

行動認識モデルを用いた柔道試合動画からの一本判定

陸 煒[†]・琉球大学大学院理工学研究科

盛 拓矢[†]・琉球大学大学院理工学研究科

北島 栄司[†]・琉球大学大学院理工学研究科

宮田 龍太・琉球大学工学部

〒903-0213 沖縄県中頭郡西原町字千原 1

TEL: 098-895-8620,

E-mail: k238442@eve.u-ryukyu.ac.jp

[†]:Equal Contribution

1 はじめに

本研究の大目標は、柔道の試合に使える video assistant referee(VAR)システムを画像認識 AI で構築することである。VAR とは、サッカーの試合でフィールドとは別の場所で複数のアングルからの映像を見ながら主審をサポートする審判員を指し、ワールドカップカタール2022での日本対スペイン戦で三笥薫選手が相手ゴール近くのライン際で外に出そうになったボールを中に折り返した場面の判定で用いられ、世に広く知られるようになった(詳細は「三苦の1mm[1]」等を参照)。

著者らが上述した大目標を立てたきっかけは、シドニー五輪柔道男子 100kg 超級決勝で起きた誤審判定問題[2]にある。この問題は日本代表の篠原信一選手とフランスのダビド・ドゥイエ選手との対戦で、ドゥイエ選手の内股を藤原選手が内股すかしで切り返したことでドゥイエ選手は背中から引っくり返ったので、本来であれば篠原選手の一本勝ちとすべきところを主審と2人いる副審のうちの1人がドゥイエ選手の有効と判定し、ドゥイエ選手に加点してしまったアクシデントを指す。この大会後からビデオ判定導入の機運が高まり、2005年に国際柔道連盟は試合での正式採用を決めた。

以上より、柔道の試合において主審および副

審が(選手と同じ目線に立っているからこそ見えるものもあるだろうが当然死角も存在するので)その一瞬の状況をより正確に判断するには、試合場から離れて俯瞰した映像に基づいて判定をサポートする仕組みが必要である。

その取っ掛かりとして本研究では、動画から特定の動きを検知する AI である行動認識(action recognition)モデルを柔道の試合終了間際3秒間の動画に適用し、一本(かつそれが内股か背負い投げか)か時間切れかを簡易的に分類する手法を開発した。ここで柔道の「一本」とは、選手のどちらかが“相手を制し”つつその“背中が大きく畳に付く”形で、“強さ”がありかつ“速い”投げを打った場合に評価される技である。ダブルクォーターションで囲った4つの条件が全て満たされて投げ技に対する「一本」(かつその時点で勝敗も)決まるが、どれも定性的な表現で、相当な熟練者でないと正確な判断が困難であることは想像にたやすい。そこで我々は挑戦的テーマとして、動画に映る人物の肩や腰といった関節位置情報を動画のみから取り出す AI である姿勢推定(pose estimation)モデルを試合動画に適用し、一本を取られる倒し・倒され方をする際の各関節の一連の大まかな動き(時系列)の特徴を捉え、それを行動認識モデルに入力することで、定量的なデータに基づく審判サポートシステムの実装をめざす。

2 提案手法

2.1 データ

本研究では、2023 年スポーツデータサイエンスコンペティション・柔道部門で提供された 2018 年から 2021 年の世界選手権の試合動画のうち、投げ技で一本となったものを抽出した.具体的には、提供データである「csv20xxWC.csv(xx には各年度が入る)」内の「EVENTNAME」で「一本」と付与されている試合のことを指す.

本研究で固め技に着目しなかった理由は、提供された動画データを確認すると、大半の投げ技は 3 秒以内の動作で完結していた一方で、固め技は 20 秒以上続行すると一本とする規定上、動作期間が全く異なることが判明したためである.

一本となった投げ技のうち、サンプル数が最も多かった「内股一本」71 件を正例として抽出した.また負例(時間切れ)としては、上記 csv ファイルの「EVENTNAME」で「一本」とも「技あり」とも付与されていない無印の 73 件を抽出した.

その後、同 csv ファイル内の「試合終了」の「映像分」と「映像秒」を取得し、それ以前の 3 秒間に相当する 73 フレームを各動画から取り出して、以降で説明する行動認識モデルの入力とした.

2.2 手順 1～3: 試合動画から審判の映り込みを除去し、両選手の領域を自動追従

予備実験により、行動認識モデルで試合中の柔道選手の動きを的確に捉えるには、同じ試合場に居る審判の映り込みを動画から先に消しておく必要があることがわかった.もしも行動認識モデルが審判の「一本」と判断した際の片手を伸ばして上方に高く上げるジェスチャーを学習すると、選手でなく審判の動作で一本判定するカンニング AI になってしまうためであ

る.

そこでまず、テキストプロンプト(単語のリスト)で画像に映っている特定の物体の領域認識(instance segmentation)を行う Language Segment-Anything model(LangSAM[3])に“one person in the black suit on the yellow floor”と入力し、図 1.(a)に示す各原動画の最初のフレームでこれから追従したい審判の注釈情報(初期マスク)を付与した.次に、その初期マスクに基づいて特定の物体の領域追従(video object segmentation)を行う XMem[4]を用いて、図 1.(b)のように動画内での審判を白抜きにした.そして、その動画の欠損部分を背景修復(video inpainting)する ProPainter[5]を用いて、図 1.(c)のように審判の映り込みを除去した試合動画を生成した.



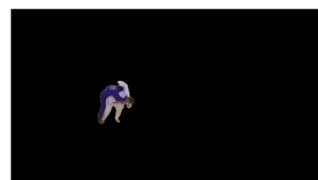
(a)元動画



(b) Xmemで審判の領域追従



(c) ProPainterで審判の映り込みを削除



(d) LangSAMで両選手に着目

図 1 手順 1～3: 試合動画から審判の映り込みを除去し、両選手のみに着目.

続いて、ここでは試合場での両選手の動きのみに注目したいので、LangSAM に“two person in white and in blue on the yellow floor.”と入力し、審判を消した試合動画から両選手の領域認識を行い、それ以外を黒塗りにした動画を生成した。これにより、次節で述べる姿勢推定モデルによる選手の関節位置情報の取得が容易になる。

2.3 手順 4: 姿勢推定モデルで推定精度が高い方の選手の関節位置情報を取得

図 1(d)に示した両選手以外が黒塗りにされた動画にトップダウン型の姿勢推定モデルを適用することで、画面に大きく映った(推定精度が高い)方の選手の関節位置情報を取得した。トップダウン型とは、先に個々の人物を検出した後で当該人物の関節位置情報を取得する(逆順がボトムアップ型)方法を指し、ここでは姿勢推定モデル DeepLabCut[6]にて取得した。図 2 に示す通り、画面手前側の選手に的確に推定した骨格線が貼り付いていることを確認した。

2.4 手順 5 : 各関節座標の系列から行動認識

手順 2 で得られた関節位置情報の一連の時系列データを行動認識モデル MMAAction2[7]に入力し、内股で一本、時間切れだったかを判定するモデルを構築した。

3 結果

3.1 内股一本と時間切れの分類

図 3(a)に内股一本と時間切れ 2 クラス分類を提案手法で行った場合の混同行列を示す。テストデータに対する正答率は 93.75 パーセントだった。内股一本の適合率と再現率も 93.75 パーセントだった。

3.2 内股一本と内股技ありと時間切れの分類

図 3(b)に内股一本と内股技ありと時間切れの 3 クラス分類を提案手法で行った混同行列を示す。テストデータに対する正答率は 63.8 パーセントだった。

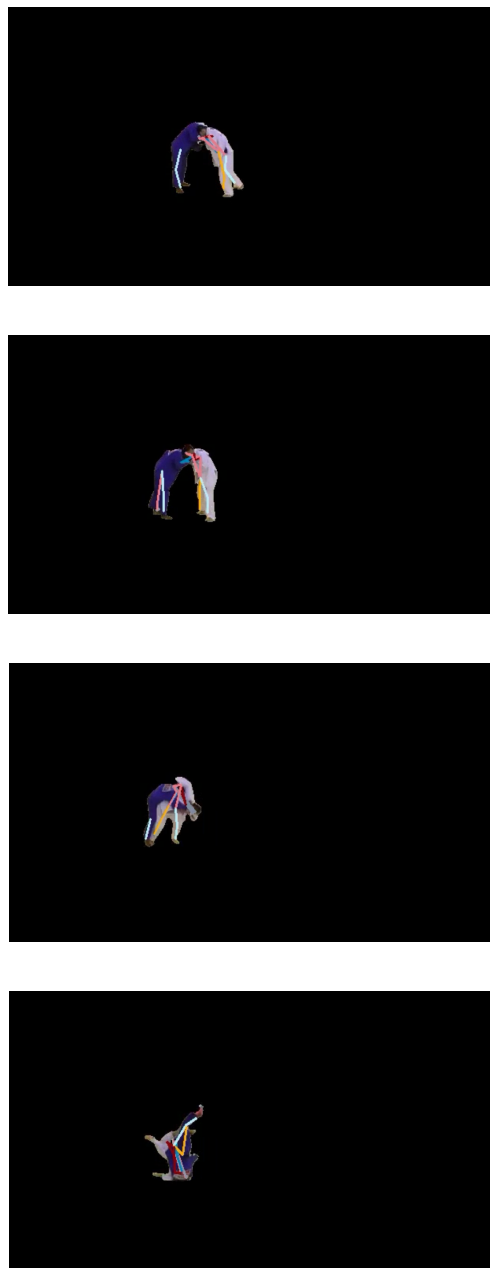
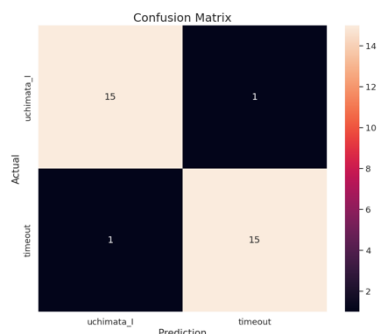
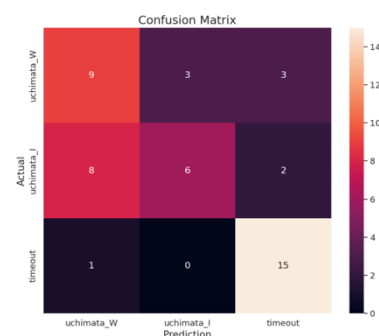


図 2 提案手法の手順 4: 姿勢推定モデルで画面手前側 (推定精度が高い方)の選手の関節位置情報を取得。



(a)内股一本と時間切れの2クラス分類



(b)内股一本と技ありと時間切れの3クラス分類

図3 混同行列

4 考察

内股一本と時間切れの2クラス分類において当てられた一本の動画は、両選手が画面中、比較的にはっきり映っていた。一方、図4(a)に示す通り、当てられなかった一本は、カメラの角度から選手のサイズは小さくて見づらい。また図4(b)に示す通り、当てられなかった時間切れは、一本と判定する状況とは異なるが、背中が畳に付いている。学習で使っていた時間切れは選手が組み合っているか離れていた動画が多くて、図4(b)はまれな状況のため、今後学習データを増やせば対処できると思っている。

そして、内股一本と技ありと時間切れの3クラス分類については、一本と技ありの判定が難しく、今後様々な状況や画角で撮影された映像を学習し、精度が上げる必要がある。



(a) 当てられなかった一本



(b) 当てられなかった時間切れ

図4 内股一本と時間切れの2クラス分類の結果に対して考察の画像

5 おわりに

本研究では、行動認識モデルにより内股の一本判定を行うシステムを構築した。今後は一本と技ありの識別精度の改善をめざす。

謝辞

本研究では、『情報・システム研究機構統計数理研究所医療健康データ科学研究センター』の支援および『公益財団法人全日本柔道連盟・科学研究部』よりデータの提供を受けた。

参考文献

- [1] ピクシブ百科事典『三笥の1mm』, <https://dic.pixiv.net/a/%E4%B8%89%E7%AC%98%E3%81%AE1mm>
- [2] Wikipedia『篠原信一: シドニー五輪決勝誤審判定問題』, <https://ja.wikipedia.org/wiki/%E7%AF%A0%E5%8E%9F%E4%BF%A1%E4%B8%80>
- [3] Medeiros L (2023) Language segment-anything, <https://github.com/luca->

medeiros/lang-segment-anything.

- [4] Cheng HK and Schwing AG (2022) XMem: Long-term video object segmentation with an Atkinson-Shiffrin memory model. ECCV 2023.
- [5] Zhou S, Li C, Chan KCK, Loy CC (2023) ProPainter: improving propagation and trans-former for video inpainting, ICCV 2023.
- [6] Jessy L et al., (2022) Multi-animal pose estimation, identification and tracking with DeepLabCut, Nature Methods:496 – 504.
- [7] MMAction2 Contributors (2020) Open-MMLab’s next generation video understanding toolbox and benchmark, <https://github.com/openmmlab/mmaaction2>.