

ステレオビジョンと深層学習を用いた  
卓球競技におけるデータ分析システムの開発  
Development of a Data Analysis System for  
Table Tennis Using Stereo Vision and Deep Learning

羽鳥 駿久<sup>†</sup>  
Toshihisa Hatori

伊藤 建一<sup>†</sup>  
Kenichi Ito

上島 慶<sup>†</sup>  
Kei Kamizima

## 1. はじめに

卓球は、200 以上の国と地域が国際大会に参加し世界的に高い普及率を誇るスポーツであり、オリンピックや世界選手権でも常に注目を集めている。特にアジアやヨーロッパを中心に競技レベルが非常に高く、近年では競技の高度化が加速している。また卓球はハイスピードなラリーが連続で展開され、わずか数秒の間に様々な戦術的要素が発生する競技である。特に競技力の高い選手ほど、ボールに多彩な回転を与え、飛翔軌道を巧みに操作することによって、相手が返球しにくい打球コースを作り出す能力に長けている。このような戦術的操作は、相手の位置、回転量、スピードなど複数の要因に基づいて選択・実行され、これらの要素を定量的に把握することは、選手の育成や戦略を立てる上で非常に重要となる。

従来の卓球における戦術分析は、指導者による主観的な観察、あるいはビデオ映像に基づく単純な記録観察から行われていた。しかし 1 秒間に 30 m を超える速度で展開される高速ラリー中のボール軌道や打球、バウンド位置を人間の視認能力で正確に捉えるには限界がある。このような背景もあって、近年ではスポーツ映像解析にコンピュータビジョン技術を適用し、試合中にリアルタイムで高精度なデータ取得と分析を可能にする研究が進んでいる。

特に深層学習 (Deep Learning) を基盤とする物体検出技術の進歩によって、選手やボールの位置推定、人間の姿勢推定など、多様な解析タスクが可能となっており、サッカー [1]、バスケットボール [2]、アイスホッケー [3] など様々な競技においても活用が進んでいる。卓球においても同様のアプローチは有効であるが、ボールの小ささや高速移動といった特性により、他競技と比べて高精度な検出と追跡が要求される。さらに、戦術的要素を正確に把握するためには、単にボールの位置を検出するだけでなく、バウンドやネットインといった重要イベントの識別、三次元的な軌道推定が不可欠である。

本研究では、卓球競技における打球の三次元軌道を推定し、そこから得られる戦術的情報を解析するシステムの構築を目的とする。得られたデータは、選手や指導者による戦術評価や技術向上のためのフィードバックに活用できることが期待される。また、収集したデータを視覚的に分かりやすく表現するかも重要な課題である。戦術情報を視覚化し、テレビ中継や観戦用インタフェースに統合することで、観戦者に対して競技の戦略性や選手間の駆け引きを直感的に伝えることが可能となる。このような取り組

みは、卓球観戦の魅力を高め、競技の新たな楽しみ方やファン層の拡大にもつながる事が期待できる。

## 2. 類似研究

卓球競技における自動映像解析技術の応用も進んでおり、深層学習を活用したボール検出や追跡技術は、軌道推定や戦術分析において有望な手法として注目されている。

Gossard ら [4] は、卓球競技における打球の三次元軌道を映像から再構成する手法「TT3D」を提案した。この手法は、単眼映像から得られる二次元軌道データに対し、カメラ視点に依存しない高精度な三次元再構成を実現するものであり、二次元追跡における視点依存性の克服に寄与している。特にサイドビューでは、映像にノイズのない条件下では平均絶対誤差 (MAE) が 8.9 cm、ノイズが含まれる場合でも 12.4 cm という精度を実現している。また、ボールスピンの推定やプレイヤー動作の三次元復元にも対応しており、放送映像を用いた戦術分析や競技支援への応用が期待される。しかし、バックビューにおいては選手による遮蔽や自動キャリブレーションの失敗の影響により精度が低下し、視点依存性や遮蔽による精度低下といった限界も述べられている。

林ら [5] は、オリンピック時の一般的な放送映像から選手の身体位置やボール軌道を抽出し、それらの情報をもとに選手の戦術と戦略を可視化する「卓球戦略獲得システム」を提案している。このシステムは、映像から対戦シーンを切り出す映像抽出モジュール、抽出されたボールの着弾位置やラリー軌道から戦術・戦略を推定する戦略獲得モジュール、監督やコーチ向けに分析結果を視覚的に表示する戦略表示モジュールの三つから構成されている。映像抽出には深層学習モデル CenterNet と白色ブロッグ検出を用い、二次元座標上でのボール軌道を推定しており、通常の場合では 10~45mm 程度の誤差に収まっている。一方で、ボールが選手の身体に隠れる場面では推定精度が低下する課題が挙げられている。また、推定される情報は二次元情報のみであり三次元的情報を活用した高度な戦略推定には対応していない。

また、ボールの軌道からプレイヤーのストローク動作を分類しようとする試みとして、Kulkarni ら [6] は、ボール軌道の二次元座標データを入力とする深層学習手法を提案した。YOLOv4 および TrackNetv2 によって抽出されたボール座標をもとに、各ストロークの時間的区間を分割し、有効なストロークを抽出した後、TCN (Temporal Convolutional Network) により 6 種類のストローク分類を行い、87.2%の識別精度を達成している。ストロークの一部 (push や topspin) は弧の軌道や速度が類似しており、識別困難なケースはあるものの将来的には複数カメラによ

<sup>†</sup> 新潟工科大学 Niigata Institute of Technology

る三次元ボール軌道推定やスピン推定への発展を視野に入れており、AI コーチングや試合解析の基盤技術としての応用可能性が示唆されている。

以上のように、既存の研究はいずれも卓球における映像解析技術の有効性を示しているが、三次元情報の不足、回転推定の実環境適用性の低さ、戦術的な意味づけの限界、そして可視化・リアルタイム性に対応した統合的なシステム設計の欠如などそれぞれに課題が挙げられる。本研究ではこれらの課題に対して、三次元打球軌道のリアルタイム推定、戦術的意味を持つイベントの抽出、ならびに直感的な可視化手法の開発・検討を通じて、実際の競技現場における即時活用を可能にする分析基盤の構築を目指す。

### 3. 提案手法

本研究では、深層学習とステレオカメラを組み合わせボールおよびバウンド・打球位置の三次元位置推定を行う。図 1 に提案する計測システムのブロック図を示す。ステレオカメラで取得した 2 視点画像に対して深層学習によるボール検出を行い、推定された左右画像それぞれのボールの二次元座標から三角測量の原理を用いてボールの三次元位置を推定する。さらにイベントの自動検出を行い、競技中の打球とバウンドイベントの位置を推定する。卓球台の側面から 6m の位置 the image source 社製カメラの DFK33UX252 を 2 台を 800mm 間隔で平行配置し、フレームレート 120fps、1920×1080 の解像度で撮影を行った。撮影の際は、2 台のカメラは発信器からの矩形波によって同期を取った。

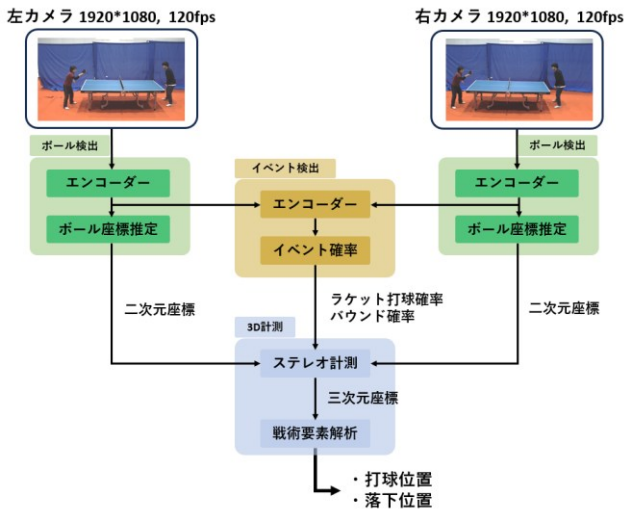


図 1 提案手法のブロック図

#### 3.1. ボール検出手法

画像上のボールおよび各フレームにおけるイベント検出には、Voeikov ら[7]の卓球の試合映像解析に特化した深層学習モデル TNet (Table Tennis Net) をファインチューニングしたものを用いた。卓球は小さなボールで高速なラリーが展開される競技であり、モーションブレンダーや背景のボールと類似性の高いオブジェクトの影響によって、検出精度が低下する。TNet では、連続する複数フレームを入力とし時間的特徴を抽出し、画像全体に対して粗い検出を

行うグローバル検出ステージ、その検出結果を基に高解像度の局所検出を行うローカル検出ステージの 2 段階構造によって高い検出精度を実現している。

#### 3.2. イベント検出手法

イベント検出は、グローバルおよびローカルのボール検出過程で得られる特徴マップを用いて行う。これらの特徴は、各フレームにおけるボール周辺の局所的な画素情報およびフレーム間の変化を含んでおり、イベントに伴う動きの差異を捉えるのに有効である。具体的には、グローバルおよびローカル検出の最終畳み込み層から出力される特徴マップをチャンネル方向に連結し、時系列にまたがる空間的特徴を統合する。これにより、イベント発生時に見られるボールの速度変化や背景との相対的な動き、輪郭変化などの変動を入力として、2 値分類によってイベントの有無を判定する。検出対象のイベントは、後述するファインチューニングによってオリジナルの TNet から変更している。

#### 3.3. ステレオカメラによる 3 次元変換

本研究におけるボールの三次元座標推定は、左右 2 台のカメラで撮影を行い、検出されたボール座標の左右の視差から三角測量の原理により、対象までの距離を測定するものである。三角測量において正確な三次元座標変換を行うためには、使用するカメラの内部パラメータ (焦点距離、主点位置、歪み係数) および外部パラメータ (カメラ間の相対位置と姿勢) を事前に高精度で推定しておく必要がある。以下ではカメラキャリブレーションの方法と歪み補正および三次座標の推定の方法について詳述する。

##### 3.3.1. パラメータ推定

本研究では、Matlab の Camera Calibration Toolbox を用いてカメラキャリブレーションを実施し、各種パラメータの推定を行った。このキャリブレーション手法は、Zhang によって提案された手法[8]に基づいており[9]、三次元構造物を必要とせず、既知の平面パターンを用いることで高精度な内部パラメータ・外部パラメータの同時推定が可能である。このツールボックスを使用してキャリブレーション結果として、カメラの内部パラメータおよび外部パラメータ、レンズ歪み係数が推定される。

推定された内部パラメータ行列  $K$  は以下の様に表現される。

$$K = \begin{bmatrix} f_x & \gamma & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix}$$

ここで、

$f_x, f_y$ : カメラの焦点距離 (画素単位),  
 $c_x, c_y$ : カメラの主点座標 (光軸中心)  
 $\gamma$ : 画像座標系の傾きを示すスキュー係数

である。

外部パラメータは、ワールド座標系 (実空間上の座標系) における三次元座標をカメラ座標系へ変換するためのカメラの位置と姿勢を表し、回転行列  $R$  および並進ベクトル  $T$  として次のように表される。

$$R = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix}, \quad T = \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix}$$

### 3.3.2. DLT 法に基づく三角測量による三次座標の推定

キャリブレーションによって得られた左右カメラの内部パラメータ  $K_l$ ,  $K_r$  および外部パラメータ  $R$ ,  $T$  を用いて、左右各カメラの射影行列を次のように表す。

$$P_l = K_l \cdot [I | 0]$$

$$P_r = K_r \cdot [R | T]$$

ここで左カメラをワールド座標系の原点とし、右カメラは左カメラに対して回転・並進していると表せる。図 2 に標準ステレオ幾何に基づく各座標の関係を示す。

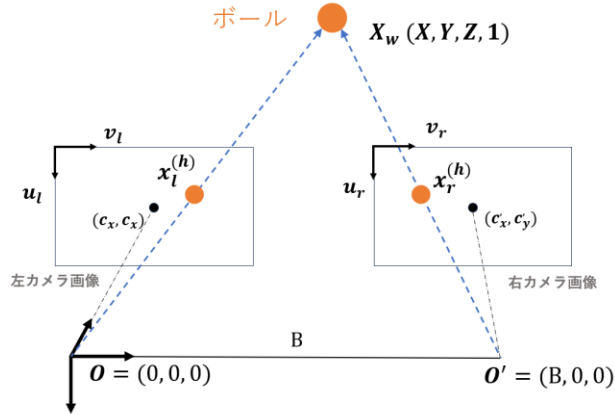


図 2 ステレオ画像の座標系

実空間上での三次元座標を  $X_w = (X, Y, Z, 1)^T$  とすると、カメラの射影モデルに従い以下の式を得る：

$$x_l^{(h)} = P_l \cdot X_w, \quad x_r^{(h)} = P_r \cdot X_w$$

$x_l^{(h)} = (u_l, v_l, 1)$ ,  $x_r^{(h)} = (u_r, v_r, 1)$  は、それぞれ左右カメラ画像上における推定対象の同次座標を表す。

これらから画像点と投影点が同一直線上にあるという射影幾何の条件

$$x_l^{(h)} \times (P_l \cdot X_w) = 0$$

$$x_r^{(h)} \times (P_r \cdot X_w) = 0$$

を適用することで各カメラに対して 2 本、合計 4 本の独立な線形方程式を得ることができる。

これにより、以下のような斉次線形方程式系が構成される：

$$A \cdot X_w = 0$$

ここで  $A \in \mathbb{R}^{4 \times 4}$  は以下の様に構成されている

$$A = \begin{bmatrix} (u_l)P_l^{(3)} - P_l^{(1)} \\ (v_l)P_l^{(3)} - P_l^{(2)} \\ (u_r)P_r^{(3)} - P_r^{(1)} \\ (v_r)P_r^{(3)} - P_r^{(2)} \end{bmatrix}$$

この線形方程式の解を特異値分解 (SVD) を用いて、最小特異値に対応する右特異ベクトルを求め、得られた

$(X, Y, Z, 1)^T$  が推定された三次元座標となる。

## 4. 推定実験

### 4.1. ファインチューニング

公開されているオリジナルの TTNNet は、「バウンド」と「ネット通過」のイベント検出を行っていたが、追加データを用いてファインチューニングを行い、検出対象を「バウンド」と「ラケットによる打球」の 2 種のイベントに変更した。学習には、オリジナルの TTNNet で用いられたオープンデータセット「OpenTTGames」に加え、追加で本学で撮影した卓球のラリー映像から、ボール座標およびイベント発生時刻をフレーム単位でラベル付けした計 6,532 枚の画像データを使用した。実装環境は PyTorch 1.2.0 を用い、GPU は NVIDIA RTX 3080 Ti を使用した。最適化関数は Adam, パラメータ  $\beta_1=0.9$ ,  $\beta_2=0.999$ ,  $\epsilon=10^{-8}$ , 初期学習率のみオリジナルの 0.001 から 0.0001 に変更し学習を行った。

また、データ拡張を行い TTNNet と同様、幅・高さともに最大 15% のランダムクロップ、 $\pm 15$  度の範囲での回転、左利き選手への対応を意図した左右反転、および輝度・コントラスト・色相のランダム変動を行う手法を使用した。

### 4.2. 検出精度の評価

#### 4.2.1. ボール検出の精度評価

ボール検出精度の評価 (グローバル・ローカル両段階) では、ボール存在判定の正解率 (Accuracy) を算出する。さらに、ボール存在が正しく検出されたフレームに対して、モデルの予測座標とラベル付けされた正解データとの画像上でのユークリッド距離に基づく誤差 RMSE (Root Mean Square Error) を計算し、検出位置の精度を評価した。

表 1 に、オリジナルの TTNNet(original) モデルとファインチューニング後の TTNNet(Fine-tuned) モデルのボール検出結果を示す。結果よりファインチューニングによって、元々高かったボール存在判定の正解率がさらに向上した。加えてグローバル RMSR も 6.79 px から 1.50 px へと改善された。元画像 (1980×1080) に対する最終的な推定誤差は、2.8px であった。一方で、ローカル RMSR は 1.97 px から 3.40 px へとやや悪化している。また、推論時間は 6.0 ms から 17.8 ms と 3 倍程度に増加した。

#### 4.2.2. イベント検出の精度評価

イベントの発生判定は各フレームにおけるバウンド・ラケットによる打球、それぞれのイベントが生起された確率値により判定を行っている。評価は PCE (Percentage of Correct Events) および SPCE (Smooth Percentage of Correct Events) の 2 指標を用いて行った。PCE は、各イベントタイプについて、予測確率と正解データのいずれも 0.5 でしきい値処理した上で、それらが一致した割合として定義される。一方、SPCE はより柔軟な評価基準を用いる。具体的には予測値と正解データの絶対誤差が 0.25 未満であれば正解とみなす。この定義により惜しい予測も捕捉することが可能である。

表 1 より、ファインチューニング済みモデルはそれぞれ 0.967, 0.969 とわずかに低下しているが、両モデルで大きな差は見られず、ファインチューニング後も高精度でイベント検出を行えている。なお、TTNet (original) は、バウンド・ネット通過を検出対象としているため、本研究で

用いたファインチューニング後の TTNNet (Fine-tuned) と検出対象が異なる。

表 1 各検出精度の比較

	TTNet(original)	TTNet(Fine-tuned)
グローバルRMSR[px]	6.79	<b>1.5</b>
ローカルRMSR[px]	<b>1.97</b>	3.4
元画像に対するRMSR[px]	-	2.8
グローバルAccuracy	0.975	<b>0.995</b>
ローカルAccuracy	0.978	<b>0.997</b>
推定時間[ms]	<b>6.0</b>	17.8
PCE*	<b>0.977</b>	0.967
SPCE*	<b>0.970</b>	0.969

### 4. 3. ステレオシステムの評価

#### 4. 3. 1. キャリブレーション結果

3.3.1 で述べた Matlab の Camera Calibration Toolbox を用いて左カメラを実空間上の原点として以下の内部パラメータ  $K_l$ ,  $K_r$  および外部パラメータを求めた。

なおキャリブレーション時に使用した既知の平面パターンはチェッカーボードパターン (図 3) である。

$$K_l = \begin{bmatrix} 2411.06 & 0.63 & 938.61 \\ 0 & 2405.99 & 524.40 \\ 0 & 0 & 1 \end{bmatrix},$$

$$K_r = \begin{bmatrix} 2409.59 & 0.66 & 926.32 \\ 0 & 2404.25 & 523.40 \\ 0 & 0 & 1 \end{bmatrix}$$

$$R = \begin{bmatrix} 1.00 & -0.01 & 0 \\ 0.01 & 1.00 & 0 \\ 0 & 0 & 1.00 \end{bmatrix}, T = \begin{bmatrix} 79.83 \\ 0.39 \\ -0.14 \end{bmatrix}$$

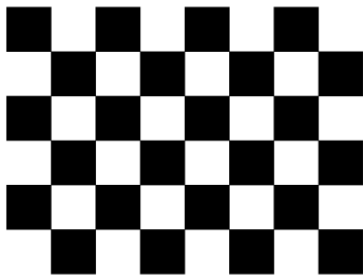


図 3 使用した平面パターン

#### 4. 3. 2. 測位精度の評価

ステレオシステムの測位精度を検証するために、卓球台上に等間隔で点群が描かれている平面パターンを手動で配置していき (図 4)、各位置における実測値と測位結果の誤差を計測した。なお、実際の計測は青枠で囲まれた下 3 行のみで測位精度評価を行った。図 5 はステレオシステムによる測位結果を三次元グラフ上の赤点で示している。実験の結果、平均誤差は x 軸方向で 4.0mm, y 軸方向で 0.5mm, z 軸方向で 11.1mm であった。また、誤差は z 軸方向 77mm が最大となった。

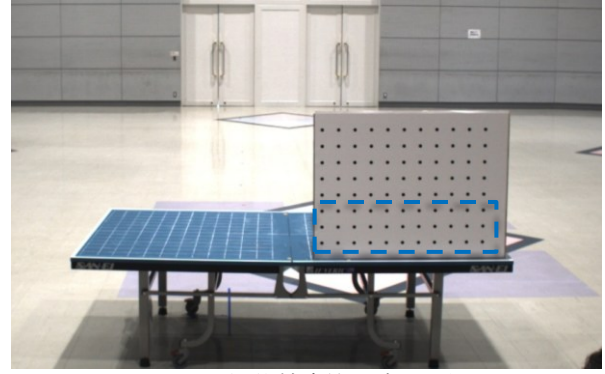
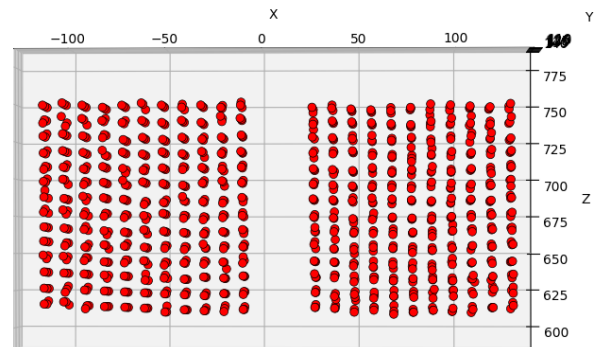
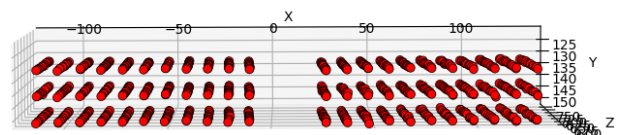


図 4 測位精度検証時の一例



(a) 見下げ図



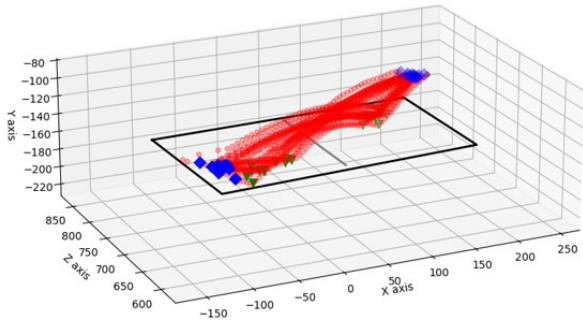
(b) 正面図

図 5 ステレオシステムの測位精度検証結果

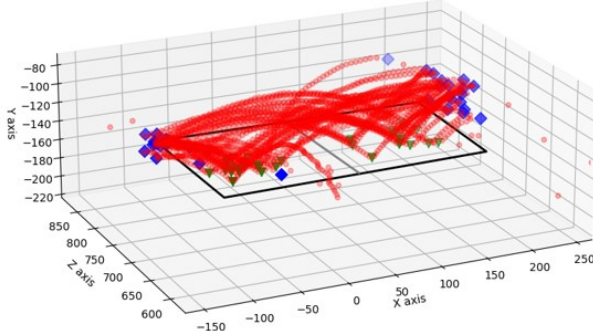
### 4. 4. 三次元座標推定実験

TTNet によるボール・イベント検出およびステレオカメラによる三次元測位を組み合わせ、ラリーと模擬戦の映像をそれぞれ 3 回撮影し、ボール軌道とイベント検出位置を三次元グラフとして描画した。イベント発生位置は、TTNet の出力する各フレームのクラス確率のうち、バウンドまたは打球のスコアが 0.5 以上であるフレームを対象とし、その時点の三次元座標をイベント発生点と定義した。また、推定された三次元座標には推定誤差や推定のばらつきに起因する高周波ノイズが含まれるため、これを除去する目的でローパスフィルタを適用した。

図 6 は推定結果の一例である。三次元空間における全フレーム間のボール軌道を赤色、イベント発生点を青 (ラケットによる打球位置)、緑 (バウンド位置) で示している。全ての試技において三次元空間におけるボール軌道とイベント発生点の推定を行えた。推定結果より、ラリーと模擬戦それぞれにおいて、ラケットによる打球位置とボール軌道の傾向が確認できる。



(a)ラリー中の推定結果



(b)模擬戦中の推定結果

図6 三次元座標推定結果の一例

## 7. 結論

本研究では、卓球競技における「バウンド」および「ラケットによる打球」イベントの三次元的な検出と可視化を目的として、イベント検出モデル TNet のファインチューニングと、ステレオカメラを用いた三次元測位を統合したシステムを構築・評価した。

イベント検出については、ファインチューニングを行い、ラケットによる打球イベント検出を実現させ、PCE で 0.967, SPCE で 0.969 の精度を記録した。これにより、モデルは高い識別性能を維持しつつ、新たなイベント定義に対応できることが示された。ボール座標検出においては、ファインチューニングによりグローバル検出精度が向上した一方で、ローカル検出では RMSE が 1.97 px から 3.40 px とやや悪化した。推論時間は 6.0 ms から 17.8 ms と増加しており、リアルタイム推定の実現には、モデルの軽量化や推論処理の高速化手法の検討が必要であることが明らかとなった。

三次元測位の評価では、平面上での xy 平面方向の精度が十分であったことから、カメラキャリブレーションおよびフレーム同期が適切に機能していると考えられる。一方で z 軸方向での誤差はやや大きく、視差に基づくステレオ計測特有の誤差傾向が確認された。また、三次元座標推定実験では各フレームの推定結果を時系列に接続することで、三次元空間上におけるボール軌道およびイベント発生点の再構成が可能となり、従来の二次元ベースでは困難であった動作解析や戦術理解への応用の可能性が見えてきた。

以上の結果から、提案システムは卓球におけるイベントやボール軌道の三次元的な可視化と解析に有効であることが示された。今後は、z 軸方向の測位精度向上、推論速度の最適化、ネットインやスマッシュなど他イベント検出に向けたモデル拡張が課題として挙げられる。また、実戦環境における長時間・多数ラリーへの適用とその信頼性の検

証も今後の研究において重要となる。

### 謝辞

本研究の一部は JSPS 科研費 24K14542 の助成を受けたものです。記して感謝いたします。

### 参考文献

- [1] 内田 郁真, スコット アトム, 大西 正輝, 藤井 慶輔, 亀田 能成. "深層学習による高速かつラベルフリーなサッカーシーン検出." 人工知能学会全国大会 (第 37 回) 講演論文集, (2023)
- [2] 中山陽太, 岡本浩行, 太田健吾, LSTM を利用したバスケットボールにおけるシュートフォームの自己学習システム検討. 計測自動制御学会論文集 vol.59, No.6, 305/307, (2023)
- [3] 山口 大輝, 杉本 充, 岡本 佳子. "アイスホッケー動画を用いた戦略分析フレームワークの提案." 人工知能学会全国大会論文集 JSAI2020
- [4] Thomas Gossard, Andreas Ziegler, Andreas Zell, "TT3D: Table Tennis 3D Reconstruction", CVSport 2025, (2025).
- [5] 林 勲, 馮 楊蘊, 入江 穂乃香, "深層学習を用いて卓球映像から獲得するボール軌道と戦術知表現" 第 12 回横幹連合コンファレンス予稿集, 12 巻, C-3-3, 2021
- [6] Kaustubh Milind Kulkarni, Rohan S Jamadagni, Jeffrey Aaron Paul, Sucheth Shenoy, "Table Tennis Stroke Detection and Recognition Using Ball Trajectory Data." Robotics and Automation Letters, arXiv:2302.09657 [cs.CV], (2023).
- [7] Voeikov, R., Falaleev, N., & Baikulov, R. Ttnet: Real-time temporal and spatial video analysis of table tennis. In Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 884–885, (2020)
- [8] Zhang, Z. "A Flexible New Technique for Camera Calibration." *IEEE Transactions on Pattern Analysis and Machine Intelligence*. Vol. 22, No. 11, 2000, pp. 1330–1334.
- [9] The MathWorks, Inc. "Camera Calibration Toolbox." *MATLAB Documentation*, <https://www.mathworks.com/help/vision/ug/camera-calibration.html>, accessed 2025.