

# サッカー選手視点映像生成による戦術的意図の推定への取り組み

黒田堅仁<sup>1</sup> 亀田能成<sup>1</sup>

**概要:** サッカーチームの戦術はピッチ上の全選手の個別の意思決定から構成されると考える。各選手の戦術的意図を解明することは、意思疎通のための効率的なトレーニングやチーム戦術理解において重要である。戦術的意図推定モデルを構築するには、試合中の選手の意図の真値が必要である。実際にプレー中の選手の意図をリアルタイムに記録することは、物理的な制約から困難である。我々は、VR空間で生成した選手視点映像を用いて試合を体験させることで意図のリアルタイムでの記録を可能にし、推定モデルの構築に活用できると考えている。本研究では、実際の試合データからVR空間上に試合状況を再現し、選手視点映像を生成する。この映像を視聴した選手が追体験的に記録した試合中の意図を真値として、戦術的意図を推定する深層学習モデルを構築する。

**キーワード:** サッカー, 戦術的意図, 選手視点映像, 深層学習

## Toward Tactical Intent Estimation via First-Person Video Generation of Soccer Players

KENTO KURODA<sup>†1</sup> YOSHINARI KAMEDA<sup>†1</sup>

### 1. はじめに

サッカーをはじめとする団体球技において、チームの戦術は個々の選手の意思決定の集積として記述される[1]。ピッチ上の同一チームに所属する11人の選手は、試合中のある時点で必ずしも一貫した戦術目標を共有しているとは限らず、各々が独自の戦術的意図を保持している。例えば、一部の選手がカウンター攻撃を意図する一方で、他方がビルドアップを選択するといった戦術的意図の不一致が生じることがある。全選手の意図を同期させ、チームの戦術を一つにできるように、実際の指導現場では反復的な練習やミーティングが実施されている。このような背景から、選手個々の戦術的意図を推定する技術への期待が高まっている。個別の意図を定量化してフィードバックできれば、練習の効率化や戦術理解の深化が期待できる。しかし、従来の戦術推定に関する研究の多くはチーム全体の戦術を単一のものとして推定する手法に留まっており[2][3][4]、選手間での意図の乖離や揺らぎを考慮した詳細な推定には至っていない。

戦術的意図推定モデルを構築するには、試合中の選手の意図の真値が必要である。しかし、実際にプレー中の選手の意図をリアルタイムに記録することは、物理的な制約から困難である。これに対し、選手視点映像を被験者に提示し、映像内の選手の立場から当時の意図を追体験的に記録する手法が一般的である。先行研究では、パスやトラップなどの意思決定を分析するために、特定のシーンを切り出した選手視点映像が用いられている[5]。また、プレー中の選手の注視箇所を分析するために、選手に取り付けたカメ

ラで撮影した映像が用いられている[6]。これらの研究で用いられた映像は、継続的な意図の推移を記録するには短尺であり、かつ11人分を網羅するバリエーションの確保も困難である。また、VR空間で選手視点映像を提示するためのシステムの開発も報告されている[7]。しかし、実際の試合状況の再現はできていないため、状況に合わせて変化する意図を記録することが困難である。したがって、選手の試合中の戦術的意図を記録するための方法は未確立である。

我々は、実際の試合データからVR空間で試合を再現し、生成した選手視点映像を用いて試合を体験させることで意図のリアルタイムでの記録を実現できると考えている。さらに、記録した戦術的意図を真値として用いることで各選手の意図を推定する深層学習モデルを構築できると考える。

本研究では、実際の試合データからVR空間上に試合状況を再現し、ピッチ上の全選手の一人称視点映像を生成する。この映像を用いて選手に試合を体験させることで戦術的意図をリアルタイムに記録することを可能にする。記録した戦術的意図を真値として試合状況を表すデータから予測するように深層学習モデルを訓練することで、意図を推定するモデルの構築を実現する。本論文では研究の第一歩として、選手視点映像の構築手法と、戦術的意図の記録及び推定手法について議論する。研究の概要を図1に示す。

### 2. 選手視点映像生成

本節では、選手視点映像生成方法について説明する。VR空間内にサッカースタジアム、両チーム合計22名の選手エージェント、ボールの3Dオブジェクトを構築する。各オブジェクトの空間座標は、実試合から取得された時系列

<sup>1</sup> 筑波大学  
University of Tsukuba

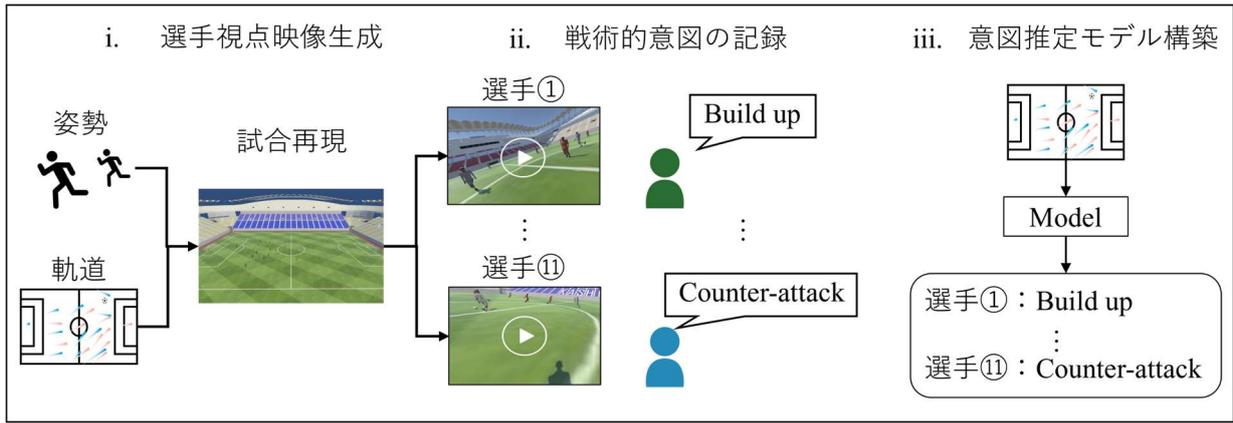


図 1 研究概要. 実際の試合データを用いて VR 環境で試合を再現する. 全選手の一人称視点映像を作成し, その映像を視聴した選手が追体験的に戦術的意図を記録する. その結果を用いて推定モデルを構築する.

トラッキングデータに基づき, 動的に制御される. 選手エージェントの姿勢についても同様に計測データに従って動的に制御される. これらにより, VR 空間における試合の再現を可能にする. さらに, 任意の選手エージェントの頭部位置に仮想カメラを配置し, 選手の移動や回転に同期させて撮影する. これにより, 試合中に各選手が取得していた視覚情報を一人称視点映像として生成する.

### 3. 戦術的意図の記録

本節では, 本研究で扱う戦術的意図の定義について説明する. さらに, VR 空間で生成した選手視点映像を用いて, 試合を体験し, 戦術的意図を記録する方法を説明する.

表 1 戦術的意図の記号. 本研究では, FIFA が定義したプレーフェーズを選手の試合中の戦術的意図とする.

戦術的意図	記号
Build up	$p_1$
Progression	$p_2$
Final third	$p_3$
Counter-attack	$p_4$
High press	$p_5$
Mid block	$p_6$
Low block	$p_7$
Counter-press	$p_8$
Recovery	$p_9$

#### 3.1 戦術的意図の定義

本研究では, 個々の選手が保持する戦術的意図の定義として, FIFA が公表した 9 種類のプレーフェーズを採用する [8]. これは指導現場の戦術概念に基づき, 試合状況を体系化したものである. 本研究では, 各選手は試合中の任意の局面において, 自身の置かれた状況をこれら 9 種類のフェーズのいずれかとして認識していると仮定する. 解析対象

とする戦術的意図の集合を  $\mathcal{Y} = \{p_1, \dots, p_9\}$  とし, 各記号と定義の対応を表 1 に示す.

#### 3.2 選手視点映像を用いた意図の記録

生成した選手視点映像を用いた戦術的意図の記録方法について定式化する. 各選手  $i \in \{1, \dots, 11\}$  に対して, 生成した一人称視点映像をアノテーターに提示する. アノテーターは映像を視聴し, 対象選手の立場から各時刻  $t$  における戦術的意図  $y_t^i \in \mathcal{Y}$  を逐次的に選択する. これにより, 実際のプレー中の主観に近い意図を真値として取得することが可能となる.

#### 4. 戦術的意図の推定

本節では, 周辺状況から注目選手の戦術的意図を予測するモデルを構築する方法について議論する.

まず, 時刻  $t$  におけるピッチ上の全オブジェクトの状態  $X_t$  を定義する. 注目選手のピッチ上二次元座標を  $\mathbf{x}_{tgt}^t \in \mathbb{R}^2$ , 味方チームの  $a$  番目の選手座標を  $\mathbf{x}_{1,a}^t$ , 相手チームの  $b$  番目の選手座標を  $\mathbf{x}_{2,b}^t$ , ボールの座標を  $\mathbf{x}_{ball}^t$  とし, これらを統合した集合として, 以下の式で構成する.

$$X_t = \{\mathbf{x}_{tgt}^t\} \cup \{\mathbf{x}_{ball}^t\} \cup \{\mathbf{x}_{1,a}^t\}_{a=1}^{10} \cup \{\mathbf{x}_{2,b}^t\}_{b=1}^{11}.$$

推定モデルへの入力  $\mathcal{X}$  は, 推定対象時刻  $T$  を中心とした前後  $W = 10$  秒間の時系列シーケンスとする.

$$\mathcal{X} = (X_{T-W}, \dots, X_T, \dots, X_{T+W}).$$

構築する推定モデル  $f_\theta$  は, 入力シーケンス  $\mathcal{X}$  から注目選手の推定意図  $\hat{y}_T^{tgt}$  を出力する.

$$\hat{y}_T^{tgt} = f_\theta(\mathcal{X}).$$

ここで,  $\hat{y}_T^{tgt}$  は戦術的意図の集合  $\mathcal{Y}$  上の確率分布として出力される. モデルのパラメータ  $\theta$  の最適化には, 記録された真値  $y_T^{tgt}$  と予測値  $\hat{y}_T^{tgt}$  の間のクロスエントロピー誤差を損失関数  $\mathcal{L}$  として用いる.

$$\mathcal{L} = -\sum_{p \in \mathcal{Y}} y_{T,p}^{tgt} \log(\hat{y}_{T,p}^{tgt}).$$

## 5. 実装

本節では、戦術的意図の推定に向けた具体的な実装について説明する。5.1 節では VR 空間を用いた選手視点映像生成の手順について説明する。5.2 節では映像を視聴する選手が戦術的意図を記録する環境について説明する。5.3 節では戦術的意図を推定する深層学習モデルの構築手順について説明する。

### 5.1 選手視点映像生成

#### 5.1.1 実試合データ

試合状況の再構築には、全 22 名の選手の三次元座標と姿勢データ、及びボールの三次元座標データが必要となる。本研究では、選手データとして FIFA ワールドカップ 2022 の 8 試合から抽出された WorldPose を使用する[9]。本データセットでは、1 シーケンス約 30 秒で構成される計 53 シーケンスを含み、ピッチ上の全選手の動的な位置、姿勢、及び体格情報を保持している。一方で、本データセットにはボールの軌道情報が含まれていないため、PFF FC 社のトラッキングデータを統合する。WorldPose のタイムスタンプに基づき、対応するボールの三次元軌道を同期して切り出し、統合データセットとして構築する。

#### 5.1.2 VR 空間における試合再現

構築したデータに基づき、Unity を用いて動的な試合環境を再構築する。各シーケンスにおいて、個々の選手の体格データを 3D エージェントに反映させ、WorldPose の姿勢と位置データを適用することで、実試合における選手の挙動を忠実に再現する。同様に、同期された軌道データをボールオブジェクトに適用し、環境内での物理的な相互作用を可視化する。

#### 5.1.3 選手視点映像の生成

各選手アバターの頭部に仮想カメラを配置し、頭部の回転角に連動して視野が変化する一人称視点映像を生成する。カメラのフレームレートは実データに基づき 30fps とする。全 53 シーケンスにおいて、両チーム全 22 名、合計 1,166 名分の選手視点映像を撮影する。



図 2 生成した選手視点映像。

### 5.2 戦術的意図の記録

アノテーターとしてサッカー経験者 11 名を選出する。各アノテーターは、生成された一人称視点映像を視聴し、対象選手の立場から戦術的意図の記録を行う。記録には Bepro 社のアプリケーションを使用する。キーボードの 1 から 9 に表 1 の戦術的意図を割り当てることで、リアルタイムな打鍵による記録を可能とする。作業中、アノテーターは定義資料を常時参照可能とし、他のアノテーターとの独立性を確保した環境下で実施する。

### 5.3 深層学習モデルによる戦術的意図の推定

#### 5.3.1 学習データセットの構築

注目時刻  $T$  を 0.2 秒間隔で移動させるスライディングウィンドウ形式で入力シーケンスを取得する。入力シーケンスは注目時刻の前後 10 秒間で構成されるため、30 秒間の各シーケンスのうち開始 10 秒後から 20 秒後の間で注目時刻をスライディングさせる。元の一つのシーケンスから 50 個の入力シーケンスが取得可能である。全 53 シーケンス、両チーム 22 名を対象とした結果、入力シーケンスの総数は、 $50 \times 53 \times 22 = 58300$  となる。元のシーケンス単位で、訓練用、検証用、テスト用データに 8 : 1 : 1 の比率で分割する。

#### 5.3.2 深層学習モデルのアーキテクチャ

本研究では、空間的な相互作用と時間的な文脈を同時に捉えるため、Graph Neural Networks (GNN) と Transformer を組み合わせたモデルを構築した。先行研究において、不規則な位置関係を持つスポーツ選手間の特徴抽出には GNN が有効であることが示されている[10]。本モデルでは、各時刻において選手及びボールをノード、チーム内の相関をエッジとした静的なグラフ構造を定義し、空間的特徴量を抽出する。各ノードには二次元座標に加え、注目選手、味方、相手、またはボールといったオブジェクトの属性を示す特徴量を付加した。その後、抽出された時系列特徴量を Transformer エンコーダーに入力し、アテンション機構を介して長期的な時間依存性を統合することで、最終的な戦術的意図の分類を行う。

## 6. 考察

本稿では提案手法の定式化及び実装の設計に焦点を当てた。本節では、現時点における技術的な制約事項及び今後の課題について議論する。

第一の課題は、生成された一人称視点映像における視覚情報の再現性である。本手法では WorldPose の姿勢推定データに基づき選手エージェントの姿勢を再現し、その頭部姿勢から視界を生成している。WorldPose の姿勢データは映像からの推定に基づいており、オクルージョン発生時などに推定精度が低下する。そのため、撮影された映像が試合中の実際の視界を再現できていない可能性がある。

第二の課題は、戦術的意図の記録の品質である。現状の

映像解像度やボールの視認性の不足は、アノテーターの状況判断に影響を及ぼし、実際の試合状況と乖離した意図が記録される要因となり得る。今後は HMD を活用し、自由な視点操作を可能にすることで、没入感及び状況理解の精度を向上させる必要がある。

第三の課題は、記録された意図の不確実性がモデル訓練の収束に与える影響である。本研究では戦術的意図を 9 種類のプレーフェーズとして定義したが、意図の記録のばらつきが学習を不安定にさせる可能性がある。これに対し、推定タスクの複雑性を低減するため、攻撃・守備・攻守の切り替えの 4 局面へとクラスを再定義することが有効なアプローチとして考えられる。

## 7. おわりに

本研究では、実際の試合データから VR 空間上に試合状況を再現し、ピッチ上の全選手の一人称視点映像を生成した。この映像を用いて選手に試合を体験させることで戦術的意図をリアルタイムに記録することを可能にする。さらに、記録した戦術的意図を真値として試合状況を表すデータから予測するように深層学習モデルを訓練することで、意図を推定するモデルの構築を実現する。

**謝辞** 本研究は、科研費 21H03476 の助成を受けたものである。

## 参考文献

- [1] Araújo, D., Davids, K.: Team Synergies in Sport: Theory and Measures, *Frontiers in Psychology*, Vol. 7 (2016).
- [2] Bauer, P.: Automated Detection of Complex Tactical Patterns in Football—Using Machine Learning Techniques to Identify Tactical Behavior, Ph.D. dissertation, University of Tuebingen (2022).
- [3] Kuroda, K., Uchida, I., Fujii, K., Kameda, Y.: Estimation of Overlapped Tactical Actions from Soccer Match Video, In *Proceedings of the 12th International Conference on Sport Sciences Research and Technology Support*, pp. 257–264 (2024).
- [4] Suzuki, G., Takahashi, S., Ogawa, T., Haseyama, M.: Team Tactics Estimation in Soccer Videos Based on a Deep Extreme Learning Machine and Characteristics of the Tactics, *IEEE Access*, Vol. 7, pp. 153238–153248 (2019).
- [5] Helsen, W., Pauwels, J. M.: The Relationship between Expertise and Visual Information Processing in Sport, *Advances in Psychology*, Vol. 102, pp. 109–134 (1993).
- [6] 伊佐稜, 堀尾郷介, 平川翼, 山下隆義, 藤吉弘亘: サッカーのトレーニングにおける中盤選手の注視点比較と視線推定 ～プロサッカー選手と大学生サッカー選手を対象として～, *映像メディア学会技術報告*, Vol. 46, No. 39, pp. 35–38 (2022).
- [7] 湯村翼: 一人称視点映像を提示する VR サッカー作戦盤システムの検討, *エンタテインメントコンピューティングシンポジウム 2022 論文集*, pp. 99–101 (2022).
- [8] FIFA.: Enhanced Football Intelligence: Explanation Document, High Performance (TSG) Football Performance Analysis & Insights, <https://www.fifatrainingcentre.com/media/native/world-cup-2022/Enhanced%20Football%20Intelligence%20EN.pdf> [accessed 2025-10-21] (2022).
- [9] Jiang, T., Billingham, J., Müksch, S., Zarate, J., Evans, N.,

Oswald, M. R., Pollefeys, M., Hilliges, O., Kaufmann, M., Song, J.: WorldPose: A World Cup Dataset for Global 3D Human Pose Estimation, *Computer Vision – ECCV 2024*, Vol. 15077, pp. 343–362 (2025).

- [10] Anzer, G., Bauer, P., Brefeld, U., Fassmeyer, D.: Detection of tactical patterns using semi-supervised graph neural networks, In *Proceedings of the 19th MIT Sloan Sports Analytics Conference* (2022).