

密集領域での動作を理解するための ハイブリッド型映像解析

Image Analysis Integrating Characteristic - Distinguishing and Learning Technologies to Understand Movement in Closed-Packed Areas

大内 一成

小林 大祐

中洲 俊信

青木 義満

■ OUCHI Kazushige

■ KOBAYASHI Daisuke

■ NAKASU Toshiaki

■ AOKI Yoshimitsu

近年、スポーツ界ではICT（情報通信技術）を活用したトレーニングや戦術分析の導入が進んでおり、画像認識を用いた戦略分析の試みも行われている。しかし、1チームが15人のラグビーのように、試合に出場する選手の数が多く、接触・密集プレーが頻繁に発生するようなスポーツでは、画像による分析は技術的なハードルが高く、これまで積極的に取り組まれてこなかった。

そこで、東芝は、強化スポーツの一つであるラグビーをターゲットとして、一つのカメラ映像から、特徴量設計方式によるボールの検出・追跡と、ディープラーニング（深層学習）方式による選手の検出・追跡を行うハイブリッド型映像解析システムを開発した。これにより、密集領域での選手の検出性能が大幅に向上し、ボールや選手の移動軌跡を精度良く仮想2次元フィールド上にマッピングできる。また、これまで人手で行われていた主要プレーのタグ付け作業を省力化するために、ディープラーニング技術を用いた自動的なプレー分類も可能である。この技術は、ラグビーに限らず様々なスポーツへ活用できるだけでなく、監視カメラを用いた作業分析など産業分野への応用も可能である。

Training and tactical analysis utilizing information and communication technology (ICT) has been progressing in the sports world in recent years, and strategy analysis using image recognition has also been attempted. However, as in the case of rugby with 15 players in a team, in a game where the number of players participating is large and contact and close-packed play occurs frequently, game analysis using image recognition is technically difficult and has therefore not been actively utilized.

In response to this situation, Toshiba has developed a hybrid type video analysis system that detects and tracks a ball according to its distinguishing characteristics and detects and tracks players using deep learning technologies with video images from a single camera. Targeted at rugby, one of the sports being specifically promoted in Japan, this system realizes improved detection of players in close-packed areas and allows the trajectory of the ball and movements of the players to be mapped onto a virtual two-dimensional (2D) rugby field. The system can also perform automatic play classification using a deep learning technique, making it possible to save labor in the task of tagging major plays that has conventionally been carried out manually. This system can be applied not only to rugby but also to various other sports, as well as to industrial fields such as work analysis using surveillance cameras.

1 まえがき

ICTの発展により、スポーツ界ではICTを活用したトレーニングや戦術分析の導入が進んでいる。特に我が国では、2019年のラグビーワールドカップ2019日本大会をはじめ、世界的なスポーツのビッグイベントが相次いで開催される。このため、これまでスポーツには関連が少なかった様々な企業も含め、産業界でもスポーツへの関心が大いに高まってきている。

東芝は、スポーツを対象とした、映像理解技術による戦略分析システムの開発を進めており、当社の強化スポーツの一つであるラグビーを最初のターゲットとした。ラグビーは、選手の数が1チーム15人と多く、接触・密集プレーが頻繁に発生するため、映像による分析には技術的な課題が多く、これまで積極的に取り組まれてこなかった。ラグビーを対象として、映像を精度良く解析できる基盤技術を開発することで、ほかのスポーツ、更にはスポーツ以外の産業分野などへの展開も期待できる。

2 スポーツへのICT活用事例

ボールを用いたチーム競技では、選手個人のデータだけでなく、試合におけるチームとしてのフォーメーションや戦術などを分析した統計データに基づく戦略立案も求められている。複数台のカメラをフィールド全体が撮影できるように設置し、選手やボールの動きを追跡してデータ化する製品がサッカーで実用化されているが、非常に高価で、システムを設置した競技場でしか利用できない。また、追跡結果やフォーメーションの特徴を用いたプレーの認識などには対応しておらず、アナリストが、映像を見ながらスポーツ専用分析ソフトウェアを使って必要なタグを手動で入力するか、あるいは手動入力したデータを購入して戦術分析を行っているのが現状である。

対象としているラグビーでは、主にトップチームで装着型デバイスを活用したチーム強化の試みが進んでいる。選手がGPS（全地球測位システム）や加速度センサーを内蔵したデバイスを背中上部に身に付けてプレーし、その選手の走行距離

や、加速・減速の回数などを記録して、フィットネスに関する分析を行っている。しかし、GPSの位置精度が十分でなく、また、データを取得するにはデバイスを取り外してクレードル経由で取り込む必要があるため、チーム全体の戦略分析には活用できていない。一方、映像による分析は、デバイスを身につける必要はないが、1章で述べたように、技術的な課題が多い。

3 ラグビー映像解析システム

3.1 機能概要

強豪の一つである当社のラグビーチーム 東芝ブレイブルーバスのコーチやアナリストなどへのニーズヒアリングに基づいて機能を検討し、センサーを装着せずに、映像だけでラグビーのプレーを解析して戦術分析に役立てることができる、映像理解技術を用いたラグビー映像解析システムを開発している(図1)。入力は、画面左上に表示されている1台のカメラで撮影した試合映像だけで、ほかのセンサーは使用しない。今回は、一般的なチームでも使用されている家庭用ビデオカメラで撮影したHD (High Definition) 映像 (1,280×720画素, 30フレーム/s) を用いた。

入力映像から、両チームの選手(どちらのチームかも識別)とボールの検出・追跡を行い、その位置を推定して画面の右上にある仮想2次元フィールド上にマッピングすることで、プレー中の両チーム選手とボールの動きを連続的に記録することを可能にした。

ボールを保持している(攻撃している)チームの推定も行う。これにより、“赤チームが、この辺りをこの方向に攻めたシーン”のような直感的な検索を可能にした(図2)。画面右上のように、検索条件として仮想2次元フィールド上に場所と方向を手書きで入力すると、これにマッチするシーンが画面の右下に表示される。この機能は、手書きタブレット向けの手書きス



図1. ラグビー映像解析システム — 1台のカメラで撮影した試合映像からボールや選手の検出・追跡を行い、仮想2次元フィールド上にマッピングすることで選手とボールの動きを連続的に記録する。

Video analysis system for rugby games



図2. ボール移動軌跡に基づくプレーシーンの手書き検索 — 場所と方向を手書きで入力することで、ボールの移動軌跡から目的のシーンを直感的に検索できる。また、プレー推定による自動タグ付けも行う。

Search of play scenes based on ball trajectory by handwritten query

ケッチによる資料検索技術¹⁾を活用した。

また、入力映像から各時点でのプレー内容を推定し、推定結果の確信度が高い場合は自動でタグを記録する。プレー推定処理は1回/sで行い、画面左下に推定結果の上位2候補を表示している。このタグを手掛かりに所望のプレーを検索できる。これまでは、タグ付け作業を手動で行うか、手動入力したデータを購入するしかなかったため、現場で試合中のプレー内容を使った分析を行うことは困難であったが、自動化により手間を省くだけでなく、例えばハーフタイムに前半の気になるプレーを重点的に確認し、後半の戦略に活用することなども可能になる。

3.2 ハイブリッド型映像解析

入力映像から選手やボールの検出・追跡を行い、その位置を推定して仮想2次元フィールド上にマッピングする機能は、特徴量設計方式によるボールの検出と、ディープラーニング方式による選手・密集領域の検出を組み合わせたハイブリッド型映像解析により実現した。全体の処理の流れを図3に示す。

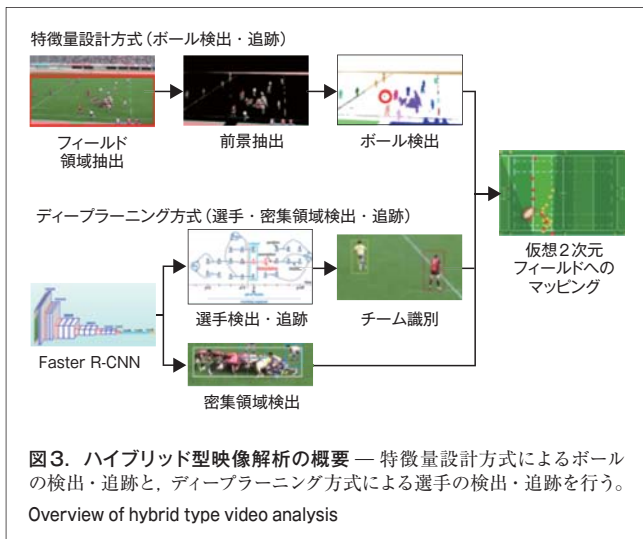
3.2.1 ボール検出

ボールは、大きさ、形状、色が決まっているので、その検出に適した特徴量を設計して検出する。

まず、観客席などのフィールド外での誤検出を防ぐため、フィールド領域の抽出を行う。具体的には、フィールド領域はほぼ同じ色であることから、あらかじめ取得したHSV色空間^(注1)における色相と彩度のしきい値によりフィールドに近い色を取り出し、その中で最も大きい面積を選択することでフィールド領域としての矩形(くけい)を取得する。

次に、フィールド領域外をマスクし、フィールド内でフィールドの色以外を前述のしきい値で抽出して得たボールや選手な

(注1) 色の種類を表す色相 (Hue)、鮮やかさを表す彩度 (Saturation)、明るさを表す明度 (Value) の三つの成分から成る色空間。

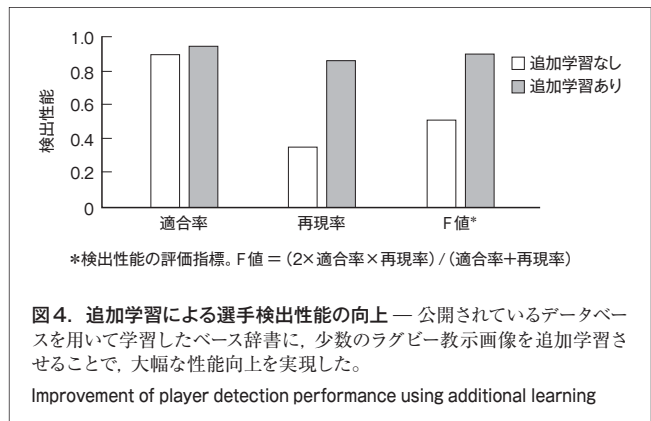


どの前景領域に対してボールとの類似度を計算する。大きさは、事前に複数の映像でラグビーボールが取り得る大きさを調査し、30～300画素を探索対象とした。形状は、ラグビーボールは楕円(だえん)球であるが、映像には円に近い形状でも映り込むことから、形状の単純さを表す指標²⁾を用いた。色情報は、HSV色空間の色相と彩度から64次元のヒストグラムによる比較を行う。事前に複数のラグビー映像から取得しておいた、代表的なボールの色相と彩度のヒストグラムと、前景のヒストグラムとの類似度を計算する。これらの大きさ、形状、色情報に関する特微量を組み合わせてボールを検出する。しかし、ラグビーではボールが選手や密集領域内に隠れることが多いため、ボール検出だけでは連続的にボール位置を推定することは困難である。そこで、ボールを検出できないシーンに対しては、補間処理やボール保持者・密集領域推定を用いてボール位置を推定する。

3.2.2 選手・密集領域検出 選手・密集領域は、大きさ、形状、姿勢、色などが様々で異なるため、事前にある程度の数の選手・密集領域の教示画像を学習してディープラーニングにより検出を行う。具体的には、ディープラーニングを用いた物体検出手法の一つである、Faster R-CNN (Region-Based Convolutional Neural Networks)³⁾を用いた。

選手の検出には、まず公開されているデータベース⁴⁾を用いて、11,540枚の画像から27,450オブジェクト(人や、動物、乗り物などの20クラスはあらかじめ教示済み)を学習したベース辞書を作成した。次に、複数のラグビー映像から500枚の画像を任意に選択し、それらの画像に対してラグビー選手を人手で教示したデータを追加学習させることにより、検出性能の向上を試みた。追加学習なしの場合と、追加学習ありの場合とで、ラグビー映像中の選手の検出性能を評価した結果、**図4**に示すように大幅な性能向上効果が確認できた。

大量のデータで汎用のベース辞書を作成しておき、ターゲット



トのドメインに応じ、そのドメインに関連が深い教示データを人手で教示できる程度の数だけ追加学習させることで、検出性能が大きく向上することが示された。この知見は、スポーツ以外にも広く活用できると考えられる。

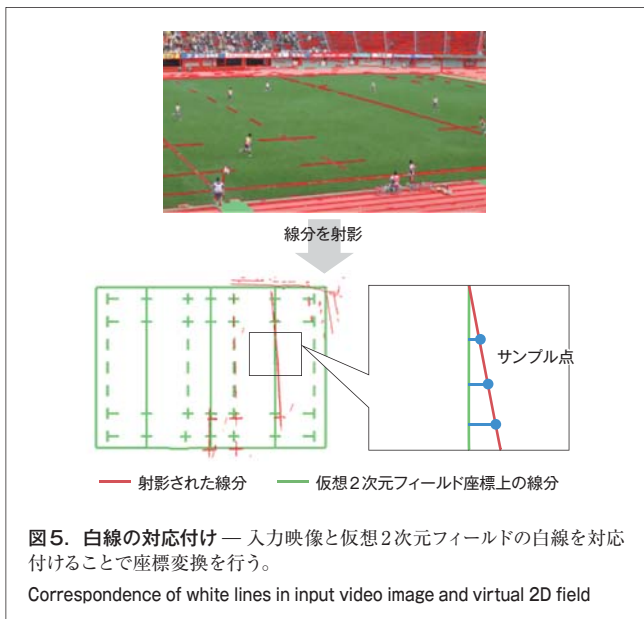
選手が密着し重なり合う密集領域で各選手を検出することは困難であるため、密集領域全体を一つの領域として検出する。“スクラム”、“ラインアウト”、“モール”、“ラック”、“その他”の5クラスの密集領域を定義し、Faster R-CNNで検出する。密集領域の学習はクラス別に行うが、少ない教示コストで多くのデータを学習するため、左右反転した画像や矩形を拡大した画像を生成して学習データの増強を行った。

選手の検出結果に対しては、重なりにロバストな人物追跡手法であるDIET⁵⁾を用いて追跡を行い、矩形内の色ヒストグラムの類似度を用いてチーム識別を行う。

3.2.3 2次元フィールド座標系への変換 ボールや選手・密集領域は、入力映像の座標系から、フィールドを真上から見た仮想2次元フィールド座標系へ射影変換して、試合中の選手やボールの移動軌跡を2次元フィールド座標系で記録する。

入力が1台のカメラからの映像であり、パンやズームにも対応させる必要があるため、射影変換を逐次実施する。まず、画像中の4点の対応付けを行うことで最初の射影変換行列を推定し、その後、カメラ移動による誤差量を入力映像と2次元フィールドの白線の対応関係から推定することで、射影変換行列を補正する。画像からの白線検出には、高速かつ検出の少ない線分検出器⁶⁾を用いている。

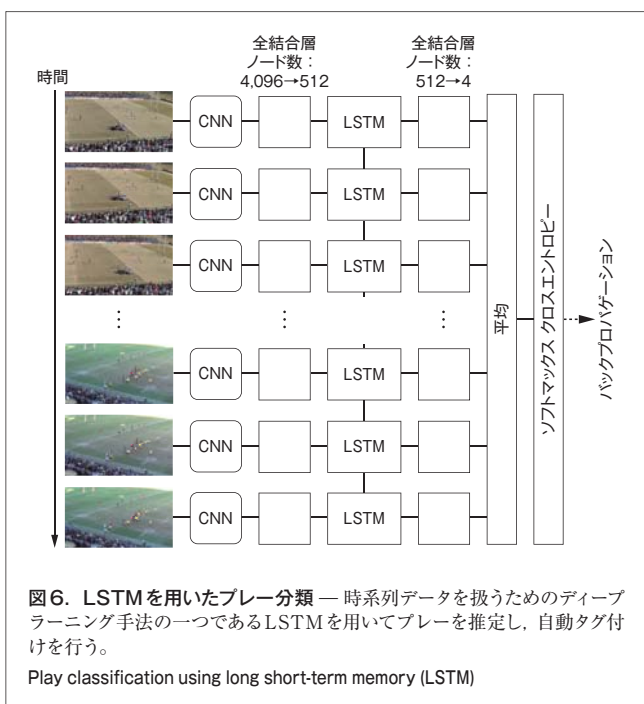
両座標系の白線対応付けの概要を**図5**に示す。緑の線が基準となる仮想2次元フィールド座標上の線分、赤の線が入力映像の現フレームにおける線分である。まず、前のフレームで推定された射影変換行列を用いて、仮想2次元フィールド座標系へ線分を射影する。次に、射影された線分に対して等間隔にサンプル点を取り、サンプル点と仮想2次元フィールド座標上の線分との距離がしきい値以下となる直線を探査する。そして、射影された線分と2次元フィールド座標上の線分間の角度差が20°以下となる点とを対応付け、複数の対応点から射影



変換行列を推定する。この対応付けと射影変換行列の推定を繰り返し行うことで誤差を最小化する。

3.3 プレー推定

タグ付け自動化のためのプレー推定には、時系列データを扱うためのディープラーニング手法の一つである、LSTM (Long Short-Term Memory)⁽⁷⁾を使用した。図6に示すとおり、切り出した静止画を1枚ずつCNNで特徴量にし、全結合層で512ノードにしてLSTMに入力する。そして、LSTMの出力を全結合層で4ノードにした後、この4ノードの平均により、スクラムやラインアウトなどのタグ付けの確率分布を算出



する。各タグ付けの確率をソフトマックス関数で表し、タグ付けの確率分布と正解との差をクロスエントロピーで示す。静止画を次々にCNN経由でLSTMに入力し、正解との差が小さくなるようにバックプロパゲーションで学習することでプレーを推定し、タグ付けを行う。

4 ほかのスポーツへの適用

映像解析の技術的な課題が多いラグビーを対象として、精度良く解析できる基盤技術を開発することにより、ほかのスポーツへの活用が期待される。ラグビーと同様にフィールド上の球技であるサッカーを対象に、今回開発した各要素技術を適用した。選手とボールの検出に際し、サッカー選手の新たな教示は行わず、3章で述べたラグビー向けのボール検出手法と、ラグビー選手を追加学習させた辞書をそのまま適用した例を図7に示す。

サッカー映像に対しても、精度良くボールと選手の検出・追跡が行われ、位置推定できていることが分かる。サッカーでは、複数カメラを使用した映像解析システムが既に導入され始めているが、今回開発したシステムは、1台のカメラからの映像だけで同様の解析が実現できる点が特長である。

5 あとがき

1台のカメラで撮影したラグビー映像を対象に、特徴量設計方式とディープラーニング方式を組み合わせたハイブリッド型映像解析を用いることで、選手やボールの高精度な検出・追跡を実現した。また、真上から見た仮想2次元フィールドにそれらの移動軌跡をマッピングすることにより、フィールド上での連続した動きの分析を可能にした。また、これまで人手で行われていた主要プレーのタグ付け作業を省力化するために、ディープラーニングによるプレー推定を用いたタグ付けの自動



化を検討した。

今回開発した映像理解技術は、4K (3,840×2,160画素) などの高解像度映像や複数カメラからの映像ではなく、一般の家庭用ビデオカメラ1台からの映像だけを解析対象としているため、今後の試合に限らず、過去に撮影した大量の試合映像をデータ化することにも活用できる。

今後は、それぞれの要素技術の更なる精度向上や処理の高速化などを行うとともに、試合現場での実証を進めながら実際に使えるシステムとしての完成度を高めていく。また、ほかのスポーツへの展開、更には監視カメラを用いた作業分析など、スポーツ以外の産業分野への幅広い応用を検討していく。

文 献

- (1) 中洲俊信, ほか. 手書きスケッチによる直感的な資料検索システム. ヒューマンインタフェース学会論文誌. 2016, 18, 3, p.141-152.
- (2) Li, W. et al. An efficient measure of compactness for two-dimensional shapes and its application in regionalization problems. International Journal of Geographical Information Science. 2013, 27, 6, p.1227-1250.
- (3) Ren, S. et al. "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks". Advances in Neural Information Processing Systems 28 (NIPS 2015). Cortes, C. et al., eds. Neural Information Processing Systems Foundation, 2015, p.91-99.
- (4) PASCAL Visual Object Classes Challenge. "Visual Object Classes Challenge 2012 (VOC2012)". The PASCAL Visual Object Classes. <<http://host.robots.ox.ac.uk/pascal/VOC/voc2012/index.html>>, (accessed 2017-06-20).
- (5) Pham, Q. V. et al. "DIET: Dynamic Integration of Extended Tracklets for Tracking Multiple Persons". Proceedings of 22nd International Conference on Pattern Recognition (ICPR2014), Stockholm, Sweden, 2014-08, International Association of Pattern Recognition. IEEE, 2014, p.1206-1211.
- (6) Grompone, R. et al. LSD: A Fast Line Segment Detector with a False Detection Control. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2010, 32, 4, p.722-732.

- (7) Hochreiter, S.; Schmidhuber, J. Long Short-Term Memory. Neural Computation. 1997, 9, 8, p.1735-1780.



大内 一成 OUCHI Kazushige, Ph.D.

技術統括部 研究開発センター
インタラクティブメディアラボラトリー
博士(工学) 情報処理学会・人間情報学会会員
Interactive Media Lab.



小林 大祐 KOBAYASHI Daisuke

技術統括部 研究開発センター
インタラクティブメディアラボラトリー
映像情報メディア学会会員
Interactive Media Lab.



中洲 俊信 NAKASU Toshiaki, Ph.D.

営業統括部 デザインセンター デザイン第一部
博士(学際情報学) 映像情報メディア学会・ヒューマンインタフェース学会・日本顔学会会員
Design Center



青木 義満 AOKI Yoshimitsu, Ph.D.

慶應義塾大学 理工学部教授
博士(工学) IEEE・電子情報通信学会・画像電子学会・映像情報メディア学会・日本顔学会会員
Faculty of Science and Technology, Keio Univ.