

スポーツ映像から抽出した時系列多次元データのアラインメント

横井 真也[†] 石川 孝明[‡] 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

〒169-0072 東京都新宿区大久保 3-14-9 早大シルマンホール 401

^{‡ † †} 早稲田大学国際情報通信研究センター 〒169-8555 東京都新宿区大久保 3-4-1

Email: [†] shinya-0storm@uri.waseda.jp [‡] takaxp@ieee.org ^{††} hiroschi.watanabe@waseda.jp

あらまし スポーツ映像からイベントを検出する手法として、Dynamic Time Warping(DTW)による身体座標データのアラインメントを利用した手法を提案する。DTWによって時系列データ間の距離を算出することで、モデルデータと入力データの類似度を測ることができる。しかし入力データに、検出したいイベント以外の動作が含まれると、正しい類似度を算出することができない。そこで入力データを分割し、入力データからイベント以外の動作を取り除き、より正確な類似度を算出する手法を示す。

キーワード スポーツ映像解析, Dynamic Time Warping, アラインメント

Sequence Alignment of Time-Series Multidimensional Data Extracted from Sports Video

Shinya Yokoi[†] Takaaki Ishikawa[‡] and Hiroshi Watanabe^{††}

^{† † †} Communication and Computer Engineering, Graduate School of Fundamental Science and Engineering, Waseda University 3-14-9 Ookubo, Shinjuku-ku, Tokyo, 169-0072 Japan

^{‡ † †} Global Information and Telecommunication Institute, Waseda University
3-4-1, Ookubo, Shinjuku-ku, Tokyo, 169-8555 Japan

Email: [†] shinya-0storm@uri.waseda.jp [‡] takaxp@ieee.org ^{††} hiroschi.watanabe@waseda.jp

Abstract Event recognition in sports video is proposed in this research. Motion similarity is calculated from time-series data of body part coordinates using Dynamic Time Warping (DTW). However, some time-series data includes unnecessary motions to perform matching using DTW. Thus, removing those motions are realized by separating DTW range.

Keyword Sports video analysis, Dynamic Time Warping, Alignment

1. 研究背景

競技者や視聴者に向けた、中継映像などのスポーツ映像データの解析、活用が進んでいる。映像からの特定イベントの抽出は、データベース作成など映像活用の下地となる作業で利用される。人間の動作などの時系列データに対し、モデルと同じ動作を判別する手法として、Dynamic Time Warping(DTW)による類似度の算出が存在する。DTWは二つの時系列データ間のアラインメントを行い、DTW距離を求める手法である。DTW距離が小さいほど二つの時系列データが類似していることを示す。しかしDTWでは時間方向の伸縮により、類似していないイベントについても距離を短く算出する、時系列データの長さにより計算量が増大するなどの問題も存在する。また入力データ側には動作の個人差などにより、モデルデータに含まれない動作が含ま

れることがあり、DTWによるマッチングに影響する。

本研究ではDTWを利用して野球中継映像からイベントを認識する手法を提案する。DTWを分割して計算することにより、計算量を削減する。また入力側のデータからモデルデータに含まれない部分を除去することで、より正確なDTWによるマッチングを実現する。

2. 関連研究

野球中継映像におけるイベント認識の1手法として、スコアテロップの情報利用が存在する。しかしスコアテロップは映像の提供者や、試合が実施された年代によって異なるという問題がある。複数種類のスコアテロップに対応した認識手法[1]も存在するが、すべてにおいて高い精度で認識を行うことは難しい。

本研究では、まずスポーツ映像から身体座標を取得

する．そしてモデルデータと入力データには座標の時系列データを使用し，DTWによってイベントの発生を認識する．映像から身体座標を取得するため，Caoらの手法[2]を利用する．

3. 提案手法

まず映像から身体座標データを取得する．そこから身体各部位の動作の時系列データを作成する．今回は，首，両肩，両肘，両腰，両膝，両足首の11部位のすべての組み合わせのベクトルの長さ，角度をモデルデータと入力データとして使用する．

入力データには，個人差や状況によって，モデルデータに含まれない動作の存在が想定される．入力データに存在するモデルデータにない動作として，入力データのイベントの長さが短いことによって発生する，入力データの始点と終点付近に存在するイベント以外の動作が存在する．これを範囲外動作と定義する．また動作の個人差によって発生する，モデルデータでは行われていない動作を個人依存動作と定義する．

入力データに存在する範囲外動作を，入力データが含むイベント部分の始点，終点を求めることによって除去する．

DTWを行って類似度を求める場合，モデルデータと入力データの長さが長くなるほど，計算量が多くなる．またモデルデータの特定の範囲に対して入力データの大部分がマッチングされるなどの間違ったマッチングが行われる可能性がある．そこで入力データの特定の範囲と類似するモデルデータを事前に探索し，モデルデータ，入力データをそれぞれ分割する事で，計算量を減らし，より正確なマッチングを行う．

さらに個人依存動作を除去するため，DTWによって時間方向の伸縮が行われた入力データから，相互相関によって個人依存動作を検出し，除去する．

最後に再度DTWを行うことで最終的な部分DTW距離を取得し，三つの部分DTW距離の合計をモデルデータと入力データのDTW距離とする．

3.1. Dynamic Time Warping

DTWは時系列データ間の各時刻における距離を合計して算出される，二つの時系列データ間の累積距離(以下DTW距離)を2データ間の距離と考え，それを最小化するように時間方向の伸縮を行う手法である．つまりDTW距離が小さいほど二つの時系列データが類似していることを表す．

今回の時系列データは，11部位のすべての組み合わせについての長さ r と角度 θ を持つ110次元のベクトル v の n フレームである．ベクトル v を以下に示す．

$$\begin{aligned} v_j &= \{R_j, \theta_j\} \\ R_j &= \{r_1, r_2, \dots, r_{55}\} \\ \theta_j &= \{\theta_1, \theta_2, \dots, \theta_{55}\} \end{aligned}$$

DTWではモデルデータと入力データの特定時刻間の距離を算出する距離関数を設定する必要がある．本研究においては二つの v の距離を1次元で出力する距離関数が必要である．そこで二つのベクトルのなす面積が，2ベクトルの距離を表していると考え，その合計が v 同士の距離であると考えた．モデル側のベクトル v^m と入力側のベクトル v^i の距離関数 D を以下に示す．

$$D(v^m, v^i) = \sum_{p=1}^{55} (r_p^m - r_p^i)^2 |q_p^m - \theta_p^i| \quad (1)$$

3.2. 時系列データの分割

DTWによって求めた類似度によって，時系列データを分割する．分割を行うため，モデルデータを解析して分割点を決定する．モデルデータの特定のフレーム f フレーム目から k フレームと，その前後の $f-1, f+1$ フレーム目から k フレームとの相互相関の平均を求める．平均が最も低いフレームを，前後のフレームと最も差異があるフレームとした．このフレームをモデルデータの前半部分，後半部分から一つずつ求め，分割点とした．

入力側の分割点を探索するため，まずモデル側の始点，終点から k フレームについて，入力データと最も類似する部分を探索する．そしてモデル側の分割点から k フレームについても同様に最も類似する部分を入力データから探索する．類似度はDTW距離を用い，DTW距離が最も小さい部分を類似範囲とし，分割点を決定する．

この結果によってモデルデータ，入力データを3分割する．

3.3. 相互相関による個人依存動作除去

入力データから個人依存動作を除去するために，モデルデータとの類似度を求める必要がある．そこでDTWによって伸縮を行ったモデルデータと入力データの相関を算出することで，個人依存動作を検出する．

DTWを行った後の入力データは，モデルデータと同じ時刻に，同じ動作が行われるように時間方向に伸縮されている．よってモデルデータと入力データの相関が低い部分には，モデルデータとマッチングする部分が無いにも関わらず，マッチングが行われた部分，つまり入力データに固有の動作が存在すると考えられる．この部分を除去することによって，同じイベントにおける個人差による動作の違いを除去することになり，モデルデータと入力データが同じイベントを表しているならば，その類似度を高めることになる．

部分ごとの相互相関結果によって，除去すべき部分に投票し，一定の値を超えた場合，入力データから該当部分を除去する．

4. 実験

4.1. 実験概要

提案手法を用いて求めた距離によって、イベントが判別できるか確認する。実験では右投手の投球時の映像 70 フレームをモデルデータとした。入力データはモデルデータと同じイベントを 19 本、モデルデータと違うイベントとして直立時、牽制時の映像を 10 本使用した。またモデルデータ、入力データに分割などを行わず、直接 DTW を行った結果も示す。

4.2. 実験結果

表 1 提案手法 DTW 実行結果

入力映像	投球	投球以外
DTW 距離最小値	1017.758	56471.448
DTW 距離最大値	240352.695	207546.453
DTW 距離平均値	27372.230	129658.962
DTW 距離中央値	15592.573	122577.302

表 2 直接 DTW 実行結果

入力映像	投球	投球以外
DTW 距離最小値	1025.264	56914.240
DTW 距離最大値	65279.103	141026.564
DTW 距離平均値	22152.259	86346.178
DTW 距離中央値	16814.222	83209.115

DTW 距離を求めた結果を表 1 に示す。表 1 の結果では投球と投球以外の DTW 距離には十分な差が現れている。しかし入力データの一つに、モデルデータと異なる複数フレームに渡る動作が含まれるため、正しくアラインメントが行われず、結果に悪影響を与えている。この入力データを取り除いた場合の結果を下に示す。

表 3 入力データから一つ取り除いた DTW 結果

入力映像	投球	投球以外
DTW 距離最小値	1017.758	56471.448
DTW 距離最大値	48096.471	207546.453
DTW 距離平均値	15539.982	129658.962
DTW 距離中央値	11693.370	122577.302

表 4 入力データから一つ取り除いた直接 DTW 結果

入力映像	投球	投球以外
DTW 距離最小値	1031.034	56914.240
DTW 距離最大値	47563.821	141026.564
DTW 距離平均値	20021.390	86346.178
DTW 距離中央値	16891.880	83209.115

表 3 の結果より、投球イベントを入力した場合に、投球以外のイベントを入力した場合の DTW 距離の最小値を超えた入力データは一つだけであり、DTW によってイベントの判別が可能だといえる。また表 3 と表 4 を比較すると、直接 DTW を行う場合に比べて、提案

手法は同じイベントとの DTW 距離をより小さく、違うイベントとの DTW 距離をより大きく算出している。これは範囲外動作と個人依存動作を除去することによって、モデルデータと入力データに含まれるイベント同士だけの距離を算出することになり、より正確なモデルデータと入力データのイベントの距離が求められた結果だと考えられる。

5. まとめ

スポーツ映像からイベント検出を行うため、部位座標時系列データを用いて DTW を行う手法を示した。また入力データに含まれる範囲外動作の除去を行うため、モデルデータの始点、終点付近と類似する箇所を探索し、入力データを分割する手法と、個人依存動作を相互相関によって検出する手法を示した。しかしすべての同一動作に対して、正しくアラインメントを行うことはできず、余計な動作の検出、削除をより正確に行うことが必要だと考えられる。

文 献

- [1] 貞元, 黒木, 廣瀬, 沼, "野球中継映像におけるスコアテロップのレイアウト認識", FIT, No.3, pp123-124, Sep.2012
- [2] Z. Cao, T. Simon, S. Wei, and Y. Sheikh, "Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields", In CVPR, pp7291-7299, Jun, 2017.