

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

増田 賀^{1,a)} 永見 智行² 井尻 敬¹

概要：スポーツの練習において、自身の動きを客観的に知るために練習風景を撮影し視聴することが広く行われている。しかし、スポーツの反復練習を撮影した動画はプレイ以外のインターバルも多く含むため、その視聴作業は冗長なものになる。そこで我々は、スポーツ反復練習動画の効率的な視聴・解析を目的とし、深層学習による姿勢追跡法を活用することで冗長な動画から反復練習部分のみを抽出し提示できるシステムを提案する。また、特定姿勢が表れるタイミングを考慮し複数の反復部分を重ね合わせるツールや、重要なランドマークの軌跡を提示するツールも提供する。提案システムの精度確認のため、野球・サッカー・バドミントンといった3種の異なるスポーツ練習動画に対し反復動作検出を行った。結果、それぞれの動画において閾値を調整すれば正しい検出を行えることを確認した。また、提案システムの有用性を示すため、提案手法により得られる複数動作の比較例やランドマーク軌跡の解析例を紹介する。

1. はじめに

スポーツの練習において、正しい姿勢の習得や筋力増強のため、そのスポーツ固有の動作を繰り返す反復練習が広く行なわれている。効率的に正しい姿勢を習得するには客観的に自身の動きを観察することが重要であり、この目的のため練習の様子を動画で撮影し観察することも行われている。しかし、スポーツの練習風景を撮影した動画は観察したい動作以外のインターバルも含むため、その観察作業は冗長なものになることが多い。例えば、野球の投球練習を撮影した動画では、投球動作自体は数秒で完結するのに対し、投球後の返球や投球前の準備動作といったインターバルに数十秒の時間がかかる。このような動画の確認には、数秒の投球動作部分の観察とインターバル部分の早送りを繰り返す必要がある。

スポーツの練習支援のため高速度カメラを用いた動画解析の研究が多くなされている。例えば、野球の投球の様子を高速度動画にて撮影し回転解析を行うもの[1,2]、スイング動作を撮影してその精度を評価するもの[3]などが挙げられる。しかし、これらの手法は、事前にインターバル部分を除去し対象動作のみが切り出された動画を入力に仮定している。また、動画の効率的な視聴のための動画再生システムの研究が行われている。動画全体をサムネイルにより一覧表示する手法[4]や、動画内の色・動き情報を利用

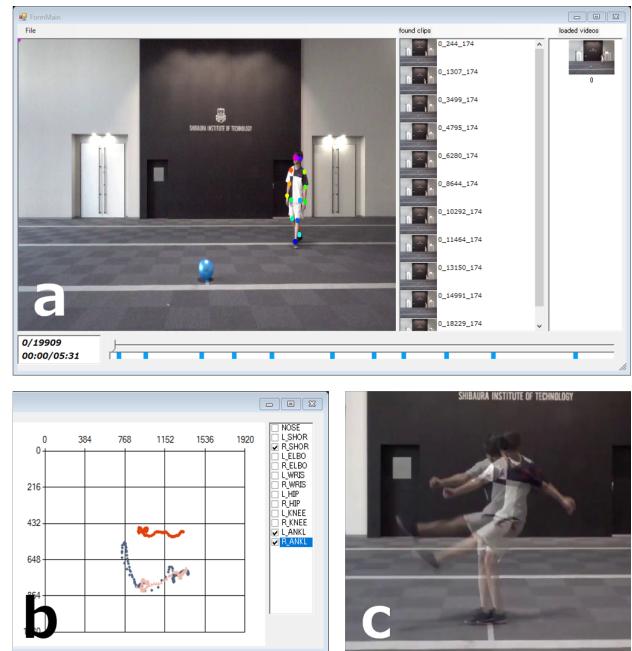


図 1 提案システムによる反復動作視聴の様子。提案システムでは、反復動作部分のみの抽出(a), 身体部位の軌跡を可視化(b), 複数動作を重ね合わせた出力(c)などが行える。

し類似シーンを検出する手法[5], 動画内の情報を利用し不要なシーンを早送りする高速再生法[6-8]などが提案されている。しかし、これらの手法はスポーツ練習動画の視聴に特化したものではない。

本研究では、スポーツ反復練習動画を効率的に視聴・解

¹ 芝浦工業大学

² 北里大学

a) tsuyoshi.masuda244@gmail.com

析できるシステムの実現を目指す。目標とするシステムに必要な機能は、(i) 反復動作の検出、(ii) 動作の比較、(iii) 動作の解析の 3 種である。反復動作の自動検出機能 (i) を実現することで、ユーザはインターバル部分を無視し、関心のある動作のみを観聽できる。また、検出した複数の動作を比較する機能を提供すれば、練習における動きの変化や選手間の差を可視化できる。さらに、特定動作の解析機能 (iii) を提供すれば、各選手の動きを数値データとして出力できる。

本研究では、前述の 3 機能を持った動画の観聽・解析システムを提案する（図 1）。提案システムの入力は、スポーツの反復練習動画と動画内の競技者の姿勢情報である。姿勢情報は、既存の深層学習による姿勢追跡法 [9, 10] により前計算しておくものとする。提案システムにおいて、ユーザが特定動作の開始・終了フレームを指定すると、システムは自動的に動画全体から類似動作を検出しユーザに提示する（図 1a 右）。ユーザは、ある動作におけるランドマーク（目・肩・肘・足など）の軌跡をグラフとして提示すること（図 1b）や、複数の動作を重ね合わせた動画を出力すること（図 1c）ができる。

提案システムの動作検出法の精度を確認するため、3 種の異なるスポーツの反復練習動画を用意し精度検証を行った。結果、今回用意した練習動画に対しては、閾値を調整すればすべての反復動作を正しく検出できることを確認した。また、提案システムの有用性を示すため、複数動作の比較やランドマーク軌跡の可視化を行った例を紹介する。

2. 関連研究

2.1 効率的な動画観聽システム

動画の内容を効率的に把握するための動画観聽法に関する研究がなされている。Boreczky ら [4] は、色情報を用いて動画を複数ショットに分割し、各ショットのキーフレームを一覧表示することで動画の要約を行った。栗原 [6] は、字幕の有無で再生速度を変化させる動画高速再生システムを提案した。また、Schoeffmann ら [5] は、動画内の局所領域の色情報や、動画全体のオプティカルフローに基づく特徴ベクトルを利用し、動画から類似シーンを自動検出してユーザへ提示するシステムを提案した。これらの手法は一般的な動画を対象とするものである。一方我々は、対象をスポーツ動画に限定し人物の姿勢情報を活用することで、堅固な動作検出やランドマーク解析を実現する。

特定のシーンを検出することでスポーツ放送動画を要約する試みが行われている。河村ら [7] はコートが映ったシーンを特定することでテニスなどラケットスポーツの放送動画の要約を行い、板摺ら [8] も同様にコートを利用してバレーボール動画の要約を行った。我々が対象とする練習動画では、選手を中心撮影することが多いため、コート等の背景情報ではなく選手の姿勢情報を利用し関心

のあるシーンを検出する。

2.2 深層学習による姿勢追跡

人物を撮影した RGB 画像からその人物の姿勢を推定する技術には大きな需要があり多くの研究が発表されている。Cao ら [9] は、人物のランドマーク位置の推定と並行して、隣接ランドマーク間のつながりを表す Part Affinity Fields を推定することで、多人数の姿勢推定を高速に行える手法を実現した。Papandreou らは、多人数が撮影された画像から姿勢とマスクを同時に output できる PersonLab を提案した [11]。Dan らは、PersonLab を応用し、TensorFlow.js を利用することで Web ブラウザ上で高速に姿勢推定を行なえる PoseNet を実装し、利用しやすい形で公開した [7]。本研究では、これらの姿勢追跡技術を活用して、反復動作検出を実現する。

3. スポーツ反復練習動画の効率的な観聽システム

3.1 提案システムの概要

本研究では、冗長なスポーツ反復練習動画を効率的に観聽・解析できる環境の実現を目的とし、人物の姿勢情報を活用することで反復動作の検出・動作の比較・ランドマーク位置の解析を行える観聽システムを提案する。提案システムの入力はスポーツ反復練習動画とその動画から抽出した姿勢情報である。ここで、動画内には主に競技者のみが映っているものとし、姿勢情報は OpenPose [9] または PoseNet [10] により抽出することとする。

動画の入力後、ユーザは、関心のある特定動作（以下、クエリ動作）の開始・終了フレーム番号を指定する。するとシステムは自動的に動画像全体から似た動作を行った部分（以下、動作クリップ）をすべて検出しユーザに提示する。その後ユーザは、各動作クリップを観聽するだけでなく、複数動作クリップを重ねた比較動画の作成や、ある動作クリップにおけるランドマークの軌跡をグラフ化・数値出力することが可能である。例えば、サッカーのシュート練習に提案システムを適用すると、シュート動作のみを抽出し観聽したり、気になる複数のシュート動作を重ね合わせて比較したり、蹴り足の軌跡をグラフ化したりすることが可能である。

3.2 ユーザインターフェース

提案システムのスクリーンショットを図 2 に示す。提案システムは、動画表示ウインドウ (a)，抽出した動作クリップサムネイル表示部 (b)，読み込み動画サムネイル表示部 (c)，操作用ダイアログ (d)，相違度グラフ (e)，ランドマーク軌跡グラフ (f) より構成される。システムを起動し動画と姿勢情報を入力すると動画像が表示され、ユーザはシークバーを動かすことで動画を観聽できる。次に、クエリ動

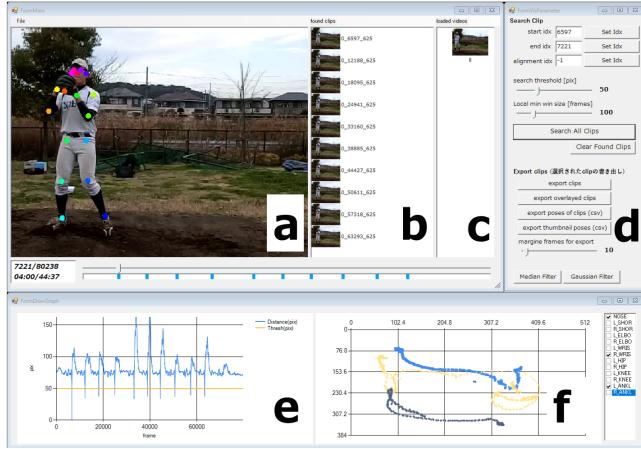


図 2 提案システムのスクリーンショット。

作の開始・終了フレームをダイアログより指定し検出開始ボタンを押すと、システムは動画全体から類似動作を検出し、そのサムネイルを提示する。この時、各動作クリップに対応するフレームがシークバー上で青色に強調表示される(図 3a)。詳細は後述するが、この検出処理では、クエリ動作と動画内の任意の動作との相違度が計算され、これが閾値以下となる部分が特定される。相違度と閾値は相違度グラフ(図 3e, 図 3b)に提示され、必要に応じてユーザは閾値を修正できる。

動作クリップの検出後、ウインドウに表示されたサムネイルをクリックすることで各クリップを視聴できる。また、複数の動作クリップを選択し、ダイアログの動画出力ボタンを押すと、選択したクリップを重ね合わせた動画を作成できる。この時、アラインメントフレームを追加指定すると、特定姿勢のタイミングを合わせた重ね合わせ動画作成が可能である。このアラインメントフレーム指定により、例えば投球動作における足の着地のタイミングや、シュート動作におけるインパクトのタイミングを合わせた比較動画生成が可能になる。

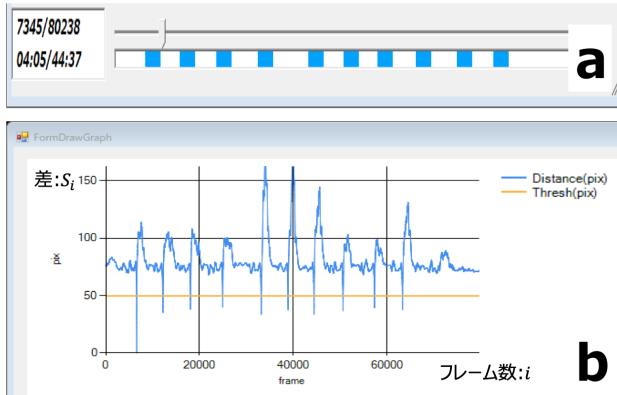


図 3 シークバーの拡大図 (a) と、相違度グラフ (b)。

提案システムは、動作解析の一例として、ランドマーク位置の軌跡を提示する機能を提供する。検出された動作ク

リップをひとつ選択し、ダイアログから観察したいランドマークを指定すると、そのランドマークの軌跡が提示される(図 4)。他のソフトウェアで詳細に解析するため、この軌跡情報をテキストデータとして出力することも可能である。

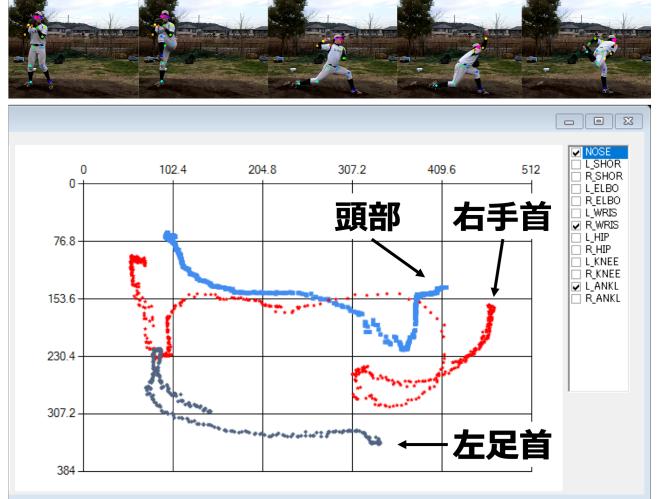


図 4 ランドマークの軌跡グラフ。

3.3 反復動作の検出アルゴリズム

提案システムにおいて、姿勢情報は、図 5 に示す全身 17箇所のランドマークにより定義される。ユーザが指定したクエリ動作の開始フレーム番号を q 、クエリ動作のフレーム数を L とする。本研究では入力動画の i 番目から $i+L-1$ 番目までのフレームにおける動作とクエリ動作との相違度 S_i を、

$$S_i = \frac{1}{17 \cdot L} \sum_{l=1}^L \sum_{k=1}^{17} \|\mathbf{p}_k^{i+l} - \mathbf{p}_k^{q+l}\| \quad (1)$$

と定義する。ここで、 \mathbf{p}_k^i は i 番目のフレーム・ k 番目のランドマークの 2 次元座標である。提案システムは、この相違度がユーザの指定した閾値以下であり、かつ、前後 L フレームにおいて局所最小となるフレーム i^* を、クエリ動作に類似した動作クリップとして検出する。

4. 結果と考察

提案システムの精度を確認するため、実際の練習動画に対し反復動作検出を行った。野球・サッカー・バドミントンといった異なるスポーツの、投球練習・シュート練習・素振り練習動画を用意し、各動画における最初の動作をクエリ動作として登録し、残りの反復動作を検出した。得られた相違度グラフを図 8 左に示す。それぞれの練習動画において反復動作部分に明確なピークが確認でき、うまく閾値を調節することで 3 種の練習動画内の全 26 回の動作を正しく検出できた。また、クエリ動作と検出された動作ク

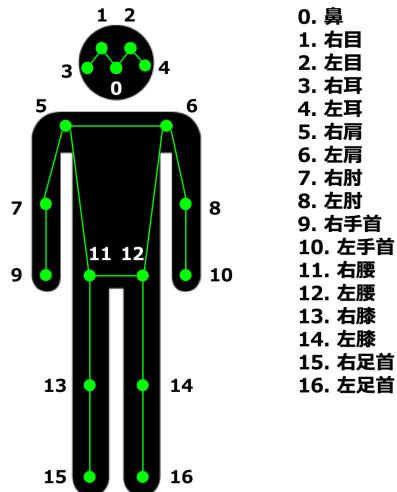


図 5 姿勢情報として利用する 17 個のランドマーク。

リップの代表例を図 8 右に示す。提案システムによってケイリ動作と類似した反復動作部分が正しく検出出来ていることが確認できる。

提案システムの有用性を示すため、サッカーのシュート練習動画から複数動作の比較やランドマーク軌跡の可視化を行なった例を紹介する。今回用意した動画には、ゴールの上部または下部を狙った複数のシュート動作が含まれている。この動画からシュート動作のみを抽出し、ゴール上部とゴール下部を狙った 2 本のシュート動作を重ね合わせた(図 6a-c)。重ね合わせ動画(図 6c)より、二つのシュートでは、体の傾きが異なることを一目で確認できる。また、サッカーのシュート練習を横から撮影した動画より、ゴール上部とゴール下部を狙った 2 本のシュート動作を抽出したもの(図 7(上))と、各シュート動作の左足のランドマーク軌跡を図 7(下)に示す。ふたつの軌跡データより、ゴール上部を狙った場合のほうが大きく蹴り足を振り上げていることが確認できる。このように、提案システムを利用すれば、非常に少ない編集作業のみで複数動作を検出・比較することが可能である。

5. まとめ

本研究では、スポーツ反復練習動画を効率的に視聴・可視化できるシステムを提案した。提案システムは、競技者の姿勢情報を活用することで、反復動作の自動検出、類似動作間の比較、ランドマーク軌跡の可視化といった機能を提供する。提案システムの反復動作検出法の精度確認のため、3種の異なるスポーツの動画に対して検出実験を行い、高精度に反復動作を検出できることを確認した。

提案システムの課題として、反復動作検出において閾値を手動で調節しなければならないことが挙げられる。この冗長な作業を省略するために自動で閾値を決定できる手法を検討している。また、今回は単一動画内の類似動作を利用し検出・比較・解析を行なった例を示した。しかし、異

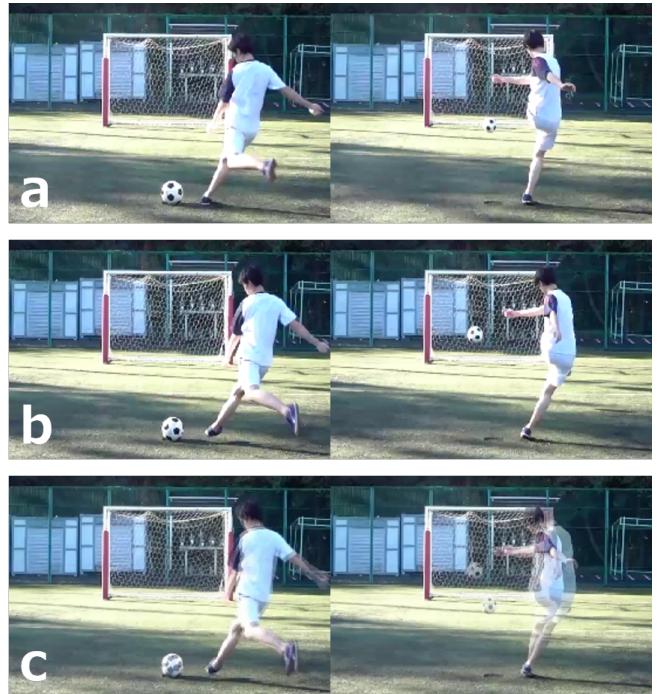


図 6 重ね合わせによる比較の例。(a) 下部を狙った動作クリップ A, (b) 上部を狙った動作クリップ B, (c) 動作クリップ AB を重ね合わせたもの。

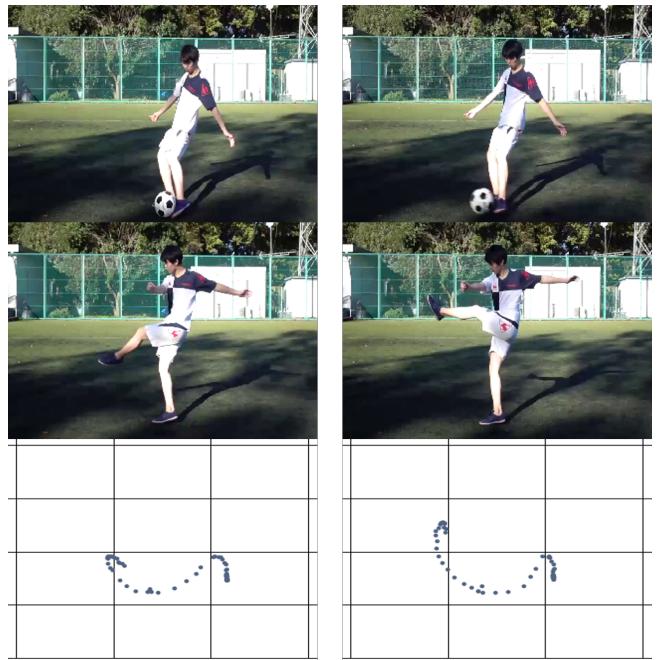


図 7 横から撮影したシュート動作の解析例。ゴール下部を狙ったシュート動作(左上)とその動作の左足の軌跡(左下)、ゴール上部を狙ったシュート動作(右上)とその動作の左足の軌跡(右下)。

なる角度から撮影された動画や別の施設で撮影された動画同士を比較したいという要望も考えられる。今後、3次元の姿勢推定が可能な手法を用いることで、視点の異なる動画間でも類似動作を検出でき比較が行なえるシステムを実

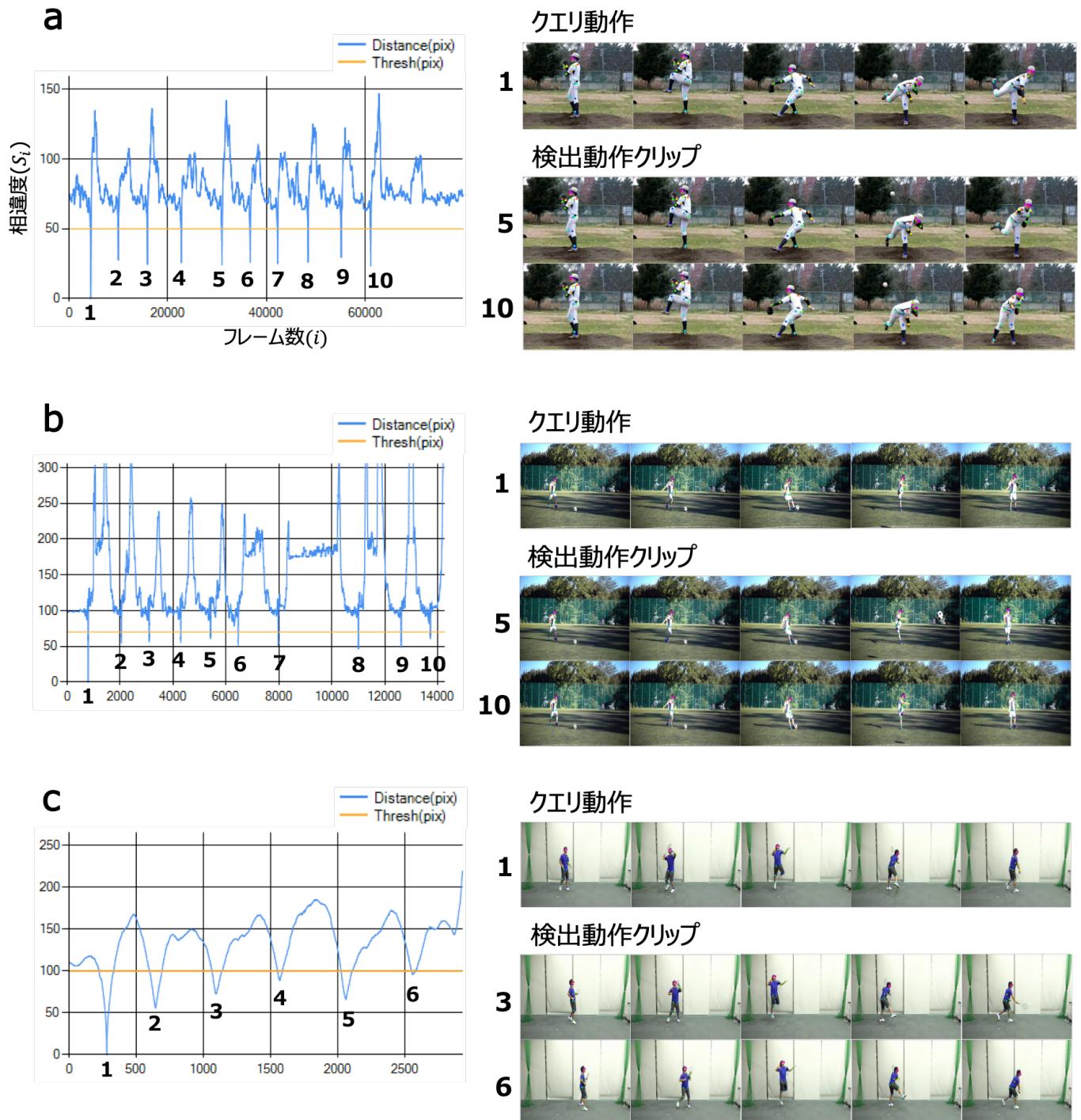


図 8 検出結果。野球の投球練習 (a), サッカーのシュート練習 (b), バドミントンの素振り練習 (c) 動画に対して検出を行った。左図は相違度グラフであり、右図はクリア動作と検出された動作クリップを等間隔に 5 枚抽出した画像である。

現したい。

謝辞 本研究進めるにあたり、練習動画の撮影に協力していただいた皆様に感謝の意を表する。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。

参考文献

- [1] Shum, H. and Komura, T.: Tracking the translational and rotational movement of the ball using high-speed camera movies, Vol. 3, pp. 1084–1087 (2005).
- [2] Ijiri, T., Nakamura, A., Hirabayashi, A., Sakai, W., Miyazaki, T. and Himeno, R.: Automatic spin measurements for pitched Baseballs via consumer-grade high-speed cameras, *Signal, Image and Video Processing*, pp. 1–8 (2017).

- [3] 新野大輔, 木坂綺花, 平林晃, 井尻敬: 民生用高速度カメラを用いた素振り練習の精度可視化, エンタテインメントコンピューティングシンポジウム 2017 論文集, pp. 38–45 (2017).
- [4] Boreczky, J., Gergensohn, A., Golovchinsky, G. and Uchihashi, S.: An Interactive Comic Book Presentation for Exploring Video, *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '00*, pp. 185–192 (2000).
- [5] Schoeffmann, K., Taschner, M. and Boeszoeremenyi, L.: The video explorer: A tool for navigation and searching within a single video based on fast content analysis, pp. 247–258 (2010).
- [6] 粟原一貴: CinemaGazer:動画の極限的な高速鑑賞のためのシステムの開発と評価, コンピュータソフトウェア, Vol. 29, No. 4, pp. 293–304 (2012).
- [7] 河村俊哉, 福里司, 平井辰典, 森島繁生: ラリーシーンに着目した映像自動要約によるラケットスポーツ動画鑑賞システム, 情報処理学会論文誌, Vol. 56, No. 3, pp. 1028–1038 (2015).
- [8] 板摺貴大, 福里司, 山口周悟, 森島繁生: コート情報に基づくバレーボール映像の鑑賞支援と戦術解析への応用の検討, 情報処理学会研究報告, Vol. 2017-CVIM-205, No. 37, pp. 1–8 (2017).
- [9] Cao, Z., Simon, T., Wei, S.-E. and Sheikh, Y.: Real-time Multi-Person 2D Pose Estimation Using Part Affinity Fields, *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2017).
- [10] Dan, O. and Irene, A.: Real-time Human Pose Estimation in the Browser with TensorFlow.js, Google Creative Lab (online), available from <<https://medium.com/tensorflow/real-time-human-pose-estimation-in-the-browser-with-tensorflow-js-7dd0bc881cd5>> (accessed 2019-11-29).
- [11] Papandreou, G., Zhu, T., Chen, L.-C., Gidaris, S., Tompson, J. and Murphy, K.: PersonLab: Person Pose Estimation and Instance Segmentation with a Bottom-Up, Part-Based, Geometric Embedding Model, *The European Conference on Computer Vision (ECCV)* (2018).