となる分数の分母として与えられる. また, 3 つの分数のうち最小値として $D(j,k)$ の計算に用いられた値の添字情報 $DI_{(j,k)}$ を記録する. 例えば, $D(j-1, k)$ が最小値として計算に用いられた場合は, $DI_{(j,k)A'} = j-1, DI_{(j,k)B'} = k$ と記録される.

式(3.25)により行列 D の全成分を算出後、添字情報 $DI_{(j,k)}$ を利用しフレーム間の対応付けを決定する。まず、 $D(P,k)$ ($1 \leq k \leq Q$)に基づき $D'(P,k)$ の値を算出する。 $D'(P,k)$ は式(3.27)により定義される。

$$D'(P,k) = \frac{D(P,k) \times L(P,k)}{P} \quad (3.27)$$

式(3.27)により算出された、 $D'(P,k)$ の値をすべて比較し、 $D'(P,k)$ が最小となる際の添字 k とおく。

次に、 $D(P,k')$ の計算に用いられた値の添字情報 $DI_{(P,k')}$ を参照する。この時、 $DI_{(P,k')_{A'}} = P - 1$ である場合は、系列 B' の k' 番目のフレーム $b'_{k'}$ が系列 B' の対応付けの終点となる。なお、フレーム $b'_{k'}$ が系列 B' の対応付けの終点となることは、 $b'_{k'}$ よりも時間的に後のフレームが系列 A' のいずれのフレームとも対応付けされないことを指す。一方、 $DI_{(P,k')_{A'}} \neq P - 1$ である場合は、フレーム $b'_{k'-1}$ を対応付けの終点の候補とし、 $D(P,k'-1)$ の計算に用いられた値の添字情報 $DI_{(P,k'-1)}$ を参照する。この操作を、 $DI_{(P,k'-e)_{A'}} = P - 1$ となるまで、 e の値を1から1ずつ増加させながら繰り返す。その後、対応付けの終点の候補に含まれるフレーム $b'_{k'-e'}$ ($1 \leq e' \leq e$)および、 $b'_{k'}$ について、系列 A' の最後のフレーム a'_p との姿勢類似度を比較する。複数フレーム $b'_{k'}$ および $b'_{k'-e'}$ ($1 \leq e' \leq e$)のなかで、フレーム a'_p との姿勢類似度が最も小さいフレームを対応付けの終点とする。

対応付けの終点決定後、添字情報に基づき、系列 A' のすべてのフレームについて対応付け先を決定する。

まず、系列 A' の最後のフレームと系列 B' の対応付けの終点であるフレームを対応付ける。ここで、対応付けた2フレームを $a'_{j'}$ および $b'_{k'}$ とおく。

次に、 $D(j',k')$ の計算に用いられた値の添字情報 $DI_{(j',k')}$ に基づき、系列 A' のフレーム $a'_{DI_{(j',k')_{A'}}$ と系列 B' のフレーム $b'_{DI_{(j',k')_{B'}}$ を対応付ける。なお、添字情報の性質上、対応付けされるフレームのペアは、 $(a'_{j'-1}, b'_{k'})$, $(a'_{j'}, b'_{k'-1})$, $(a'_{j'-1}, b'_{k'-1})$ の三つのいずれかである。

続いて、対応付けた2フレームを $a'_{j'}$ および $b'_{k'}$ とおき、一つ前のパラグラフの処理を再度行う。この操作を系列 A' の最初のフレーム a'_1 の対応付け先が決定するまで、再帰的に繰り返す。

最後に、系列 A' の最初のフレーム a'_1 と対応付けされた系列 B' のフレームのなかで、 a'_1 との姿勢類似度が最小であるフレームを系列 B' の対応付けの始点とする。これは、系列 B' に含まれる、対応付けの始点よりも時間的に前のフレームが、系列 A' のいずれのフレームとも対応付けされないことを表す。

系列 A' に含まれるすべてのフレームの対応付け先を決定後、第3.5節の処理により削減されたフレームを対応付ける。削減されたフレームの対応付け先は、始点終点自由なDTWにより求めた各対応付けに基づき、決定される。系列 A において削減されたフレームは系列 B

の少なくとも一つ以上のフレームと対応付けされる。一方，系列 B において削減されたフレームは，系列 B' の対応付けの始点より時間的に後，かつ対応付けの終点よりも時間的に前である場合のみ，系列 A に含まれる少なくとも一つ以上のフレームと対応付けされる。

3.9 むすび

本章では，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提案手法について述べた。

第4章 評価実験および考察

4.1 まえがき

本章では，提案手法に対する評価実験の結果を示す．また，結果をもとに提案手法の有効性を考察する．

4.2 実験の概要

4.2.1 データセット

本研究では，野球の投球動作もしくはゴルフのスイング動作が含まれる映像を用いた．実験にて使用した映像の詳細を表 4.1 にまとめる．

表 4.1 実験にて使用した映像の詳細

	投球動作を含む映像	スイング動作を含む映像
総動画数 [本]	124	85
右投げ or 右打ちの動画数 [本]	122	83
左投げ or 左打ちの動画数 [本]	2	2
フレームレート [fps]	30	30
動作の回数 [回]	1	1

本研究で使用した映像のうち，右投げの選手による投球動作が含まれる映像 122 本を投球データセット BP1，右打ちの選手によるスイング動作が含まれる映像 83 本をスイングデータセット GS1 とする．各映像には，動作の開始から終了までを指す動作区間と動作開始前および終了前を指す非動作区間が含まれている．なお，投球動作の開始は軸足でない足が地面から離れる瞬間，終了は蹴り上げた軸足が地面に着く直前とした．右投げの選手の場合，軸足は右足となり，左投げの選手の場合，軸足は左足となる．一方，ゴルフスイング動作の開始はテイクバックが始まる瞬間，終了はボールを打ってから体の回転が静止するまでの間で利き手側の肘が最も高い位置にある瞬間とした．ここで，テイクバックとは，構えた位置からゴルフクラブを後方に引く動作を指す．また，利き手側の肘は，右打ちの選手の場合は右肘，左打ちの選手の場合は左肘とする．

投球データセット BP1 およびスイングデータセット GS1 に含まれる映像の詳細を表 4.2 にまとめる．なお，動画フレーム数が最長である動画と，動作区間または非動作区間のフレーム数が最長である動画は一致しているとは限らない．また，最短についても同様である．

表 4.2 投球データセットBP1およびスイングデータセットGS1の詳細

		BP1	GS1
動画数 [本]		122	83
動画フレーム数	最長	343	2460
	最短	76	120
	平均	172.1	340.5
動作区間の フレーム数	最長	107	88
	最短	39	49
	平均	69.2	65.7
非動作区間の フレーム数	最長	289	2405
	最短	22	52
	平均	102.9	274.8

投球データセット BP1 に含まれる各映像から目視により投球動作区間のみを切り出した動画 122 本を投球データセット BP2 とする。また、スイングデータセット GS1 に含まれる各映像から目視によりゴルフスイング動作区間のみを切り出した動画 83 本をスイングデータセット GS2 とする。投球データセット BP2 およびスイングデータセット GS2 に含まれる各映像について、最初のフレームが動作の開始時、最後のフレームが動作の終了時とそれぞれ一致する。

なお、投球データセット BP2 に含まれる 122 本の動画から無作為に選んだ 2 本の動画をそれぞれモデル投球動画 1 およびモデル投球動画 2 とする。また、スイングデータセット GS2 に含まれる 83 本の動画から無作為に選んだ 2 本の動画をそれぞれモデルスイング動画 1 およびモデルスイング動画 2 とする。

さらに、本研究で使用した映像のうち、左投げの選手による投球動作が含まれる映像 2 本をモデル投球動画 3 およびモデル投球動画 4、左打ちの選手によるスイング動作が含まれる映像 2 本をモデルスイング動画 3 およびモデルスイング動画 4 とする。なお、各モデル動画には動作区間のみが含まれており、最初のフレームが動作の開始時、最後のフレームが動作の終了時とそれぞれ一致するものとする。

各モデル投球動画の詳細を表 4.3 に、各モデルスイング動画の詳細を表 4.4 にそれぞれまとめる。

表 4.3 各モデル投球動画の詳細

	動画 1	動画 2	動画 3	動画 4
右投げ or 左投げ フレーム数	右投げ 84	右投げ 80	左投げ 69	左投げ 77

表 4.4 各モデルスイング動画の詳細

	動画 1	動画 2	動画 3	動画 4
右打ち or 左打ち	右打ち	右打ち	左打ち	左打ち
フレーム数	59	65	48	57

なお、本研究で用いた投球動作が含まれる映像およびスイング動作が含まれる映像は、それぞれ図 3.2 および図 3.3 と同一のカメラアングルで撮影されたものである。

4.2.2 評価項目

本論文では、以下に述べる 3 つの評価項目により提案手法の有効性を考察する。

まず、第 3.4 節で述べた OpenPose による左右反転検出に関する判定精度を求める。第 3.4.2 項に示した処理の結果、OpenPose によって推定された腕もしくは足が左右逆であると判定されたフレームについて、元映像と OpenPose による選手の推定結果を比較し、腕または足の検出が誤っているかを確認する。

続いて、第 3.7 節に示した姿勢類似度の算出手法について、ケンドールの順位相関係数 [13]を用いて定量的に評価する。なお、評価指標として用いたケンドールの順位相関係数は第 4.4 節で説明する。

最後に、第 3.8 節で述べた始点終点自由な DTW について、動作区間の抽出精度を確認する。提案手法により求めた動作の開始フレームおよび終了フレームと、元映像から目視により決定した動作の開始フレームおよび終了フレームとの誤差を算出する。

4.3 左右反転検出の修正に関する評価

投球データセット BP1 およびスイングデータセット GS1 に含まれる映像に対して、第 3.4.2 項で述べた OpenPose による左右反転検出の判定処理を適用した。処理の結果、OpenPose による選手の腕もしくは足の検出が左右逆であると判定されたフレームについて、元映像と OpenPose による選手の推定結果を比較し、腕または足が左右逆に検出されているかを確認した。左右反転検出に関する評価実験結果を表 4.5 に示す。

表 4.5 OpenPoseによる左右反転検出の判定精度

	腕の左右反転	足の左右反転
判定フレーム数	39	105
左右反転フレーム数	30	105
判定精度 [%]	76.9	100

表 4.5 における判定フレームとは、第 3.4.2 項で述べた処理により、OpenPose による腕もしくは足の検出が左右逆であると判定されたフレームを指す。また、左右反転フレーム数と

は、判定フレームのうち実際に OpenPose による検出が左右逆になっていることが確認されたフレーム数である。

表 4.5 に示した実験結果より、足の左右反転検出においては、すべての判定フレームにおいて OpenPose により推定された足の位置が左右逆、すなわち右膝や右足首が左膝や左足首として、左膝や左足首が右膝や右足首として検出されていた。一方、腕の左右反転検出の判定精度は足の左右反転検出と比べ低く、約 77%であった。誤判定フレームの原因について、次の段落で考察する。なお、誤判定フレームとは判定フレームのうち、OpenPose による腕の検出が正確、すなわち左右逆でないフレームを指す。

第 2.2.2 項で示したように、OpenPose は画像のみから人物の骨格を推定するため、身体部位が画像上に映らない場合には検出精度が低下するという問題点が存在する。誤判定フレームの直前のフレームを確認すると、選手の腕と胴体が画像上で重なっており、OpenPose による腕の推定位置が不正確であった。本研究では、直前フレームとの身体部位の座標差に基づき、OpenPose による左右反転検出を判定する。そのため、判定に用いる肘や手首の 2 次元座標値が実際の画像上での肘や手首の位置とずれている際に、腕の左右反転検出を誤って判定するケースが発生すると推察した。

4.4 ケンドールの順位相関係数による評価

4.4.1 ケンドールの順位相関係数

第 3.7 節に示した姿勢類似度の算出手法について、ケンドールの順位相関係数[13]を用いて定量的に評価する。映像 $A = \{a_1, a_2, \dots, a_M\}$ と映像 $B = \{b_1, b_2, \dots, b_N\}$ を例に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、映像 A 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、映像 B 内で最も類似するフレーム $b_p, b_q (1 \leq p, q \leq N)$ をそれぞれ求める。本研究において、最も類似するフレーム b_p, b_q は、第 3.7 節に示した選手の姿勢類似度に基づき決定される。

続いて、求めた類似ペア (a_i, b_p) および (a_j, b_q) を用いてフレームの順序を比較する。なお、 (a_i, b_p) とは、映像 B に含まれるすべてのフレームのうち最も a_i と類似しているフレームが b_p であることを示す。フレーム a_i および a_j に最も類似するフレームとしてそれぞれ選択された b_p, b_q について、 $p < q$ 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。ここで、 $p < q$ とはフレーム 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ことを表す。また、順序一致数 U および順序不一致数 V の初期値は 0 である。

最後に、すべての類似ペアについて上記の処理を適用後、ケンドールの順位相関係数 τ を算出する。ケンドールの順位相関係数 τ は順序一致数 U および順序不一致数 V を用いて式(4.1)で定義される。

$$\tau = \frac{U - V}{\frac{M \times (M - 1)}{2}} \quad (4.1)$$

ケンドールの順位相関係数 τ は-1 から 1 の範囲の値をとる. 1 に近いほど, 類似フレーム系列 $B' = \{b'_1, b'_2, \dots, b'_M\}$ に含まれる任意の 2 フレームについて, 系列 B' 内での順序と映像 B における時間的な前後関係との相関が強いことを表す. ここで, 類似フレーム系列 B' に含まれるフレーム $b'_i (1 \leq i \leq M)$ は, 映像 B に含まれるすべてのフレームのうち最も a_i と類似しているフレームである.

4.4.2 投球動作が含まれる映像に対する評価実験および考察

従来手法および提案手法に対して, 表 4.3 に示した 4 本のモデル投球動画のうちの 1 動画, および投球データセット BP1 に含まれる任意の 1 動画の計 2 動画を入力し, ケンドールの順位相関係数を算出した. 各モデル投球動画について, 投球データセット BP1 に含まれる動画の本数分実験し, ケンドールの順位相関係数の平均値を求めた. なお, 4 本のモデル投球動画のうち動画 1 および動画 2 は投球データセット BP1 に含まれる 2 本の映像からそれぞれ動作区間のみを切り出した動画である. そのため, モデル投球動画 1 およびモデル投球動画 2 については, 切り出し元動画を除く 121 本の動画を用いて実験した.

野球の投球動作が含まれる映像に対して従来手法および提案手法を適用した際の, ケンドールの順位相関係数の平均値を表 4.6 にまとめる. なお, 従来手法の一つである TCC[13] は, 入力する 2 本の映像には共に動作区間のみが含まれていることを想定した手法である. そこで, 投球データセット BP1 に含まれる, 非動作区間が存在する各映像から目視により投球動作区間のみを切り出した投球データセット BP2 についても同様の実験を行った. また, もう一つの従来手法[12]は右投げと右投げ, もしくは左投げと左投げの 2 選手の投球動作を同期させることが可能な手法である. 右投げと左投げの 2 選手の動作を対応付けることは研究目的として想定されていないため, 左投げの選手による投球動作が含まれるモデル投球動画 3 およびモデル投球動画 4 については実験を行っていない.

表 4.6 投球動作映像に対し各手法を適用した際のケンドールの順位相関係数

手法	データセット	ケンドールの順位相関係数			
		モデル 1	モデル 2	モデル 3	モデル 4
TCC[13]	BP1	0.620	0.567	0.605	0.419
	BP2	0.797	0.714	0.754	0.764
横井[12]	BP1	0.738	0.711	—	—
提案手法	BP1	0.855	0.831	0.838	0.816

表 4.6 より, 投球データセット BP1 を用いた際のケンドールの順位相関係数を比較する

と、各列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。これは、モデル投球動画と BP1 に含まれる 1 動画（以下、対象投球動画）の計 2 動画に対して提案手法を適用した際に、モデル投球動画に含まれる任意の 2 フレーム a_i, a_j の時間的な前後関係と、第 3.7 節で述べた選手の姿勢類似度に基づく類似フレーム b_p, b_q の時間的な前後関係が一致している割合が従来手法を適用した場合よりも高いことを示す。すなわち、提案手法で用いた姿勢類似度は動作中の選手の細かな動きを従来手法よりも高精度で区別することが可能であるといえる。なお、類似フレーム b_p, b_q とは対象投球動画に含まれるすべてのフレームのうち、選手の姿勢類似度に基づきそれぞれ a_i および a_j と最も類似していると判定されたフレームを指す。

また、左投げの選手による投球動作が含まれるモデル投球動画 3 およびモデル投球動画 4 における提案手法での結果を、右投げの選手による投球動作が含まれるモデル投球動画 1 やモデル投球動画 2 を用いた際の結果と比較すると、ケンドールの順位相関係数の値に大きな差はないことがわかる。投球データセット BP1 内の動画に含まれる投球動作はすべて、右投手によるものであるため、右投げの選手と左投げの選手の投球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。

4.4.3 スイング動作が含まれる映像に対する評価実験および考察

従来手法および提案手法に対して、表 4.4 に示した 4 本のモデルスイング動画のうちの 1 動画、およびスイングデータセット GS1 に含まれる任意の 1 動画の計 2 動画を入力し、ケンドールの順位相関係数を算出した。各モデルスイング動画について、スイングデータセット GS1 に含まれる動画の本数分実験し、ケンドールの順位相関係数の平均値を求めた。なお、4 本のモデルスイング動画のうち動画 1 および動画 2 はスイングデータセット GS1 に含まれる 2 本の映像からそれぞれ動作区間のみを切り出した動画である。そのため、モデルスイング動画 1 およびモデルスイング動画 2 については、切り出し元動画を除く 82 本の動画を用いて実験した。

ゴルフのスイング動作が含まれる映像に対して従来手法および提案手法を適用した際の、ケンドールの順位相関係数の平均値を表 4.7 にまとめる。なお、従来手法の一つである TCC[13]は、入力する 2 本の映像には共に動作区間のみが含まれていることを想定した手法である。そこで、スイングデータセット GS1 に含まれる、非動作区間が存在する各映像から目視によりスイング動作区間のみを切り出したスイングデータセット GS2 についても同様の実験を行った。また、もう一つの従来手法[12]は右打ちと右打ち、もしくは左打ちと左打ちの 2 選手のスイング動作を同期させることが可能な手法である。右打ちと左打ちの 2 選手の動作を対応付けることは研究目的として想定されていないため、左打ちの選手によるスイング動作が含まれるモデルスイング動画 3 およびモデルスイング動画 4 については実験を行っていない。

表 4.7 スイング動作映像に対し各手法を適用した際のケンドールの順位相関係数

手法	データセット	ケンドールの順位相関係数			
		モデル 1	モデル 2	モデル 3	モデル 4
TCC[13]	GS1	0.576	0.473	0.465	0.582
	GS2	0.596	0.568	0.513	0.649
横井[12]	GS1	0.687	0.705	—	—
提案手法	GS1	0.882	0.897	0.911	0.917

表 4.7 より、スイングデータセット GS1 を用いた際のケンドールの順位相関係数を比較すると、各列において提案手法の値が最も大きく、1 に近いことがわかる。特に、左打ちの選手によるスイング動作が含まれるモデルスイング動画 3 およびモデルスイング動画 4 を用いた際のケンドールの順位相関係数は 0.9 を超えており、提案手法の有効性が確認できる。

表 4.2 に示したように、スイングデータセット GS1 に含まれる動画の平均フレーム数や非動作区間の平均フレーム数は投球データセット BP1 の約 2 倍である。表 4.6 および表 4.7 を比較すると、従来手法では全体的に投球映像よりもスイング映像に対するケンドールの順位相関係数が小さい値であることがわかる。一方、提案手法ではスイング映像に対するケンドールの順位相関係数の方が大きく、1 に近い。したがって、非動作区間のフレーム数が多い映像に対しても提案手法は有効であるといえる。

4.5 対応付け結果

表 4.3 に示した各モデル投球動画と投球データセット BP1 に含まれる任意の 1 動画（以下、対象投球動画）に対して本手法を適用し、映像内のフレームを対応付けた。この際、始点終点自由な DTW を用いて、対象投球動画に含まれる投球動作の開始フレームおよび終了フレームを選手の姿勢類似度に基づき決定した。始点終点自由な DTW により求めた各対象投球動画の動作開始フレームおよび動作終了フレームと、各対象投球動画から目視により決定した投球動作の開始フレームおよび終了フレームとの誤差の平均値を表 4.8 にまとめる。以下、始点終点自由な DTW により求めた各対象投球動画の動作開始フレームおよび動作終了フレームをそれぞれ、開始判定フレームおよび終了判定フレームとよび、各対象投球動画から目視により決定した動作の開始フレームおよび終了フレームをそれぞれ開始目視フレームおよび終了目視フレームとよぶこととする。なお、比較として、第 2.6 節で述べた従来手法[12]でも同様の実験を行った。ただし、従来手法[12]は右投げと左投げの 2 選手の動作を対応付けることは想定されていないため、左投げの選手による投球動作が含まれるモデル投球動画 3 およびモデル投球動画 4 については実験を行っていない。

表 4.8 投球動作の開始フレームおよび終了フレームの誤差

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[12]	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

また、表 4.4 に示した各モデルスイング動画とスイングデータセット GS1 に含まれる任意の 1 動画に対しても同様の実験を行った。スイング動作の開始判定フレームおよび終了判定フレームと開始目視フレームおよび終了目視フレームの誤差の平均値を表 4.9 にまとめる。ただし、従来手法[12]は右打ちと左打ちの 2 選手の動作を対応付けることは想定されていないため、左打ちの選手によるスイング動作が含まれるモデルスイング動画 3 およびモデルスイング動画 4 については実験を行っていない。

表 4.9 スイング動作の開始フレームおよび終了フレームの誤差

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[12]	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

表 4.8 および表 4.9 より、各動作におけるモデル動画のうちモデル 1 またはモデル 2 を用いた場合、モデル投球動画 1 を用いた際の開始フレーム誤差を除き、提案手法が従来手法よりも誤差が小さいことが確認できる。モデル投球動画 1 を用いた場合の開始フレーム誤差についても、差はほとんどなく、全体として見ると提案手法の方が優れているといえる。また、提案手法において左投げもしくは左打ちの選手による動作が含まれるモデル 3 またはモデル 4 で実験した場合の誤差を確認すると、モデル 1 やモデル 2 における誤差より小さくなっているケースも存在する。したがって、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場合でも、提案手法は有効であると考えられる。

表 4.8 において、モデル投球動画 1 を用いた場合の終了フレームの誤差において、従来手法よりも提案手法の値が非常に小さくなっていることが確認できる。従来手法では、BP1 に含まれる複数の対象投球動画において、終了目視フレームよりも 30 フレーム以上後のフレームが終了判定フレームとなっていたことが確認された。これは、モデル投球動画 1 の最後のフレーム、すなわち投球動作の終了時における選手の姿勢が直立姿勢に近いことが影

響していると推察した。ここで、直立姿勢とは足を肩幅程度に開いた状態で真っ直ぐ立っている姿勢を指す。モデル投球動画 1 に含まれる投球動作の終了時における選手の姿勢と、終了フレームの誤差が大きかった対象投球動画に含まれる投球動作の終了目視フレームおよび終了判定フレームにおける選手の姿勢を図 4.1 に示す。

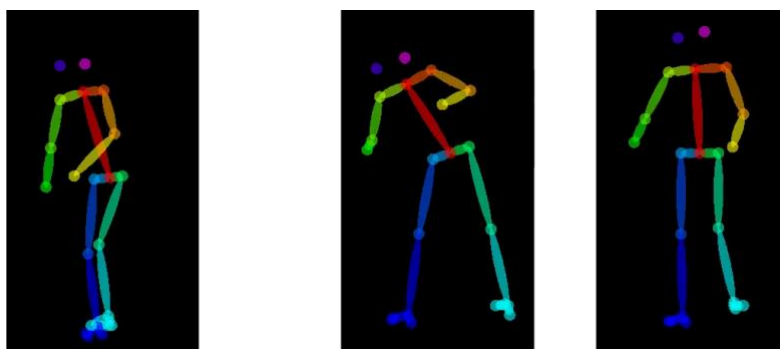


図 4.1 投球動作終了時の姿勢比較

図 4.1 では選手の姿勢を OpenPose により推定された骨格で表している。図 4.1 において、左の画像に映る骨格がモデル投球動画 1 に含まれる投球動作の終了時における選手の姿勢である。右の 2 枚の画像に映る骨格は、ある対象投球動画における選手の姿勢を表しており、左から終了目視フレームにおける姿勢、および終了判定フレームにおける姿勢である。なお、図 4.1 に示した例において、終了判定フレームは終了目視フレームの約 1 秒後である。

図 4.1 より、左の画像における選手の姿勢と右の画像における選手の姿勢は共に直立姿勢に近く、2 枚の画像に映る骨格は類似していることがわかる。一方、中央の終了目視フレームにおいては、左右の足がやや開いている。従来手法はフレーム間の選手の姿勢が似ているフレーム同士を対応付けるため、図 4.1 に示したケースなどでは終了判定フレームと終了目視フレームに大きな差が生じる可能性がある。

一方、提案手法ではフレーム間の姿勢類似度を算出する際に、第 3.6.2 項で述べた空間ベクトルと第 3.6.3 項で述べた時間ベクトルを用いている。すなわち、各身体部位の時間的な動きを考慮して、選手の姿勢類似度を求めている。そのため、似たような姿勢であっても、動作中の姿勢なのか静止中の姿勢なのかを区別することが可能となる。その結果、表 4.8 に示したように、従来手法よりも終了フレームの誤差が小さくなったと考えられる。

モデル投球動画 1 と投球データセット BP1 に含まれる 1 動画（以下、対象投球動画）に対して提案手法を実行し、投球動作のタイミングを対応付けた結果を図 4.2 に示す。

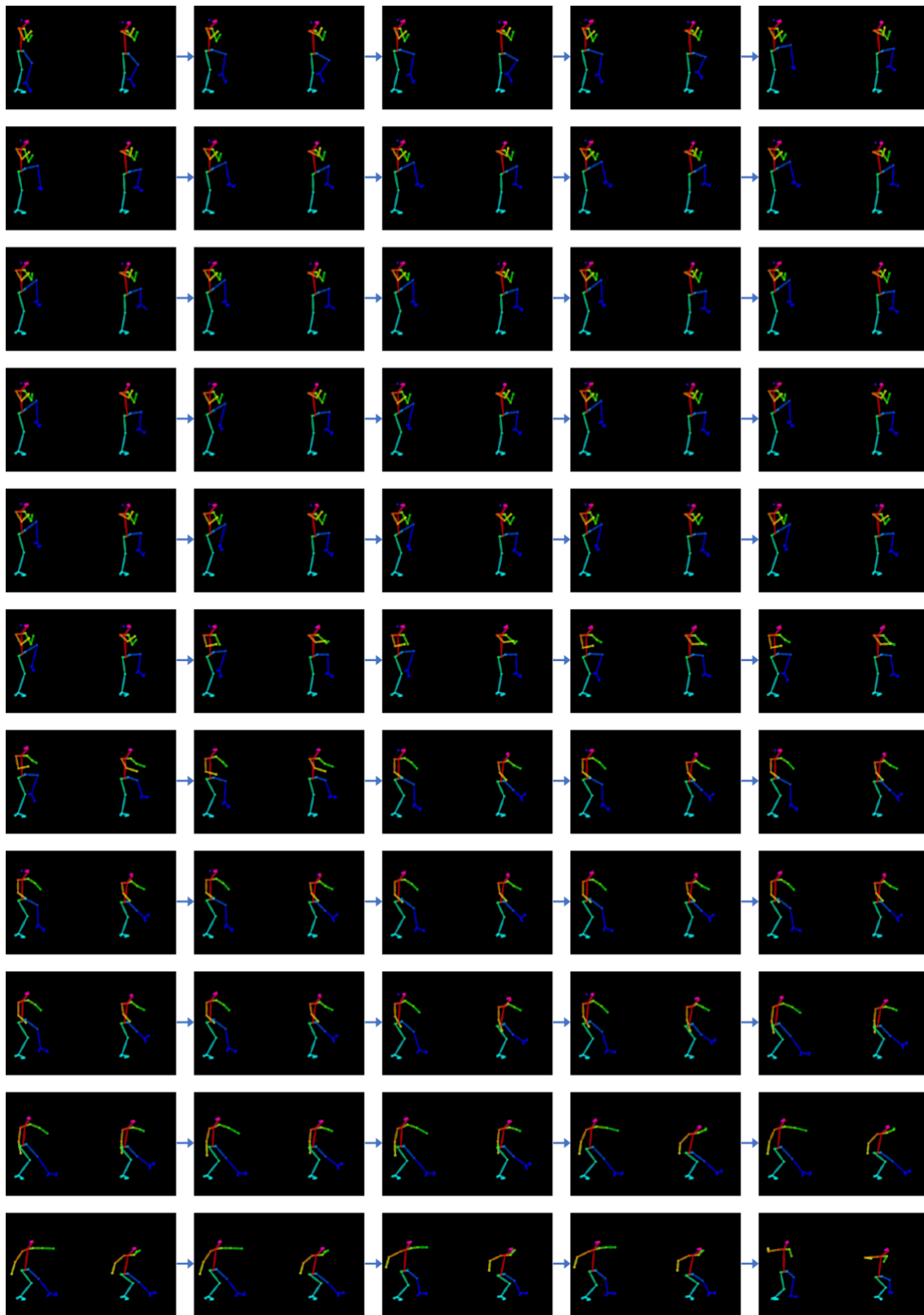


図 4.2 野球の投球動作の対応付け結果 (その1)



図 4.2 野球の投球動作の対応付け結果（その2）

図 4.2 に含まれる各画像は、提案手法により対応付けされたフレームを左右に並べた結果である。ただし、選手の姿勢を比較しやすくするために、OpenPose により推定された選手の骨格のみを表示している。各画像において、左側がモデル投球動画 1 にて投球動作を行っている選手の骨格、右側が対象投球動画にて投球動作を行っている選手の骨格である。一番上の行の左の画像が最初の対応付け、一番下の行の右の画像が最後の対応付けを表す。

また、モデルスイング動画 1 とスイングデータセット GS1 に含まれる 1 動画（以下、対象スイング動画）に対して提案手法を実行し、ゴルフスイング動作のタイミングを対応付けた結果を図 4.3 に示す。



図 4.3 ゴルフのスイング動作の対応付け結果 (その1)

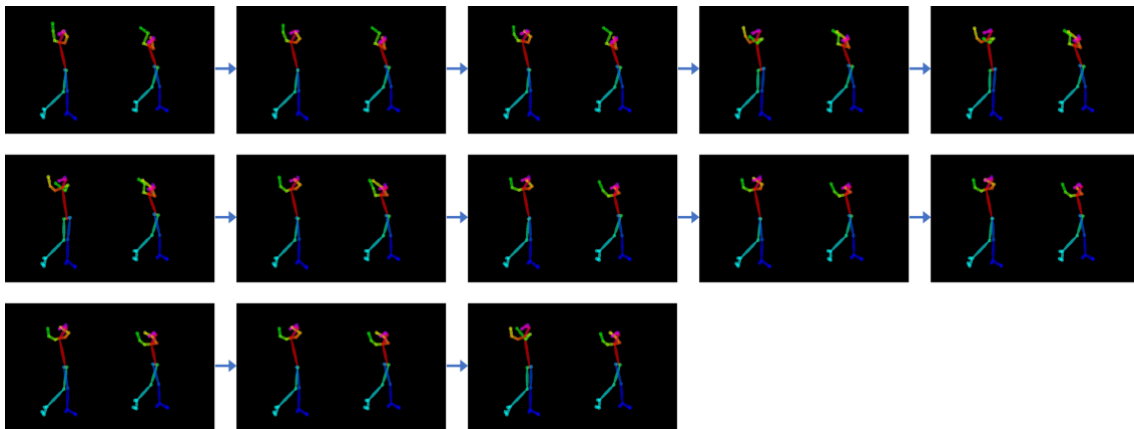


図 4.3 ゴルフのスイング動作の対応付け結果（その2）

図 4.3 に含まれる各画像は、提案手法により対応付けされたフレームを左右に並べた結果である。ただし、選手の姿勢を比較しやすくするために、OpenPose により推定された選手の骨格のみを表示している。各画像において、左側がモデルスイング動画 1 にてゴルフのスイング動作を行っている選手の骨格、右側が対象スイング動画にてスイング動作を行っている選手の骨格である。一番上の行の左の画像が最初の対応付け、一番下の行の右の画像が最後の対応付けを表す。

図 4.2 および図 4.3 より、選手の姿勢が類似するフレームが正確に対応付けされていることが確認できる。

4.6 むすび

本章では、第 3 章で示した提案手法に対する評価実験を行った。また、ケンドールの順位相関係数などを用いて、従来手法と提案手法を比較し、提案手法の有効性を確認した。

第5章 結論と今後の課題

5.1 結論

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した。具体的な手法の手順として、まず映像から選手の身体部位に関する2次元座標値を取得する。その後、得られた2次元座標値から各身体部位の時間変化を考慮した姿勢データをフレームごとに算出する。最後に、姿勢データから求めたフレーム間の類似度を基に、2本の映像のフレームを対応付ける。

提案手法では、空間ベクトルと時間ベクトルを用いて算出される選手の姿勢類似度に基づき、フレーム間の類似度を算出した。ケンドールの順位相関係数を用いた評価実験により、フレーム間の類似度算出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、従来手法よりも有効であることを確認した。

また、始点終点自由なDTWによる、一方の映像に含まれる動作区間の抽出精度を実験により確認した。抽出された動作区間の開始フレームおよび終了フレームについて、目視により決定した動作の開始フレームおよび終了フレームとの誤差を比較すると、従来手法よりも抽出精度が高いことがわかった。

5.2 今後の課題

提案手法には、OpenPoseによる腕もしくは足の検出が左右逆となっているフレームを特定し座標値の修正を図る処理が含まれている。しかし、評価実験により腕の左右判定検出の判定が誤っているケースが存在することを確認した。そのため、腕の左右判定検出を高精度で判定可能な手法を考案する必要がある。

本手法は投球動作やゴルフスイング動作等のスポーツ動作が1回のみ含まれている映像を対象としている。そのため、映像に複数回の動作が含まれている場合は、1回分の動作しか抽出できない。そこで、映像に複数回の動作が含まれている場合でもすべての動作を抽出できるように、処理を改良することが求められる。映像に含まれる動作の回数が多いほど、映像の長さは長くなると考えられる。また、本手法に含まれる姿勢類似度の算出処理や始点終点自由なDTWによる処理は映像のフレーム数に比例して計算量が増加する。そのため、手法の処理速度向上も今後の課題として挙げられる。

謝辞

本研究の実験環境を整えてくださり、研究の方向性や問題点などに関して丁寧かつ熱心にご指導くださった渡辺裕教授に深く感謝申し上げます。

また、本研究や論文執筆において適切なアドバイスを数多くくださり、丁寧かつ熱心にご指導くださった早稲田大学国際情報通信センターの石川孝明様に深く感謝申し上げます。

さらに、日頃から貴重なご意見をくださった研究室の皆様に心より御礼申し上げます。

最後に、ここまで私を育ててくださり、学業に専念できるようあらゆる面で支えてくださった家族に深く感謝いたします。

参考文献

- [1] TRACKMAN, “TrackMan Baseball,” (最終閲覧日：2021年11月30日), <https://trackmanbaseball.com/>
- [2] 永見智行, “三次元計測に基づくスポーツ研究・応用事例：野球の場合,” 計測と制御, Vol. 59, No. 10, pp. 733-788, Oct. 2020.
- [3] 村上早苗, “福岡ソフトバンクホークス4年連続日本一達成にひと役買った最先端システムとは?,” GOETHE, 2021年4月12日公開 (最終閲覧日：2021年11月30日), <https://gotheweb.jp/person/article/20210412-laiblitz>
- [4] 三宅隆, 滝田勝紀, “〈野球×デジタル〉『福岡ソフトバンクホークス』が iPhone&iPad で試合に勝つ方法【デジタルでスポーツの勝利をつかむ #2】,” MELOS, 2017年8月4日公開 (最終閲覧日：2021年11月30日), <https://melos.media/hobby/5934/>
- [5] ソニー株式会社, “ホークアイ (Hawk-Eye) 可視化のテクノロジーでスポーツの感動を支える,” 2021年1月20日公開 (最終閲覧日：2021年11月30日), <https://www.sony.com/ja/SonyInfo/technology/stories/Hawk-Eye/>
- [6] 遠藤修哉, “ヤクルトが導入した新兵器“ホークアイ”はチーム浮上の力となるか? 文春野球コラム ペナントレース 2021,” 文春オンライン, 2021年4月25日公開 (最終閲覧日：2021年11月30日), <https://bunshun.jp/articles/-/44432>
- [7] 東京ヤクルトスワローズ, “ホークアイデータ項目解説,” (最終閲覧日：2021年11月30日), <https://www.yakult-swallows.co.jp/pages/info/players/hawkeye>
- [8] 湯本勝大, “最下位からVヤクルト高津再生工場の秘密 最先端“鷹の目”で全動作数
字化,” 日刊スポーツ, 2021年10月27日公開 (最終閲覧日：2021年11月30日), <https://www.nikkansports.com/baseball/news/202110220000342.html>
- [9] 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 (CVPR), pp. 1302-1310, July 2017.
- [10] W. Zhang, M. Zhu, and K. G. Derpanis, “From Actemes to Action: A Strongly-supervised Representation for Detailed Action Understanding,” International Conference on Computer Vision (ICCV), pp. 2248-2255, Dec. 2013.
- [11] P. F. Felzenszwalb and R. Zabih, “Dynamic programming and graph algorithms in computer vision,” IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 33, No. 4, pp. 721-740, Apr. 2011.
- [12] 横井真也, “スポーツ映像から取得した身体座標データに対するアラインメント,” 早稲田大学大学院基幹理工学研究科情報理工・情報通信専攻修士論文, Feb. 2019.

- [13] D. Dwibedi, Y. Aytar, J. Tompson, P. Sermanet and A. Zisserman, “Temporal Cycle-Consistency Learning,” IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1801-1810, June 2019.

図一覧

図 2.1	入力画像[10].....	3
図 2.2	出力画像.....	4
図 2.3	出力結果（特徴点と骨格のみ）.....	4
図 2.4	要素数の異なる 2 系列に対する DTW の適用結果.....	6
図 2.5	DTW による不正確な対応付けの例.....	7
図 2.6	始点終点自由な DTW による対応付け結果.....	9
図 3.1	提案手法の処理フロー.....	12
図 3.2	投球動作を含む映像に対する OpenPose の推定結果.....	13
図 3.3	スイング動作を含む映像に対する OpenPose の推定結果.....	14
図 3.4	OpenPose による腕の左右反転検出例.....	15
図 3.5	空間ベクトルの長さおよび角度.....	22
図 3.6	時間ベクトルの長さおよび角度の例.....	24
図 4.1	投球動作終了時の姿勢比較.....	39
図 4.2	野球の投球動作の対応付け結果.....	40
図 4.3	ゴルフのスイング動作の対応付け結果.....	42

表一覧

表 3.1	各 l の値と身体部位の対応関係	16
表 3.2	対称部位の組み合わせ	17
表 3.3	各身体部位と kp の値の対応関係	20
表 3.4	構成部位の身体部位番号と添字 v の値の対応関係	21
表 4.1	実験にて使用した映像の詳細	31
表 4.2	投球データセット BP1 およびスイングデータセット GS1 の詳細	32
表 4.3	各モデル投球動画の詳細	32
表 4.4	各モデルスイング動画の詳細	33
表 4.5	OpenPose による左右反転検出の判定精度	33
表 4.6	投球動作映像に対し各手法を適用した際のケンドールの順位相関係数	35
表 4.7	スイング動作映像に対し各手法を適用した際のケンドールの順位相関係数	37
表 4.8	投球動作の開始フレームおよび終了フレームの誤差	38
表 4.9	スイング動作の開始フレームおよび終了フレームの誤差	38

研究業績

- [1] 大澤遼平, 兼近紘之, 武井淳也, 山川敦也, 渡辺裕, “OpenPose を用いたピッチングフォームの解析,” 映像情報メディア学会スポーツ情報処理時限研究会, Feb. 2019.
- [2] 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, Aug. 2019.
- [3] 稲田健太郎, 大澤遼平, 山川敦也, “OpenPose,” 映像情報メディア学会誌「私の研究開発ツール 一第 107 回一」, Vol. 73, No. 5, pp.932-934, Sep. 2019.
- [4] 大澤遼平, 石川孝明, 渡辺裕, “異なるピッチングスタイルに対するシーン照合の一検討,” 映像情報メディア学会スポーツ情報処理時限研究会・メディア工学研究会 ME2019-117, Vol.43, No.39, pp.13-14, Nov. 2019.
- [5] 大澤遼平, 石川孝明, 渡辺裕, “OpenPose と Dynamic Time Warping を用いた投球シーン照合の一検討,” 映像情報メディア学会冬季大会, 11A-3, Dec. 2019.
- [6] 大澤遼平, 石川孝明, 渡辺裕, “OpenPose を用いた投球シーン照合の一検討,” 情報処理学会第 82 回全国大会, 1R-06, Mar. 2020.
- [7] R. Osawa, T. Ishikawa, and H. Watanabe, “Pitching Motion Matching Based on Pose Similarity Using Dynamic Time Warping,” 2020 IEEE 9th Global Conference on Consumer Electronics (GCCE), pp.104-108, Oct. 2020.