

2018 年度 修士論文

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

Alignment Method for Body Parts Coordinates
Obtained from Sport Video

指導教員 渡辺裕

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

5117F095-6

横井真也

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
2.3	Dynamic Time Warping	4
2.3.1	Dynamic Time Warping によるイベント検出	5
2.3.2	Dynamic Time Warping によるアラインメント	6
2.3.3	Dynamic Time Warping の欠点	6
2.3.4	Prefix and Suffix Invariant Dynamic Time Warping	8
2.4	最長共通部分列	9
2.5	むすび	9
第 3 章	提案手法	10
3.1	まえがき	10
3.2	提案手法概要	10
3.3	スポーツ映像データに対する Dynamic Time Warping	12
3.3.1	時系列多次元データ	12
3.3.2	時系列多次元データに対する Dynamic Time Warping	12
3.4	始点終点検出	13
3.4.1	始点と終点自由な DTW の実現	14
3.5	個人依存動作検出	16
3.6	身体座標情報を利用した DTW によるイベント認識	17
3.7	欠損フレーム予測	17
3.7.1	欠損フレームの推定	19
3.7.2	欠損フレーム構築	20

3.8	アラインメント結果を利用した誤検出部位の修正	22
3.8.1	検出失敗部位の特徴	22
3.8.2	欠損部位座標予測	23
3.9	むすび	24
第 4 章	評価実験	25
4.1	まえがき	25
4.2	個人依存動作除去の評価	25
4.3	始点終点自由な Dynamic Time Warping の評価	26
4.3.1	Prefix and Suffix Invariant Dynamic Time Warping の評価	26
4.3.2	アルゴリズム 3 の評価実験	36
4.3.3	アルゴリズム 4 の評価実験	41
4.4	DTW によるイベント検出	50
4.5	欠損フレーム予測の実験	51
4.6	部位予測についての評価実験	54
4.7	むすび	56
第 5 章	結論と今後の課題	57
5.1	結論	57
5.2	今後の課題	57
謝辞		58
参考文献		59
研究業績		60

第 1 章

序論

1.1 研究背景

スポーツ映像を解析し、競技者や視聴者、競技の進行者に対して役立つコンテンツを提供する試みが行われている。競技者に対して提供されるコンテンツには、競技者自身や相手選手、チームの映像を解析し、自身の能力を高める、もしくは相手選手やチームの弱点を発見するなどの目的が存在する。例としては映像比較を指導に利用するための研究 [1] などが存在し、また映像の比較は、映像が入手できれば誰でも容易に行える基本的な映像活用の手法でもある。一方視聴者に対しては、競技者や競技に関わる物体の動きを、時系列を明確に視覚化し、映像として提示することで、競技に対する深い理解がなくとも楽しむことができるコンテンツが提供されている。例としては 2.5 次元マルチモーション映像 [2] などの、時系列順に映像を提示する手法が存在する。進行者に対しては、競技の進行に関わる役割を、映像解析を行うことで自動化や高精度化を行い、進行の手助けを行うシステムが提供されている。例としてはテニスの The Hawk-Eye Officiating System[3] や、野球におけるリクエストシステムなど審判の判定を補助し、高精度化する試みが存在する。

野球はテレビやネット配信など様々なデバイスによって映像配信が行われる競技であり、映像データが多量に存在するスポーツである。近年では様々な形で映像が活用されている。例としては映像解析により試合の各種データを数値化し収集する STATCAST、Major League Baseball におけるチャレンジ制度や、日本プロ野球におけるリクエスト制度などの判定補助システム、iCube[4] など収集した映像を選手のトレーニングに利用するシステムが存在する。これらのシステムは収集した映像を解析し、数値や他の形の映像に変換することで実現されている。また職業野球以外に、高校野球やリトルリーグなど競技として野球を行う場が幅広く存在するため、野球に対する映像活用は今後さらに行われていくことが予想される。

野球に限らずスポーツ競技における映像の活用として、最も基本的な手法は、映像の比較である。複数の同じ動作を行う人物が映る映像を比べることにより、動作を比較する。自身の動作と優秀な選手の動作を比較することで、自身の動作に存在する問題点を発見することができ、自身の映像同士を比較することで、動作が行われた時の状況が、自身の動作にどのような影響を与えるかを検討できる。また映像を比較するだけならば、解析用の器具や、複数の高性能カメラなど、高価な機材を導入する

ことなく、スマートフォンなどの比較的安価で容易に調達可能な撮影機材や、テレビ映像のような配信されている映像を活用することで、実行可能である。

1.2 研究目的

動作の比較において求められることは、時系列に沿って行われる動作のうち、同じ部分を比較することである。つまり系列 A と系列 B の動作を比較する際に、映像として比較すべきものは、比較元の系列 A の特定のタイミングと、同じ動作である系列 B に存在する同じ動作のタイミングである。2 動作のすべてのタイミングを対応付けることで、2 系列の動作を比較可能である。

本研究では動的時間伸縮法 (Dynamic Time Warping, DTW) によって動作のアラインメントを行う。アラインメントを行うデータの対象として、野球における投手の投球映像から、OpenPose によって取得した身体部位座標データを利用する。個人差が存在する投手の投球時の身体座標データに対し、正確にアラインメントを行う手法を提案する。

1.3 本論文の構成

本論文は 5 章で構成されている。

第 1 章「序論」は本章であり、本研究の背景と目的を述べている。

第 2 章「関連研究」では、本研究で用いる身体部位座標の取得手法である OpenPose、基本となるアラインメント手法である Dynamic Time Warping を説明する。

第 3 章「提案手法」では、投球映像に対するアラインメント手法を示す。

第 4 章「評価実験」では、第 3 章で提案した手法に対する評価実験について述べる。

第 5 章では、まとめを述べる。

第 2 章

関連研究

2.1 まえがき

本章では本研究で利用する関連研究や従来手法について述べる。また従来手法のアルゴリズムと、その特徴を示す。

2.2 OpenPose

身体の映像上の部位座標を取得する手法として、Cao らが提案した Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields (OpenPose) [5] が存在する。この手法により映像中に存在する複数の人物の身体部位座標を取得可能である。Convolutional Neural Network (CNN) を学習する際に使用するモデルによって、身体の主要部位のうち 15, 18 もしくは 25 部位の位置を推定し画像上の座標を取得できる。加えて映像に十分な大きさと手や顔が映されていれば、手の関節や顔のパーツの輪郭なども検出可能である。検出結果は画像上での各身体座標の x 座標, y 座標, そして検出の信頼度が出力される。25 部位を検出可能なモデルでは、足の構造をより詳しく検出可能なものの、本研究では必要としないため、本研究では 18 部位が検出可能なモデルを使用した。以下に本研究で使用したモデルで検出を行った例を示す。



図 2.1 OpenPose で検出された身体座標

検出可能な部位は、両目、両耳、鼻、首、両肩、両肘、両手首、両腰、両膝、両足首の 18 部位である。

OpenPose では、身体部位の存在確率を示す Confidence Maps と部位間の接続を表す Part Affinity Fields を学習することで、身体部位座標の検出を実現している。通常の人物検出では、映像内の人体の存在領域を検出した後に、部位などの細かい検出を行う。しかし OpenPose では存在領域検出を行うことなく、直接身体座標を検出している。直接身体座標を検出可能なため、人体の一部が映像上に存在すれば、その映像に映る部位の種類、そして画像上での座標値を推定できる。また人体が他の物体に重なり、映像上に映らないようなオクルージョンが発生している場合でも、精度は低下するもののその位置を推定することが可能である。しかし OpenPose は時系列のデータを身体部位座標推定に用いず、各フレームごとに検出が独立している。そのため時系列で位置が推定可能なオクルージョンの場合、人の目で見て推定が可能な場合でも、正しい座標推定が行えない場合がある。

2.3 Dynamic Time Warping

Dynamic Time Warping (以下 DTW) とは、2 系列間の合計距離 (DTW 距離) を最小化するような対応付け (アラインメント) を探索する手法である。DTW 距離とは、系列がどれだけ類似しているかを示し、DTW 距離が小さいほど系列が類似していることを表す。特に時系列データに対して用いられる場合が多く、時間方向の伸縮によって最適なアラインメントを探索する。時間方向の伸縮とは、アラインメントにおいて複数の時間の要素が、特定の一つの時間に対応付けされることを表す。

時系列データ A, B に対して DTW を行う場合を考える。時系列データ A の時間 p の要素 A_p について、 B_q から B_{q+r} のデータが対応付けられたとすれば、それは A_p が r の大きさだけ時間の上で伸ばされたことに等しい。DTW はこのような要素の伸長を行うことで、身体動作など時間方向に

個人や状況による差がつく事象についても、正しくアラインメント可能である。

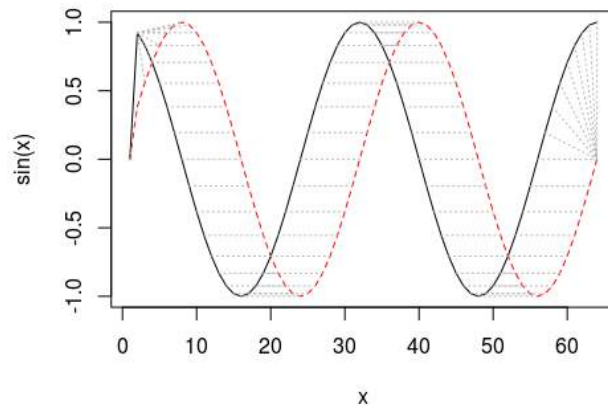


図 2.2 位相差の存在する 2 系列に対する DTW 図

図 2.2 は同じ変化をするが、位相に差が存在する 2 系列に対して DTW を適用した例である。同じ時間で 2 系列の要素を対応させるのではなく、それぞれの系列中における位置として同じ、もしくは近いものが対応付けられていることがわかる。

2.3.1 Dynamic Time Warping によるイベント検出

DTW は 2 系列間の DTW 距離を求めることができる。この DTW 距離は、2 つの系列が類似しているほど小さくなる値となる。そのため、検出するイベントの系列と、入力系列の間の DTW 距離を求めることで、入力系列がイベントの系列にどれだけ類似しているかがわかる。したがって入力系列においてそのイベントが発生したか判定することが可能である。例として、DTW を利用した車両カウントシステム的设计 [6] があげられる。車両走行音を機器に受信するまでの時間差を示すサウンドマップに対して DTW を用いることで、車両の通過を検出する。

DTW 距離を求めることでイベントを検出するためには、検出対象の各イベントについて、モデル系列を用意し、そのすべてとの DTW 距離を求めて、最も DTW 距離が小さくなるモデル系列を調べるか、閾値を設定し、DTW 距離がその値以下となるモデル系列を求める必要がある。しかし DTW 距離を正規化することは難しいため、基本的にすべてのモデル系列との DTW 距離を求め、相対的にどのモデル系列との DTW 距離が最小となるか調べる必要がある。

2.3.2 Dynamic Time Warping によるアラインメント

アラインメントとは、2つの系列の要素を対応付けること、もしくは対応付けられた結果を示す。アラインメントは主に DNA の塩基の検索などに用いられ、高速化の研究 [7] などが行われている。また遺伝子工学の分野以外にも、対応付けを必要とするような研究は多数存在し、訳文を対応付ける手法 [8] などに用いられている。

2つの系列をそれぞれ M , I とし、その長さをそれぞれ p , q とする。この2系列間の DTW 距離とアラインメントを求める。まず $(p+1) \times (q+1)$ の範囲を持つすべての要素が無限大であるコスト行列 C について、 $C(0,0) = 0$ とする。また2系列の要素間の距離を出力する距離関数を D とし、一方の系列の要素 M_x と、もう一方の系列の要素 I_y の距離を $D(M_x, I_y)$ とする。そして下に記述するアルゴリズムによってコスト行列を更新する。

Algorithm 1 オリジナル DTW1

```
 $C \leftarrow \text{infinity\_matrix}(p+1, q+1)$ 
 $C(0,0) = 0$ 
for  $i = 1 \dots p$  do
  for  $j = 1 \dots q$  do
     $C(i,j) = D(M_i, I_j) + \min(C(i-1, j-1), C(i, j-1), C(i-1, j))$ 
  end for
end for
```

系列 M , I 間の DTW 距離は、上記のアルゴリズムでコスト行列が更新された後の、 $C(p, q)$ の値となる。

この時系列間のすべての対応付けから、DTW 距離を最小化するアラインメントとして適した組み合わせを探索することになる。つまり $C(p, q)$ を最小にする対応付けをすべて並べたものが、DTW によってもとまる系列 M , I 間のアラインメントとなる。この時 DTW の計算量はモデル系列の長さを入力系列の長さの積に比例する。

2.3.3 Dynamic Time Warping の欠点

DTW の欠点として、以下があげられる。

1. 計算量が多い
2. ノイズや余計な要素の混入に弱い
3. DTW 距離が累積である

まず計算量については、アルゴリズム 1 に示したコスト行列の更新を行う際、行列のすべての部分について系列の要素間の距離を求める必要がある。そのためモデル系列と入力系列の 2 系列の長さに比例した計算が必要になる。DTW によって動作のアラインメントを行う場合、モデル系列よりも長い入力系列が想定されるため、計算量はモデル系列の長さの 2 乗を超える。したがってモデル系列に含まれる動作などのイベントが長くなる、つまり系列に含まれる要素数が多くなると、計算量が急激に増加する。

基本的に DTW は一方の系列のすべての要素を、もう一方の系列の要素のどれかに対応付けなければならない。そのため動作の個人差などデータの大まかな形を変えない場合には正しくアラインメント可能である。しかし計測時点での誤差や、動作そのものに違いがある場合など、データ自体に余計な要素やノイズが発生する場合には、不正確なアラインメントが発生する確率が高い。データに対してノイズが加わると、アラインメントを行うもう一方の系列の要素との、距離関数によって算出される距離が変化することになる。これにより本来対応すべき要素との距離が増大することや、それ以外の要素との距離が減少することが考えられる。この距離の変化がアラインメント結果や最終的な DTW 距離に影響を与える。

余計な要素の混入もアラインメントや DTW 距離に影響を与える。余計な要素が混入することで、実際にはもう一方の系列に対応付けるべきではない要素が存在することになる。この要素が対応付けられることで、全体のアラインメント結果や DTW 距離に影響する。

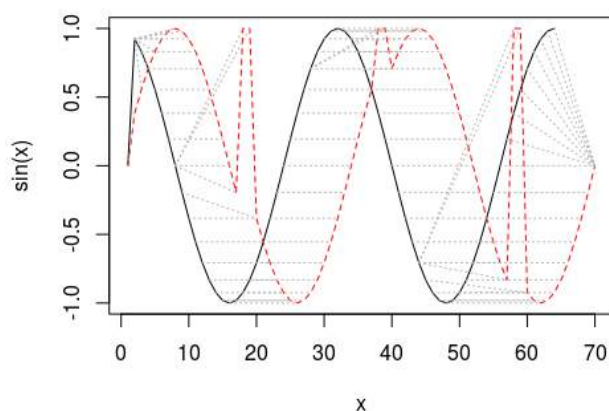


図 2.3 不要な要素が挿入された系列に対する DTW 図

図 2.3 は二つの \sin 関数に対して DTW によってアラインメントを行う際、一方に本来存在しない要素を挿入したアラインメント結果である。対応付けにおいて縦方向に変化が多い対応付け、つまり他の対応付けに比べて距離関数の値が大きくなる対応が存在している。DTW 距離の結果は累積であるため、この余計な要素に対して対応付けられた距離が、そのまま最終的な DTW 距離の出力に加算

される結果になる。この2系列間のDTW距離は、片方の要素に余計な要素が混入した分だけ大きくなるべきではあるが、一部の誤差が全体の結果に大きな影響を与えることになり、正しい結果とはいえない。

第三の欠点であるDTW距離が累積であることは、上述したように一部の結果が全体の結果に影響を与えることである。そして対応付けにおいて、対応付けの数を減らすことがDTW距離を小さくすることに繋がる可能性がある。DTWは二つの系列の1要素ごとにもう一方の系列に対応させ、その対応の距離の合計を値が大きくなるほど類似していないという指標となるDTW距離として出力する。この時DTWでの対応付けの数は、最小で2系列のうちの長い方の系列の長さとなる。この最小となる場合は、系列を「伸ばす」という変換が行われなかった時である。系列の要素を複製することで、系列を「伸ばす」という処理が実行される場合は、複数の要素をもう一方の一つの系列に対応させた方が、距離関数の合計値が小さくなる場合である。しかし「伸ばす」処理によって対応付けの数が増加し、DTW距離に加算される回数が増加することになる。このため対応付けとして「伸ばす」処理を行うことが正しいとしても、DTW距離の最小が条件であるDTWでは、この「伸ばす」処理が行われない可能性がある。

2.3.4 Prefix and Suffix Invariant Dynamic Time Warping

Prefix and Suffix Invariant Dynamic Time Warping (以降PSI-DTWとする) [9]は、始点終点自由なDTWである。始点終点とは、アラインメントのうち最初と最後の対応付けされた要素の、各系列内における時間を表す。

入力系列が有限である場合、その系列に含まれるアラインメント対象のイベントの始点と終点は、系列自身の始点と終点から0以上離れた時間に存在する。この系列の始点、終点と、イベントの始点、終点の時間差の最大値を T とする。この T の位置を始点と終点の候補としてDTWを行えば、始点より前、終点より後ろの時間のデータを含めないアラインメントが可能となる。

第2.3節で述べたとおり、DTWはコスト行列を更新することで、最適なアラインメントを探索する。系列の始点を S 、終点を E とし、 $E - S > R > T$ となる R を考える。コスト行列の更新の際、 S から $S + R$ まで、そして $E - R$ から E までを始点、終点の時間の候補として、最適なアラインメントを探索することが、始点、終点を検出することと同義となる。モデル系列の長さを p 、入力系列の長さを q 、コスト行列をサイズが $(p + 1) \times (q + 1)$ の、すべての要素が無限大で初期化してある行列として、以下にPSI-DTWのアルゴリズムを示す。

Algorithm 2 Prefix and Suffix Invariant Dynamic Time Warping

```
 $C \leftarrow \text{infinity\_matrix}(p + 1, q + 1)$   
 $C([0, R], 0) \leftarrow 0$   
 $C(0, [0, R]) \leftarrow 0$   
for  $i = 1 \dots p$  do  
  for  $j = 1 \dots q$  do  
     $C(i, j) = D(M_i, I_j) + \min(C(i - 1, j - 1), C(i, j - 1), C(i - 1, j))$   
  end for  
end for  
 $DTWdis = \min(\min(C([p - R, p], q)), \min(C(p, [q - R, q])))$ 
```

DTW 距離はアルゴリズム 2 の $DTWdis$ である。

2.4 最長共通部分列

最小共通部分列 (Longest Common Subsequence, LCS) は複数系列の最も長い共通部分列を表す。LCS の例を以下に示す。

$$\begin{aligned} X &= (a, b, c, d, e, a, b, c, d, e) \\ Y &= (a, c, e, b, d, a, c, e, b, d) \\ LCS(X, Y) &= (a, c, e, b, d, e) \end{aligned}$$

LCS は系列の順番を完全に保持し、二つの系列に共通して存在しない要素だけを除去するため、時系列データの比較の前処理として有効である。DTW のようなノイズや余計な要素の混入に弱い処理を行う場合、LCS を抽出し、LCS に対して DTW などの処理を行うことで、ノイズや余計な要素の混入が、処理に与える影響を低減可能である。

2.5 むすび

本研究に関連する研究や、従来手法について示した。また各手法のアルゴリズムや欠点を提示し、改善すべき点を示した。

第 3 章

提案手法

3.1 まえがき

本章では、提案手法の具体的内容を示す。そして従来手法の存在する手法については提案手法によって、従来手法の問題点をどのように解決するかを示す。

3.2 提案手法概要

スポーツ映像から抽出した身体座標データに対して、DTW によってアラインメントを行う。モデル系列を用意し、入力系列との DTW 距離を求めることで、そのモデル系列に設定したイベントが発生したか DTW 距離により判定できる。

スポーツ競技における特定の動作は、特定の型が存在し、似たような動きになるものの、個人や状況によって動作に変化が起こる可能性がある。この動作にそのまま DTW を行った場合、不正確な対応付けが行われ、類似度の算出やアラインメントの正確性などに影響する。共通の動作に対してアラインメントを行う際、影響を与えるのは以下の要素である。

1. 動作の前後に存在する、アラインメントを行う動作部分ではない要素である対象外動作
2. 個人や状況によって変化する、動作内の個人に依存する動作である個人依存動作

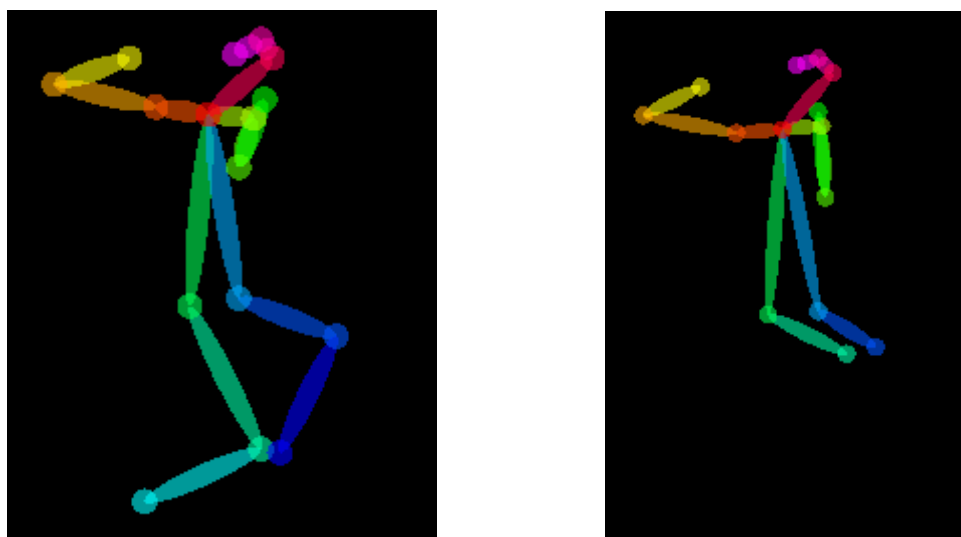
比較元の系列をモデル系列、比較対象の系列を入力系列とする。まず対象外動作について、入力系列を用意する際、正確に動作部分を切り出せないことで発生する部分である。本来は対応付けがされるべきではない要素が対応付けの候補となることで、全体のアラインメント結果に影響を与える。また始点と終点に連続する部分であるため、始点、終点付近と類似する要素となり、類似部分がもう一方の系列に存在する可能性が高くなる。その結果正しくないアラインメントが発生する。モデル系列は正確に用意する事が求められる上、基本的に 1 動作に対して一つであるため、動作部分だけを切り出すことにコストはかからない。したがってモデル系列には対象外動作は存在しないものとして考えることが可能である。そこで入力系列から対象外動作を除去することが必要となる。対象外動作を除

去するためには、対象となる動作の始点と終点を検出し、その間以外の部分を除去すれば良い。

個人依存動作は、片方の系列にのみ存在する動作である。そのため対応付け可能な要素がもう一方の系列に存在しない。DTW のようにすべての要素を対応付けする手法の場合、個人依存動作のような対応付けが存在しない要素に対応付けを試みることで、全体のアラインメント結果に影響する。また時系列データからノイズを取り除くような手法の多くは、オリジナルのデータに加算する形でノイズが含まれていることを想定している。そのような手法の場合、依然として対応付け不能な要素が系列に存在しているため、正しいアラインメントを行うことはできない。このような場合には対応付け不可能な要素を取り除くことが必要となる。そこでモデル系列に存在しない要素を入力系列から取り除く。

正しいアラインメントを実現するためには、除去だけではなく、動作データを再構築する必要がある。これは映像情報には、データの不連続性に起因する欠点が存在するためである。それは映像の質の問題である。動作を映像に残した場合、カメラの性能によって、欠損する部分が発生する。特にスポーツにおける動作は高速であることが多いため、撮影時のフレームレートによる動作の欠損が発生する可能性が高い。したがってアラインメントにおいてもその欠損が影響する可能性が存在する。前述の不要な部分が混在することによる対応付け不能要素の存在とは逆に、本来なら対応付け可能な要素が、一方の系列で欠損しているため対応付けを行えないという場合である。この欠損部分を再構築することで、より正確な動作映像に対するアラインメントが実現できる。

同様に本研究でのアラインメントに影響を与える要素として、身体部位座標の誤検出がある。OpenPose を用いて映像から身体座標を検出しているが、オクルージョンや、映像の質の問題などにより、部位が検出できない可能性や、間違った場所を検出する可能性がある。



[1] 検出成功フレーム

[2] 検出失敗フレーム

図 3.1 OpenPose 検出成功フレームと失敗フレーム

図 3.1 は OpenPose によってすべての部位の検出に成功したフレームと、失敗した部位が存在するフレームを示した結果である。特に動作が大きく高速な部位に対して誤検出は発生する可能性が高く、動作のアラインメントへの影響は大きい。そこで欠損部位を短い区間で等速に変化する、という条件だけではなく、DTW に利用する距離関数の値から予測する。モデル系列の身体座標データが正確なデータであるという前提の元、モデル系列の要素を入力系列の欠損部位の予測に用いることで、より正確に欠損部位を予測する。

3.3 スポーツ映像データに対する Dynamic Time Warping

スポーツ映像から取得した時系列多次元データに対して、DTW を適用する手法を示す。

3.3.1 時系列多次元データ

本研究ではスポーツ映像から OpenPose を利用して身体部位座標を取り出した後、座標データを変換し、2 部位間の長さや角度に置き換える。この変換後のデータを時系列多次元データとして DTW を行う。本研究では安定して検出可能な部位として、首、両肩、両肘、両腰、両膝、両足の 11 部位を使用する。この 11 部位の部位間の長さ、角度データに置き換えると、角度と長さの 2 次元と、11 部位すべての組み合わせ 55 次元を合わせたデータとなる。映像データを n フレーム、部位間の長さを r 、角度を θ とした時の時系列データ V を以下に示す。

$$\begin{aligned} V_t &= \{v_1, v_2, \dots, v_n\} \\ v_i &= \{P_1, P_2, \dots, P_{55}\} \\ P_j &= \{r_j, \theta_j\} \end{aligned}$$

長さに変換した場合、選手ごとの身長などの差や、映像ごとのカメラの距離などの差によって距離関数の値などに影響を与える。そこで正規化を行うことで、個人差やカメラの差の影響を低減する。まず選手が直立姿勢となっているフレームを特定する。そして首から右腰部分までの長さが 1 となるような係数を求める。すべてのフレームのデータに対し求めた係数をかけることで、正規化を行う。

3.3.2 時系列多次元データに対する Dynamic Time Warping

DTW を行うためには、各フレームの対応付けにおけるフレーム間の距離を決定する距離関数が必要となる。入力は各系列の 110 次元データ v_i であり、出力は 1 次元の値となる。動作のアラインメントのための距離関数であるため、動作の各タイミングにおける身体の位置関係に近いほど、距離が小さくなるような距離関数が必要となる。また映像上の動作に対するアラインメントであり、人間の

目で画像を見て似ている場合を、距離が小さくなるように出力する必要がある。そこで部位間のベクトルの角度の影響が大きいと考え、2ベクトル間の角度差に対して係数をかける形の距離関数とした。モデル系列の多次元データを v_i^M 、入力系列の多次元データを v_j^I 、モデル系列の部位間の角度、長さの組みを P_i^M 、入力系列の部位間の角度、長さの組みを P_j^I とし、本研究で使用する距離関数 D を以下に示す。

$$d(P_a^{M_i}, P_a^{I_j}) = (r_a^{M_i} - r_a^{I_j})^2 |\theta_a^{M_i} - \theta_a^{I_j}| \quad (3.1)$$

$$D(v_i^M, v_j^I) = \sum_{b=1}^B (d(P_a^{M_i}, P_a^{I_j})) \quad (3.2)$$

11 部位すべての組み合わせの数は 55 個のため、本研究では $B = 55$ とする。関数 d はある部位間の長さや角度の組 P 間の距離を示す。距離関数 D は関数 d の値の合計となる。

3.4 始点終点検出

始点終点とは、動作の始点、終点である。モデル系列の始点、終点は、動作の始点、終点と一致しているものとする。始点終点自由、つまり始点終点の決定を行うと同時に DTW による最適なアラインメント探索を行う手法として、第 2.3.4 項で述べた PSI-DTW が存在する。アラインメントの始点終点を、モデル系列と入力系列の始点と終点に限定せず、特定の範囲から最適な点を決定可能にすることで、始点と終点を事前に決めずに DTW を行うことを可能とした手法である。しかしこの手法には以下の問題点がある。

1. 始点と終点の探索範囲 R が必要である。
2. 始点と終点が自由であり、累積距離の最小値を探索するため、どちらかの系列の終点の範囲までアラインメントが求められた時点で、以降の対応付けが探索されない。

まず問題点 1 については、第 2.3.4 項で述べたように、それぞれの系列に含まれるイベントの始点と終点の時間によって、始点終点の探索範囲 R が必要となる。この R は入力系列と、それに含まれるイベントの時間的な存在位置の関係に依存して有効な値が決まる。しかし有効な値を正確に決定するためには、イベントの始点、終点を入力系列ごとに求める必要がある。それでは PSI-DTW を行う利点がないため、 R の値は、各系列に含まれるイベントの想定される長さや、各系列自体の長さから、有効な値を推測し、使用することになる。探索範囲として最適な値を T とすると、 $R > T$ となるように R を設定すれば、 R は探索範囲として有効だといえる。しかし、確実に $R > T$ となるように、系列の長さの値を R とすると、始点と終点の拘束条件が不十分であることが影響し、不正確なアラインメントが発生する可能性がある。

問題点 2 については、DTW が累積距離によって、最適なアラインメントを探索することに原因が存在する。終点が自由であるということは、アラインメント対象の各系列の要素のうち、どちらかの系列の終点の探索範囲に存在する要素が対応付けされた時点で、終了条件を満たすことになる。そし

て探索範囲 R は前述したように、可能な限り大きい値に設定することが、始点、終点を探索範囲に含めるために必要である。そのため終点の時間より前にも探索範囲が存在する。それにより終点到達する前にアラインメントが終了する場合がある。したがって探索範囲 R の値や、アラインメント対象のイベントの性質によっては、イベントの終点までアラインメントが行われない可能性が存在する。

上記の二つの事象について確認するために事前実験を行った結果、不正確なアラインメントや、終点を含まないアラインメントの発生が確認できた。実験の結果については第 4 章で述べる。

3.4.1 始点と終点自由な DTW の実現

PSI-DTW で問題であったのは、 R の決定方法、終了条件による終点検出の障害である。しかしモデルとする系列の始点、終点が、イベントの始点、終点と一致していた場合、入力系列における始点と終点自由な DTW は可能である。

まず探索範囲 R については、 R を入力系列の始点、終点の探索範囲とする。モデル系列の始点と終点は求まっている前提のため、入力系列のみ始点と終点を求める必要があるためである。そして R の値は、入力系列の長さの値とする。つまり始点、終点共に入力系列のどこに存在してもよい、ということである。この場合、イベントが入力系列のどこに含まれていたとしても、始点と終点を決定し、アラインメントが可能になる。

終点検出の問題については、累積距離を最も小さくするという手法が影響しているため、コスト行列の更新方法を変更する必要がある。対応間の距離の合計を最小化する場合、終了条件を満たした時点での DTW 距離だけが、DTW 距離最小値の候補となる。これは終了条件を満たした時点から、新たな対応をアラインメントに含めると、その対応の距離が加算されることで、必ず加算前より DTW 距離が大きくなるためである。そこで対応付けの探索の際、累積距離が短くなるような合計距離の更新方法を取り入れれば良いと考えた。対応間の距離が短い、つまり対応付けとして正しい可能性が高い組み合わせが、アラインメントに加わると、累積距離が小さくなるような累積距離の更新手法である。モデル系列の長さを p 、入力系列の長さを q 、距離関数を D 、モデル系列の時間 i 、入力系列の時間 j の時の距離関数の値を $D(i, j)$ とし、以下にコスト行列 C の更新手順を示す。

Algorithm 3 始点終点自由な DTW1

```
 $C \leftarrow \text{infinity\_matrix}(p+1, q+1)$ 
 $L \leftarrow \text{zero\_matrix}(p+1, q+1)$ 
 $C(0, [0, q]) = 0$ 
for  $i = 1 \dots p$  do
  for  $j = 1 \dots q$  do
    if  $(D(i, j) + C(i-1, j)) / (L(i-1, j) + 1) < \min((D(i, j) + C(i, j-1)) / (L(i, j-1) + 1), (D(i, j) + C(i-1, j-1)) / (L(i-1, j-1) + 1))$  then
       $x = -1, y = 0$ 
    else if  $(D(i, j) + C(i, j-1)) / (L(i, j-1) + 1) < (D(i, j) + C(i-1, j-1)) / (L(i-1, j-1) + 1)$ 
then
       $x = 0, y = -1$ 
    else
       $x = -1, y = -1$ 
    end if
     $C(i, j) = D(i, j) + C(i+x, j+y)$ 
     $L(i, j) = L(i+x, j+y) + 1$ 
  end for
end for
```

上記のアルゴリズムは、DTW において累積距離を比較して評価する際に、累積距離ではなくその平均値を用いて評価するように変更した DTW である。この手法の場合、累積距離の評価ではなく、対応ごとの距離の平均によって最適なアラインメントが探索される。そのため終了条件は距離関数 D の値が大きい対応しかアラインメントに追加する事ができない場合となる。つまり距離関数 D の値が、それまでの対応間の距離の平均値を下回る限り、対応付けを続けることになる。

しかし上述の手法は、距離が短い対応なら必ずアラインメントに追加することになる。そのため最終的な対応の数が多くなり、最適なアラインメントとはならない状況が発生する。アラインメントの数を減らすためには、常に増加し続ける値に対して最小値を探す、という条件が必要である。つまり累積距離が小さくなるように対応付けを行うことで、過剰に対応付けを行うことを回避可能である。しかし累積距離で評価することは、終点が自由な DTW の場合は適切ではない。そこで始点と終点が決まっているモデル系列に対しては累積距離で評価し、始点と終点が決まっていない入力系列に対しては平均値で評価する必要がある。アラインメントに含まれる入力系列の最初の要素からの時間に対する DTW 距離の平均値を求める関数が、 $IA(\text{DTW 距離})$ で表されるとして、このアルゴリズムを以下に示す。

Algorithm 4 始点終点自由な DTW2

```
 $C \leftarrow \text{infinity\_matrix}(p+1, q+1)$ 
 $C(0, [0, q]) = 0$ 
for  $i = 1 \dots p$  do
  for  $j = 1 \dots q$  do
    if  $IA(C(i-1, j)) < \min(IA(C(i, j-1)), IA(C(i-1, j-1)))$  then
       $x = -1, y = 0$ 
    else if  $IA(C(i, j-1)) < IA(C(i-1, j-1))$  then
       $x = 0, y = -1$ 
    else
       $x = -1, y = -1$ 
    end if
     $C(i, j) = D(i, j) + C(i+x, j+y)$ 
  end for
end for
 $\text{minIndex} \leftarrow 1$ 
for  $j = 2 \dots q$  do
  if  $IA(C(p, j)) < IA(C(p, \text{minIndex}))$  then
     $\text{minIndex} = j$ 
  end if
end for
 $\text{DTWdis} = C(p, \text{minIndex})$ 
```

上記のアルゴリズムは入力系列の新たな要素の対応付けを行う時、場合により累積距離が減少するように、アルゴリズム 3 を変更した DTW である。累積距離を比較する際に、対応付けた入力系列の始点からの時間に対する累積距離の平均値を比較するように変更することで、より正確にアラインメントが行われるようになる。

しかし上記のアルゴリズムでは、モデル系列の始点、終点に対応付けられる入力系列の点が、過剰になる可能性がある。そこでモデル系列の始点、終点の含まれる対応のうち、最も距離が小さい対応を始点、終点として採用し、始点ならばその点以前、終点ならばその点以降の対応を除去する。これにより過剰な対応付けや、不正確な始終点検出、対応付けのないアラインメントが可能となる。

3.5 個人依存動作検出

個人依存動作を検出するためには、一方の系列にしか存在しない要素を検出することが必要である。それには 2 系列の共通部分だけを取り出せばよい。関連研究で触れた LCS は、この目的に合致して

いる。時系列の拘束は保ちつつ、共通部分だけを取り出すためである。しかし LCS を適用するためには、時系列データを完全一致が可能な記号列へ変換する必要がある。本研究で扱うような数値データを、完全一致が可能な記号列へ変換することは困難である。そこで時系列多次元データから取得したベクトルの角度データを、類似度算出可能な記号列へ変換し、記号列間の類似度によって共通部分を抽出する。

まずモデル系列の角度データから最大値と最小値を取り出す。そして取り出した最小値から最大値の範囲を 4 分割し、それぞれに番号を与える。その後入力系列の角度データを、分割した 4 範囲のどこに属するか判定し、属する範囲に与えられた番号を、入力系列の角度データに対応する番号とする。範囲を超える数値に対しては、その数値に最も近い範囲の記号を対応させる。

その後番号に変換した角度データの類似度によって、入力系列に存在する要素から個人依存動作を検出する。類似度は番号列の一致度とし、閾値として設定した値を超えた場合、類似フレーム (個人依存動作ではない部分) とする。

3.6 身体座標情報を利用した DTW によるイベント認識

スポーツのイベント検出手法として、映像上の試合情報を利用する手法 [10] や、スポーツ映像に付随する実況の音声データを利用する手法 [11] が存在する。これらの手法はスポーツの映像が放映されることに伴って、試合映像に追加される情報を利用したものであり、常に利用できるとは限らない。また映像そのものを解析し、イベントを識別する手法として、映像の遷移を隠れマルコフモデルにより学習する手法 [12] が存在する。野球中継映像の各シーンをシンボル列として各イベントごとに学習することで、映像からのイベント分類を実現している。しかしこの手法では、ホームランやヒットなど映像上で変化が大きいイベントに関しては高い精度で識別可能なものの、打席が完了しないイベントについては識別が不可能である。そこでスポーツ映像のすべてのイベントについて識別するためには、スポーツ映像中の選手の動作を利用することが必要だと考えた。

そこで身体座標情報に対して DTW を用いることで、DTW 距離を求め、その値によってイベントを識別する。しかし識別したいイベントを含む入力系列に対して、そのまま DTW を適用しても、対象外動作の影響で正しく類似度を求めることはできない。そこで第 3.4 節で述べたような始点と終点を検出して DTW を行う手法によって、正しい類似度を求める。モデル系列と入力系列が同じイベントである場合と、異なるイベントである場合の DTW 距離は、離れた値になる。この DTW 距離の差によってイベントを識別する。

3.7 欠損フレーム予測

時系列データは、完全に連続した数値データではなく、離散している場合がある。つまり現実の動作の場合は完全に繋がっており、アナログ信号であるといえるが、現実の動作から映像などのデータ

として取得した場合，離散的な値しか取得できず，欠損しているデータが存在する．本研究に使用した映像データにおいても，人間の目で確認した場合に欠損したフレームが存在することを確認している．



図 3.2 映像フレーム 1

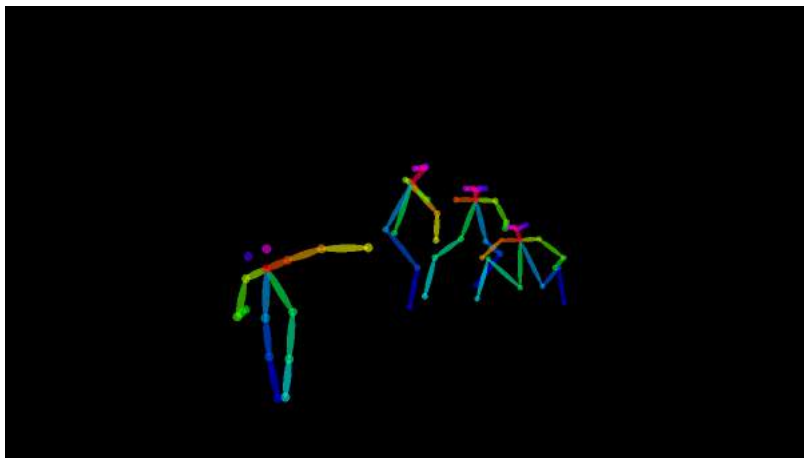


図 3.3 映像フレーム 2

図 3.2 と図 3.3 に示した画像は，欠損フレームが存在する映像部分を OpenPose の出力画像で示した画像である．2 画像は連続したフレームの画像であるが，カメラの性能などの影響により，欠損している映像部分が存在することが確認される．特にスポーツ動作は高速である場合が多く，欠損フレームの影響が大きくなる可能性が高い．アラインメントにおいても，欠損フレームの存在によって影響を受ける．欠損フレームの発生によって片側にしか存在しない動作部分が存在する場合，正しいアラ

インメントが行われない可能性が高い。また DTW によって出力される類似度も、欠損フレームに対応するフレームが、本来対応付けされるべきではないフレームに対応付けられることになるため、距離関数の出力も本来の値より大きい値となり、正しい値を求めることができない。そこで欠損フレームの発生を検出し、欠損フレームに対応する身体位置座標を推定することで、動作の時系列データを補間し、より正確なアラインメントを実現する。

3.7.1 欠損フレームの推定

欠損フレームを検出するために、アラインメントの結果を利用する。欠損フレームが存在する場合のアラインメントは、欠損フレームに最も近い、存在しているフレームに対応付けられていると考えられる。

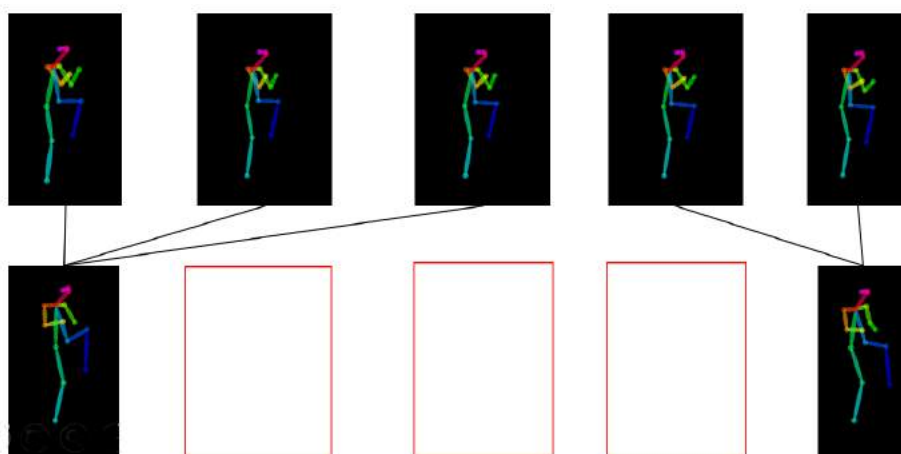


図 3.4 欠損フレームと対応付け

図 3.4 に、欠損フレームによって対応付けに影響が与えられたと推測される部分の OpenPose による検出結果を示した。上のモデル系列に対して、下の入力系列では、欠損フレームが存在するため、複数のモデル系列のフレームが、入力系列の 1 フレームに対して対応付けられている。このように、特定フレームに対して複数のフレームの対応付けが行われている場合は、欠損フレームが存在する可能性が高い。しかし動作の個人差などの関係で、対応付けが正しいにも関わらず、複数フレームの対応付けが行われる可能性がある。そこで複数のフレームが対応付けされた区間から、DTW 距離によって欠損フレームの発生を判定する。

欠損フレームが存在することにより、本来対応付けが存在するはずのフレームに対応付けが行えなくなり、異なるフレームに対応付けされているとする。この時 DTW 距離は、対応付けされるべきではないフレームに対応付けが行われている分、値が大きくなると考えられる。図 3.4 において本来対

対応付けられるべきフレームが左端の2フレーム、右端の2フレームだとする。この時 DTW 距離は、対応付けが存在している両端の2フレームよりも、その間に存在する対応付けされるべきフレームが欠損しているフレームの方が大きくなるはずである。したがって複数フレームが対応付けられている連続した2フレームから、それぞれのフレームに対応付けられたフレームのうち、DTW 距離が最小になる対応を探索することで、欠損フレームを検出する。

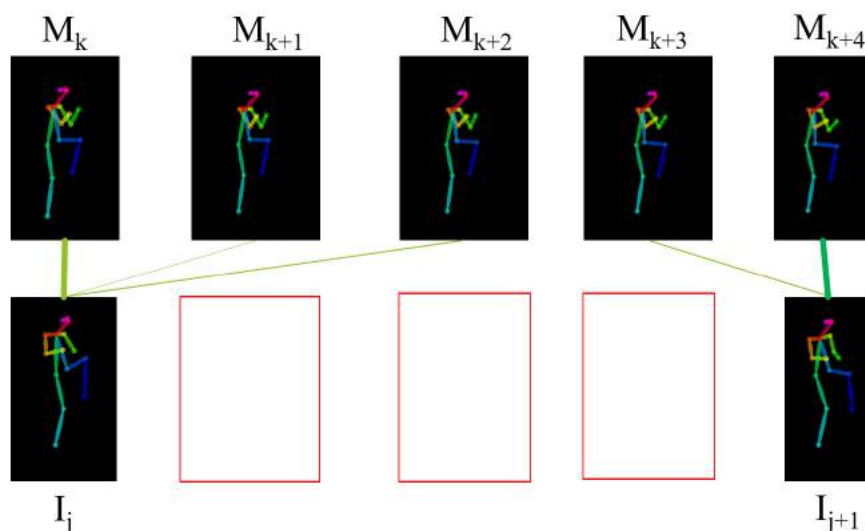


図 3.5 欠損フレーム区間の検出

図 3.5 の状況での欠損フレーム検出の流れを示す。上側の画像が入力系列，下側の画像がモデル系列であり，画像を結ぶ線は DTW によって対応付けられた結果とする。太線は複数対応付けが存在する中で，DTW 距離が最小の対応を示す。DTW によって出力されたアラインメントから，複数のフレームが対応付けられたフレーム I_j を探索する。そしてそのフレームに対応付けられたフレームのうち，距離関数 D において距離が最小であるフレームを求める。同様にその次フレームである I_{j+1} と DTW 距離が最小となるフレームを求める。図 3.5 で太線となっている $D(M_k, I_j)$, $D(M_{k+4}, I_{j+1})$ が距離最小である。この時 M_k と M_{k+4} の間にはフレーム M_{k+1} , M_{k+2} , M_{k+3} が存在する。この M_{k+1} , M_{k+2} , M_{k+3} に対応するべきフレームが欠損フレームである。

3.7.2 欠損フレーム構築

検出した欠損フレーム部分の身体座標を予測する。時系列データから値が存在しない時間の身体座標の値を予測し，適する値を探索する。

まず欠損フレームにおける対応フレームとの距離関数の値を予測する。欠損フレーム部分の前後のフレームでは2フレームが対応付けられており，距離関数の値が算出可能である。図 3.5 では，距離

関数を D として, $D(M_k, I_j)$ と $D(M_{k+4}, I_{j+1})$ が欠損フレーム前後の距離関数の値となる. また欠損フレームが存在したとすれば, 対応付けは図 3.6 のようになる.

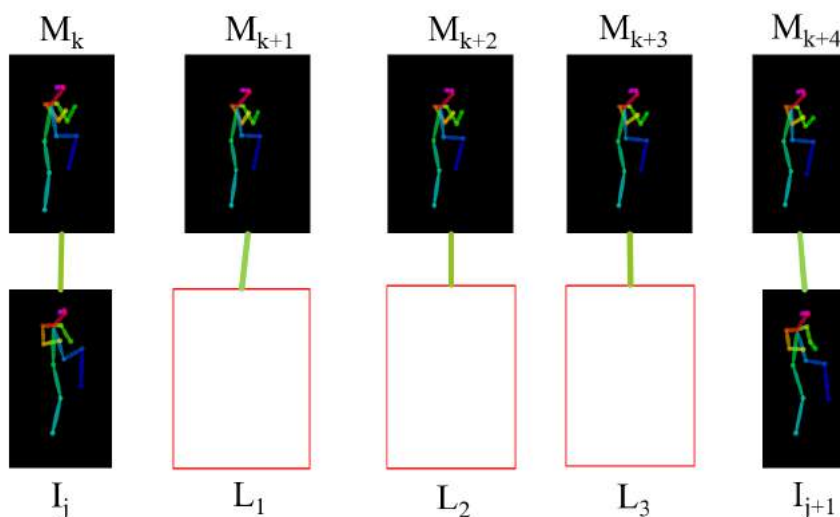


図 3.6 欠損フレーム区間の対応

上図のフレーム L_t は, 欠損フレームを表す. この時, 欠損フレームに対する距離関数 D の値を下式のように推測する.

$$D(M_k + p, L_p) = D(M_k, I_j) + \{D(M_k, I_j) - D(M_{k+4}, I_{j+1})\} \times p/4 \quad (3.3)$$

この距離に最も近い予測フレームが, 欠損フレームではないかと考えられる.

しかし上述の手法では, 予測フレームと欠損フレームの合計距離の誤差を最小化することになる. この場合, どの部位間にどれだけ誤差が存在するか評価できない. つまり距離関数 D の値にはすべての部位の組み合わせの距離が合計されているため, この値によって正しい部位座標を予測することは難しい. そこで各部位の組ごとの距離を算出する距離関数 d (式 3.1) の値を予測する. P をある 2 部位間の長さや角度のデータとして, 予測フレームの各 P と, 欠損フレームと対応付けられるべきモデル側のフレームの各 P との距離を関数 d によって算出する. 各 d の値の予測値との誤差の合計を最小化する予測フレームの身体座標の値を欠損フレームの身体座標とする. モデル系列のフレームを M_a , 予測フレームを E , 予測される距離関数 d の値を F_x とすると, 下式のように誤差合計を算出する.

$$S = \sum_{x=1}^{55} |F_x - d(P_x^{M_a}, P_x^E)| \quad (3.4)$$

この誤差の合計 S が最小となるような予測フレーム E を求める.

予測フレームは、時系列身体座標データの値を元に算出する。まず時系列身体座標データのフレーム間において、身体座標が等速で動いていると仮定する。そして欠損フレーム区間の値を、フレーム間を 10 等分した時間ごとに、等速で動いたと仮定した値によって算出する。図 3.6 の状況であれば I_j と L_1 の間、 L_1 と L_2 の間、 L_2 と L_3 の間、 L_3 と I_{j+1} の間をすべて 10 等分するような時間での身体座標の予測値を、等速移動したという仮定の下算出する。そして欠損フレームに対応付けられるべきモデル側の各フレームにおいて、式 3.4 の S を最小化する予測フレームの時間を求める。求めた予測フレームの時間における身体座標が、欠損フレームでの身体座標である。

3.8 アラインメント結果を利用した誤検出部位の修正

3.8.1 検出失敗部位の特徴

本研究では、OpenPose を用いて映像から身体座標を取得している。しかし上述したように OpenPose のオクルージョンに対する頑強性にも限界があり、どのような身体座標取得手法においても映像に映らない部位の座標を正確に予測することは難しい。そこでこのような時系列データを利用するようなシステムでは、正しく検出できなかった部位を予測値で埋めることによって、検出が完全には行えない場合でも、正しい結果に近い結果を取得可能にしている。この検出失敗部位を予測する手法として最も容易なものは、時系列データならばその前後の値の中間値を、検出失敗部位の予測値とする手法である。しかしこの手法の場合には、誤検出フレームにおいて映像上で直線に部位座標が移動していない場合、不正確な結果を得ることになる。

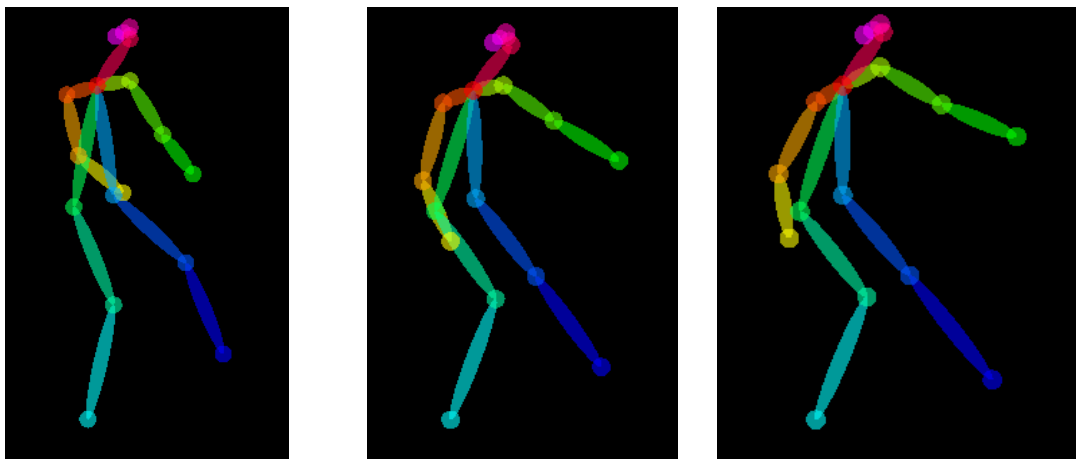


図 3.7 OpenPose によって取得した動作の連続した 3 フレーム

図 3.7 にその例を示した。三つの画像は時系列で連続しており、右腕部分に注目すると、各図の間で等速直線運動してはいることがわかる。

そこで時系列上での変化をより正確に得るために、検出失敗部分の前後のフレームを取得する範囲

を拡大し、複数フレームにおける変化から検出失敗部位の移動法則の関数を作成することで、より正確に予測することが可能である。しかし予測に使用する範囲を広げることで、他の検出失敗部分を含む可能性もあり、またスポーツのような高速で短い時間に行われる動作においては正確に予測関数を得ることは難しい。特に高速な動作については、カメラの性能にも影響されるものの、時系列データとして映像に捕らえられていない、欠損した部分が多く存在する可能性が高い。その状況で予測関数を求めたとしても正確なものになる可能性は低い。



図 3.8 カメラ性能と動作速度による動作部分の欠損

図 3.8 は高速動作によって大きく欠損した部分が存在する映像に対しての OpenPose による検出結果の例である。図の 2 画像は連続した 2 フレームの画像であるが、右腕部分が大きく動いており、高速な動作によって、動作部分の多くが欠損したと考えられる。

そこでアラインメントされ、対応付けられたフレームのデータを利用することで、短い区間のデータから検出失敗部位の座標を予測する手法を示す。

3.8.2 欠損部位座標予測

欠損部位の座標を、アラインメントの結果から予測する。本研究ではモデル系列のデータは正確にとれる、もしくは修正可能であるという前提であるため、入力系列に欠損部位が存在するとし、予測手法を示す。

最初に入力系列における欠損部位と欠損部位以外との間の長さや角度の組 P_i^I について、そのフレームに対応付けられたモデル系列の長さや角度の組 P_j^M との距離の予測値を求める。これは第 3.7.2 項で述べたように、全体の距離の値の合計では、各部位ごとの誤差が吸収され、誤った結果が正しいと判定される可能性があるためである。

欠損部位の存在する入力系列の特定のフレーム番号を f とする。入力系列の $f-1$, f , $f+1$ フレームに対応付けられたモデル系列のフレーム番号を、各 $g-1$, g , $g+1$ フレームとする。そして f フレーム目の対応についての、ある長さや角度の組 P 間の距離の予測値 Pd_a を、 $f-1$ フレーム目の対応の距離関数の値と $f+1$ フレーム目の距離関数の値から求める。

$$Pd_a(P_a^{M_g}, P_a^{I_f}) = \{d(P_a^{M_{g-1}}, P_a^{I_{f-1}}) + d(P_a^{M_{g+1}}, P_a^{I_{f+1}})\} / 2 \quad (3.5)$$

式 3.5 は距離 Pd_a の予測値を求める式である。同じ動作のアラインメントに対する距離関数の値は、急激に変化する可能性が低く、前後のフレームから値の予測が可能だといえる。

その後誤差算出関数の結果を用いて最適な部位座標を判定する。誤差算出関数は、上述した P 間の距離の予測値と、部位座標を仮の値を入力することで、入力した仮の座標での P と、モデル系列の P 間の距離と、予測距離の値との誤差を合計し、出力する関数である。欠損部位に関わる P の番号を lp_1, lp_2, \dots, lp_b とした時、誤差算出関数 ES を以下に示す。

$$ES = \sum_{k=1}^b |d(P_{lp_k}^M, P_{lp_k}^I) - pd_{lp_k}| \quad (3.6)$$

式 3.6 の値を、特定の範囲の座標値を入力することで求める。そして特定の範囲すべての座標値での誤差算出関数の結果から、最も小さい ES を出力する座標を、欠損部位の予測座標とする。

3.9 むすび

各提案手法について具体的な内容を示した。特に従来手法が存在する手法について、提案手法と比較し、どのように従来手法を改良するか示した。

第 4 章

評価実験

4.1 まえがき

本章では、アラインメントについての評価実験を行った結果を示す。実験に用いるデータセットは、野球中継映像における右投手の投球映像 28 本である。このうち 1 本をモデル系列とし、同じイベントである、その他の 27 本の映像を入力系列とする。またモデル系列の映像から OpenPose によって取得した身体座標データは、誤検出が発生している場所を修正し、正しいデータが取得できたことと同じ状況にする。これはモデル系列については正しいデータが取得できていることが前提の手法であるためである。誤検出された身体座標データを目視で確認し、正しい値へ変更している。入力系列については誤検出についても一切の修正を行っていない。アラインメント結果については、対応付けられたフレームを左右に並べた画像結果と、実験によっては目視によりアラインメントの成否を確認した数値データの結果を加えて示す。画像結果については、OpenPose によって取得した身体座標をプロットし、背景色をすべて黒にした画像によって、アラインメントの結果を示す。対応付けられた各系列のフレームを、モデル系列の画像を画像左側、入力系列の画像を画像右側に貼り付けることによって、対応付けの画像列として結果を示す。目視の結果判定については、対応付けされるべきフレームが他に存在する場合、誤った対応付けとしてカウントする。

4.2 個人依存動作除去の評価

第 3 章第 3.5 節で述べた個人依存動作除去の結果を示す。モデル系列を固定し、個人依存動作を除去した系列の DTW 距離と、除去していない系列の DTW 距離を求め、その結果を表 4.1 に示す。

表 4.1 個人依存動作除去の DTW 距離に対する影響

個人依存動作の除去	無し	有り
DTW 距離最小値	0.121	0.111
DTW 距離最大値	9.365	1.700
DTW 距離平均値	1.786	0.585
DTW 距離中央値	1.284	0.453

DTW 距離の比較実験から，個人依存動作除去によって，DTW 距離が小さくなり，系列がより類似するように変化したといえる．それは表 4.1 に示されている．イベント検出に DTW を利用する場合には，この手法によってモデル系列と異なるイベントについても DTW 距離を小さくしてしまうため，有効ではない．しかし同じイベントに対してアラインメントする場合には，この手法によって系列間の距離を近付ける処理は，正しい対応付けを行うために有効である．

4.3 始点終点自由な Dynamic Time Warping の評価

始点終点自由な DTW についての評価実験の結果を示す．

4.3.1 Prefix and Suffix Invariant Dynamic Time Warping の評価

従来手法である，PSI-DTW について性能を確認した実験の結果を示す．この実験では，PSI-DTW の実行に必要なパラメータである，始点と終点の探索範囲 R について，理想値に近い値が求まっていることを仮定した．理想値が求まっていることとほぼ同じ状況で実験を行うため，モデル系列についてはイベントの始点と終点が既知であり，探索範囲 R は，入力系列全体とした．モデル系列のイベントの始点と終点を既知とすることで，その要素すべてについて対応付けを行わなければならない．そのため探索範囲 R が理想値より大きい値であることによって発生する，系列に含まれるイベントの一部しかアラインメントが行われないという問題を回避できる．

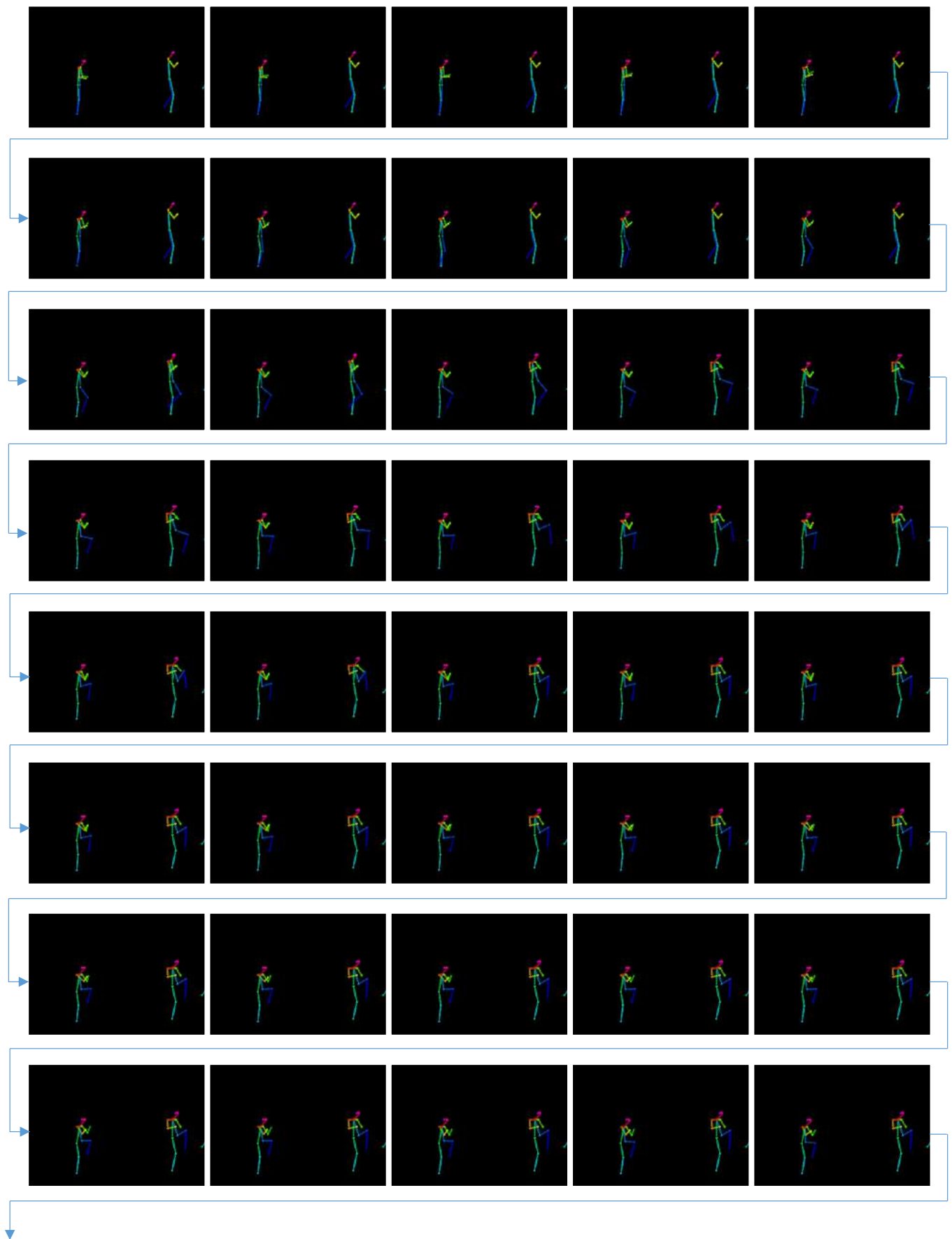
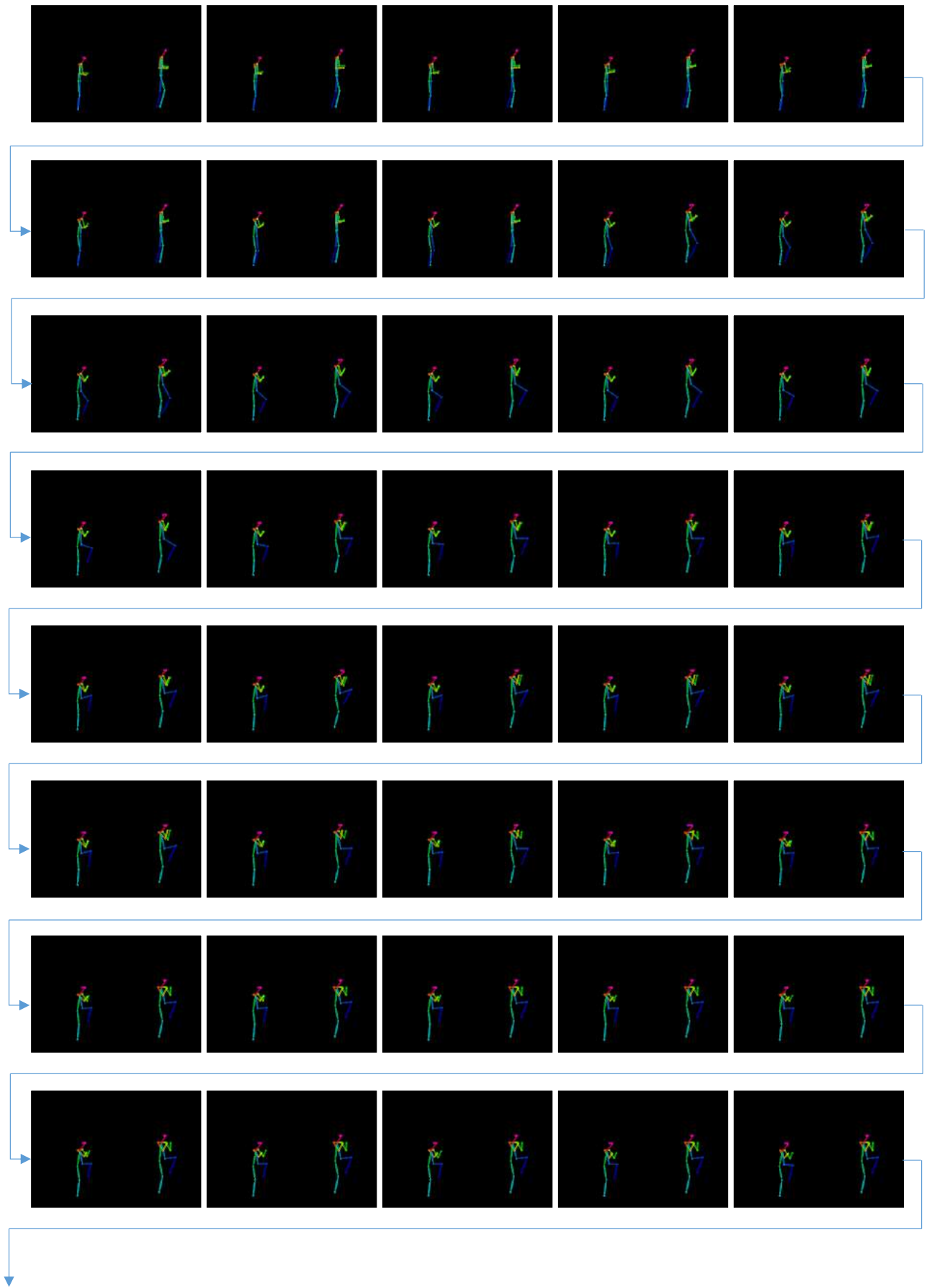




図 4.1 PSI-DTW のアラインメント成功例

図 4.1 はアラインメントの結果を表した図である。図中の各画像は、PSI-DTW によって対応付けられたフレーム同士を左右につなげて並べた結果である。各画像の左側がモデル系列の身体座標を検出した OpenPose の結果であり、右側が入力系列の身体座標を検出した OpenPose の結果である。左上の画像が最初の対応付けであり、各行で右方向に連続している。

そして図 4.1 はアラインメントに成功した例である。ほとんどの対応付けが正しく行われており、また始点と終点についても、投球動作の開始地点から終点までを失うことも、余計な部分を含むこともなく、アラインメントに成功している。



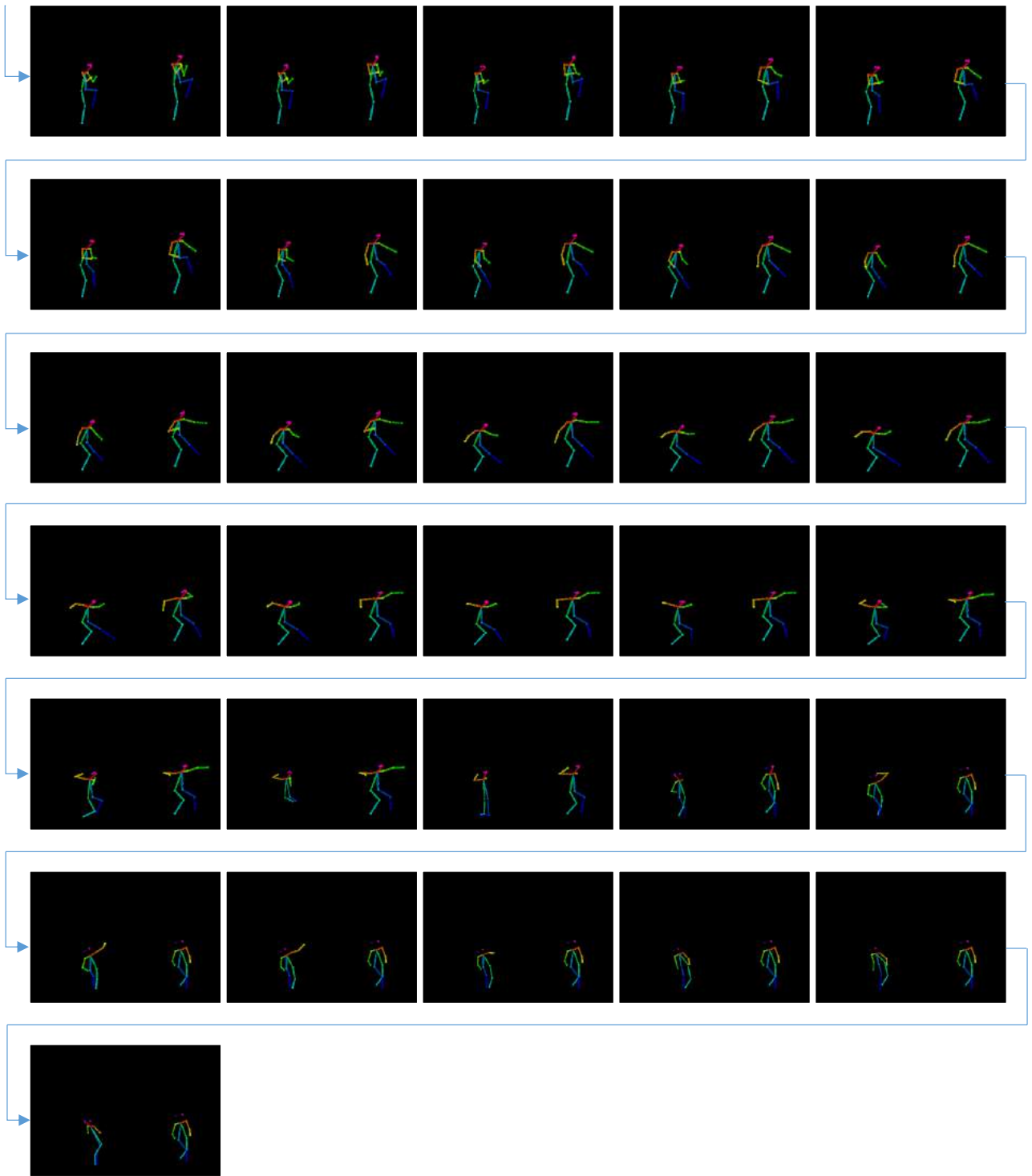


図 4.2 PSI-DTW のアラインメント失敗例 1

図 4.2 は PSI-DTW の失敗例の結果である．成功例の対応付け画像と同じく，投球動作の開始地点からほとんどの対応付けは正しい．しかし終点付近において，正しく対応付けが行われていない組が存在する．

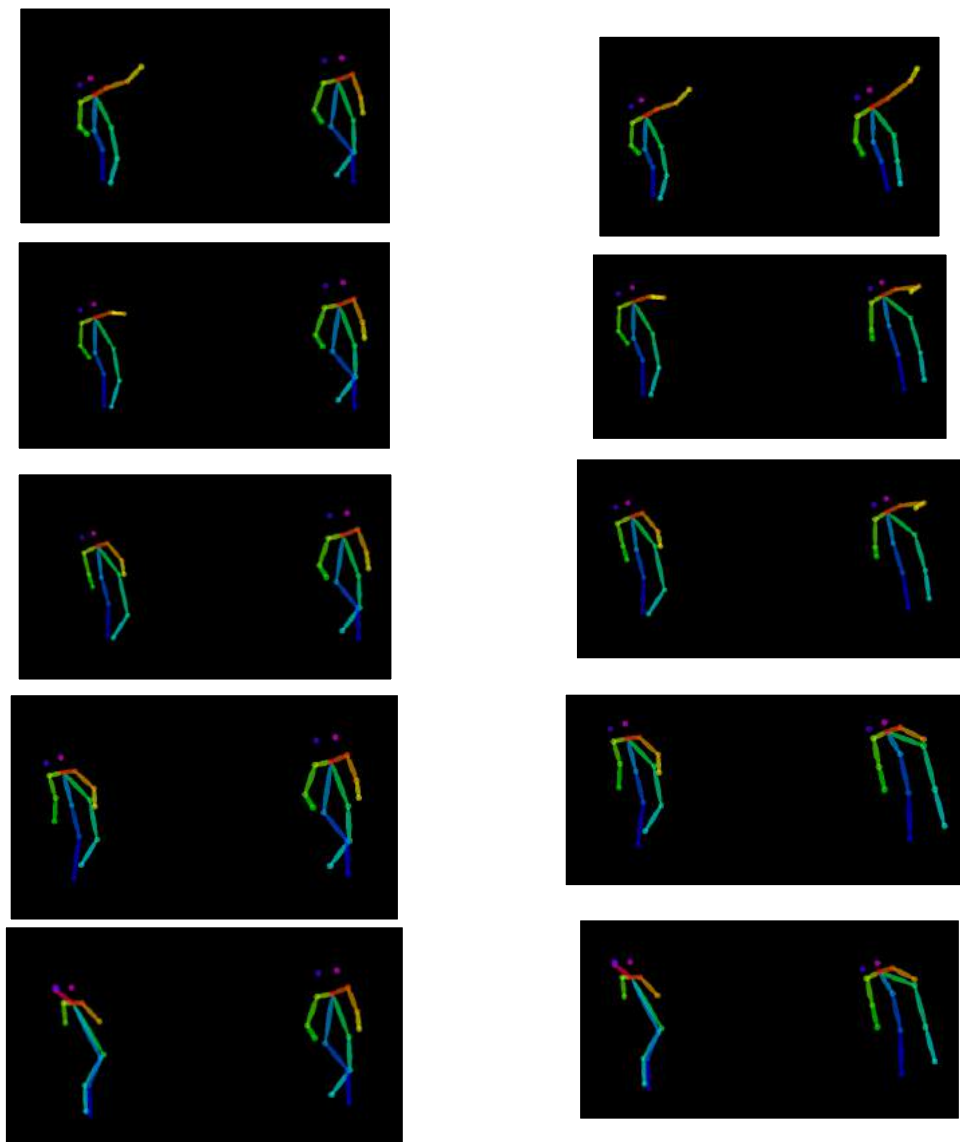


図 4.3 正しいアラインメントとの比較

図 4.3 は左に図 4.2 の終点付近の対応付け五つを示した結果，右が終点付近についての正しいアラインメントを示した結果である．元の映像を黒に加工しており，情報が失われているが，図 4.2 の

終端付近 5 フレーム分の対応付けは、正しい結果ではない。図 4.2 の最後の五つの対応付けを示した図 4.3 の左の縦列は、画像右側の入力系列のフレームがすべて同じ画像であるが、これは誤検出が発生したフレームである。オリジナルの映像を確認すると、このフレームは投球動作の終点ではない。したがって図 4.3 の右の縦列に示した対応付けが正しいアラインメントとなる。間違った対応付けが発生した原因は、PSI-DTW の特徴である、(始点と) 終点が自由な状態でアラインメントを行ったこと、そして累積距離によって最適なアラインメントを探索したことにありといえる。PSI-DTW は終点位置が自由であるため、アラインメントの終了条件は、終点の探索領域に含まれる要素を対応付けることである。実験条件より、モデル系列のイベントの始点と終点は既知であるため、アラインメントの終了条件は、モデル系列のイベントの終点を対応付けることである。これはモデル系列のイベントの終点が対応付けされれば、入力系列はどこが終端になっても良いということである。図 4.3 の左の縦列に示されているように、誤検出されたフレームは、モデル系列の終点付近のフレームと似ている状態である。この誤検出されたことによる近似フレームの発生によって、入力系列のさらに先を対応付けるよりも、この近似フレームに対してモデル系列のフレームを複数対応付け、終了条件を満たした場合、累積距離では小さい値になったと考えられる。通常の DTW ならば入力系列の終端までアラインメントを行うため、このような対応付けは発生しないが、PSI-DTW では終点が自由であるため、累積距離を短くするために、入力系列のイベントの途中でアラインメントを終了する対応付けが発生しうる。

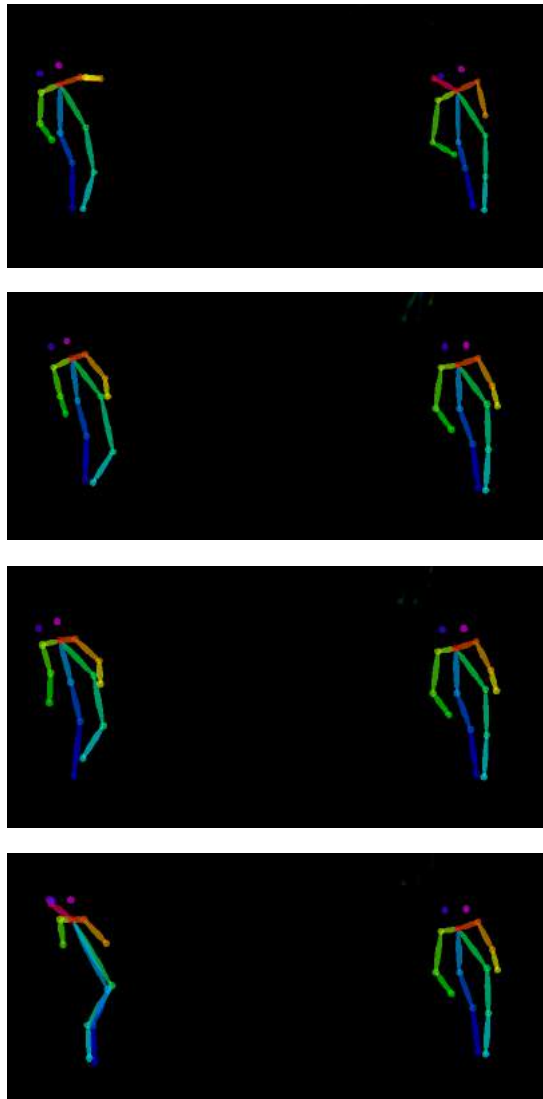


図 4.4 PSI-DTW のアラインメント失敗例 2

図 4.4 も PSI-DTW によるアラインメントが失敗した例である。図 4.2 と同様に終点付近のアラインメントが正しくないため、その部分のみを抽出し、画像で対応付け結果を示した。これは累積距離を最小にするという条件では正確にアラインメントされた結果である。この対応付け間違いは、距離関数の上でも、目視で確認してもわずかな差でしかなく、OpenPose による欠損や誤検出の発生した座標データでは、そのわずかな差をアラインメントに反映できなかったものと考えられる。

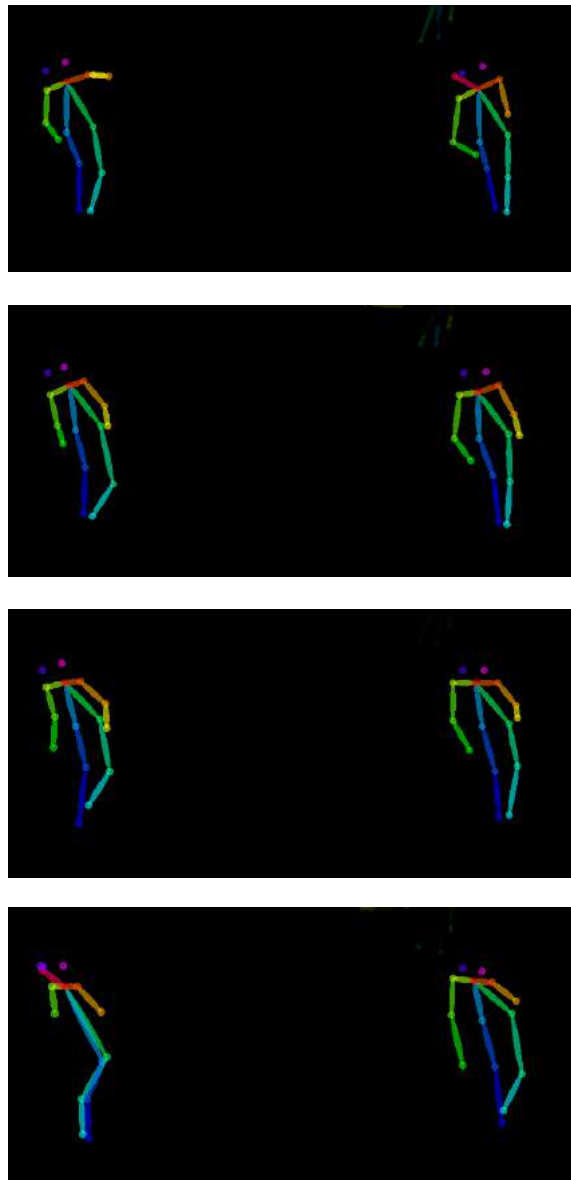


図 4.5 図 4.4 の正しいアラインメント

上図は図 4.4 の誤ったアラインメントに対する正しいアラインメントを示した結果である。元の映像では右腕部分の動きを確認可能であり、図 4.5 が正しいアラインメントであると確認した。

目視により対応付けごとに正しいアラインメントが行われた回数と間違ったアラインメントが行われた回数を確認し、まとめた結果を以下に示す。

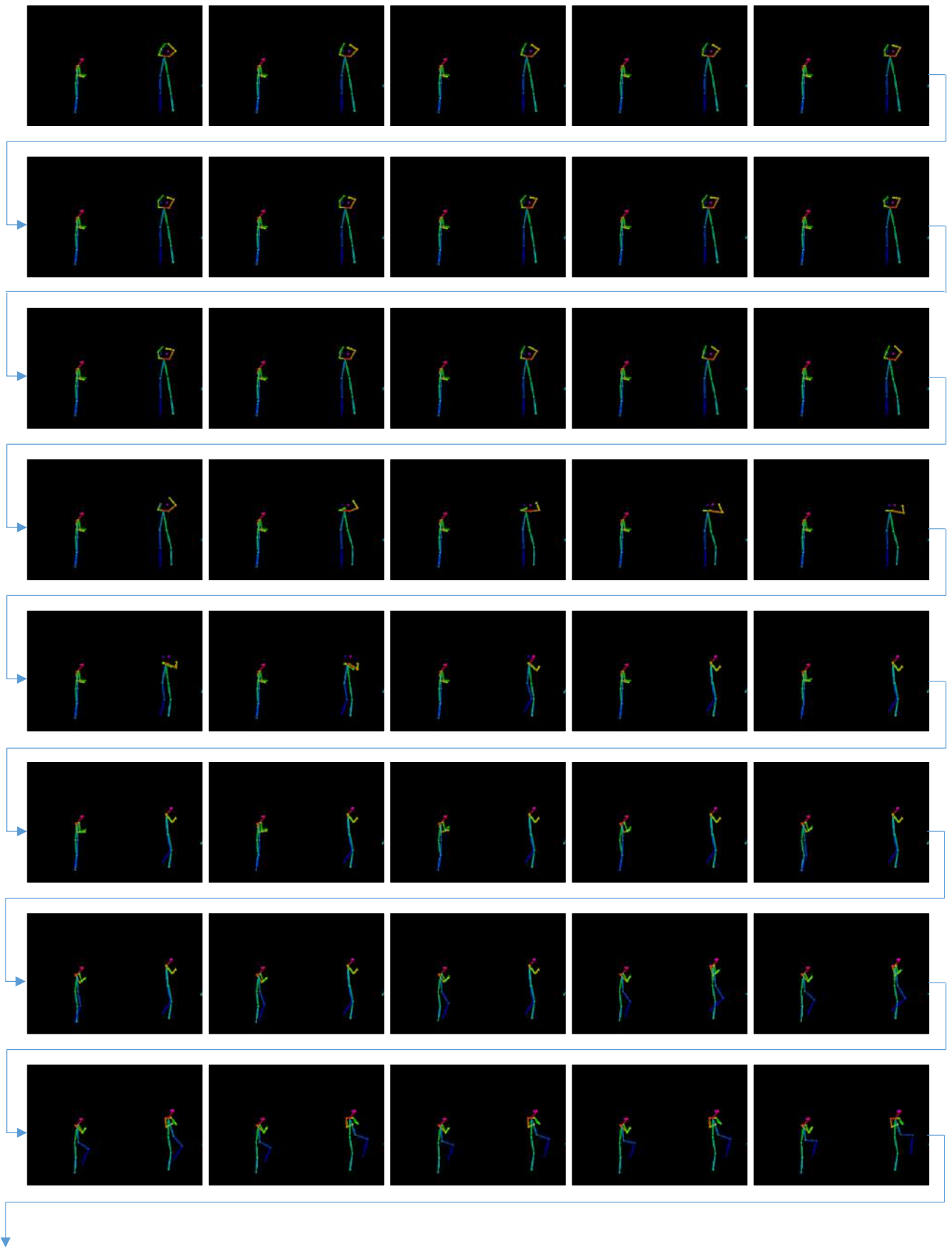
表 4.2 PSI-DTW アラインメント結果

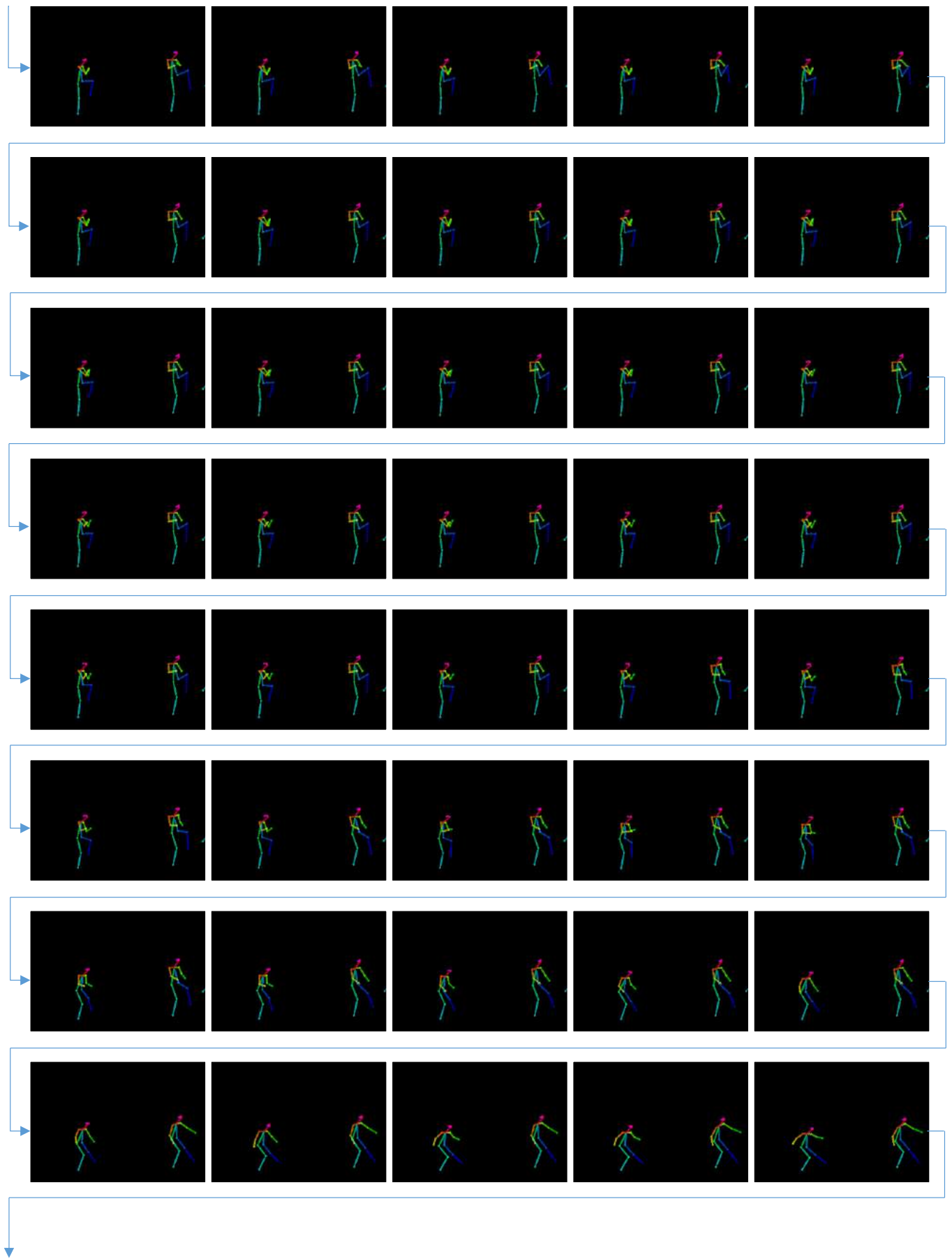
	成功フレーム	失敗フレーム
フレーム数	1836	139
割合 [%]	92.96	8.04

終点付近以外は正確にアラインメントできている部分が多いため、全体のアラインメント精度は良い結果が得られた。

4.3.2 アルゴリズム 3 の評価実験

第 3 章のアルゴリズム 3 についての評価実験の結果を示す。





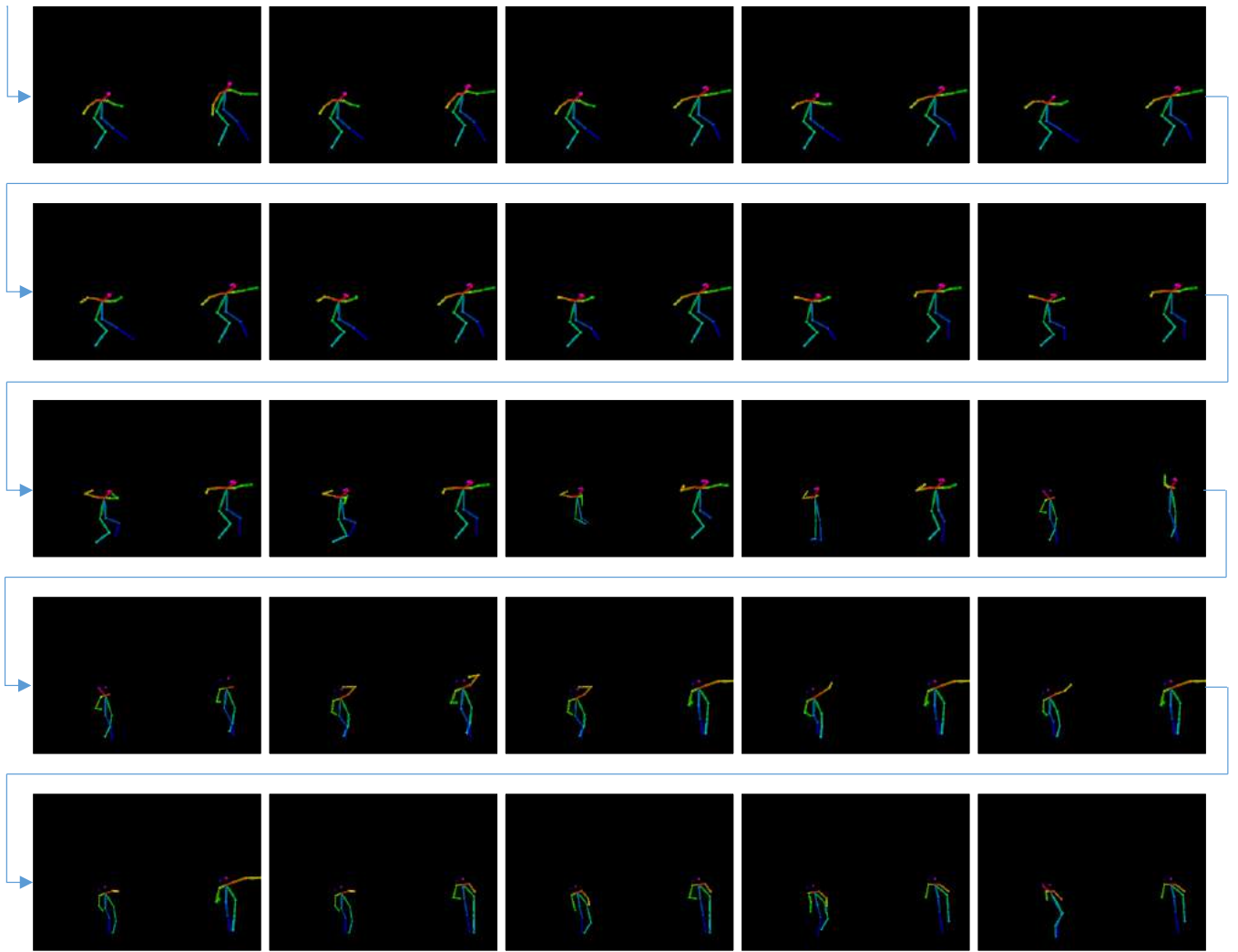


図 4.6 アルゴリズム 3 によるアラインメント

図 4.6 は PSI-DTW で正しくアラインメントが行われた入力系列について、アラインメントを行った結果である。アルゴリズム 3 の、各対応付けの距離の平均値を最小化する性質上、対応付けの数が増加した。似たようなフレームを対応付けているものの、よりよい対応付けが存在する部分が多く、不正確なアラインメントである。

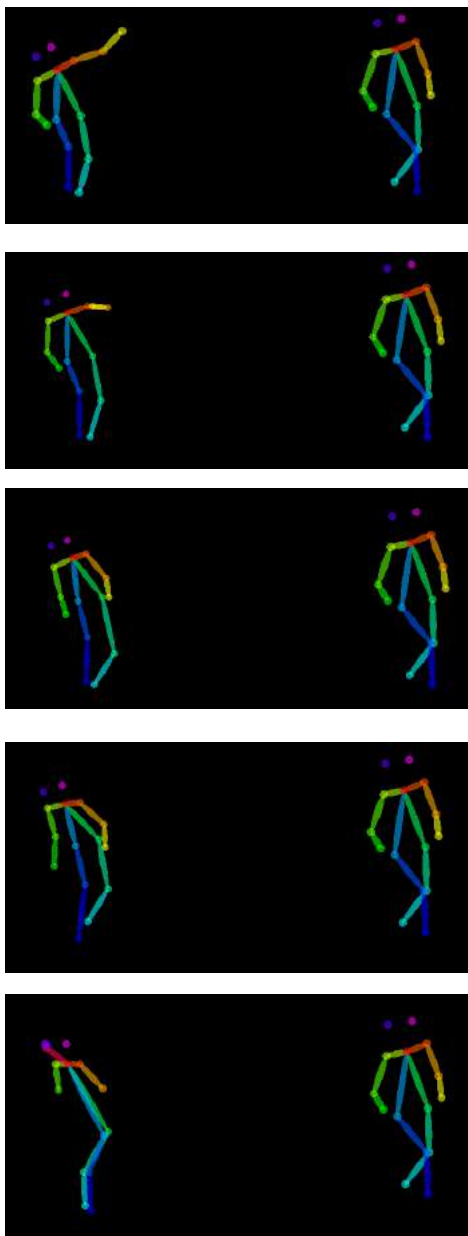


図 4.7 アルゴリズム 3 による終点付近のアラインメント

またこの手法においても終点を正確に対応付けることはできなかった。図 4.7 は、アルゴリズム 3 によるアラインメントのうち、終点付近を示した結果である。アルゴリズム 3 は対応付けの平均距離が小さくなるようにアラインメントを行う手法であるが、誤検出によって距離そのものが不正確な状況においては不十分なアルゴリズムとなっている。図 4.7 に示した結果についても、誤検出によって入力系列の特定のフレームが、モデル系列の終点付近のフレームに近づいたため、平均距離の上でも正しいアラインメントを行うことができなかったと考えられる。

最後に、このアルゴリズムによる対応付けの個数と、PSI-DTW による対応付けの個数の差を示す。

表 4.3 アルゴリズム 3 と PSI-DTW のアラインメント長さ比較 (その 1)

入力系列映像種類		1	2	3	4	5	6	7	8	9	10	11	12	13	14
対応付け個数	PSI-DTW	72	91	72	71	74	70	70	75	77	71	74	83	76	71
	アルゴリズム 3	113	130	128	119	100	113	112	110	121	105	106	128	112	90

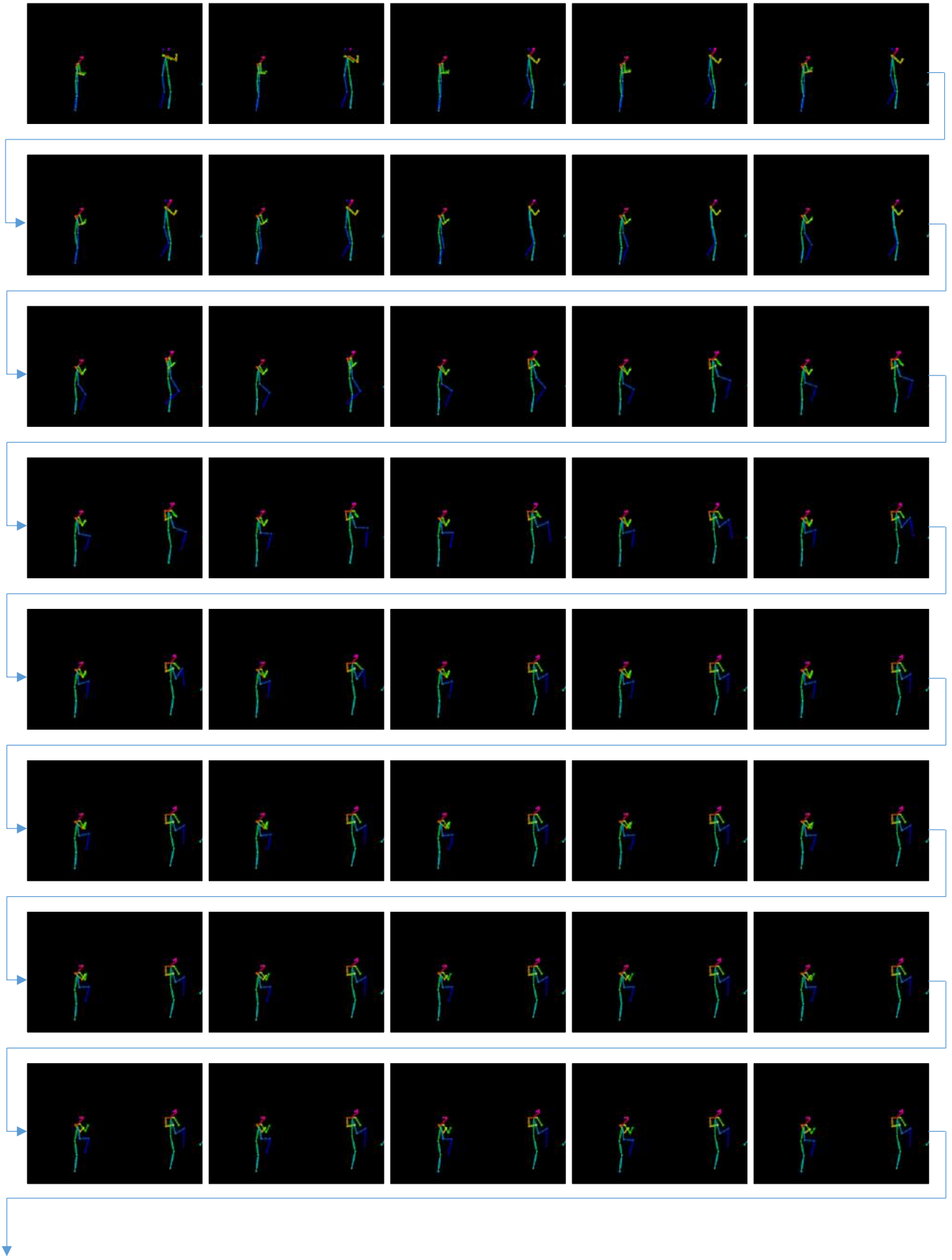
表 4.4 アルゴリズム 3 と PSI-DTW のアラインメント長さ比較 (その 2)

入力系列映像種類		15	16	17	18	19	20	21	22	23	24	25	26	27
対応付け個数	PSI-DTW	72	72	71	70	70	72	70	77	70	72	71	71	70
	アルゴリズム 3	105	92	93	106	111	112	102	126	109	102	89	108	130

表 4.3, 表 4.4 に、対応付け数の比較結果を示す。アルゴリズム 3 の対応付け数が、PSI-DTW に比べて増加していることが示されている。アルゴリズム 3 の場合、近似するフレームに対して対応付けを行うことはできているものの、他により適した対応付けが存在する場合でも、アラインメントに含める結果となった。不必要な対応付けが含まれることで、対応付けの数が増加しており、適切なアラインメント結果ではない。

4.3.3 アルゴリズム 4 の評価実験

第 3 章のアルゴリズム 4 についての評価実験の結果を示す。



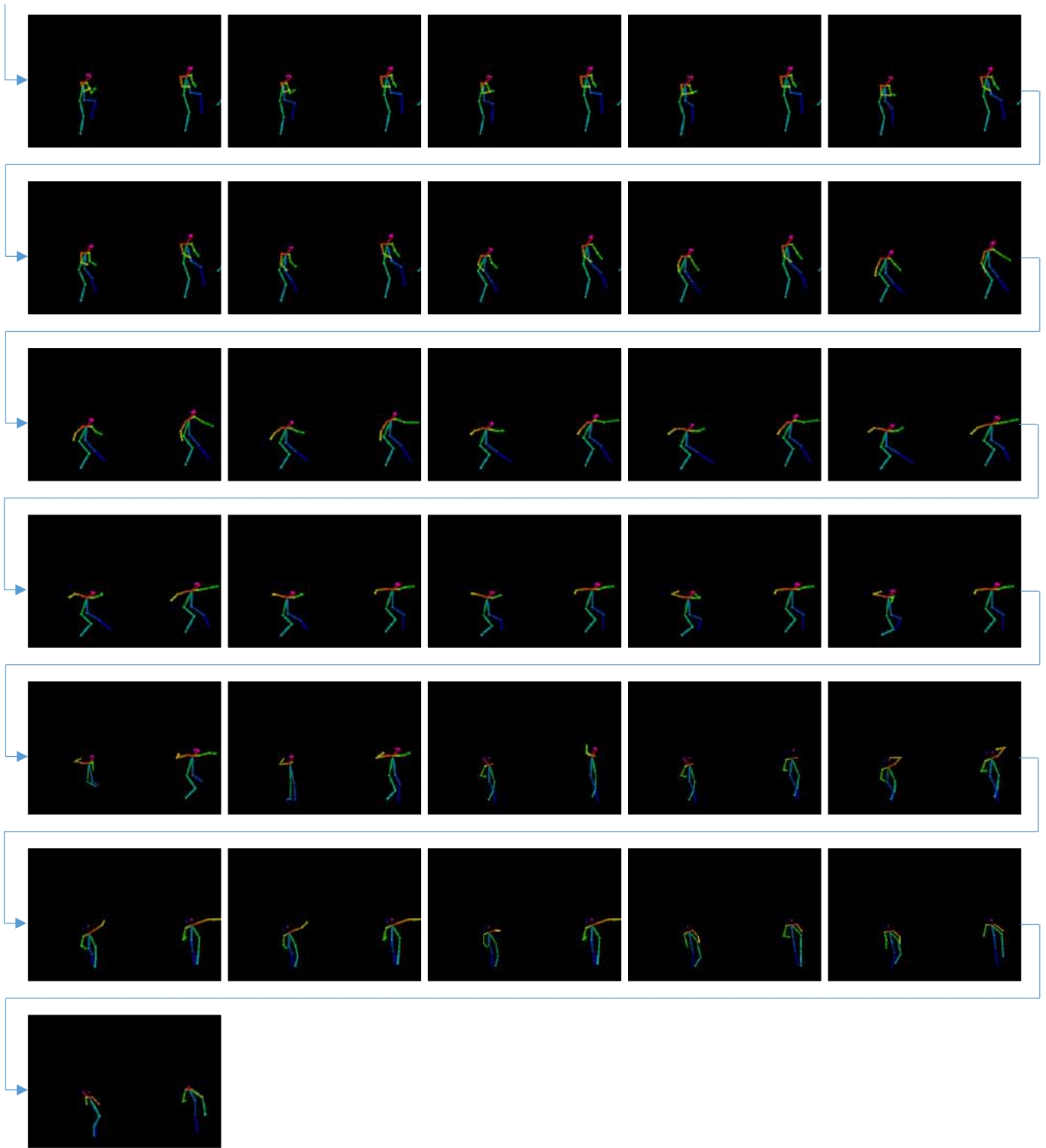


図 4.8 アルゴリズム 4 のアラインメント成功例

まず PSI-DTW によってアラインメントに成功した入力系列についてのアラインメント結果を示す。図 4.8 に示したように、他の手法でアラインメントに成功していたデータについては同様に正しいアラインメントを得た。特にアラインメントの大部分については PSI-DTW の結果と同様の結果を得た。DTW のコスト行列の更新部分について変更を加えたものの、PSI-DTW によって正しくアラインメント可能なデータについては、同様にこのアルゴリズムにおいても正しくアラインメントが可能だといえる。

次に PSI-DTW ではアラインメントに失敗していたデータについてアラインメントを実行した結果を示す。

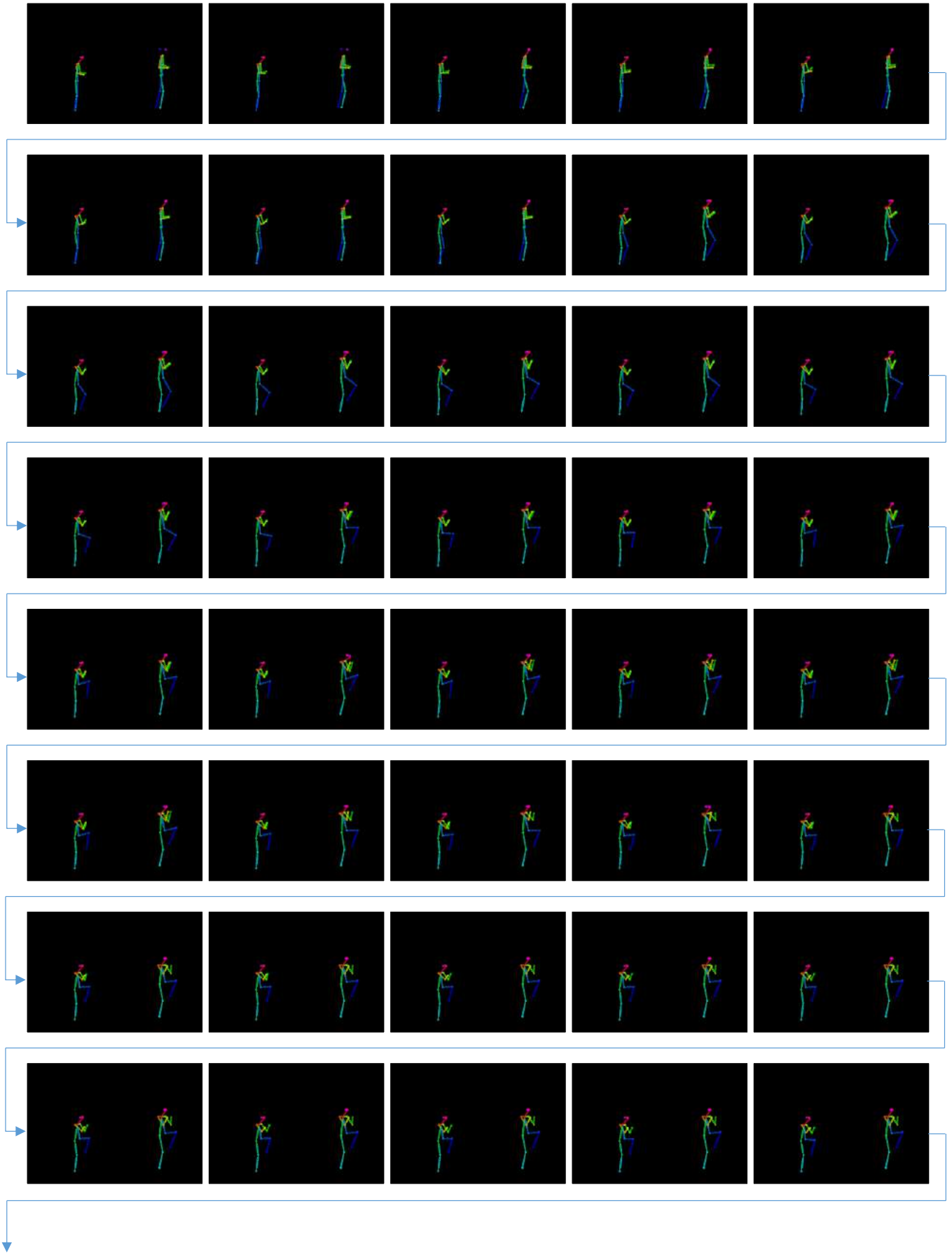




図 4.9 アルゴリズム 4 のアラインメント成功例 2

図 4.9 はアラインメントの画像結果である。始点と終点付近以外の部分については PSI-DTW のアラインメントと同様の結果である。しかし終点付近の、PSI-DTW で失敗していた誤検出フレームを含む終点付近のアラインメントについて、正しいアラインメントを得ることができた。モデル系列の終点付近と類似している入力系列の誤検出部分についても、そのフレームを終点とすることなく、入力系列における投球動作の終点までアラインメントに含まれている。またアルゴリズム 3 に近い形のアルゴリズムであるものの、余計な対応までアラインメントに含めてしまうという問題も発生せず、対応付けの数はほとんどの場合で PSI-DTW を実行した際の対応付けの数と同じとなった。これは単純な平均距離の比較としなかったことで、累積距離を評価する性質がこの手法に存在し、余計な対応を含めない結果になったと考えられる。

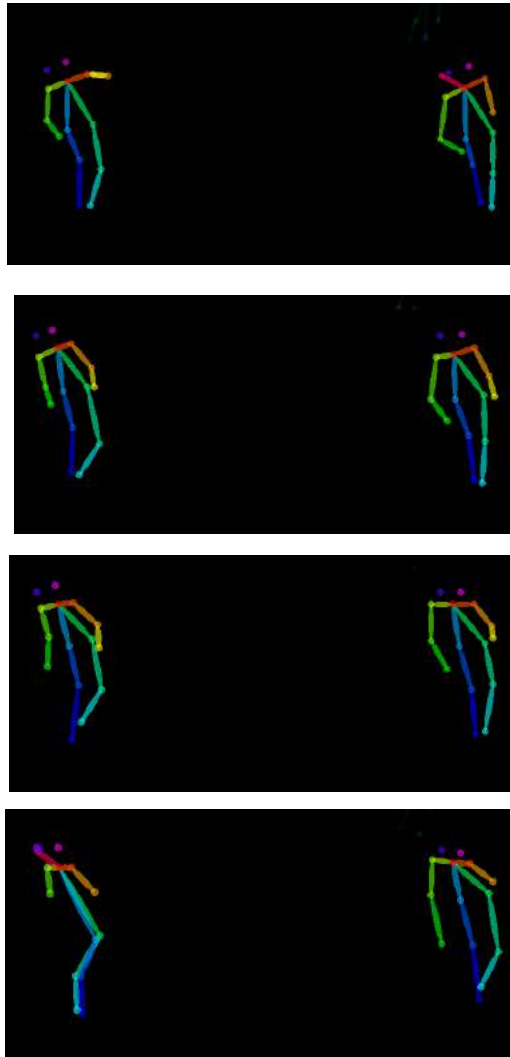


図 4.10 アルゴリズム 4 のアラインメント成功例 3

図 4.10 も PSI-DTW では終点付近のアラインメントに失敗していた入力系列に対するアラインメント結果であり，終点付近の四つの対応付けについて示した図である．図 4.5 に示した成功例と同様の結果を得ており，終点付近の似たようなフレームが連続するアラインメントについて，終点自由な入力系列側の終点を正確に決定することができている．

目視により対応付けごとに正しいアラインメントが行われた回数と間違ったアラインメントが行われた回数を確認し，まとめた結果を以下に示す．

表 4.5 アルゴリズム 4 アラインメント結果

	成功フレーム	失敗フレーム
フレーム数	1972	119
割合 [%]	94.31	5.69

終点付近が正確にアラインメント可能になったことで、PSI-DTW から精度が向上している。またそのことによって他のアラインメントが不正確にはならず、精度が単純に向上したという結果が示されている。

しかし本手法においてもアラインメントに成功していないフレームは存在する。その原因は OpnePose による身体座標の誤検出にある。本手法では誤検出によって、間違った身体座標がデータ系列の一部に存在する場合でも、正しくアラインメント可能な場合が存在することを示した。しかしすべての誤検出したデータについて、正しくアラインメント可能ではない。特に連続して誤検出が発生している身体座標データに対しては、正しい対応付けは困難である。以下に、連続して OpenPose による身体座標の誤検出が発生した例を示す。



図 4.11 連続して OpenPose による誤検出が発生したフレーム

図 4.11 は連続したフレームに対しての OpnePose による身体座標検出結果である。右足部分について、足を下ろしているような検出結果が示されているが、実際には膝を腰の高さまであげている。このような連続した誤検出には DTW によるアラインメント自体が困難なため、データの連続性などから身体座標の誤検出部分を特定し、修正する必要があると考えられる。

4.4 DTW によるイベント検出

DTW によるイベント検出が可能かを実験により確認する。DTW によって算出可能な DTW 距離は正規化が困難であり、閾値を設定してモデル系列と同じイベントが入力されたか、異なるイベントが入力されたかを判別する必要がある。しかし同じイベントが入力された場合と異なるイベントが入力された場合とで DTW 距離に明確な差が存在すれば、DTW 距離によってイベントを判別することが可能だといえる。そこでモデル系列と同じイベントと異なるイベントを入力した場合での DTW 距離を示し、DTW によるイベント検出が可能かを確認する。

使用するイベントのうち、モデル系列に右投手の投球映像を使用し、モデル系列と同様のイベントの映像を 27 本使用する。またモデル系列と異なるイベントとして、投手の直立映像と牽制動作映像を合計 10 本使用する。また用いる DTW は 2 章のアルゴリズム 1 と、提案手法に示したアルゴリズム 4 とし、結果を表 4.6、表 4.7 に示す。

表 4.6 アルゴリズム 1 の DTW 距離

入力イベント	投球映像	投球以外の映像
DTW 距離最小値	0.121	4.603
DTW 距離最大値	9.365	10.144
DTW 距離平均値	1.786	7.017
DTW 距離中央値	1.284	6.462

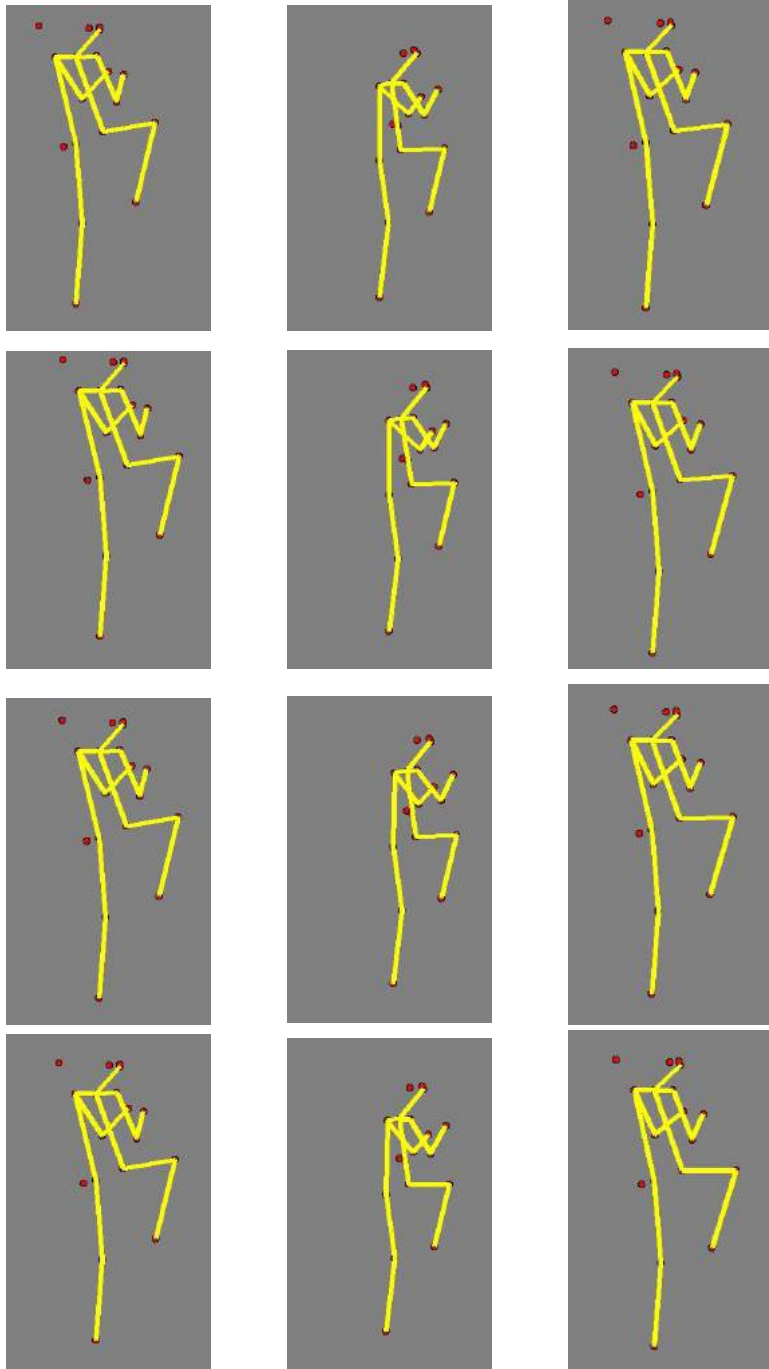
表 4.7 アルゴリズム 4 の DTW 距離

入力イベント	投球映像	投球以外の映像
DTW 距離最小値	0.089	4.587
DTW 距離最大値	1.724	9.632
DTW 距離平均値	0.549	6.431
DTW 距離中央値	0.415	6.186

モデル系系列と同様のイベントを入力した場合と、異なるイベントを入力した場合とで、明確な差が現れていることが表 4.6、表 4.7 に示されている。DTW 距離中央値、DTW 距離平均値ともに差が存在し、DTW 距離によってイベントを判別することが可能だといえる。また DTW 距離最小値と最大値の結果に注目すると、表 4.7 より、アルゴリズム 4 ではモデル映像と同じ映像を入力した場合の DTW 距離の最大値が、異なる映像を入力した場合の DTW 距離の最小値を下回っているが、表 4.6 よりアルゴリズム 1 の場合はそうではない。したがってアルゴリズム 4 はよりイベントを判別するために適した手法だといえる。

4.5 欠損フレーム予測の実験

アラインメントを行った後，提案手法によって欠損フレーム部分を予測した結果を図で示す．アラインメント全体のうち，欠損フレーム予測が行われた部分のみを示す．



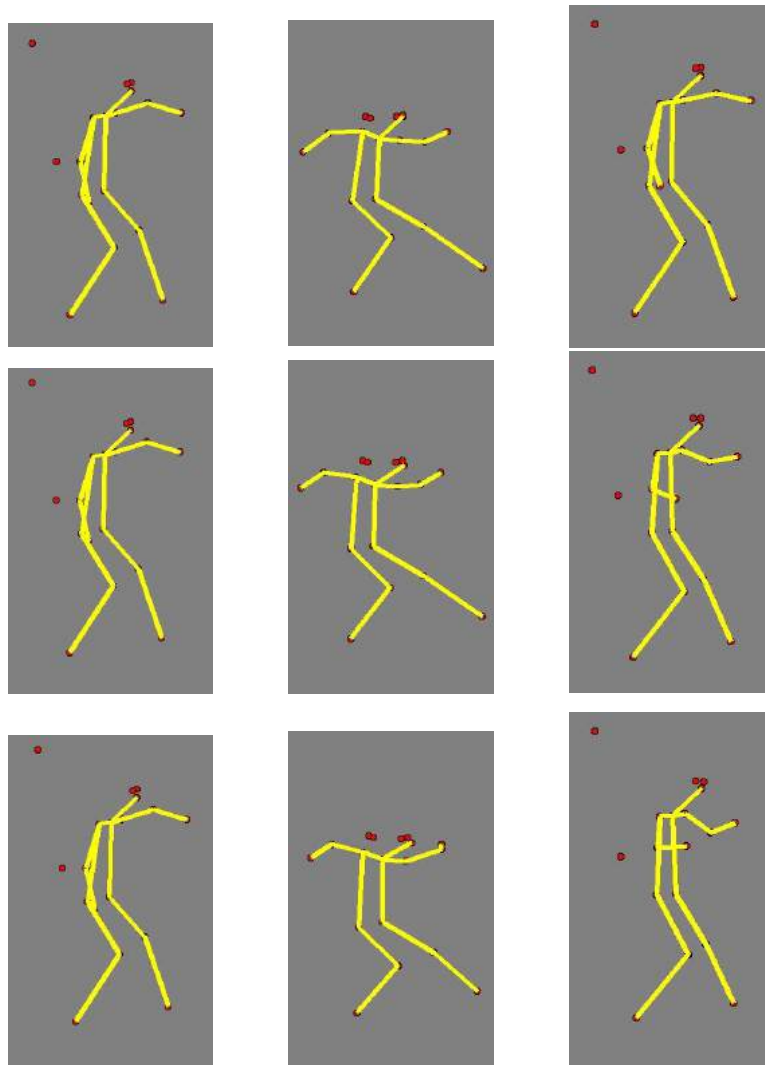


図 4.12 欠損フレーム予測結果

図 4.12 は、欠損フレームが発生した部分の ALIGNMENT 結果と、欠損フレームに対応する予測フレームを作成した結果である。左側の縦列が、入力系列の映像から OpenPose によって取得した身体座標をそのまま示した結果である。中心の縦列はモデル系列の身体座標を示した結果である。右側の縦列は提案手法によって予測した身体座標によって、入力系列の欠損フレームを予測した結果である。左右の縦列の画像は、中心のモデル系列と ALIGNMENT された結果であり、時間方向に伸縮している。各画像のうち、赤点が身体部位を示し、黄線が隣接した身体部位を結んだ線を表す。画像を比較すると、右側の縦列の画像は、より中心の縦列の画像に似た身体座標の構成を示している。そして時系列として不自然な身体座標の変化が発生していない。以上のことから提案手法による欠損フレーム

の予測は有効だといえる。

4.6 部位予測についての評価実験

部位予測についての精度評価実験と結果を示す。部位予測の精度評価実験は、69 フレーム 11 部位について予測しその結果を示す。真の値が判明しているフレームについて、提案手法に基づいてそのフレームの 1 部位が欠損していると仮定し、部位座標の予測を行い、部位座標の予測値を求める。予測値について、真の値との誤差の平均値を、提案手法の性能として示す。また実際に部位座標予測を行った結果の図を示す。予測する部位座標は首、両肩、両肘、両腰、両膝、両足とし、OpenPose によって検出した結果 (真値) と、提案手法によって予測した部位座標を画像として並べて示す。

表 4.8 部位予測の平均誤差

	x 座標	y 座標
平均誤差 [ピクセル]	16.032	23.549

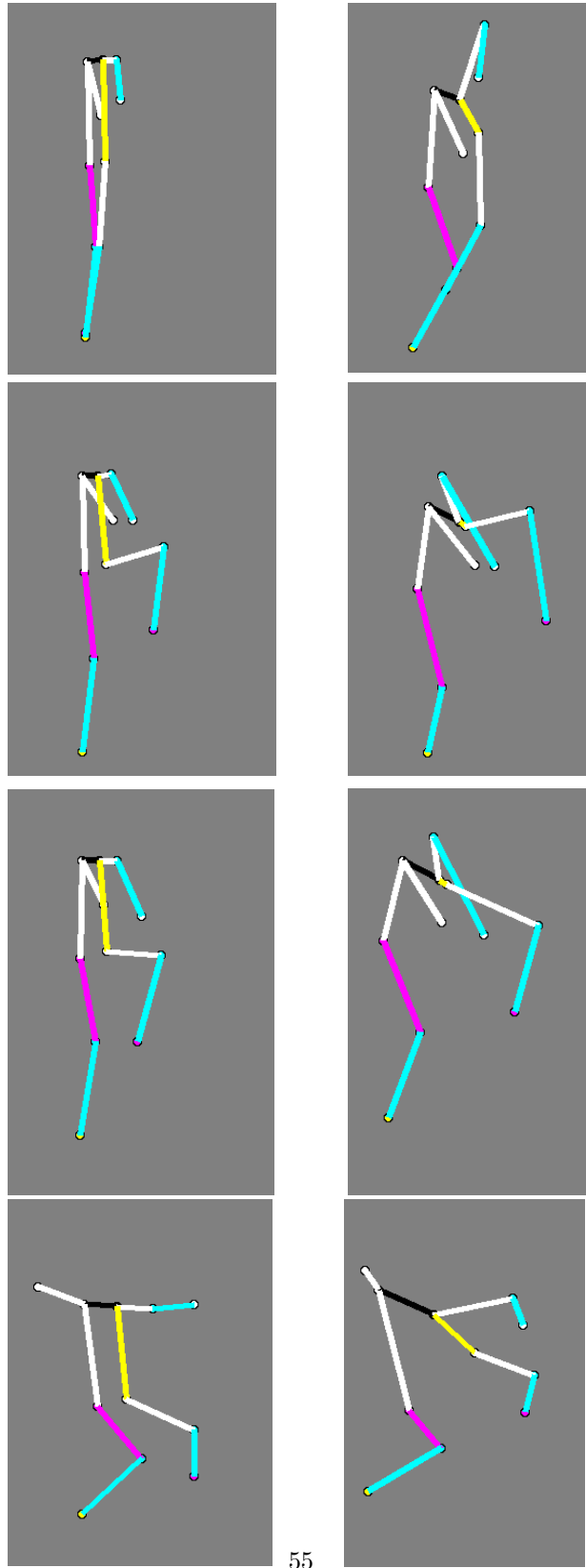


图 4.13 欠損部位予測結果

図 4.13 は，左に OpenPose によって検出した部位座標，右に提案手法によって予測した部位座標を示した図である．表 4.8 に示したように，部位予測の結果については誤差の値が大きく，良い結果を得ることができなかった．予測結果については特に右肩と右腰部分を結んだ黄線の差が大きく，正しい身体構造を示していない．この原因は，提案手法では DTW 距離を予測しているが，その精度が十分ではないことが考えられる．

4.7 むすび

本章では，提案手法の有効性を確認するための，評価実験の結果を示した．提案手法は概ね良好な結果を示し，有効性を確認することができた．

第 5 章

結論と今後の課題

5.1 結論

本研究では、スポーツ映像から取得した身体座標データに対して DTW によって、正確なアラインメントを行う手法を示した。

本手法の元となった Prefix and Suffix Invariant Dynamic Time Warping のアラインメント結果には、終点付近に誤りが発生することを実験により示した。原因として、DTW における累積距離を最小化するアルゴリズムが、始点と終点の検出精度を低下させている可能性を指摘した。その解決策として、DTW のアルゴリズムを変更する手法を提案した。具体的には、コスト行列の更新を平均距離の評価によって行うように変更した。その結果、従来手法の正しいアラインメントが行われた部分はそのままに、終点付近についてより正確にアラインメントが可能であることを確認できた。

5.2 今後の課題

本研究ではモデル系列の始点、終点を既知としているが、終点については未知でもアラインメントは可能だと考えられる。したがってモデル系列の終点が未知の場合の、DTW における新たなコスト行列更新アルゴリズムを検討する必要がある。

謝辞

本研究を行うにあたり，素晴らしい研究環境とアドバイスをくださった渡辺教授に感謝いたします。
また研究にあたり，研究活動から論文執筆に至るまで，様々な面で助けていただきました石川孝明氏に感謝いたします。

最後に，研究生生活を支えてくださった渡辺研究室の皆様，そして家族に感謝の意を表します。

参考文献

- [1] 神田弘行, “映像比較を用いた動きの指導について”, 日本コーチング学会第 22 回大会, Mar.2011
- [2] 高橋正樹, 盛岡, 池谷, 横澤, 今井, 三ッ峯, 三科, “進化するスポーツ中継向け放送技術 ～新映像表現と新しい考え方の解説放送「音声ガイド」～”, 放送研究と調査, 9月号臨時増刊, pp.94-103, Sep. 2017
- [3] Hawk-Eye Innovations Ltd. Hawk-Eye. <https://www.hawkeyeinnovations.com>
- [4] EON Sports, EON Icube, <https://www.eonreality.com/systems/eon-icube/>
- [5] Zhe Cao, T. Simon, S. Wei, and Y. Sheikh, “Realtime Multi-Person 2D Pose Estimation using Part Affinity Field,” In 2017 IEEE Conference on Computer Vision and Pattern Recognition, pp1302-1310, July 2017
- [6] 石田繁巳, 劉, 三村, 田頭, 福田, “Dynamic Time Warping を用いた路側設置マイクロフォンによる車両カウントシステムの設計と初期的評価”, IPSJ SIG Technical Report, Vol.2016-ITS-64 No 6, pp1-6, Mar. 2016
- [7] 尾風, 山田, 宮城, 市川, “NW アルゴリズム拡張による配列アラインメントの高速化についての検討”, 電気学会論文誌 D (産業応用部門誌), Vol.136 No.10, pp686-691, June 2016
- [8] 西野正彬, 鈴木, 梅谷, 平尾, 永田, “集合分割問題に基づく系列アラインメントのモデル化”, 自然言語処理, Vol.23 No.2, pp176-193, Mar. 2016
- [9] D.F.Silva, E.Keogh, “Prefix and Suffix Invariant Dynamic Time Warping,” IEEE 16th International Conference on Data Mining, pp1209-1214, Dec. 2016
- [10] 貞元太志, 黒木, 廣瀬, 沼, “野球中継におけるスコアテロップのレイアウト認識”, FIT2012, 第 3 分冊, pp123-pp124, Sep. 2012
- [11] 荒澤孔明, 服部, “野球映像の自動タギングのための Web 速報の更新時間の誤差補正アルゴリズム”, IEICE Technical Report, pp31-36, Nov. 2016
- [12] 望月貴裕, 藤井, 八木, 篠田, “投球の次ショットに重きを置いたシーンのシンボル列化による野球放送映像プレー種分類”, NHK 技研 R&D No.121, pp42-55, May 2010

研究業績

- [1] 横井真也, 浅見, 石川, 渡辺, “ボロノイ図に基づく3次元優勢領域によるパスコース評価について”, 映像情報メディア学会年次大会, 13D-3, Sep. 2017
- [2] 横井真也, 石川, 渡辺, “野球選手のスイング解析による打撃成績予測”, 電子情報通信学会総合大会, D-12-45, Mar. 2018
- [3] 横井真也, 石川, 渡辺, “スポーツ映像における身体座標情報を利用したイベント認識”, PCSJ/IMPS2018, P-2-10, Nov. 2018
- [4] 横井真也, 石川, 渡辺, “スポーツ映像における動作類似度を利用したイベント認識”, 映像情報メディア学会冬季大会, 13D-1, Dec. 2018

目次

2.1	OpenPose で検出された身体座標	4
2.2	位相差の存在する 2 系列に対する DTW 図	5
2.3	不要な要素が挿入された系列に対する DTW 図	7
3.1	OpenPose 検出成功フレームと失敗フレーム	11
3.2	映像フレーム 1	18
3.3	映像フレーム 2	18
3.4	欠損フレームと対応付け	19
3.5	欠損フレーム区間の検出	20
3.6	欠損フレーム区間の対応	21
3.7	OpenPose によって取得した動作の連続した 3 フレーム	22
3.8	カメラ性能と動作速度による動作部分の欠損	23
4.1	PSI-DTW のアラインメント成功例	28
4.2	PSI-DTW のアラインメント失敗例 1	31
4.3	正しいアラインメントとの比較	32
4.4	PSI-DTW のアラインメント失敗例 2	34
4.5	図 4.4 の正しいアラインメント	35
4.6	アルゴリズム 3 によるアラインメント	39
4.7	アルゴリズム 3 による終点付近のアラインメント	40
4.8	アルゴリズム 4 のアラインメント成功例	43
4.9	アルゴリズム 4 のアラインメント成功例 2	46
4.10	アルゴリズム 4 のアラインメント成功例 3	48
4.11	連続して OpenPose による誤検出が発生したフレーム	49
4.12	欠損フレーム予測結果	53
4.13	欠損部位予測結果	55