

# 物体検出と姿勢推定を活用した後方伸身宙返りの練習支援システム

岡部研究室 524C-5008 高橋 洸汰

## 1. はじめに

器械体操は空中での回転や複雑な動きを伴うため、演技中は自分自身の動きを視覚的に確認することはできない。そのため、演技を評価するためには動画撮影やコーチ・他者からのフィードバックが必要となる。しかし、部活動の練習現場では、指導者の数が限られているため、一人ひとりに十分な指導を行うことが難しい状況がある。また、初心者にとっては、自身の演技を動画で確認した際に「どこを修正すればよいか」や「どの部分が適切にできていないか」を判断することが難しく、これが技術向上の障壁となっている。

近年のスポーツ現場では、ICT(Information and Communication Technology)の導入が進められており、データを用いたパフォーマンス分析やトレーニングの効率化が注目されている。器械体操の分野においても ICT を活用した研究が行われている。例えば、ドローンを用いて技や演技を俯瞰的に撮影する手法[1]や映像遅延再生ツールを用いて演技映像をすぐに見返せるシステム[2]が提案されている。しかし、これらのシステムは主に映像を確認するための撮影方法の提案や、動画を見返す機能のみに留まっており、「どのように演技を改善すればよいか」を具体的に指摘する機能を備えた技術は未だ存在していない。

そこで、器械体操の練習支援システムを提案する。本研究では、後方伸身宙返りの練習を支援することに焦点を当てている。後方伸身宙返りは器械体操の床種目の技であり、空中姿勢の体幹コントロールやひねり技などの高度な技の発展において重要な基礎技の一つである。本システムでは、ユーザの演技を細かく確認することができる機能や、演技を分析して具体的な改善点をコメント形式で提示する機能を提供する。このコメントは、システムが演技を分析した上で「高さが低いです」「腰が曲がっています」といった改善点を提示する。このような、演技の可視化や具体的な改善点を把握できる環境を提供することで、より効果的な練習プロセスの実現や技術向上を目指す。

## 2. 提案手法

提案するシステムは大きく分けて二つある。

### ① 演技の動きを可視化し、分析をサポートするシステム

これは、演技中の動作を直感的に理解できるよう、連続写真の形式で動きの流れを可視化するものである。

### ② 具体的な改善点を提示するシステム

これは、演技中の体の各関節位置を推定して得られる座標データを用いてその演技がどのような演技であるかを分析する。その分析結果をもとに改善点をコメント形式で提示してくれるものである。

## 2.1. 連続写真

連続写真は、演技の各ステップを一度に確認することや、特定の姿勢をピンポイントで見ることができ、さらに、器械体操の教本にも用いられているため、手本となる姿勢との比較が容易になるというメリットがある。連続写真の簡単な作り方として、動画のフレーム画像を一枚ずつトリミングし、つなぎ合わせるという方法がある。しかし、この方法ではトリミングを手動で行なっているため時間と手間がかかってしまう。そこで、本研究ではこのトリミングを自動で行う手法を提案する。

まず、人物部分のトリミングを行うために物体検出技術を用いた。物体検出技術とは画像や動画中に含まれる物体を特定し、その位置を示す技術である。提案手法では、リアルタイムでの物体検出が可能なほどの高速処理と高い検出精度を兼ね備えたYOLOv8[3]で図1左のように演技動画から人物部分の検出を行う。その後、検出した範囲を利用して図1右のようにトリミングする。

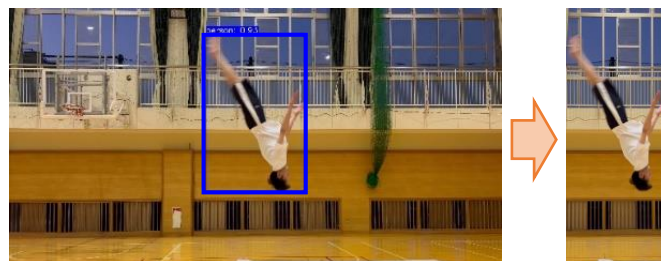


図1：左) 人物部分の検出 右) 検出範囲のトリミング

動画のフレーム毎にこのトリミング作業を行い、それらをつなぎ合わせることで連続写真を作成する。しかし、全フレームをつなぎ合わせると見にくくなるため、3フレーム毎に間隔を空けてピックアップし、それらを連続写真として構成している。提案手法により、高速かつ自動に図2のような連続写真の作成が可能になった。



図2：連続写真

## 2.2. 姿勢推定

演技を分析するためには、映像データから人物の関節位置を抽出し、それを座標データとして解析可能な形式に変換する必要がある。本研究では、2次元空間の姿勢推定を用いて関節の座標データを取得している。姿勢推定とは、人物の身体各関節点を特定し、その位置を座標として推定する技術である。現在、研究に用いることができる映像データの数には限りがあるため、姿勢推定にはRTMPose[4]とViTPose[5]の二つのモデルを採用して解析を行っている。複数のモデルを用いることで、データ不足を補うとともに、多様な解析を可能にしている。RTMPoseはリアルタイム性に優れており、高速な推論が可能である。一方、ViTPoseは高精度な解析が可能で、細かな動作の特徴を詳細に捉えることに適している。また、姿勢推定の対象となる関節点は鼻と左右の目、耳、肩、肘、手首、腰、膝、足首の計17カ所である。姿勢推定のアルゴリズムとしてはトップダウン方式を採用した。トップダウン方式とは、図3左のように画像の中の物体(人間)を検出した後、図3右のように検出された領域に対して姿勢推定を行うものである。この方式は、検出した領域に集中して姿勢推定を行うため、精度が高く、複数人が写る画像や動画に対しても安定した推定結果を得ることができる。



図3：トップダウン方式

### 2.2.1. 姿勢推定の誤り

後方伸身宙返りのような複雑な動作においては、動きが速い

部分や視覚的に遮られる部分で、各関節の位置が正しく推定されないことがある。図4に、RTMPoseを用いて姿勢推定をした結果10,11フレーム目において誤りが生じた演技を示す。図4上は10フレーム目の関節点の位置を描画したものである。黄色枠で各関節点の位置の全体を囲っているが、すべての関節点において明らかに誤った位置に描画されていることが確認できる。また、図4下は同演技の左足首において姿勢推定を行った結果をグラフ化したものである。左はx座標、右はy座標をそれぞれフレーム数に対してプロットしている。10フレーム目の姿勢推定結果のグラフは赤枠で囲まれた部分に該当する。x座標、y座標いずれにおいても他のデータ点に比べて明らかに異常な値を示していることがわかる。今回、RTMPoseによる姿勢推定結果の誤りの例を紹介したが、ViTPoseにおいても同様に推定結果の誤りは生じた。

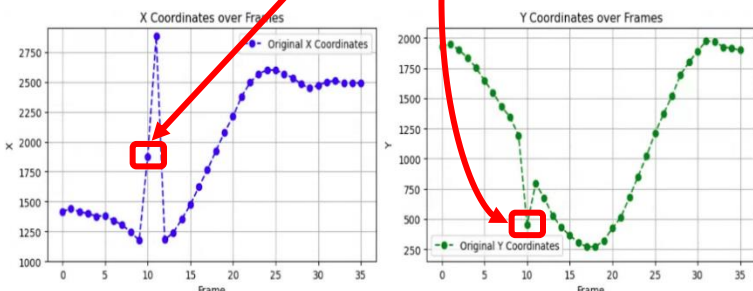
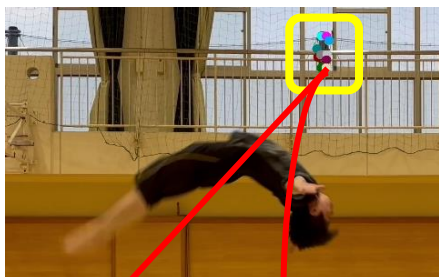


図4：上) 姿勢推定結果の描画 下) 姿勢推定結果のグラフ

適切な分析を行うためには誤りとなる外れ値を修正する必要がある。こうした外れ値を検出するため、1次微分を計算し速度を求めた。外れ値になっている部分では負の速度から正の速度になっている、もしくはその逆になっていることが多いことが分かった。さらに、2次微分を計算し加速度を求めたところ、加速度の絶対値が大きい部分で外れ値になっていることが多いことが分かった。したがって、速度の符号が変化する部分と加速度の絶対値が大きい部分をもとに外れ値の検出を行い、外れ値となったフレームにおいて補間を行う。図4の演技の左足首に対する姿勢推定結果の場合、外れ値として検出されたフレームは9,10,11,12フレームの4つであり、誤りが生じていた部分を検出できていることがわかる。この外れ値検出部分に対して補間を行う。提案手法では、補間にスプライン補間を用いた。スプライン補間とはデータ点を滑らかに結ぶ曲線を作る手法である。図5は外れ値を補間した後のグラフである。補間前を示した図4下のグラフと比較しても正しい推移が行われていることがわかる。また、実際の演技動画に各関節点の位置を描画したところ、補間された部分は正しい関節点の位置に修正されることが確認することができた。

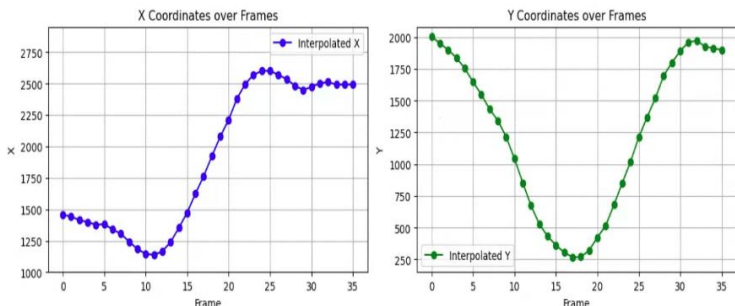


図5：外れ値修正後の姿勢推定結果のグラフ

### 3. 現状と今後の展望

具体的な改善点を提示するシステムの概要を図6に示す。現状では図6の赤枠部分である外れ値を考慮した姿勢推定の座標データを得ることができた。今後は青枠部分である比較モデルの構築に取り組む。このモデルは姿勢データから演技の特徴を分析し、それがどのような特徴を持った演技であるのかを判別するものである。最終的には、姿勢データを入力すると比較モデルが上手い演技と比較して「高さが低い」「腰が曲がっています」といった具体的な改善点をコメント形式で提示するシステムを目指している。

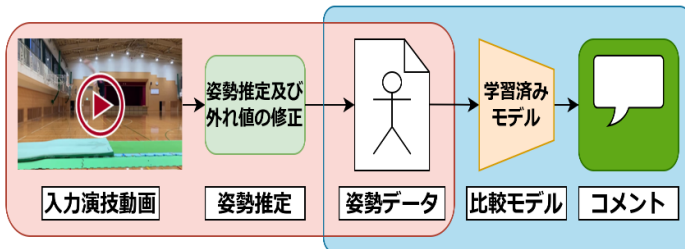


図6：改善点提示システムの概要

比較モデルの学習プロセスを図7上に示す。図6の赤枠の取り組みにより得られた姿勢データに対して「上手い演技」や「高さが低い」など演技に基づく特徴をラベルとして付与する。このラベル付けは研究者自身が演技の特徴を観察して判断する形で行う。このラベル付きデータを用いてモデルを学習させることでネットワークは、入力された姿勢データがどのラベルに該当するか分類できるようになる。学習後は、図7下のように新しい姿勢データを入力すると、そのデータがどのラベル(演技の特徴)に該当するかを判定できるようになる。

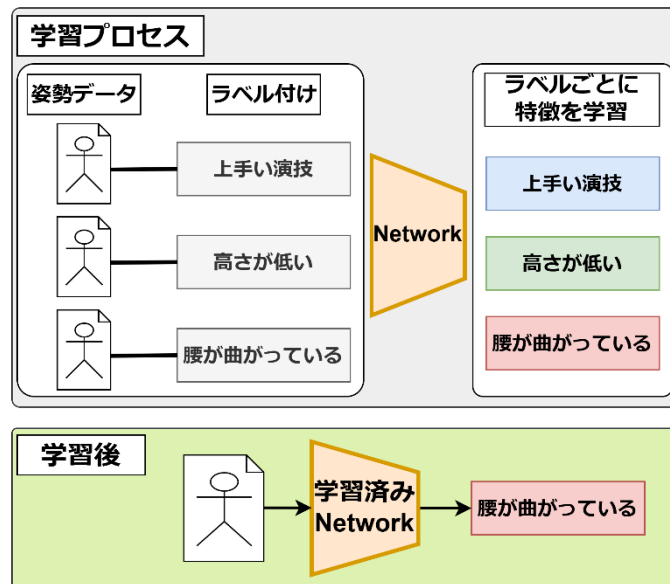


図7：上) 比較モデルの学習プロセス 下) 学習後の流れ

#### 参考文献

- [1] 安田純輝, 安井政樹, 樋口善英, “体育授業におけるドローンの活用及び展開例”, 札幌国際大学 教師・教育実践研究, 6巻, pp.68-75, 2023年.
- [2] 宮崎彰吾 “ICTを活用した器械運動の授業における学習者への効果の検討”, 第69回日本体育学会, 2018年.
- [3] Jocher Glenn. YOLOv8 release v8.1.0. <https://github.com/ultralytics/ultralytics/releases/tag/v8.1.0>, 2024. 3, 7.
- [4] Tao Jiang, Peng Lu, Li Zhang, Ningsheng Ma, Rui Han, Chengqi Lyu, Yining Li, Kai Chen, “RTMPose:Real-Time Multi-Person Pose Estimation based on MMPose”, arXiv:2303.07399.
- [5] Yufei Xu, Jing Zhang, Qiming Zhang, Dacheng Tao, “ViTPose:Simple Vision Transformer Baselines for Human Pose Estimation”, NeurIPS, 2022.