

スポーツ映像から取得した骨格座標データ に対するアラインメント

Alignment for Skeleton Coordinates Data Obtained from Sports Video

横井真也*1 石川孝明*2 渡辺裕*1
Shinya Yokoi Takaaki Ishikawa Hiroshi Watanabe

*1 早稲田大学大学院基幹理工学研究科

*1 Graduate School of Fundamental Science and Engineering

*2 早稲田大学国際情報通信センター

*2 Global Information and Telecommunication Institute,
Waseda University

1. まえがき

スポーツ競技者の映像利用において、動作の比較は基本的な映像活用である。複数のデータ系列について、データを対応づけることをアラインメントと呼ぶ。スポーツ映像でのアラインメントは、動作の各タイミングについて対応付けを行うことである。動作はそれを行う個人によって速度や形に違いがある。時間方向の伸縮を行いつつアラインメントを行う手法として Dynamic Time Warping (DTW) が存在する。しかし DTW は検討対象外のデータ混入に弱く、正確なアラインメントが行われないという問題がある。

本研究では、スポーツ映像から抽出した骨格座標データについて、DTW を利用してより正確なアラインメントを行う手法を提示する。入力系列から不要な部分を除去することで、共通部分動作だけのアラインメントを実現する。

2. Dynamic Time Warping

DTW は、距離関数 D に基づいて、2 系列間の対応付けを行い、各対応の距離の合計を最小化するようなアラインメントを行う手法である。

本研究では Cao らの手法[1]を利用して取得した骨格座標データに対してアラインメントを行う。利用する部位は首、両肩、両肘、両腰、両膝、両足の 11 部位とし、各部位間の長さや角度を時系列のデータとして使用した。時系列骨格座標データ V と距離関数 D を以下に示す。

$$v_j = \{R_j, \theta_j\}$$

$$R_j = \{r_1, r_2, \dots, r_{55}\}$$

$$\theta_j = \{\theta_1, \theta_2, \dots, \theta_{55}\}$$

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

3. 不要部分除去

DTW を行う系列の内、比較元となる系列をモデル系列、比較対象の系列を入力系列とする。不要部分とは、比較対象の動作の前後に存在する、対象外動作と、入力系列の動作を行った個人の特徴である、入力系列だけに存在する個人依存動作である。この二つの動作部分を検出し、入力系列から除去した後 DTW を行う。

3.1 対象外動作の検出

対象外動作は、比較対象の動作の前後に存在するため、比較対象動作の始点と終点を検出することで除去すること

ができる。そこでモデル動作の始点から n フレーム、終点から n フレームに対応する部分を、DTW によって DTW 距離を算出することで検出する。DTW 距離が小さいほど二つの系列は類似しているといえる。したがってモデル系列の始点及び終点付近のデータに対して、DTW 距離が最小となる入力系列側の時間を求め、その地点を動作の始点及び終点とする。

3.2 個人依存動作の検出

個人依存動作を除去するためには、モデル系列と入力系列の共通部分だけを抽出すれば良い。複数の系列の最長の共通部分を、最長共通部分列(Longest Common Subsequence, LCS)と呼ぶ。LCS を抽出するためには、対象とするデータ系列を完全一致が可能な記号列に変換する必要がある。骨格座標データを完全一致可能な記号列へ変換することは困難なため、正規化可能な類似度を算出できる記号列に変換し、閾値を設定することでデータの一致を判定する。まずモデル系列の各角度データの最小値から最大値の範囲を 4 分割し、各範囲に記号を与える。そしてそれに基づき、入力系列の各角度データを記号列化する。最後に入力系列の各データについてモデル系列と一致するものが存在するかを判定し、共通部分を抽出する。

4. 実験

提案手法によるアラインメントの結果と、提案手法にある処理を行わず直接 DTW を行った結果を表 1 に示す。

右投手の投球データ 1 本をモデルデータとし、13 本の投球データとのアラインメントを行い映像に戻した後、目視によって正誤を判定した。

表 1 アラインメント結果

	提案手法	直接 DTW
正解フレーム数	926	970
不正解フレーム数	46	194
正解率	0.953	0.833

5. むすび

スポーツ映像から得た骨格座標データのアラインメントをより正確に行う手法を示した。

文 献

[1] Z. Cao, T. Simon, S. Wei, Y. Sheikh, "Realtime Multi-Person 2D Pose Estimation using Part Affinity Field", In CVPR, pp7291-7299, Jun, 2017.