

テニス選手の姿勢情報を用いた打球方向推定

清水 友博^{1,a)} 吉川 隆士² Lee Chonho² 斎藤英雄¹

概要: 本論文では OpenPose を用いて得られる姿勢情報を用いて、実際のテニスの試合映像からテニス選手の打球の方向推定を行う手法を提案する。提案手法は、学習用データ作成部と、学習用データによる学習・認識部に分けられる。データ作成部では、YouTube から得られたプロテニスの試合の動画を用いて、手前側選手のみ映した矩形画像、選手のコート内の位置、打球方向を学習データとして生成する。学習データによる学習と認識部では、時系列データの機械学習手法の 1 つである LSTM を用いて、選手の姿勢情報と選手のコート内位置を入力データ、打球方向を正解データとして学習を行う。ここで、姿勢情報とは OpenPose によって得られた 18 点の関節位置を指す。このように学習されたネットワークに、学習データを入力することにより、打球方向の推定結果を得る。評価実験として、選手の矩形画像から特徴量の異なる 4 つのデータを作成し、提案手法の優位性を示した。その後、時間軸をずらしたデータセットを用い、最も打球方向推定精度が高くなるデータセットを確認した結果、66.8%の精度を得た。

Shot Direction Estimation using Posture Information of Tennis Players

1. 背景

2020 年東京オリンピック・パラリンピックに向けて、スポーツ分野におけるファン層の拡大、エンターテインメント性の向上、選手能力の向上等の目的のため、データ活用や映像処理技術の需要は高まっている。すでに実用化されている映像処理技術として、水泳競技におけるコース上への順位表示や、メジャーリーグベースボールにおけるボールトラッキングシステムなどがあげられる。データ活用は、スポーツの様々な分野で戦略分析に用いられている。

一方、深層学習に代表される機械学習技術の飛躍的な進歩に伴い、スポーツの分野でも機械学習を利用する研究に注目が集まっている。特に、データサイエンスの進歩にともない大量のデータを取得することが容易になったことは、スポーツ分野への機械学習利用に大きな役割を果たしている。例えば Takahashi ら [1] は、機械学習を用いて試合映像からサッカーボールを頑健に検出・追跡する手法を提案している。

従来、テニスを対象とした統計データを用いる研究やボールトラッキングに関する研究が行われてきた。テニスの国際大会に導入されている The Hawk-Eye Officiating System [2] は大量の統計データをとることを可能にした。Michal ら [3] は、過去の選手の勝率やコートサーフェスなどを利用し、テニスの試合の勝者を推定する手法を提案している。一方、フォームの解析を利用し、テニス選手のパフォーマンス向上やけが防止などに役立てる研究などは十分になされていない。

テニスの試合において、テニス選手は対戦相手の打球方向を予測すること、そして対戦相手の予測と反対方向にボールを打つことで試合を有利に進められる。打球方向の予測には、これまでの対戦相手の打球方向の統計のほか、対戦相手のフォームを見て予測をしている。一方で、ネット越しに見える対戦相手は小さく、打球時のラケットの面まで見る事が出来ない。そこで、テニス選手は手首のスナップを活かして球を打つことで、対戦相手に自身のフォームから打球方向を予測させずに球を打つことが出来る。そのため、どの程度対戦相手のフォームから打球方向を予測されているか知ることはパフォーマンス向上のために重要である。

そこで本研究では、テニス選手のパフォーマンス向上のため、選手のフォームがどの程度対戦相手に打球方向を予

¹ 慶應義塾大学理工学研究科
Hiyoshi, Kouhokuku, Yokohamashi, Kanagawa 3-14-1, Japan

² 大阪大学サイバーメディア先進高性能計算機システムアーキテクチャ共同研究部門
Mihogaoka, Ibarakishi, Oosaka 5-5-1, Japan

a) tomy1201@keio.jp

測されやすいかを定量化する手法を提案する。OpenPoseを用いて得られる姿勢情報を用いて、実際のテニスの試合映像からテニス選手の打球の方向推定を行う。なお本研究では、テニス選手の後方の映像のみを使用する。

2. 関連手法

Johannes ら [4] は、テニスのフォアハンドのフォームについて技術レベルのレベル分けを定量的に行う手法を提案している。この手法では、フォームの肘や手の位置関係や、インパクト時のラケットスピードをもとにフォームの分析を行っているが、選手の姿勢情報を得る手段として、ラケットと体にマーカーをつける必要がある。

一方、Multi-Person Pose Estimation (以降 OpenPose) [5] は単眼画像から人の姿勢推定を行うことを可能にしている。OpenPose は畳み込みニューラルネットワーク (Convolutional Neural Network, CNN) を用いて、画像に映る人物の肩・肘・目など関節位置の位置推定を行う。”COCO” Model [6] を用いることで 18 点の関節位置の位置推定が行える。出力結果として、画像内の人の関節位置の 2 次元座標だけでなく、ヒートマップも出力として得ることが出来る。Hossein ら [7] は、OpenPose の姿勢推定結果を利用し、水泳選手のクロールのフォームにおいて水をかく動作を推定することを可能にしている。

本研究では打球方向の推定に、テニス選手の姿勢情報の時系列データを入力として用いる。時系列の予測モデルは一般に定常波と非定常波に分かれ、前者は ARIMA モデル、後者はカルマンフィルタが使われることが多い。しかし、これらのモデルは逐次的に観測された説明変数と目的変数を使う。一方、テニス選手の打球方向推定は一連の姿勢データから最後に目的変数が得られると考えられるので特殊な時系列モデルとなる。このようなモデルとしては、LSTM(long short-time memory) [8] による一連の文章から肯定的か否定的かを予測する感情分析 [9] や株価の動向から短期的に上昇下降を予測する SCW [10] がある。テニスにおいては、Cai ら [11] の研究が挙げられる。彼らは、テニス選手の後方からの映像を、CNN による特徴抽出によって得られた特徴量を入力として、LSTM を用いることでテニス選手の行動をする手法を提案している。

3. 提案手法

3.1 提案手法の概略

図 1 に提案手法の流れを示す。提案手法は、データセット作成部と、学習用データによる打球方向推定部に分けられる。データセット作成部では、YouTube から得られたプロテニスの試合の動画を用いて、手前側選手のみ映した矩形画像、選手のコート内の位置、打球方向をデータセットとして作成する。打球方向推定部では、時系列データの機械学習手法の 1 つである LSTM を用いて、選手の姿勢情報

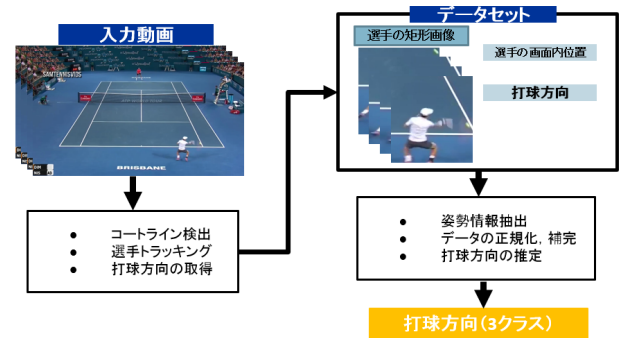


図 1 提案手法の流れ

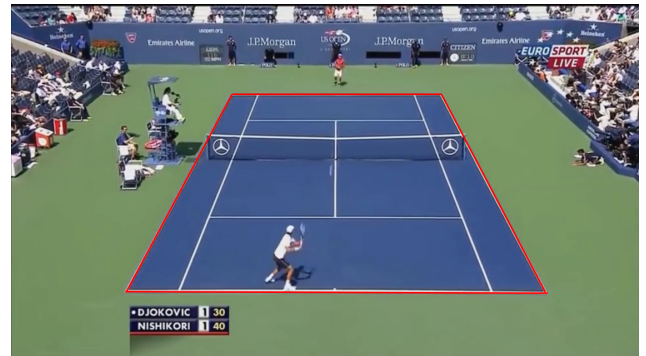


図 2 コートライン検出結果

と選手のコート内位置を入力データ、打球方向を正解データとして学習を行う。ここで、姿勢情報とは OpenPose によって得られる 18 点の関節位置を指す。このように学習されたネットワークに、学習データを入力することにより、打球方向の推定結果を得る。

3.1.1 コートライン検出

テニスの試合のビデオでは、テニスコートが画面に占める割合が最も大きく、またコートの平面とカメラの視線がなす角も固定されている。そのため、テニスコートを検出できれば、観客領域のようなコート以外の様々な領域を容易に識別することが出来る。例えばコートラインを検出することで、ボールボーイなど選手以外の人間を誤って選手と検出することを避けることが出来る。

プロのテニス試合で使用されるテニスコートは、長さ 23.78 メートル、幅 10.97 メートルである。RGB 画像フレーム内でテニスコートのラインの位置を特定するために、初めに画像 2 値化により、白線を検出する。その後、ハフ変換を適用してコートラインの候補となる直線を検出する。検出した直線の中から、直線の相対位置と直線の長さを利用し、コートの最も外側の 4 線を選択しコートラインを得る。4 線を得ることで、選手のコート内位置を求める際の線形変換 (Direct Liner Transformation, DLT) [12] に必要な 4 つの交点を求めることが出来る。図 2 にコートラインの検出結果を示す。

3.1.2 選手トラッキング

本手法では、選手トラッキングにパーティクルフィルタ

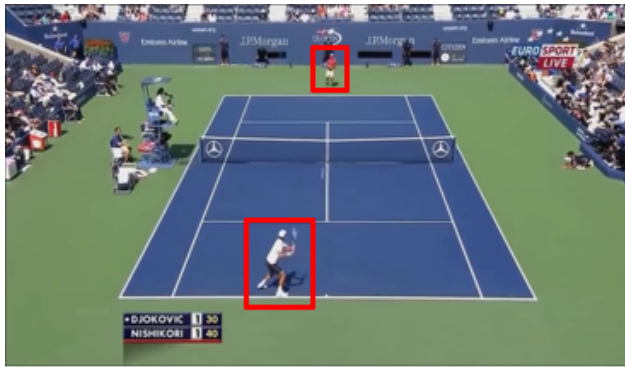


図 3 選手トラッキング結果

[13] を使用する。出力は、それぞれの選手を囲む二つの矩形領域である。

パーティクルフィルタは局所探索法の一つで、短時間のオクルージョンにも頑強に対象を追跡することが出来る。確率分布に基づく時系列データのトラッキング手法で、前フレームの結果をもとに次フレームの位置を予測モデルに基づいて推測して対象物体の探索を行う。しかし、対象物体を見失ったときや、最初のフレームでは手動で対象物体を指定する必要がある。

選手トラッキングの手順としては以下のとおりである。追跡の精度を高めるために初めにフレーム差分による動物体抽出を行う。前フレームの対象物体の位置から特定の予測モデルに基づいてパーティクルを配置する。今回は、選手が等速直線運動すると仮定して、前フレームの対象物体の位置から一定の範囲内にランダムにパーティクルを配置した。このとき、それぞれのパーティクルには尤度関数に基づいて重みづけがされている。尤度関数は、追跡対象の RGB 値へのユークリッド距離である。その後、それぞれのパーティクルの重みつき平均を求めることで、その位置を対象物体の追跡結果とする。手前の選手は追跡結果を中心に $x=180, y=220$ の大きさの矩形、奥側の選手は追跡結果を中心に $x=100, y=140$ で囲み、出力結果である矩形画像を得る。図 3 に選手トラッキング結果を示す。

3.2 データセット作成

本項では、姿勢情報を利用して、テニス選手の打球方向を推定する前段階として、学習に用いるデータセットの作成方法を述べる。今回は YouTube に投稿されているプロテニスの試合の動画 (1080×720 pixel, 25fps) を用いる。動画を入力とし、テニス選手のインパクトタイミングから、インパクトタイミング 20 フレーム前までの選手の矩形画像、選手のコート上での位置、そのインパクトから放たれた打球方向 (右方向へのクロス、左方向へのクロス、ストレートの 3 クラス) をデータセットとして得る。手前側の選手を対象としているため、1 つの動画から手前側のテニス選手のインパクトの数と同じ数のデータセットを得るこ

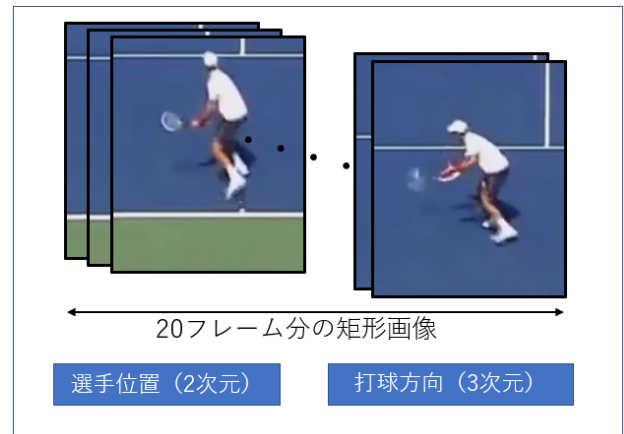


図 4 作成するデータセットの例

とが出来る。作成するデータセットの例は図 4 のとおりである。

3.2.1 打球方向の取得

本手法では、選手のインパクトタイミングは目視で、打球方向の取得は選手トラッキング結果を利用する。

打球方向はボールをトラッキングして得る方法が考えられるが、テニスにおけるボールトラッキングはボールが小さいこと、ボールのスピードが速く変形して映像に残ること、オクルージョンによりボールが映像に映っていないことが多いという問題点がある。これまでに様々な研究 [14] [15] が行われているが、精度は高々 85% 程度と常にボールをトラッキングすることは出来ていない。そこで、本研究では選手トラッキングにより打球方向を判断する。テニスの試合は、コートを含んで常に二人で行われる。そのため、ボールを打った選手の位置 (X_1) と、そのボールをリターンした選手の位置 (X_2) を利用することで打球方向を得る。図 5 のように、 X_1 の x 座標と X_2 の x 座標の差を求め、その値と閾値を比べ、その結果により打球方向を 3 クラス (右方向へのクロス、ストレート、左方向へのクロス) に分類する。

3.3 打球方向推定

初めに、打球方向推定の概要を図 6 に示す。3.2 節で得られた選手の矩形画像に OpenPose を利用し、姿勢情報を得る。その後、正規化を行う。また、選手の画面内位置とコートライン検出結果を利用し、選手のコート内での位置を求める。得られた二つの特徴量を利用し、モデルに基づいて学習を行う。

3.3.1 データの正規化

3.2 節で得られた矩形画像から、OpenPose を利用することで姿勢情報を抽出する。姿勢情報は図 7 で表される 18 点の関節位置で得られる。18 点のうち、5 点は頭周辺の関節位置 (両目、鼻、両耳) であるため、5 つの平均値を頭の関節位置とし、14 点の座標データを姿勢情報として用い

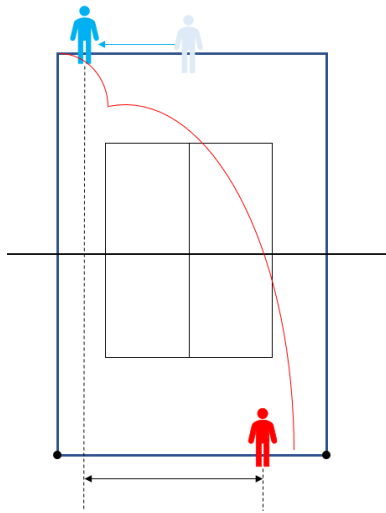


図 5 打球方向の決定方法 (赤い人がボールを打った選手、青色の人がリターンをした選手、薄い青色の人が赤い人のインパクトタイミング時の奥側の選手の位置、赤い線は打球の軌道)

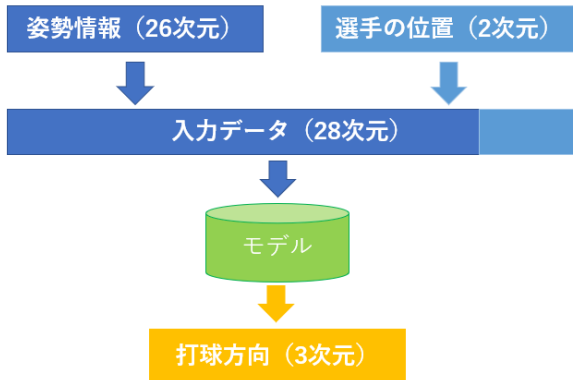


図 6 打球方向推定の概要

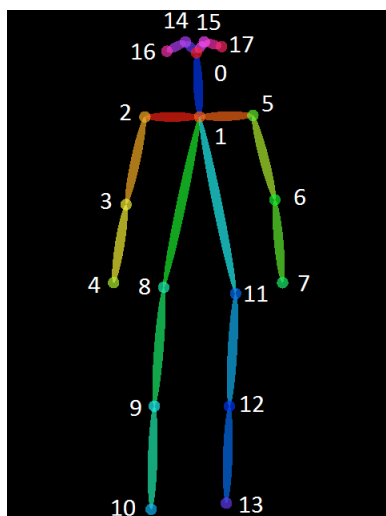


図 7 OpenPose の出力関節位置

る。得られた姿勢情報を打球方向推定に用いるため、データの正規化を行う。

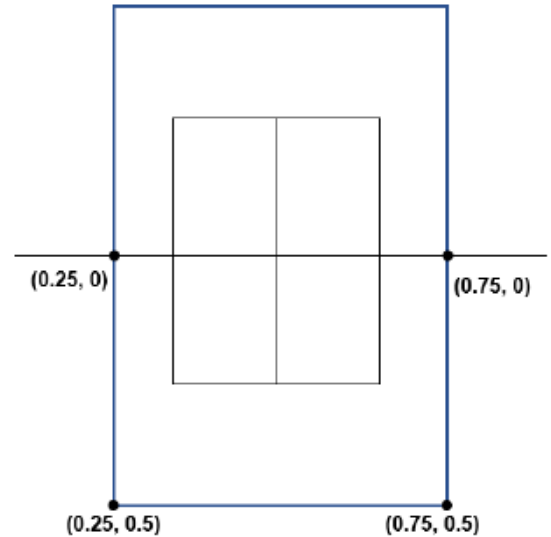


図 8 OpenPose の出力関節位置

A, コート内選手位置の正規化

本研究で用いる試合映像は、試合によってカメラとコートの位置関係が変わるため、フレーム内の選手の位置座標 (x_{frame}, y_{frame}) はコート内の位置ではない。そのため、選手の位置をそれぞれの試合でコート内の位置座標に変換する必要がある。3.2 節で得たフレーム内のコートライン 4 隅の座標を用いて射影変換行列を求めることで、射影変換により選手のコート内での位置 $X_{court}=(x,y)$ を求めることができる。また、その際 (x,y) を 0 以上 1 以下で表すため、コート内座標を以下の図 8 のように設定する。対応点を利用し、射影変換行列を H を求める。求めた H を用いて X_{court} は以下で表せる。

$$X_{court} = H \begin{bmatrix} x_{frame} \\ y_{frame} \\ 1 \end{bmatrix} \quad (1)$$

B, 姿勢情報の正規化

本手法で用いる選手の画像は、選手トラッキングで得られた選手位置を中心に 180×220 画素の画像である。OpenPose によって得られる姿勢情報は画像中における姿勢の関節位置情報 $Y_j=(x_j,y_i)$ であるため、対象選手の頭の位置が原点となるように座標変換する必要がある。変換関数を f 、頭の座標 $Y_{head}=(x_{head},y_{head})$ をとすると、各パーツの位置座標は次のように補正される。

$$f(Y) = Y_j - Y_{head} = (x_j - x_{head}, y_j - y_{head}) \quad (2)$$

ここで、 j は 1 以上 13 以下の自然数である。

C, 欠損値補間

OpenPose によって得られた姿勢情報は、姿勢推定が失敗し姿勢情報を得られない場合がある。この場合、座標が与えられず 0 となり、このような異常値が訓練データに含

まれると識別精度に影響を与えてしまう。そこで、これらの姿勢推定に失敗している情報を補間することで影響を低減させる。一つの試合で、一つの打球方向にクラス分けされる姿勢情報は一定であると推定される。そのため、フォアハンドで打つとき、バックハンドで打つときに場合分けをし、0となっていない関節位置座標の平均値を求め、その値を欠損値に当てはめることで補間する。

3.3.2 姿勢情報による打球方向推定

3.3.1 節で正規化した姿勢情報を用いて、機械学習を利用して打球方向を推定する。本手法では、リカレントニューラルネットワーク (Recurrent Neural Network, RNN) の一種である LSTM[8] を利用する。時系列データを学習する深層学習アルゴリズムとして RNN があり、ニューラルネットワークの中間層の出力値を次のデータの入力とするモデルであるが、学習するデータの時刻が増えていくことで中間層の深さが増大していく順伝搬型ニューラルネットワークと同じようになり、勾配消失や勾配爆発の問題が発生する。

それらを解決するために改良を加えたものが LSTM であり、入力ゲート、出力ゲート、忘却ゲートを取り入れることによって情報の取捨選択を行うことにより時系列データの学習精度が向上している。LSTM は、人間の短期記憶と長期記憶を模倣するような機構を有している点に特徴がある。

本研究では、打球方向推定に用いるデータとして、前節までで作成したデータを正規化したものを用いる。具体的には、入力データを姿勢情報 26 次元とコート内位置情報を連結した 28 次元、正解データを 3 次元のデータとする。

4. 実験

4.1 実験の概要

本章では提案手法を評価するために行った実験及びその結果と考察について述べる。まず、3.2 節で述べた学習データ作成に従い、実際のテニスの試合映像 (1080 × 720pixel, 25fps) から、データを収集した。なお、試合は YouTube にあるテニスの試合映像から作成し、合計 1482 フォーム分の画像を収集した。テニスの試合映像は、ハードコート、クレイコートの異なる 8 試合分を用い、錦織圭、ジョコビッチ、マレー、その他選手のデータをそれぞれ 5:2:2:1 の割合である。なお、選手は両手でバックハンドを打つ選手を対象とした。

本実験では、新たに特徴量の異なる 4 つの学習データを作成し、提案手法と比較することで提案手法の優位性を確認する。その後、時間軸をずらしたデータセットを用い、最も打球方向推定精度が高くなる学習データを確認する。なお、得られた学習データは 8:1:1 の割合で訓練データ、検証データ、テストデータに区別した。

4.2 評価実験 1

評価実験 1 では、本手法の学習データに優位性があることを確認するため、同じデータセットから特徴量の異なる 5 つの学習データを作成、学習を行い結果を比較した。作成した学習データは以下のとおりである。なお、対象フレームはインパクトタイミングから 5 フレーム前までの姿勢情報とする。

A, 本手法の学習データ

前節まで述べたように、姿勢情報を関節位置の 2 次元座標として OpenPose を用いて取得し、データの正規化、整形を行った。

B, VGG16[16] を用いた学習データ

VGG は、ILSVRC2014 に参加したチーム名であり、VGG モデルはローカライゼーション課題で 1 位、分類課題では 2 位であったモデルで、VGG16 は 16 層の CNN である。Imagenet [17] に代表されるデータセットを用いることで、画像の分類精度は上記のとおり高いものになっている。今回は、Imagenet のデータセットを用い、VGG16 の全結合層を取り除くことで入力画像の特徴量のみを抽出する。得られた特徴量を入力データとして実験する。学習データ B を用いることで、OpenPose の 2 次元の関節位置情報を用いることの優位性を確認する。

C, ヒートマップを用いた学習データ

ヒートマップとは、データを可視化するため、行列型の数値データの強弱を色で視覚化した画像である。OpenPose は、関節位置を座標位置という二次元データだけでなく、部位の位置と部位間の関連度の行列を出力するため、ヒートマップとして出力することもできる。入力画像のヒートマップを作成し、それぞれのヒートマップを頭の関節位置を基準に移動し位置をそろえることで正規化したものを入力データとして用いる。得られたヒートマップは自前で用意した CNN を用い、学習の際に LSTM の重みと同様に CNN の重みも学習しながら実験する。図 9 を入力画像としたときのヒートマップの出力は図 10 の通りである。データセット C を用いることで、姿勢情報は関節位置の 2 次元座標で得ることの優位性を確認する。

D, 補間を行わない学習データ

Openpose で得られた姿勢情報の関節位置の 2 次元座標は、3.3.1 項で述べた通り、検出していない場合もある。学習データ D を用いることで、欠損値補間の優位性を確認する。

E, コート内の位置情報を加えない学習データ

コート内の位置情報を加えず、姿勢情報のみの 26 次元の学習データを作成した。学習データ E を用いることで、コート内位置情報を学習データに加えることの優位性を確認する。

エポック数を 50 としたときのそれぞれの学習データの実験結果を表 1 に示す、本手法である学習データ A を用い

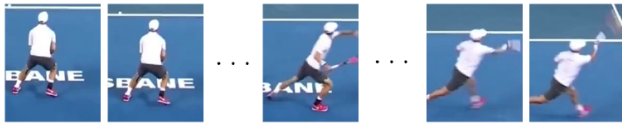


図 9 入力画像



図 10 ヒートマップの出力画像

表 1 各学習データにおける実験結果

データの型	Accuracy(%)	Precision(%)	Recall(%)
学習データ A	64.2	61.0	64.1
学習データ B	41.5	55.0	55.0
学習データ C	41.1	34.8	34.7
学習データ D	62.8	59.7	61.1
学習データ E	56.2	54.9	59.5

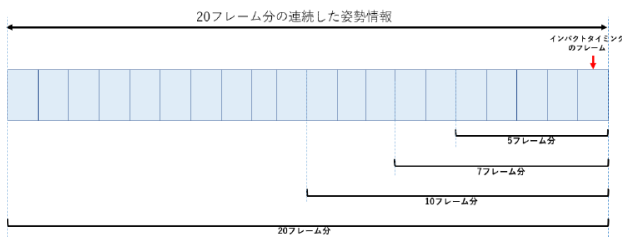


図 11 用いる学習データ

表 2 各学習データにおける実験結果

入力データのフレーム数	Accuracy(%)	Precision(%)	Recall(%)
5 フレーム	64.2	61.0	64.1
7 フレーム	66.8	74.2	67.7
10 フレーム	61.3	56.3	64.2
20 フレーム	60.2	59.4	60.6

た結果がいずれの評価指標も最も高い結果となった。

4.3 評価実験 2

評価実験 2 では、提案手法の学習データを用いて、図 11 のように姿勢情報のフレーム数を変化させた時の提案手法の優位性の比較実験を行う。それぞれインパクトタイミング前から 20 フレーム、10 フレーム、7 フレーム、5 フレームで比較を行う。

それぞれの実験結果は表 2 のとおりである。

4.4 評価実験 3

評価実験 3 では、提案手法の学習データを用いて図 12 のように姿勢情報を 5 フレームとし、インパクトタイミングから時間軸をずらした時の提案手法の優位性の比較実験を行う。それぞれインパクトタイミング前から 5 フレーム

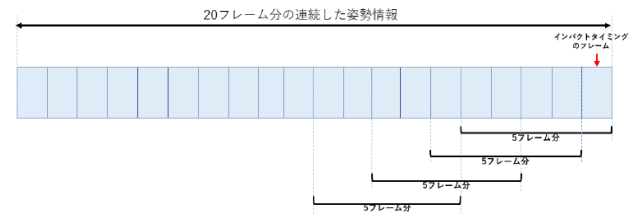


図 12 用いる学習データ

表 3 各学習データにおける実験結果

インパクトタイミングからのフレーム数	Accuracy(%)	Precision(%)	Recall(%)
0 フレーム	64.2	61.0	64.1
1 フレーム	63.3	60.4	60.7
3 フレーム	62.5	57.3	57.5
5 フレーム	60.1	52.0	62.7

ム、3 フレーム、1 フレーム、0 フレームで比較を行う。それぞれの実験結果は表 3 のとおりである。

4.5 考察

評価実験 1 より、提案手法が最も精度が高く打球方向推定ができていることが分かる。データセット B の結果より、Imagenet は物体認識を行うデータセットであり、つまり「人である」という特徴量が得られるが姿勢情報の特徴量は得られないことが原因であると考えられる。また、クレイコートとハードコートでは選手の背景のテクスチャが異なるため、コートによる影響もあると考えられる。一方、データセット C は、自前の CNN を用意し、LSTM とともに学習させた。精度が低い理由として考えられることは、学習データが足りないことである。CNN を学習させたため、学習データはより多くなければならないであろう。データセット A とデータセット D を比較することで、補間に優位性があること、またデータセット E と比較することで、コート内の位置情報を加えることに優位性があることがわかる。そもそも、打球方向を 3 クラス分類の場合、インパクト時に選手のコート内の位置である程度 2 クラス分類と同義になる。例えばコート右にいるときに、右方向へのクロスはアウトになってしまうため打たないからだ。また、コート内の選手位置を加えることで戦術分析と被ってしまうという可能性がある。例えばある選手は右側にいるときは左方向へのクロスを打つことが多いので、左方向へのクロスと推定されるということだ。しかし、テニスはテニスコート以外でボールを打つことは考えられず、姿勢情報とコート内の選手位置は互いに独立していないと考えられる。そのため、コート内の選手位置を姿勢情報に加えることは合理性がある。

評価実験 2 より、7 フレーム分のデータが最も打球方向推定の精度が高くなる。入力画像を確認したところ、平均して 10 フレームあたりからテイクバックが開始され、7 フレームあたりでラケットを振り出している。5 フレームあ

たりはラケットを振っている最中であった。そのため、ラケットの振出からインパクトタイミングまでの姿勢情報を用いるのが最も打球方向推定の精度が高くなることが分かった。今回実験に用いた動画は 25[fps] の動画であったため、7 フレーム目あたりからラケットの振出が始まっているが、今後フレームレートの高い動画になった場合でも、ラケットの振出位置から姿勢情報の入力とすることが精度が最も高くなると考えられる。

評価実験 3 より、インパクトタイミングに近い入力データほど精度が高くなることがわかる。5 フレーム分インパクトタイミングより前の姿勢情報からでも 6 割の打球方向推定が可能となっている理由として考えられるのは、コート内の選手位置も入力データに含まれるからであると考えられる。

5. 結論

本研究では、フォーム解析によるけが防止やテニス選手のパフォーマンス向上の一環として、テニス選手の姿勢情報より打球方向推定の手法を提案した。提案手法では、学習データの作成方法を述べた後、OpenPose によって得られた姿勢情報の前処理、そして、学習の方法を述べ、実験の結果、その他画像の特徴量を得る手法よりも提案手法の優位性を示した。また、打球方向推定の精度向上のために学習の手法の比較検討を行い、インパクトタイミング前 7 フレームからインパクトタイミングのフレームの 7 フレームを用いることで、打球方向推定で 66.8% の精度を得た。

参考文献

- [1] T.M.Masaki Takahashi, Toshiyuki Nakamura, "A method of soccer ball tracking from several viewpoints using a machine learning algorithm," in 映像情報メディア学会技術報告, pp. 5-8, 2014.
- [2] N.Owens, C.Harris, and C.Stennett, "Hawk-eye tennis system", in VIE 2003 International Conference on Visual Information Engineering 2003, pp.182-185, IET, 2003.
- [3] Sipko, Michal, and William Knottenbelt. "Machine learning for the prediction of professional tennis matches." MEng computing-final year project, Imperial College London, 2015.
- [4] J.Landlinger, S.Lindinger, T.Stoggl, H. Wagner, and E. Muller, "Key factors and timing patterns in the tennis forehand of different skill levels," Journal of sports science & medicine, vol.9, no.4, p.643, 2010.
- [5] Z.Cao, T.Simon, S.-E. Wei, and Y.Sheikh, Realtime multi-person 2d pose estimation using part affinity fields," in 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp.1302-1310, IEEE, 2017.
- [6] T.-Y. Lin, M.Maire, S.Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollar, and C. L. Zitnick, "Microsoft coco: Common objects in context," in European conference on computer vision, pp.740-755, Springer, 2014.
- [7] H.Fani, A.Mirlohi, H.Hosseini, and R.Herperst, "Swim stroke analytic: Front crawl pulling pose classification," in 2018 25th IEEE International Conference on Image Processing (ICIP), pp.4068-4072, IEEE, 2018.
- [8] S.Hochreiter and J.Schmidhuber, "Long short-term memory," Neural computation, vol.9, no.8, pp.1735-1780, 1997.
- [9] D.Yogatama, P.Blunsom, C.Dyer, E.Grefenstette, and W. Ling, "Learning to compose words into sentences with reinforcement learning," arXiv preprint arXiv:1611.09100, 2016.
- [10] J.Wang, P.Zhao, and S.C.Hoi, "Exact soft confidence-weighted learning," arXiv preprint arXiv:1206.4612, 2012.
- [11] Cai, Jiaxin, and Xin Tang. "RGB Video Based Tennis Action Recognition Using a Deep Weighted Long Short-Term Memory." arXiv preprint arXiv:1808.00845, 2018.
- [12] E.Dubrofsky, "Homography estimation," 2009.
- [13] A.Blake and M.Isard, "The condensation algorithm-conditional density propagation and applications to visual tracking," in Advances in Neural Information Processing Systems, pp.361-367, 1997.
- [14] X.Zhou, Q.Huang, L.Xie, and S.Cox, "A two layered data association approach for ball tracking," in Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on, pp.2317-2321, IEEE, 2013.
- [15] F.Yan, W.Christmas, and J.Kittler, "A tennis ball tracking algorithm for automatic annotation of tennis match," in British machine vision conference, vol.2, pp.619-628, 2005.
- [16] K.Simonyan and A.Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv preprint arXiv:1409.1556, 2014.
- [17] J.Deng, W.Dong, R.Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A largescale hierarchical image database," in Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on, pp.248-255, Ieee, 2009.