

3次元姿勢推定を用いた反復練習動画の効率的な視聴システム

平栗 義久 井尻 敬

芝浦工業大学 〒135-8548 東京都江東区豊洲 3-7-5

E-mail: {a117094, ijiri}@shibaura-it.ac.jp

あらまし スポーツやジャグリングの反復練習において、フォームや姿勢を詳細に確認するため、反復の様子を動画撮影しそれを見返すことが広く行われている。しかし、一般的に、反復練習はインターバルを含むため、反復練習動画の視聴は冗長なものとなる。そこで本研究では、反復練習動画の効率的な視聴の実現を目的とし、動画から推定される姿勢情報を利用することで興味のある動作のみを自動検出できるシステムを提案する。提案システムでは、動作の検出に3次元的な姿勢情報を利用するため、異なる角度から撮影された複数動画からも類似動作の検出が可能である。提案システムの有用性を示すため、ジャグリングおよびバットスイングの反復練習を複数視点より撮影した動画から特定動作を検出する実験を実施した。結果、3次元姿勢を正しく推定できれば、高精度で関心のある動作を検出できることを確認した。

キーワード スポーツ練習, 反復練習, 動画像処理, 姿勢推定.

1. はじめに

スポーツやジャグリングでは、正しいフォームを習得するため、同じ動作を繰り返す反復練習が行われる。この反復練習において、自身のフォームを詳細にかつ客観的に観察するため、練習の様子を撮影し動画を見返す練習法が広く行われている。しかし、反復練習を撮影した動画には関心のある動作だけでなく、動作間のインターバルも含まれている。そのため、動画の視聴時に、観察したい部分をスライダーバーにより検索するという冗長な作業が必要となる。さらに、自身と他者のフォームを比較する際や、観察したい動画が異なる視点から撮影されている場合など、対象動画が複数ある場合、この検索作業はより冗長なものとなる。

スポーツの練習支援のため、練習者の姿勢を解析してフィードバックを返す研究がなされている。例えば、Zoeらは、バレエ練習において、見本姿勢とユーザ姿勢の違いをリアルタイムに表示するシステムを提案した[1]。また、Yamaokaらは、フライングディスクを投げる姿勢を解析し、評価をフィードバックするシステムを提案した[2]。ただし、これらは特定スポーツに特化した設計がなされており、多様なスポーツの反復練習動画へ適用することは難しい。増田らは、反復練習動画の効率的な視聴を目的とし、動画から特定動作を抽出し提示するシステムを提案した[3]。しかし、この研究では、2次元姿勢情報を用いて動作検出を行っているため、異なる視点の複数動画を扱えないという課題がある。

そこで本研究では、異なる視点の反復練習動画の効率的な視聴環境の実現を目的とし、動画から関心のある動作のみを自動検出できるシステムを提案する。既存研究[3]と異なり、特定動作検出に3次元姿勢情報を用いることで、異なる視点から撮影された複数の反復

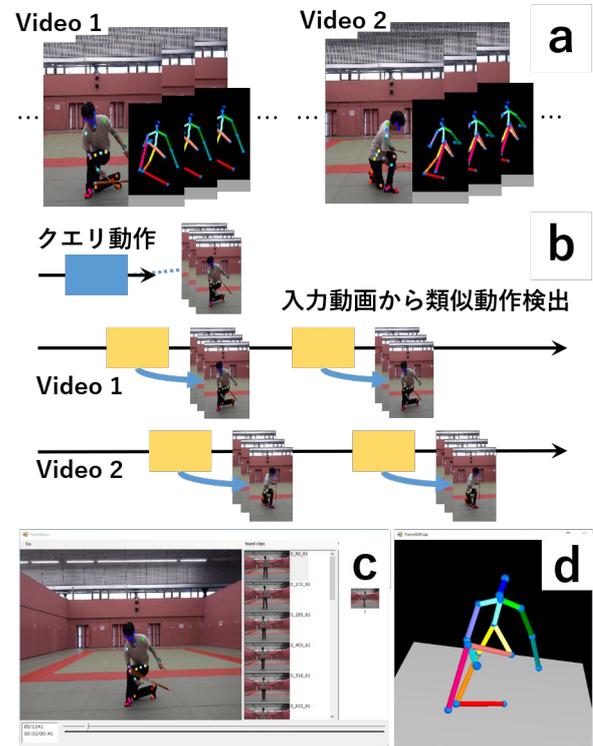


図 1.3 3次元情報を用いた反復練習動画視聴システム

練習動画に対しても動作検出が可能となる。

2. 反復練習動画視聴システム

2.1. ユーザインタフェース

提案システムの入力は、複数のスポーツ反復練習動画と、動画内の練習者の2次元姿勢情報、及び、3次元姿勢情報である(図 1a)。動画内には主に反復練習を行っている人物一人が映っているものとする。練習

者の2次元姿勢情報はOpenPose[4]により取得する。さらに、OpenPoseにより推定した2次元姿勢情報にRepNet[5]を適用することで3次元姿勢情報も取得する。

複数動画および姿勢情報の入力後、ユーザは、動画からひとつの反復動作の開始フレーム・終了フレームを指定する(以下、クエリ動作)。すると入力された全ての動画からクエリ動作と類似の動作を行った部分(以下、動作クリップ)が自動的に検出される。この時、ユーザは、クエリ動作と動画内の局所領域との相違度グラフを参照しながら閾値の調整が可能である(後述する)。検出された動作クリップは、サムネイルとしてユーザに提示され、ユーザはサムネイルをクリックすることで、各動作クリップを再生できる(図1b)。また、各クリップの再生に合わせて3次元姿勢の観察も行える(図1cd)。

2.2. 反復動作の検出

ユーザが指定したクエリ動作と類似する動作クリップを検出する方法を説明する。複数の入力動画、動画の各フレームにおける3次元姿勢情報、クエリ動作の開始・終了フレーム番号は与えられているものとする。また、姿勢情報は、16個のランドマークの3次元座標により表現される。

今、クエリ動作のフレーム数を L とする。我々は、検索対象動画の s フレームから $s+L-1$ フレームまでのクリップと、クエリ動作との相違度を以下の通り計算する、

$$S = \frac{1}{16L} \sum_{i=0}^{L-1} \sum_{k=0}^{15} \|\{R_{pq}(\mathbf{p}_k^{s+i} - \mathbf{c}^{s+i}) + \mathbf{c}^i\} - \mathbf{q}_k^i\|. \quad (1)$$

ここで、 $\mathbf{p}_k^i \in R^3$ は、検索対象動画における i 番目のフレーム、 k 番目の関節の3次元座標、 \mathbf{c}^i は i 番目のフレームの3次元関節座標の重心である。また、 $\mathbf{q}_k^i \in R^3$ はクエリ動作における、 i 番目のフレーム、 k 番目の関節の3次元座標、 \mathbf{c}^i は i 番目のフレームの3次元関節座標の重心である。 R_{pq} については後述する。提案システムは、入力動画における長さ L フレームの局所領域に対して相違度の計算を行い、閾値以下かつ前後 L フレームにおいて局所最小となる領域を動作クリップとして検出する。

我々は、3次元姿勢推定にRepNetを利用する[5]。このRepNetで推定された3次元関節座標は、姿勢ごとに異なる回転行列がかかるため、相違度計算の際に姿勢間の回転を補正する必要がある。そこで本研究では、ShapeMatching法[6]を利用する。具体的には、ふたつの間接座標点群 $\mathbf{p}_k^i, \mathbf{q}_k^i, (i=0, 1, \dots, L-1, k=0, 1, \dots, 15)$ が与えられると、点群 \mathbf{p}_k^i が点群 \mathbf{q}_k^i になるべく一致するような回転行列 R_{pq} を求め、これを相違度の計算(1)に利

用する。

動作クリップ検出に必要な閾値は、対象となるクエリ動作によって変化させる必要がある。提案システムでは、ユーザの意図通りの検索を行えるよう、閾値を自動決定するのではなく、グラフを用いた調整ツールを提供する(図2)。クエリ動作が入力されると、システムは、入力動画の任意の局所領域に対して相違度を計算し、計算結果をグラフ状に提示する。反復動作のある動画が入力されている場合、図2の通り相違度が繰り返し小さくなるような相違度グラフが得られる。ユーザはこのグラフを見ながら、閾値(図中の黄色い直線)を調整する。

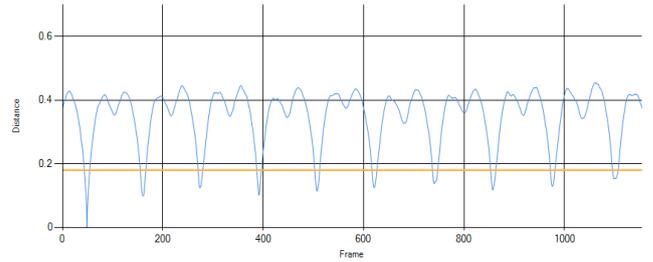


図2. 閾値調整のための相違度グラフツール。

2.3.3 3次元姿勢の可視化機能

複数動作比較の際に、より細かな違いを確認するために、提案システムは、3次元姿勢可視化機能を有する。動作クリップの検出後、複数の動作クリップを選択し3次元姿勢比較ボタンを押すと、選択した動作クリップの3次元姿勢が重ね合わせて可視化される(図3)。これにより、ユーザは異なる視点から撮影された動画でも、その3次元的な姿勢の違いを比較することが可能となる。

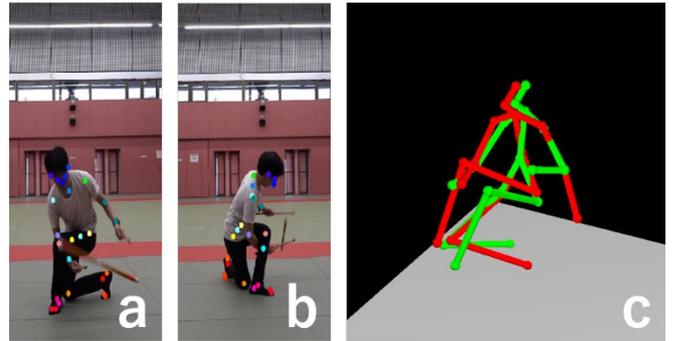


図3. 動画像(a), (b)の3次元姿勢が重ね合わせて可視化された例(c)。

3. 結果と考察

3.1. 実験手順

提案システムの検出精度を確認するため評価実験

を実施した。まず、ジャグリングの反復練習を4視点から撮影した動画を2種類、野球のバットスイング動画を3視点から撮影した動画を1種類用意した。これらを、ジャグリング動画群A、ジャグリング動画群B、スイング動画群と呼ぶこととする。

用意した各動画に対して、すべての反復動作を手作業で抽出した。ジャグリング動画群Aからは、しゃがんでから立ち上がるまでを反復動作として抽出した。ジャグリング動画群Bからは、右足を上げる動作からアイドリングに戻るまでを反復動作として抽出した。スイング動画群からは、テークバックからフォロースルーまでをクエリ動作として抽出した。続いて、各動画の最初の反復動作をクエリ動作とし、同じ反復動作が含まれる全動画に対して2次元姿勢情報による動作検出と3次元姿勢情報による動作検出を行った。最後に、手作業で抽出した動作を正解とし、検出結果の精度評価を行った。ここで、精度評価には、正解領域と検出領域の Intersection over Union(IoU)を利用した。

3.2. 検出精度

ジャグリング動画群Aに対して、動作検出を行った結果を表1、表2に示す。このジャグリング動画は、正面・右側面・背面・左側面という異なる4視点から撮影された4本の動画を含む。表の縦軸はクエリ動作を指定した動画を、横軸は動作検出を行った動画を示す。

2次元姿勢情報を用いた動作検出(表1)では、異なる視点の動画間で、高い検出精度が得られていない(正面をクエリとし、背面動画から検出した場合など)。一方、3次元姿勢情報を用いた動作検出では、異なる視点の動画間でもIoUの比較的高い検出精度が観察された。このことから、提案システムは異なる視点の動画間でも精度よく動作を検出できると考えられる。

ただし、ジャグリング動画群Bやバットスイング動画群において3次元姿勢情報を用いて動作検出を行った結果、3次元姿勢推定エラーに起因する低い検出精度が観察された。図4の通り、体の一部が遮蔽されるような姿勢をとると、OpenPoseの2次元姿勢推定に誤りが生じ、結果として不正確な3次元情報が推定される。提案システムの精度は、3次元姿勢推定結果に強く依存するため、より高精度な姿勢推定手法の導入や、姿勢推定結果の信頼度を利用した検出結果の調整などが今後の課題である。

3.3. 類似動作の検出例

提案システムの有用性を示すため、ジャグリング動画に対して、類似動作検出を行った結果を図5に示す。

この例では、正面から撮影された練習動画からクエリ動作を抽出した(図5a)。このクエリ動作との類似動作を、異なる視点から撮影された動画から検索した例が図中の(b-e)である。この通り、提案システムは、自動的に興味関心のある動作を検索・閲覧できるため、効率的な反復練習動画の視聴が可能となる。

4. まとめ

本研究では、反復練習動画の効率的な視聴を目的とし、3次元姿勢情報を活用し、動画から関心のある動作を検出する手法とを提案した。提案システムの検出精度を確認するため、異なる視点から撮影された複数の反復練習動画に対して、2次元姿勢情報および3次元姿勢情報を用いた動作検出を実施した。結果、正しく推定できた3次元姿勢情報を用いれば、提案システムにより異なる視点から撮影された動画からでも高精度に類似動作を検出できることを確認した。

今後、高精度な3次元姿勢推定法を適用することで提案システムの高精度化を行いたい。また、提案システムを多様なスポーツ動作に適用しその精度を検証することは重要な将来課題である。さらに、クエリフレームや閾値の自動指定も今後取り組みたい将来課題である。

表1. ジャグリング動画群Aからの2次元姿勢情報によるIoU.

| | | 検出の対象動画 | | | |
|-------|-----|---------|-------|-------|-------|
| | | 正面 | 右側面 | 背面 | 左側面 |
| クエリ動作 | 正面 | 0.944 | 0.750 | 0.379 | 0.949 |
| | 右側面 | 0.600 | 0.865 | 0.918 | 0.552 |
| | 背面 | 0.486 | 0.798 | 0.878 | 0.519 |
| | 左側面 | 0.885 | 0.560 | 0.534 | 0.906 |

表2. ジャグリング動画群Aからの3次元姿勢情報によるIoU.

| | | 検出の対象動画 | | | |
|-------|-----|---------|-------|-------|-------|
| | | 正面 | 右側面 | 背面 | 左側面 |
| クエリ動作 | 正面 | 0.948 | 0.847 | 0.940 | 0.943 |
| | 右側面 | 0.951 | 0.863 | 0.918 | 0.941 |
| | 背面 | 0.753 | 0.789 | 0.886 | 0.904 |
| | 左側面 | 0.901 | 0.711 | 0.911 | 0.918 |

謝辞

本研究は、日本学術振興会科学研究費 基盤C(18K11606)の支援を受けて行われたものである。

文 献

- [1] Z. Marquardt, J. Beria, N. Em, I. Paiva, S. Kox. A kinect interface for ballet dancers, *Proc. CHI '12 Extended Abstracts on Human Factors in Computing Systems*, pp.1619-1624, May. 2012.
- [2] K. Yamaoka, M. Uehara, T. Shima, Y. Tamura. Feedback of flying disc throw with Kinect and its evaluation, *Proc. Procedia Computer Science*, 22, pp.912-920, Sep. 2013.
- [3] 増田毅, 永見智行, 井尻敬. 姿勢追跡技術を活用したスポーツ反復練習動画の視聴システム, 研究報告ヒューマンコンピュータインタラクション (HCI), 2020-HCI-186, 36, pp.1-3, Jan. 2020.
- [4] Z. Cao, G. Hidalgo, T. Simon, S. Wei, Y. Sheikh. Open-Pose: realtime multi-person 2D pose estimation using Part Affinity Fields, *Proc. IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019.
- [5] B. Wandt, B. Rosenhahn. Repnet: Weakly supervised training of an adversarial reprojection network for 3d human pose estimation, *Proc. Computer Vision and Pattern Recognition (CVPR)*, pp.7782-7791, May. 2018.
- [6] M. Muller, B. Heidelberger, M. Teschner, M. Gross. Meshless deformations based on shapematching, *Proc. Association for Computing Machinery*, 24, 3, pp. 471-478, Jul. 2005.



図4. ジャグリング動画における3次元姿勢推定エラーの例.

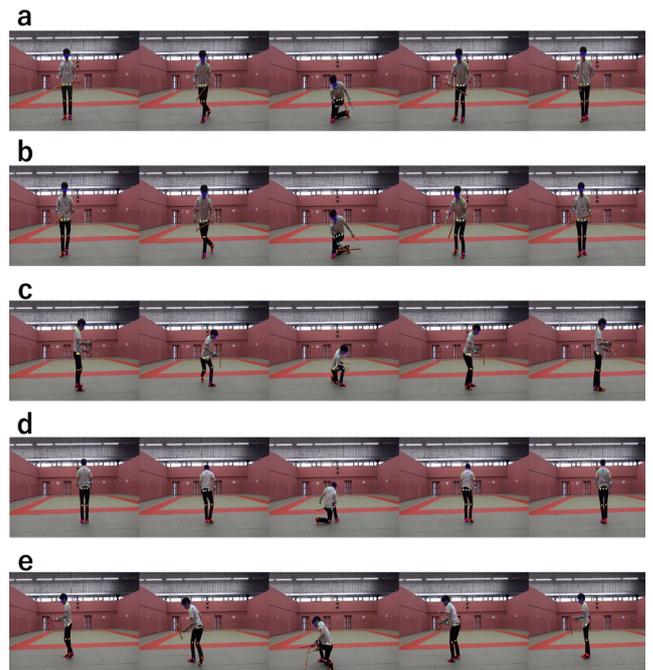


図5. 提案システムによる類似動作検出の例.