

卒業論文

審判ジェスチャに着目した
バスケットボールシュート種類
判別手法

関西学院大学 工学部 情報工学課程

2503 山形隼士

2025 年 3 月

指導教員 山本 優也 教授

内容梗概

こんにちは、アブストラクトです

目次

第 1 章 はじめに	1
第 2 章 コンピュータビジョンを用いた バスケットボールの研究	3
2.1 バスケットボール競技	3
2.2 関連研究	5
2.3 研究の目的	7
第 3 章 審判のジェスチャを利用した シュート種類判別法	9
3.1 概要	9
3.2 物体認識	9
3.3 シュート位置を用いた判別	9
3.4 骨格認識	9
3.5 ジェスチャ認識を用いた判別	9
3.6 統合手法を用いた判別	9
第 4 章 評価	10
第 5 章 まとめ	11
謝辞	12
参考文献	13

図目次

2.1	バスケットボールコートの図	4
2.2	2 ポイントシュート（左）および 3 ポイントシュート（右）のジェスチャー	4
2.3	スタッツの例	5

表目次

第 1 章

はじめに

近年、スポーツアナリティクスの分野は著しい発展を遂げており、バスケットボール競技においても、各選手のパフォーマンスを詳細に表すスタッツと呼ばれるデータを用いた戦術分析や選手評価が一般化している。これらの指標は、シュート成功率、リバウンド数、アシスト数などの定量的データに基づいており、チームの勝利に直結する重要な要素となる。

一方で、現状のスタッツ収集は主に人力で行われている点や、収集に用いるアプリケーションが競技カテゴリごとに異なり統一されていないなどの課題がある。そのため、スタッツ収集のプロセスを自動化する必要性が高まっている。

この点に関して、Versnik ら [1] は試合映像を入力とし、物体検出や追跡技術を用いてスタッツを自動生成するシステムを提案している。しかし、映像内の選手位置情報のみから、2 ポイントシュートか 3 ポイントシュートかというシュートの種類を判定する手法には技術的な限界が存在する。具体的には、足元の位置検出における微小な誤差や、カメラアングルによる射影変換の歪み、あるいはディフェンス選手によるオクルージョンが発生し、物体検出での未検出が起こった場合、3 ポイントライン際での判定精度が著しく低下するという課題がある。

この課題に対し、本研究では審判の情報に着目する。バスケットボールの審判は、シュートが放たれた際、その放たれた位置や状況に応じて、2 ポイントあるいは 3 ポイントを示す特定のハンドジェスチャを行う規定がある。すなわち、審判の動作は、シュートの種類を決定づける信頼性の高い視覚情報であると言える。

以上の背景から、本研究では、従来の選手位置座標のみに依存した判定手法に加え、シュート時における審判のジェスチャ認識を導入した新たな統合判別手法を提案する。審判のジェスチャを統合することで、上記のような判定困難な状況において、シュート種類

の判別精度がどの程度改善されるかを検証し、その有効性を明らかにする。

本論文は全 5 章で構成する。第 2 章では、関連研究と本研究の目的について述べる。第 3 章では、本研究で提案するシュート種類の判別手法の詳細を述べる。第 4 章では、評価と結果について述べる。第 5 章では、本研究のまとめについて述べる。

第 2 章

コンピュータビジョンを用いた バスケットボールの研究

2.1 バスケットボール競技

バスケットボールは、1チーム5名のプレイヤーからなる2つのチームが、コートの両端に設置されたリング（バスケット）へボールを投じ、その通過数によって得点を競う球技である。コート上には計10名の選手が存在し、攻守を激しく入れ替えながら試合が展開される。

攻撃側の選手がリングに向けてボールを投じる動作を「シュート」と呼び、その動作を行う選手を「シューター」と呼ぶ。得点の種類はシュートが放たれた位置によって厳密に区分されており、その基準となるのがゴールの中心から6.75m離れた位置に描かれた「3ポイントライン」である。図1にバスケットボールのコート図を示す。このラインを境界として、ラインの内側（2ポイントエリア）から放たれたシュートは「2ポイントシュート」、外側（3ポイントエリア）から放たれたシュートは「3ポイントシュート」と定義される。

試合の判定および進行管理は、最大3名の審判員（レフェリー）によって行われる。審判は、ファウルやバイオレーションなどの反則行為に対する判定だけでなく、試合中に発生するあらゆる事象に対して規定のハンドジェスチャを用いて視覚的に伝達を行う役割を担う。特に、本研究で着目するシュート動作に関しては、シュートが放たれた瞬間にその試行が2ポイントであるか3ポイントであるかを明確にするため、審判はそれぞれに対応した異なるジェスチャを提示する規定となっている。図2にシュートにおける審判のジェスチャを示す。

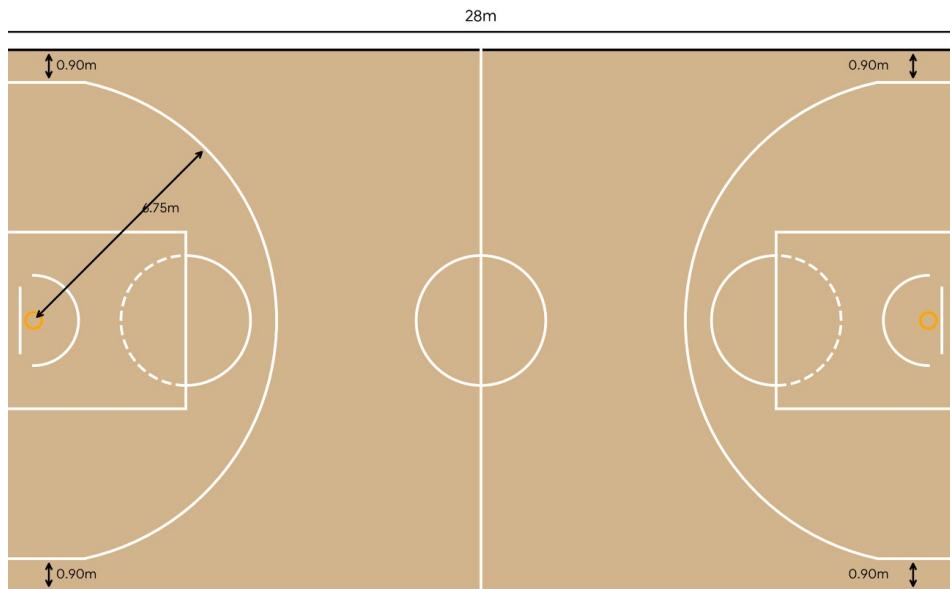


図 2.1 バスケットボールコートの図



図 2.2 2 ポイントシュート（左）および 3 ポイントシュート（右）のジェスチャ

また、現代のバスケットボール競技においては、試合の勝敗だけでなく、各選手の貢献度やパフォーマンスを詳細に可視化するために「スタッツ (Stats)」と呼ばれるデータが重要視されている。スタッツとは、その選手が何本シュートを決めたか、何回反則を犯したかなどの詳細なプレイ記録の総称である。図 2.3 にスタッツの例を示す。表中の数値は選手の成績を表しており、例えば図中の「MIN」の箇所は出場時間、「A (Attempted)」の箇所はシュートを打った数、「M (Made)」の箇所はシュートを決めた数を示している。現状、これらのスタッツ収集は、専門の記録員がシステムを用いて行っており、依然として人の目視による判断と手動入力に依存している。

MIN	PTS	FGM	FGA	FG%	2FGM	2FGA	2FG%	3FGM	3FGA	3FG%
00:00	0	0	0	0%	0	0	0%	0	0	0%
21:07	8	3	10	30.0%	1	4	25.0%	2	6	33.3%
31:23	18	7	15	46.7%	5	11	45.5%	2	4	50.0%
13:44	2	1	5	20.0%	1	2	50.0%	0	3	0.0%
28:36	9	3	10	30.0%	0	4	0.0%	3	6	50.0%
26:16	5	2	8	25.0%	1	3	33.3%	1	5	20.0%
18:52	0	0	0	0%	0	0	0%	0	0	0%
02:32	0	0	0	0%	0	0	0%	0	0	0%
12:14	4	2	2	100.0%	2	2	100.0%	0	0	0%
33:52	16	7	9	77.8%	5	6	83.3%	2	3	66.7%
11:24	5	2	3	66.7%	2	3	66.7%	0	0	0%
00:00	0	0	0	0%	0	0	0%	0	0	0%
	0	0	0	0%	0	0	0%	0	0	0%
200:00	67	27	62	43.5%	17	35	48.6%	10	27	37.0%

図 2.3 スタッツの例

2.2 関連研究

本節では、バスケットボールにおけるスタッツの自動収集および審判のジェスチャ認識に関する既存研究を概観し、それぞれの提案手法の詳細と、本研究の立場から見た課題（欠点）について述べる。

2.2.1 スタッツの自動収集

Veršnik と Šajn (2023) は、NBA の放送映像を入力とし、ディープラーニングを用いて選手やボールを検出し、試合の統計データ（スタッツ）を自動的に収集するシステムを提案している [1]。彼らの手法では、まず物体検出アルゴリズムである YOLO を用いて選手を検出し、追跡アルゴリズムである DeepSORT と組み合わせることで各選手のトラッキングを行っている。さらに、DeepSORT 単体では選手同士が交差した際などに ID が入れ替わりチーム分類を誤る問題があったため、MobileNetV2 を用いた画像分類モデルを

追加で導入し、チーム分類の精度を 99.41% まで向上させた。シュート判定においては、ホモグラフィ変換を用いて映像内の 3 次元座標をコートの 2 次元平面座標へ変換し、ボールとゴールの位置関係および選手の足元の位置に基づいて、3 ポイント、2 ポイント、フリースローの判定を行っている。

しかしながら、本手法にはシュート判別の精度が低いという重大な欠点がある。彼らの実験結果によると、シュート試投数の検出精度は高いものの、シュート成功（の判定や種別の分類において誤検出が目立った。具体的には、3 ポイントシュートの成功検出精度は 53.4% に留まり、試投の検出精度も 63.3% であった。この精度の低さは、主にホモグラフィ変換の不安定さに起因している。特にコートの片側における変換行列の計算誤差が大きく、フリースローの場面において、選手の位置が 3 ポイントラインの外側であると誤って計算され、3 ポイントシュートとして誤分類される事例が多発したことが報告されている。このように、画像座標から物理座標への変換に依存したシュート判定は、カメラアングルの変化やオクルージョンに対して脆弱であり、実用的なスタッツ収集には不十分である。

2.2.2 HOG, LBP 特徴量と SVM による審判ジェスチャ認識

Žemgulys ら (2020) は、コンピュータビジョンを用いてバスケットボールの試合映像から審判のジェスチャを認識する手法を提案している [2]。彼らは、スポーツ映像のようなノイズの多い実環境下での認識を実現するため、画像の前処理として RGB 画像をグレースケール化し、Sobel フィルタを用いてエッジ検出を行うことで、照明変動の影響を軽減している。特徴量抽出には、形状情報を捉える HOG と、テクスチャ情報を捉える LBP の 2 種類を採用し、これらを SVM およびランダムフォレストで分類する実験を行った。実験の結果、LBP 特徴量と SVM を組み合わせたモデルが最も高い性能を示し、全クラス平均で 95.6% の認識精度を達成した。特に 3 ポイントシュート成功のジェスチャに関しては、正面からの映像で 97.7% という極めて高い精度で認識できている。

しかしながら、この研究の欠点は、認識したジェスチャをスタッツの自動集計やシュート判定の結果確定に活用していない点にある。本研究の主眼は、あくまで審判と記録員間のコミュニケーションミスを減らすためのジェスチャの認識そのものに置かれている。そのため、認識された 3 ポイント成功のジェスチャが、実際の得点加算処理や、Versnik らの研究で課題とされたようなシュートが本当にに入ったかどうかの判定の補正に利用されるシステム構成にはなっていない。単独のジェスチャ認識に留まっており、試合のコンテキストと統合されたスタッツ生成システムとしては機能していない。

2.2.3 マルチスケール時空間特徴を用いた深層学習による審判ジェスチャ認識

Wang ら (2024) は、マルチスケールの時空間特徴を活用し、複雑なバスケットボールのシーンにおいて審判のジェスチャをリアルタイムで認識するシステムを提案している [3]。彼らは「GIS-ResT」と呼ばれる軽量なネットワークモデルを構築した。これは、時間領域のマルチスケール時空間畳み込みモジュールと、大域的な情報を同期させるためのチャネル注意機構を組み合わせたものである。このモデルは、従来の 3D 畳み込みと比較して計算コストを大幅に削減しつつ、ジェスチャの詳細な特徴を捉えることに成功しており、平均認識率は 92.20% に達している。また、このシステムはモバイル端末や Web ページ上での動作を想定して実装されており、選手や観客などユーザーに対して認識したジェスチャの種類を音声でブロードキャストする機能を有している。

この研究もまた、スタッツの収集やシュート判定のロジックにジェスチャを利用していない点が欠点である。本システムの目的は、聴覚や視覚に障害を持つ観客への支援や、記録員が見逃した判定の確認、あるいはコーチや選手がビデオ分析を行う際の該当シーンの検索補助 (Locate relevant video materials) に限定されている。システムはジェスチャを分類して提示する「補助機能 (Auxiliary function)」として設計されており、例えば「審判が 2 点のジェスチャをしたから、自動的にスコアに 2 点を加算する」といった、スタッツ管理の自動化に直結する機能は実装されていない。

2.3 研究の目的

前節で述べたように、Veršnik ら [1] による先行研究では、ホモグラフィ変換を用いて映像内の座標をコート平面上の座標へ変換し、その位置情報に基づいてシュート種別の判定を行っていた。しかし、この手法はカメラアングルやレンズの歪みの影響を受けやすく、特にコートの奥側や端部において変換行列の計算誤差が増大するという欠点を抱えていた。その結果、本来はフリースローである事象を 3 ポイントシュートとして誤認識するなど、シュート種類の判別精度が低下する課題が未解決のまま残されている。

一方で、バスケットボールの競技規則において、審判は 3 ポイントシュートの試投時および成功時に、2 ポイントシュートは成功時のみに、その種類を特定するための明確なジェスチャを行うことが義務付けられている。Žemgulys ら [2] や Wang ら [3] の研究が示すように、画像処理によるジェスチャ認識は高い精度を実現しているものの、これらを

シュート種類の判別ロジックに使用する試みはなされていない。

そこで本研究では、従来の位置情報に基づく判定に加え、シュート時における審判のジェスチャ認識を新たな判断パラメータとして導入する手法を提案する。本研究の目的は、先行研究 [1] のアプローチに対して審判のジェスチャ情報を統合することで、シュート種類の判別精度がどの程度向上するかを明らかにすることである。具体的には、シュート位置による判別のみを用いた場合と、ジェスチャ認識による判別を組み合わせた場合とでシュート判別の検出率を比較検証し、シュート種類判別システムにおけるジェスチャ認識の有効性を検討する。

第 3 章

審判のジェスチャを利用した シート種類判別法

設計方針（目的達成に必要な機能の一覧），全体構成と各モジュールの役割や関係，各モジュールの詳細，システムの利用例，について書く。モジュールの説明においては，入力と出力を明確にすること。自分がやったことを書く。

3.1 概要

3.2 物体認識

3.3 シート位置を用いた判別

3.4 骨格認識

3.5 ジェスチャ認識を用いた判別

3.6 統合手法を用いた判別

第 4 章

評価

評価の目的, 方法, 結果, 考察 (研究の目的がどこまで達成できたのかについて), 今後の課題について書く. 実験結果はグラフや表にまとめる. 実験結果に関しては必要に応じて, 統計的検定を行う. 考察に関してはその根拠となる実験結果を示すこと.

第 5 章

まとめ

研究の目的と結果を簡潔に書く。今後の課題について簡潔に書く。

謝辞

本研究を進めるにあたり、終始ご指導下さった XX 教授に深く感謝いたします。

また、音声認識に関して協力していただいた知能機械工学課程 YY 教授に厚くお礼申し上げます。ほかに協力してもらった人も書いておく。

北村研究室の同輩諸氏には、日頃から多くのご助言やご支援をいただき、大変感謝しています。

最後に、4年間大学に通わせてくれた両親に、心から感謝しています。

参考文献

- [1] A. Versnik and L. Sajn. Automatic tracking of NBA statistics from a live broadcast. https://www.preprints.org/frontend/manuscript/f008c9e5907813ae907d648be34e9375/download_pub. 2025 年 9 月 7 日アクセス.
- [2] B. Wang, Y. HoKun, and C. Zhou. A real-time basketball referee gesture recognition system based on multi-scale spatio-temporal features. *Applied Mathematics and Nonlinear Sciences*, 9(1):1–23, 2024.
- [3] J. Žemgulys, V. Raudonis, R. Maskeliūnas, and R. Damaševičius. Recognition of basketball referee signals from real-time videos. *Journal of Ambient Intelligence and Humanized Computing*, 11(3):979–991, 2020.