

## 3次元特徴量を利用したバスケットボールにおける選手の行動認識 Action Recognition of Basketball Players Using Three Dimensional Features

坂野 貴斗<sup>†</sup>      水野 秀之<sup>†</sup>  
Takato Sakano      Hideyuki Mizuno

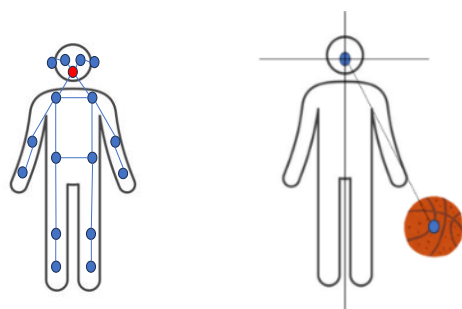
### 1. はじめに

近年、プロスポーツの分野では試合や練習における選手の行動データを用いた選手のパフォーマンスの向上やチームの戦術立案への活用が一般化しつつある[1]。バスケットボールの分野においても試合中の選手の行動を識別・分類するための研究は盛んにおこなわれており、試合中の静止画や動画から選手の行動を認識する手法が主流となっている[2][3]。しかし、選手の大まかな行動は高精度に分類可能であることが確認されているが、詳細なプレー行為については必ずしも高い精度は得られていない。一方、最近では画像中の物体の3次元座標を推定する研究も盛んにおこなわれており、高い精度で3次元座標の推定が可能となってきた[4]ことから3次元の情報を行動認識に利用することも現実的になってきている。

そこで、本研究では人物及び物体の3次元の特徴量をもとにバスケットボール選手の行動を認識する行動認識方法の提案と性能評価を行った。

### 2. 行動認識特徴量

本研究で行動認識に使用する特徴量を算出する基となる特徴点について、図1(a)に人物の17個のキーポイント、図1(b)にボールの特徴点(中心座標)を示す。人物の特徴量については、鼻のキーポイント(図1(a)の赤点)の座標を基準として、他のキーポイント16個との相対角度を特徴量として用いた。また、ボールについては、鼻のキーポイント座標とボールの中心座標との相対距離および相対角度を、特徴量とした。ただし、3次元座標上における相対角度としては、xz平面への投射角度とyz平面への投射角度の2つの角度を用いた。そのため、特徴量としては合計35個のデータを行動認識に使用する。



(a)人物の特徴点      (b)ボールの特徴点

図1 人物の特徴点及びボールの特徴点

### 3. 学習データ

本研究では、対象とする行動(プレー)を深度カメラであるIntel社RealsenseD455カメラを用いて図2のような配置で撮影した。ただし、実際の試合中の画像ではなく、指

定したプレーのみを行うよう指示し撮影を行ったため、画像内の人数は最大で2名であり、撮影時間も当該プレーの開始から終了までである。深度カメラの設定は、最大深度が20m、深度の分解能は0.01m、画素数は1280×720ピクセル、1フレーム33msである。収集した行動毎の動画数を表1に示す。特徴量の抽出手順は以下の通りである。

- ・撮影動画からIntelのpyrealsenseライブラリを利用して、1フレーム毎にRGB画像と深度画像を取得
- ・YOLOv8[5]を利用してRGB画像から画像内の人物のキーポイント及びボールの中心位置の座標を取得
- ・取得した各特徴点の位置座標と深度画像を用いて3次元座標を推定し特徴量を取得
- ・各動画のフレーム数は200に固定し、200以下の場合は200までのフレームの全特徴量の値を0とし、200以上の場合は200以上のフレームを切り捨てた。

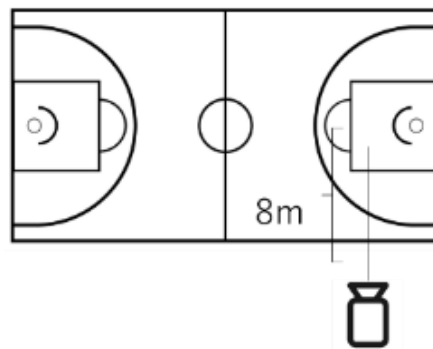


図2 撮影時のカメラ配置

表1 行動データの種類及びデータ数

行動	データ数
ブロック	1722
ドリブル	1794
スティール	1794
ボール保持	3528
移動	4116
待機	1710
パス	420
パス受け取り	420
リバウンド	1800
シュートA(ワンハンド)	3594
シュートB(レイアップ)	3588
シュートC(ツーハンド)	3522

<sup>†</sup> 公立諏訪東京理科大学 Suwa University of Science

## 4. 行動認識モデル

本研究では近年、時系列データに対する識別・回帰モデルとして盛んに利用されている Transformer[5]をベースとし、系列識別ではないことから Encoder 部のみを利用したモデルを行動認識に用いた。

### 4.1 今回使用するネットワーク構成

ネットワーク構成を図 3 で示す。Multi-Head Attention と線形層で構成される残差ブロックを複数層重ねた Transformer Encoder 層から出力された特徴量を線形層に入力し線形層から推定結果を出力する。また、Transformer Encoder 層の残差ブロックの構成を図 4 に示す。

予備検討として Transformer Encoder 層から出力される特徴量のうち、分類に用いる特徴量として CLS トークンのみを利用する場合と、CLS トークンを含む Transformer Encoder 層からの出力ベクトル全てを利用する場合の一部データの行動認識精度を比較した。結果として、CLS トークンのみの場合 94%、全出力を用いた場合 98%と全出力を利用した方が高精度なため、線形層への入力は Transformer Encoder 層全ての出力を用いることとした。

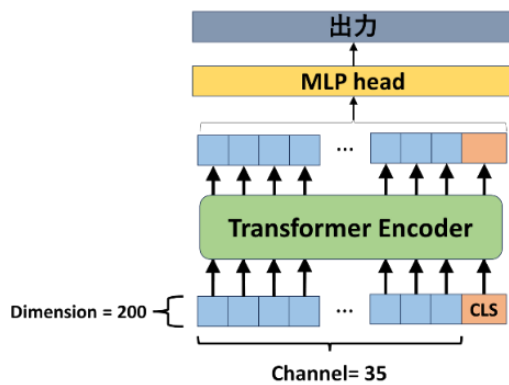


図 3 ネットワーク構成図

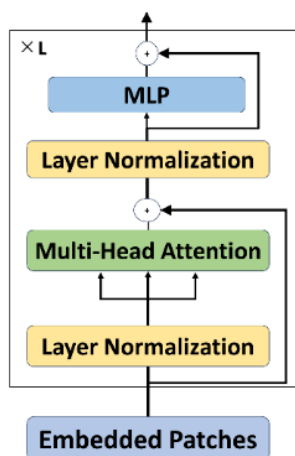


図 4 Transformer Encoder 層内の残差ブロック

## 5. 実験及び評価

### 5.1 実験内容

提案する行動認識モデルを、表 1 で示したデータを用いて 10 交差検証法により評価した。またボールと人物の位置関係の情報の有無による識別精度への影響を検証するため、ボールの特徴量を用いる場合と用いない場合での精度を比較した。

### 5.2 実験結果

実験の結果、ボールの特徴量を用いた場合の正解率は 94.5%、用いない場合の正解率は 92.4%となり、ボールの特徴量を用いることで精度が向上することが確認できた。詳細な結果を表 2 に示す。表 2 から全般的にボール有りのほうが高い精度が得られているものの、シュート B やパス関係の行動はボールの特徴量の使用の有無に関わらず精度が低く、さらにボールの特徴量を使用しない方が精度が高くなるのがわかった。

表 2 実験結果(ボール有り/ボール無し)

行動の種類	適合率	再現率	F 値
ブロック	0.94/0.93	0.94/0.92	0.94/0.93
ドリブル	0.88/0.85	0.94/0.96	0.91/0.90
スティール	0.94/0.87	0.86/0.94	0.90/0.90
ボール保持	0.99/0.98	0.99/0.78	0.98/0.87
移動	0.95/0.95	0.98/0.95	0.97/0.95
待機	0.96/0.88	0.97/0.99	0.97/0.93
パス	0.79/0.82	0.8/0.69	0.79/0.75
パス受け取り	0.66/0.76	0.8/0.74	0.72/0.75
リバウンド	0.96/0.89	0.96/0.96	0.96/0.92
シュート A	0.98/0.91	0.84/0.96	0.91/0.94
シュート B	0.91/0.96	0.98/0.94	0.95/0.95
シュート C	0.95/0.92	0.95/0.93	0.95/0.93

## 6. まとめ

本研究ではバスケットボールにおける人物及びボールの 3 次元座標に基づいた特徴量を活用した行動認識手法を提案した。実験の結果、ボールの特徴量を利用することで平均正解率も向上し大半の行動を高い精度で認識可能であることが確認できた。

### 参考文献

- [1] Datamix, "スポーツアナリティクスとは・スポーツ×データサイエンス(データ分析)の可能性", 2023. 11. 10, <https://datamix.co.jp/media/datascience/blog-sports-dataanalysis/>
- [2] Khobdeh, S. B et al. "Basketball action recognition based on the combination of TOLO and a deep fuzzy LSTM network", Jsupercomput 80, 2024, pp. 3528-3553
- [3] H.B.Wicaksono et al. "Spatin-Temporal Actioin Verfication of Basketball Travelling Dribble Violation Using Mediapipe-YOLO-LSTM Framework", IWIS, 2023, pp.1-6
- [4] Ce Zheng et al. "3D Human Pose Estimation With Spatial and Temporal Transformers", ICCV, 2021, pp.11656-11665
- [5] Ultralytics, "Ultralytics YOLO 資料" 2023. 11. 12, <https://docs.ultralytics.com/ja/models/yolov8/>
- [6] Ashish Vaswani et al. "Attention Is All You Need", NIPS, Advances in neural information processing systems, 2017, 30